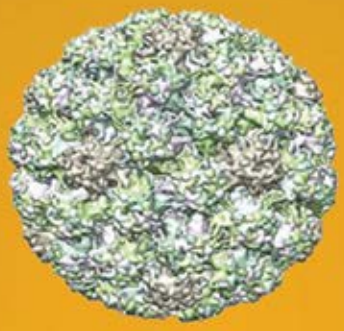
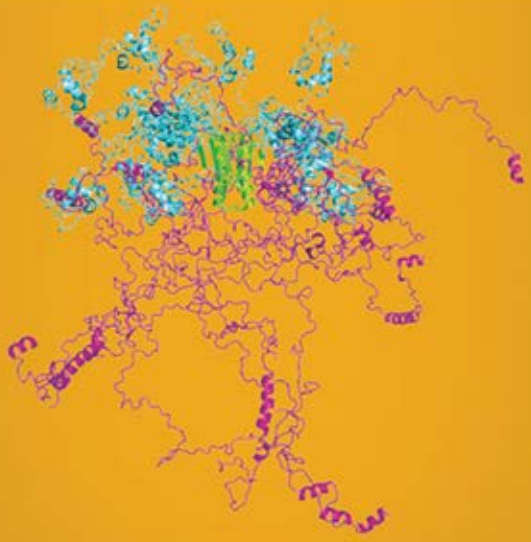
WILEY SERIES IN PROTEIN AND PEPTIDE SCIENCE
Vladimir N. Uversky, Series Editor



FLEXIBLE VIRUSES

Structural Disorder in Viral Proteins

EDITED BY Vladimir N. Uversky AND Sonia Longhi



WILEY

FLEXIBLE VIRUSES

WILEY SERIES ON PROTEIN AND PEPTIDE SCIENCE

Vladimir N. Uversky, Series Editor

Metalloproteomics · Eugene Permyakov

Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation Vladimir Uversky and Sonia Longhi

Protein Misfolding Diseases: Current and Emerging Principles and Therapies · Marina Ramirez-Alvarado, Jeffery W. Kelly, and Christopher M. Dobson

Calcium Binding Proteins · Eugene Permyakov and Robert H. Kretsinger

Protein Chaperones and Protection from Neurodegenerative Diseases · Stephan Witt

Transmembrane Dynamics of Lipids · Philippe Devaux and Andreas Herrmann

Flexible Viruses: Structural Disorder in Viral Proteins · Vladimir Uversky and Sonia Longhi

FLEXIBLE VIRUSES

Structural Disorder in Viral Proteins

Edited by

VLADIMIR N. UVERSKY

SONIA LONGHI

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Uversky, Vladimir N.
Flexible viruses structural disorder in viral proteins / Vladimir Uversky, Sonia Longhi.
p. cm. – (Wiley series in protein and peptide science ; 11)
Includes index.
ISBN 978-0-470-61831-8 (hardback)
1. Viral proteins. I. Longhi, Sonia. II. Title.
QR460.U94 2012
612′.015756–dc23

2011028226

Printed in the United States of America

oBook ISBN: 978-1-118-13557-0
ePDF ISBN: 978-1-118-13554-9
ePub ISBN: 978-1-118-13556-3
eMobi ISBN: 978-1-118-13555-6

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

The hypothesis that protein function relies on a precise 3D structure constitutes one of the central paradigms of biochemistry. According to this concept, a protein can perform its biological function(s) only being folded into a unique 3D structure, and all the information necessary for a protein to gain this unique 3D structure (in a given environment) is encoded in its amino acid sequence. However, recently, the validity of this structure–function paradigm has been seriously challenged, primarily through the wealth of counterexamples that have gradually accumulated over the past 20–25 years. These counterexamples demonstrated that many functional proteins or protein parts exist in an entirely or partly disordered state. These intrinsically disordered proteins (IDPs) lack a unique, stable 3D structure in solution, existing instead as dynamic ensembles of conformations and exerting their biological activity without a prerequisite stably folded structure.

IDPs possess a distinct set of specific features of their amino acid sequences and compositions (e.g., amino acid sequences of extended disordered proteins are characterized by the combination of a high content of charged residues and a low content of hydrophobic residues) that allows them to be distinguished from globular proteins. These peculiar sequence features have led to the development of various algorithms for disorder predictions, which allowed an estimation of the abundance of disorder in various biological systems. These studies showed that the frequency and length of disordered regions increase with increasing complexity of the organism. For example, long intrinsically disordered regions have been predicted to occur in 33% of eukaryotic proteins, and more than 10% of all eukaryotic proteins are expected to be wholly disordered. Furthermore, viruses and eukaryota were predicted to have 10 times more conserved disorder (roughly 1%) than archaea and bacteria (0.1%). Beyond these computational studies, an increasing amount of experimental evidence has been gathered in the last decade pointing

out the large abundance of intrinsic disorder within the living world: more than 625 proteins containing 1342 disordered regions have been annotated so far in the Disprot database (http://www.disprot.org).

Despite this large body of experimental evidence pointing out the abundance and biological relevance of intrinsic disorder in the living world, the notion of a tight dependence of protein function on a precise 3D structure is still deeply anchored in many scientists' mind. The reasons for this lack of awareness or even "resistance" to the concept of protein intrinsic disorder are multiple. First, the growing numbers of protein structures determined by X-ray crystallography and by NMR in the last three decades have shifted the attention of scientists away from the numerous examples of IDPs. Second, IDPs have been long unnoticed because researchers encountering examples of structural disorder mainly ascribed them to experimental errors and artifacts (e.g., failure to purify a given protein in folded biologically active form) and, as such, purged them from papers and reports. Third, structural disorder is hard to conceive and classify. Fourth, IDPs have been neglected because of the perception that a limited amount of mechanistic data could be derived from their study. Fifth, until very recently, no special techniques existed for targeted structural characterization of IDPs and information about intrinsic disorder was retrieved mainly as the lack of specific signals expected for ordered proteins. Yet, the evidence that IDPs exist both *in vitro* and *in vivo* is compelling and justifies considering them as a separate class within the protein realm.

Many IDPs undergo a disorder-to-order transition on binding to their physiological partner(s), a process termed *induced folding*. IDPs bind to their target(s) through "molecular recognition elements" (MoREs) or "molecular recognition features" (MoRFs). MoRFs are interaction-prone short segments with an increased foldability, which are embedded within long disordered regions and become ordered on binding to a specific partner. On the basis of their secondary structure in the bound form, MoRFs can be grouped into four structural classes: α-MoRFs, β-MoRFs, i-MoRFs (irregular-MoRFs), and complex-MoRFs. The conformation of MoRFs in the unbound forms can be either wholly disordered or partially preformed, thus reflecting an inherent conformational preference. In the latter case, a transiently populated folded state would exist even in the absence of the partner for a part time, thus implying that the folding induced by the partner would rely (at least partly) on conformer selection (i.e., selection by the partner of a preexisting conformation) rather than on a "fly-casting" mechanism. It has been proposed that the restriction in the conformational space of MoRFs in the unbound state could reduce the entropic cost of binding, thereby enhancing affinity. IDPs can bind their target(s) with a high extent of conformational polymorphism, with binding generally involving larger normalized interface areas than those found between rigid partners, with protein interfaces being enriched in hydrophobic residues. Thus, protein–protein interactions established by IDPs rely more on hydrophobic–hydrophobic than on polar–polar contacts. Finally, the structural plasticity of MoRFs is assumed to facilitate the binding of IDPs to multiple structurally unrelated partners. Strikingly, as a result of such one-to-many recognition, one IDP can recognize multiple binding partners and gain different types of structure being bound to these different binding

partners. In other words, when the situation necessitates it, the MoRF can "morph" into α-helix, β-strand, and irregular structure in order to accommodate different structured partners.

The protein flexibility that is inherent to disorder confers numerous functional advantages. The increased plasticity of IDPs (i) enables binding to numerous structurally distinct targets; (ii) provides the ability to overcome steric restrictions by enabling larger surfaces of interaction; and (iii) allows protein interactions to occur with both high specificity and low affinity. Accordingly, most IDPs are involved in functions that imply multiple partner interactions (e.g., one-to-many and many-to-one binding scenarios), such as molecular recognition, molecular assembly (and amyloidogenesis), cell cycle regulation, signal transduction, and transcription. As such, IDPs are implicated in the development of several pathological conditions (including cancer and cardiovascular diseases) and have been shown to be promising targets for drug development.

Intrinsic disorder is a distinctive and common feature of "hub" proteins, with disorder serving as a determinant of protein promiscuity. Intrinsic disorder also serves as a determinant of the transient nature of the interactions that IDPs can establish, by virtue of the presumed rather low affinity that typifies interactions involving IDPs. The relationship between structural disorder and regulation provides a plausible explanation for the prevalence of disorder in higher organisms, which have more complex signaling and regulatory pathways. On the other hand, the abundance of disorder within viruses likely reflects the need for genetic compaction, where a single disordered protein can establish multiple interactions and hence exert multiple concomitant biological effects. In addition, structural disorder might endow viral proteins with broader ability to interact with the components of the host and may also be related to high adaptability levels and mutation rates observed in viruses, thus representing a unique strategy for buffering the deleterious effects of mutations.

In this book, a thorough description of the current knowledge on the abundance, structural peculiarities, and functional implementations of intrinsic disorder in viral proteins is provided.

Chapter 1, by Bin Xue, Robert W. Williams, Christopher J. Oldfield, Gerard Kian-Meng Goh, A. Keith Dunker, and Vladimir N. Uversky, provides the general overview of intrinsic disorder in viral proteins. It illustrates some structural peculiarities of viral proteins and discusses the roles of intrinsic disorder in functions of different viral proteins.

In Chapter 2, Lars Liljas considers the multiple roles of intrinsic disorder in the form of flexible arms in virus capsids on the basis of the structures of several nonenveloped viruses. The covered aspects range from the roles of those flexible arms in binding to the viral nucleic acids, to controlling the assembly of capsids with quasi-equivalence, and to stabilizing the shell to be controlled by external signals for the release of the viral genome.

In Chapter 3, Johnny Habchi, Laurent Mamelli, and Sonia Longhi analyze the experimental data on the abundance of structural disorder within the nucleoprotein and phosphoprotein from the closely related measles, Nipah, and Hendra viruses.

They also describe the molecular mechanisms governing the disorder-to-order transition of the intrinsically disordered C-terminal domain of measles virus N on binding to the C-terminal X domain of the measles virus phosphoprotein.

Chapter 4, by Malene Ringkjøbing Jensen, Pau Bernadó, Rob W. H. Ruigrok, and Martin Blackledge, addresses the peculiarity of structural disorder in Sendai virus nucleoprotein and phosphoprotein. The chapter is focused on the domain organization of the phosphoprotein and nucleoprotein and the structural characterization of these proteins using different experimental techniques such as X-ray crystallography, nuclear magnetic resonance spectroscopy, and small angle X-ray scattering.

Cédric Leyrat, Francine C.A. Gérard, Euripedes A. Ribeiro Jr., Ivan Ivanov, and Marc Jamin in Chapter 5 describe the functional role of intrinsic disorder in the Rhabdoviridae replication complex comprised of three proteins, the nucleoprotein, the phosphoprotein, and the large subunit (L) of the RNA-dependent RNA polymerase. The roles of intrinsically disordered regions in the mechanism of replication/transcription are discussed, and a new model for the interaction of the large subunit of the RNA-dependent RNA polymerase with its N-RNA template is proposed.

Chapter 6, by Gerard Kian-Meng Goh, Bin Xue, A. Keith Dunker, and Vladimir N. Uversky, is dedicated to the analysis of intrinsic disorder in matrix proteins from HIV-related viruses, whereas in Chapter 7, the same authors consider various aspects of structural disorder in proteins from the influenza A virus.

Tali H. Reingewertz, Deborah E. Shalev, and Assaf Friedler dedicated Chapter 8 to elucidating the roles of intrinsic disorder in the function of the HIV-1 Vif protein, which is known to counteract the antiviral activity of the host cellular cytosine deaminase and its interactions with multiple binding partners.

In Chapter 9, Shaheen Shojania and Joe D. O'Neil introduce a small, intrinsically disordered RNA-binding protein crucial for viral replication, the HIV-1 transcriptional regulator Tat. The authors emphasize that intrinsic disorder in the polypeptide backbone can explain Tat's binding promiscuity and its ability to modulate multiple biological processes.

In Chapter 10, Smita Nair, M.R.N. Murthy, and H.S. Savithri summarize the current knowledge on biophysical, biochemical, and structural properties of the intrinsically disordered proteins, VPg (viral proteins genome-linked) and P8, and the disordered segments of coat protein from the *Sesbania* mosaic virus.

Chapter 11 is written by Jadwiga Chroboczek, Eugénie Hébrard, Kristiina Mäkinen, Thierry Michon, and Kimmo Rantalainen to address the peculiarities of intrinsic disorder in the VPgs of potyviruses. This international team provided a compelling support to the idea that intrinsic disorder is crucial for the biological activity of VPgs of potyviruses and suggests that intrinsic disorder may be a feature shared by all the VPgs of unrelated RNA viruses.

In Chapter 12, Lucía B. Chemes, Ignacio E. Sánchez, Leonardo G. Alonso, and Gonzalo de Prat-Gay discuss the roles of intrinsic disorder in a prototypic viral

oncoprotein, the E7 protein from the human papillomavirus, which is responsible for the cellular transformation behind one of the most widespread cancers in women.

Manuel Morillas, Heike Eberl, Fréderic H.-T. Allain, Rudi Glockshuber, and Eva Kuennnemann in Chapter 13 present data supporting the intriguing mechanism of the enzymatic activity of the Semliki Forest virus capsid protease, which is shown to be disordered and yet displays catalytic activity.

In Chapter 14, Roland Ivanyi-Nagy, Eve-Isabelle Pécheur, and Jean-Luc Darlix focus on the multifaceted activities of core proteins in hepatitis C virus and related viruses, and put special emphasis on the relevance of intrinsic disorder for these functions.

Chapter 15, by Xavier Hanoulle, Isabelle Huvent, Arnaud Leroy, Hong Ye, Cong Bao Kang, Yu Liang, Claire Rosnoblet, Jean-Michel Wieruszeski, Ho Sup Yoon, and Guy Lippens, discusses the evolutionary conservation of intrinsic disorder in viral proteins using the intrinsically unstructured domain 2 of the RNA-dependent RNA polymerase from the hepatitis C virus as an illustration.

In Chapter 16, Kristian Schweimer and Paul Rösch introduce the antitermination protein N from bacteriophage λ, which is disordered in its free form, but gains defined structures on interaction with the RNA recognition site *nutBoxB* and bacterial host factor NusA.

Finally, V. V. Makarov, M. E. Taliansky, E. N. Dobrov, and N. O. Kalinina dedicated their Chapter 17 to the Hordeivirus movement TGB1 proteins, which form ribonucleoprotein complex for the cell-to-cell and long-distance movement of viral genome in plants. In *Poa* semilatent virus, TGB1 contains both an N-terminal extension region, which consists of a completely intrinsically disordered extreme N-terminal domain (NTD) and an internal domain (ID) adopting a partially disordered, molten globule state, and a C-terminal NTPase/helicase domain. The functional implications of flexibility of the disordered domains are discussed in light of their role in the assembly and movement of viral RNP complexes at different stages of viral transport in the plant.

This book is intended to stimulate and inspire scientists to further extend this fascinating area of research, and we hope that in future years, it will promote research in this rather poorly explored field.

Vladimir N. Uversky
Sonia Longhi

INTRODUCTION TO THE WILEY SERIES ON PROTEIN AND PEPTIDE SCIENCE

Proteins and peptides are the major functional components of the living cell. They are involved in all aspects of the maintenance of life. Their structural and functional repertoires are endless. They may act alone or in conjunction with other proteins, peptides, nucleic acids, membranes, small molecules, and ions during various stages of life. Dysfunction of proteins and peptides may result in the development of various pathological conditions and diseases. Therefore, the protein/peptide structure–function relationship is a key scientific problem lying at the junction point of modern biochemistry, biophysics, genetics, physiology, molecular and cellular biology, proteomics, and medicine.

The *Wiley Series on Protein and Peptide Science* is designed to supply a complementary perspective from current publications by focusing each volume on a specific protein- or peptide-associated question and endowing it with the broadest possible context and outlook. The volumes in this series should be considered required reading for biochemists, biophysicists, molecular biologists, geneticists, cell biologists, and physiologists, as well as those specialists in drug design and development, proteomics, and molecular medicine, with an interest in proteins and peptides. I hope that each reader will find in the volumes within this book series interesting and useful information.

First and foremost, I would like to acknowledge the assistance of Anita Lekhwani of John Wiley & Sons, Inc. throughout this project. She has guided me through countless difficulties in the preparation of this book series, and her enthusiasm, input, suggestions, and efforts were indispensable in bringing the *Wiley Series on Protein and Peptide Science* into existence. I would like to take this opportunity to thank everybody whose contribution in one way or another has helped and supported this project. Finally, special thank you goes to my wife, sons, and mother for their constant support, invaluable assistance, and continuous encouragement.

CONTRIBUTORS

Fréderic H.-T. Allain, Institute of Molecular Biology and Biophysics, ETH-Hoenggerberg, Zurich CH-8093, Switzerland

Leonardo G. Alonso, XBio Inc., Concepción Arenal 4220, CP 1427 Buenos Aires, Argentina

Pau Bernadó, Institute for Research in Biomedicine. Baldiri Reixac, 10.08028 Barcelona, Spain

Martin Blackledge, Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF, UMR 5075, 41 rue Jules Horowitz, 38027 Grenoble, France

Lucía B. Chemes, Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, CONICET, Av. Patricias Argentinas 435, CP 1405, Buenos Aires, Argentina

Jadwiga Chroboczek, Therex/TIMC-IMAG, Université Joseph Fourier, Domaine de la Merci, 38706 La Tronche Cedex, France

Jean-Luc Darlix, LaboRetro, Unité de Virologie Humaine INSERM 758, IFR 128, ENS de Lyon, 46 allée d'Italie, 69364 Lyon, France

Gonzalo de Prat-Gay, Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, CONICET, Av. Patricias Argentinas 435, CP 1405, Buenos Aires, Argentina

Eugeny N. Dobrov, A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia

A. Keith Dunker, Center for Computational Biology and Bioinformatics, Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, IN 46202, USA; Indiana University School of Informatics, Indianapolis, IN 46202, USA

Heike Eberl, Institute of Molecular Biology and Biophysics, ETH-Hoenggerberg, Zurich CH-8093, Switzerland; Department of LR-EI, Clinical Trials Centralized Diagnostics, Roche Diagnostics GmbH, Penzberg, Germany

Assaf Friedler, Institute of Chemistry, The Hebrew University of Jerusalem, Safra Campus, Givat Ram, 91904 Jerusalem, Israel

Francine C.A. Gérard, UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions (UVHCI), 6 rue Jules Horowitz 38042 Grenoble cedex 9, France

Rudi Glockshuber, Institute of Molecular Biology and Biophysics, ETH-Hoenggerberg, Zurich CH-8093, Switzerland

Gerard Kian-Meng Goh, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA; Institute of Molecular and Cell Biology, Singapore 138673, Republic of Singapore

Johnny Habchi, Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités d'Aix-Marseille I et II, 163, Avenue de Luminy, Case 932, 13288 Marseille cedex 09, France

Xavier Hanoulle, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS, IFR 147, Université Lille1-Sciences et Technologies, 59655 Villeneuve d'Ascq, France

Eugénie Hébrard, UMR 186 Résistance des Plantes aux Bio-agresseurs, Institut de Recherche pour le Développement, BP 64501, 34394 Montpellier cedex 5, France

Isabelle Huvent, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS, IFR 147, Université Lille1-Sciences et Technologies, 59655 Villeneuve d'Ascq, France

Ivan Ivanov, UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz 38042 Grenoble cedex 9, France; Institut Laue Langevin, 6 rue Jules Horowitz 38042 Grenoble cedex 9, France

Roland Ivanyi-Nagy, Molecular Parasitology Group, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

Marc Jamin, UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble cedex 9, France

Malene Ringkjøbing Jensen, Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF, UMR 5075, 41 rue Jules Horowitz, 38027 Grenoble, France

Natalia O. Kalinina, A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia

Cong Bao Kang, Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637511, Singapore

Eva Kuennemann, Institute of Molecular Biology and Biophysics, ETH-Hoenggerberg, Zurich CH-8093, Switzerland; Prionics G, Schlieren, Switzerland

Arnaud Leroy, Université Lille1-Sciences et Technologies, Unité de Glycobiologie Structurale et Fonctionnelle, Villeneuve d'Ascq, France; Laboratoire de Biochimie Appliquée, Faculté de Pharmacie à Châtenay-Malabry (Paris XI), Châtenay-Malabry cedex, France, Nanyang Technological University, School of Biological Sciences, Singapore

Cédric Leyrat, UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions (UVHCI), 6 rue Jules Horowitz 38042 Grenoble cedex 9, France

Yu Liang, Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637511, Singapore

Lars Liljas, Department of Cell and Molecular Biology, Uppsala University, Box 596, 751 24 Uppsala, Sweden

Guy Lippens, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS, IFR 147, Université Lille1-Sciences et Technologies, 59655 Villeneuve d'Ascq, France

Sonia Longhi, Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités d'Aix-Marseille I et II, 163, Avenue de Luminy, Case 932, 13288 Marseille cedex 09, France

Valentin V. Makarov, A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia

Kristiina Mäkinen, Department of Food and Environmental Sciences, University of Helsinki, P.O. Box 27, 00014 Finland

Laurent Mamelli, Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités d'Aix-Marseille I et II, 163, Avenue de Luminy, Case 932, 13288 Marseille cedex 09, France

Thierry Michon, Thierry Michon, UMR 1332 Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France

Manuel Morillas, Institute of Molecular Biology and Biophysics, ETH-Hoenggerberg, Zurich CH-8093, Switzerland; Institut de Recerca Vall d'Hebron, Barcelona, Spain

M. R. N. Murthy, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Smita Nair, Department of Biochemistry, Indian Institute of Science, Bangalore-560012, India; NCI-NIH, HIV Drug resistance program, 1050 Boyles street, 535 Bldg, Rm# 235, Frederick, MD 21702, USA

Christopher J. Oldfield, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA; Indiana University School of Informatics, Indianapolis, IN 46202, USA

Joe D. O'Neil, Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Eve-Isabelle Pécheur, Institut de Biologie et Chimie des Protéines, UMR CNRS 5086, Université Lyon 1, IFR128 Lyon BioSciences Gerland, 69007 Lyon, France

Kimmo Rantalainen, Department of Biosciences and Nutrition, Karolinska Institute, 14157 Huddinge, Sweden

Tali H. Reingewertz, Institute of Chemistry, The Hebrew University of Jerusalem, Safra Campus, Givat Ram, 91904 Jerusalem, Israel

Euripedes A. Ribeiro, Jr., UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions (UVHCI), 6 rue Jules Horowitz, 38042 Grenoble cedex 9, France

Paul Rösch, Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

Claire Rosnoblet, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS, IFR 147, Université Lille1-Sciences et Technologies, 59655 Villeneuve d'Ascq, France

Rob W. H. Ruigrok, Unit for Virus Host Cell Interactions, UJF-EMBL-CNRS UMI 3265, 6 rue Jules Horowitz, BP 181, 38042 Grenoble, France

Ignacio E. Sánchez, Protein Physiology Laboratory, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Av, Intendente Güiraldes 2160, Ciudad Universitaria, CP 1428 Buenos Aires, Argentina

H. S. Savithri, Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India

Kristian Schweimer, Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

Deborah E. Shalev, The Wolfson Centre for Applied Structural Biology, The Hebrew University of Jerusalem, Safra Campus, Givat Ram, 91904 Jerusalem, Israel

Shaheen Shojania, Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Michael E. Taliansky, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA Scotland, UK

Vladimir N. Uversky, Center for Computational Biology and Bioinformatics, Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, IN 46202, USA; Department of Molecular Medicine, University of South Florida, FL 33612, USA; Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

Jean-Michel Wieruszeski, Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS, IFR 147, Université Lille1-Sciences et Technologies, 59655 Villeneuve d'Ascq, France

Robert W. Williams, Department of Biomedical Informatics, Uniformed Services University, Bethesda, MD 20814, USA

Bin Xue, Center for Computational Biology and Bioinformatics, Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, IN 46202, USA; Department of Molecular Medicine, University of South Florida, FL 33612, USA

Hong Ye, Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637511, Singapore

Ho Sup Yoon, Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637511, Singapore

1

DO VIRAL PROTEINS POSSESS UNIQUE FEATURES?

BIN XUE, ROBERT W. WILLIAMS, CHRISTOPHER J. OLDFIELD, GERARD K.-M. GOH, A. KEITH DUNKER, AND VLADIMIR N. UVERSKY

1.1 INTRODUCTION

Many proteins (or protein regions) are intrinsically disordered. They lack unique 3D structures in their native, functional states under physiological conditions *in vitro* (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001, 2002a,b; Tompa, 2002, 2003; Uversky, 2002a,b, 2003; Minezaki et al., 2006). The major functions of such proteins and regions are signaling, recognition, and regulation activities (Wright and Dyson, 1999, 2009; Dunker et al., 2002a,b; 2005; 2008a,b; Dyson and Wright, 2005; Uversky et al., 2005; Radivojac et al., 2007; Dunker and Uversky, 2008; Oldfield et al., 2008; Tompa et al., 2009). Owing to these crucial functional roles, intrinsically disordered proteins (IDPs) are highly abundant in all species. According to computational predictions, typically 7–30% prokaryotic proteins contain long disordered regions of more than 30 consecutive residues, whereas in eukaryotes the amount of such proteins reaches 45–50% (Romero et al., 1997, 2001; Dunker et al., 2001; Ward et al., 2004; Oldfield et al., 2005a,b; Feng et al., 2006). Furthermore, almost 70% of proteins in the PDB (which is biased to structured proteins) have intrinsically disordered regions (IDRs), which are indicated by missing electron density (Obradovic et al., 2003). Numerous disordered proteins have been shown to be associated with cancer (Iakoucheva et al., 2002), cardiovascular disease (Cheng et al., 2006), amyloidoses (Uversky, 2008a), neurodegenerative diseases (Uversky, 2008b), diabetes, and other human diseases (Uversky et al., 2008), an

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

observation that was used to introduce the "disorder in disorders" or D^2 concept (Uversky et al., 2008).

Recently, we showed also that IDPs are abundant in the human diseasome (Midic et al., 2009), a framework that systematically linked the human disease phenome (which includes all the human genetic diseases) with the human disease genome (which contains all the disease-related genes) (Goh et al., 2007). This framework was constructed from the analysis of two networks, a network of genetic diseases, the "human disease network," where two diseases are directly linked if there is a gene that is directly related to both of them, and a network of disease genes, the "disease gene network," where two genes are directly linked if there is a disease to which they are both directly related (Goh et al., 2007). Our analysis revealed that there were noticeable differences in the abundance of intrinsic disorder in human disease-related as compared to disease-unrelated proteins (Midic et al., 2009). Furthermore, various disease classes were significantly different with respect to the content of disordered proteins.

Furthermore, we have shown that intrinsic disorder is highly abundant in proteins of the parasitic protozoa (Mohan et al., 2008). Since viruses are common infectious pathogens, here, we summarize some literature data on the abundance of intrinsic disorder in viruses and explore the functional roles of intrinsic disorder in these intriguing "organisms at the edge of life."

Viruses are the most abundant living entities (Breitbart and Rohwer, 2005). For example, 1 mL of natural water contains up to 2.5×10^8 viral particles (Bergh et al., 1989), and the total number of viral particles exceeds the number of cells by at least an order of magnitude (Sano et al., 2004; Edwards and Rohwer, 2005). They are common parasitic organisms that live in the infected cells of Eukarya, Archaea, and Bacteria (or even inside other viruses) and produce virions to disseminate their genes (Breitbart and Rohwer, 2005; Edwards and Rohwer, 2005; Lawrence et al., 2009). Viruses do not have a defined cellular structure and are structurally very simple consisting of two or three parts. This includes two common components found in all viruses, DNA- (double-stranded or single-stranded) or RNA-based genes, and a protein coat protecting the genetic material (this proteinaceous coat is known as the *capsid*), and a lipid-based envelope surrounding some of the viruses when they are outside the host cells. In addition to the capsid proteins, some complex viruses also contain the so-called nonstructural proteins that assist in the construction of their capsid and viral regulatory and accessory proteins. Furthermore, enveloped viruses contain several integral membrane proteins, and matrix proteins forming the so-called matrix, another biologically active proteinaceous coat located right beneath the envelope.

Historically, there is no uniform opinion on whether the viruses are a form of life or just simple nonliving organic structures that interact with living organisms, or are yet the "organisms at the edge of life" (Rybicki, 1990). This difference in opinion originates from the facts that although viruses possess genes, evolve by natural selection, and reproduce by creating multiple copies of themselves through self-assembly, they do not have a defined cellular structure, as well as they lack

their own metabolism, require a host cell to make new products, and therefore cannot reproduce outside the host cell (Holmes, 2007).

In the evolutionary history of life, the origin of viruses is unclear. Currently, there are three major hypotheses for virus origin (Forterre, 2006):

1. Coevolution or the virus first hypothesis (here, viruses appeared simultaneously with the cells early in the history of earth and since that time are dependent on cellular life for many millions of years);
2. Cellular origin or vagrancy hypothesis (here, viruses evolved from pieces of pieces of RNA or DNA (e.g., plasmids, pieces of naked DNA that can move between, or transposons, pieces of DNA that replicate and move around to different positions within the genes) that "escaped" from the genes of a larger organism);
3. Regressive or degeneracy hypothesis (here, viruses originally were small cells that parasitized larger cells and that, with time, lost all the genes unused because of their parasitism).

It is suggested that RNA viruses may have originated in the nucleoprotein world (which followed the RNA world) by escaping or reduction from the primordial RNA-containing cells, whereas DNA viruses (at least some of them) might have evolved directly from RNA viruses (Forterre, 2006). Irrespective of the virus origin hypothesis, the facts that viruses infect cells from the three domains of life, Archaea, Bacteria, and Eukarya, share homologous features, and have probably existed since living cells first evolved (Iyer et al., 2006), clearly suggest that viruses originated very early in the evolution of life (Koonin et al., 2006). This antiquity of viruses can explain why most viral proteins have no homologs in cellular organisms or have only distantly related ones (Forterre, 2006).

Importantly, viruses are suggested to play a number of crucial roles in the general evolution of life. For example, they are responsible for the so-called horizontal gene transfer, a process by which an organism incorporates genetic material from another organism without being the offspring of that organism, which increases genetic diversity (Canchaya et al., 2003). In fact, 3–8% of the human genome is suggested to be composed of fragments of viral DNA. Furthermore, since it is believed that some DNA replication proteins originated in the virosphere and were later transferred to cellular organisms, viruses could play a vital role in the invention of DNA and DNA replication mechanisms and therefore could serve as crucial drives of the origin of the eukaryotic nucleus, and even of the formation of the three domains of life (Forterre, 2006).

Since viruses are believed to play a major role in the evolution of life, and since they are very different from all other life forms on earth, recently, a division was proposed to biological entities into two groups of organisms, namely, ribosome-encoding organisms, which include eukaryotic, archaeal, and bacterial organisms, and capsid-encoding organisms, which include viruses (Raoult and Forterre, 2008). Therefore, viruses are defined now as capsid-encoding organisms, which contain proteins and nucleic acids, self-assemble in the nucleocapsids, and use a ribosome-encoding organism for the completion of their life cycle (Raoult and Forterre, 2008).

This chapter illustrates some structural peculiarities of viral proteins and discusses the role of intrinsic disorder in their functions.

1.2 CLASSIFICATION AND FUNCTIONS OF VIRAL PROTEINS

Viral genomes are typically rather small ranging in size from 6 to 8 proteins (e.g., human papilloma virus (HPV)) to ~1000 proteins (e.g., *Acanthamoeba polyphaga* mimivirus (APMV)). Functionally, viral proteins are grouped into structural, nonstructural (NS), regulatory, and accessory proteins. For example, there are eight major proteins encoded by HPV. Proteins E1 and E2 are involved in viral replication as well as in the regulation of early transcription. E1 binds to the origin of replication and exhibits ATPase as well as helicase activity (Ustav and Stenlund, 1991; Hughes and Romanos, 1993), whereas E2 forms a complex with E1, facilitating its binding to the origin of viral replication (Mohr et al., 1990; Ustav and Stenlund, 1991; Frattini and Laimins, 1994). Furthermore, E2 acts as a transcription factor that positively and negatively regulates early gene expression by binding to specific E2 recognition sites within the upstream regulatory region (URR) (Cripe et al., 1987; Gloss et al., 1987). E4 is the most highly expressed protein in the productive life cycle of HPVs, and it plays a number of important roles in promoting the differentiation-dependent productive phase of the viral life cycle (Wilson et al., 2005; Brown et al., 2006; Davy et al., 2006). The E5 protein has weak transforming capabilities *in vitro* (Leechanachai et al., 1992; Straight et al., 1993), supports HPV late functions (Fehrmann et al., 2003; Genther et al., 2003), and disrupts MHC class II maturation (Zhang et al., 2003). Finally, L1 and L2 are the major and the minor capsid proteins, respectively.

Two early proteins (E6 and E7 oncoproteins) are mainly responsible for HPV-mediated malignant cell progression, leading ultimately to an invasive carcinoma. Proteins E6 and E7 function as oncoproteins in high risk HPVs, at least in part, by targeting the cell cycle regulators p53 and Rb, respectively. E7 has been shown to be involved in cellular processes such as cell growth and transformation (McIntyre et al., 1996), gene transcription (Massimi et al., 1997), apoptosis, and DNA synthesis, among other processes (Halpern and Münger, 1995). It interacts with many important proteins including the Rb tumor suppressor and its family members, p107 and p130 (Dyson et al., 1989), glycolytic enzymes (Zwerschke et al., 1999; Mazurek et al., 2001), histone deacetylase (Brehm et al., 1999), kinase p33CDK2, and cyclin A (Tommasino et al., 1993), as well as the cyclin-dependent kinase inhibitor $p21^{cip1}$ protein (Jian et al., 1998). Furthermore, it has been shown that E7 also binds to a protein phosphatase 2A (PP2A) (Pim et al., 2005). Formation of this complex sequesters PP2A, inhibiting its interaction with protein kinase B (PKB) or Akt (which is one of the several second messenger kinases that are activated by cell attachment and growth factor signaling and that transmit signals to the cell nucleus to inhibit apoptosis and thereby increase cell survival during proliferation (Brazil and Hemmings, 2001)), thereby maintaining PKB/Akt signaling by inhibiting its dephosphorylation.

E6 primarily promotes tumorigenesis by stimulating cellular degradation of the tumor suppressor p53 via formation of a trimeric complex comprising E6, p53, and the cellular ubiquitination enzyme E6AP (Scheffner et al., 1990, 1993). Besides this crucial role in the regulation of p53 degradation, E6 displays numerous activities unrelated to p53. These include but are not limited to recognition of a variety of other cellular proteins: transcription coactivators p300/CBP (Patel et al., 1999; Zimmermann et al., 1999) and ADA3 (Kumar et al., 2002), transcription factors c-Myc (Gross-Mesilaty et al., 1998) and IRF3 (Ronco et al., 1998), replication protein hMCM7 (Kukimoto et al., 1998), DNA repair proteins MGMT (Srivenugopal and Ali-Osman, 2002), protein kinases PKN (Gao et al., 2000) and Tyk2 (Li et al., 1999), Rap-GTPase activating protein E6TP1 (Gao et al., 1999), tumor necrosis factor receptor TNF-R1 (Filippova et al., 2002), apoptotic protein Bak (Thomas and Banks, 1999), clathrin-adaptor complex AP-1 (Tong et al., 1998), focal adhesion component paxillin (Tong and Howley, 1997), calcium-binding proteins E6BP (Chen et al., 1995) and fibulin-1 (Du et al., 2002), and several members of the PDZ protein family, including hDLG (Kiyono et al., 1997), hScrib (Nakagawa and Huibregtse, 2000), MAGI-1 (Glaunsinger et al., 2000), and MUPP1 (Lee et al., 2000). Furthermore, E6 activates or represses several cellular or viral transcription promoters (Sedman et al., 1991; Morosov et al., 1994; Dey et al., 1997; Ronco et al., 1998), such as transcriptional activation of the gene encoding the retrotranscriptase of human telomerase (Gewin and Galloway, 2001; Oh et al., 2001). In addition, it has been recently established that E6 recognizes four-way DNA junctions (Ristriani et al., 2000, 2001). The function of the low risk HPV E6 is less well studied. However, the low risk E6 lacks a number of activities that correlate with the oncogenic activity of the high risk HPV E6. For example, low risk E6 neither binds PDZ proteins (Kiyono et al., 1997) or E6TP1 (Gao et al., 1999) nor targets p53 for degradation (Scheffner et al., 1990; Li and Coffino, 1996). Like the high risk E6, low risk E6 does bind MCM7 (Kukimoto et al., 1998) and Bak (Thomas and Banks, 1999) and inhibits p300 acetylation of p53 (Thomas and Chiang, 2005).

1.2.1 Structural Proteins Form the Viral Capsid and Envelope

1.2.1.1 Capsid The capsid is the proteinaceous shell of the virus, which consists of several protomers (also known as *capsomers*), oligomeric protein subunits. Often, capsid proteins are conjugated with DNA or RNA, forming the viral nucleoprotein complex. It is important to remember that such viral nucleoproteins are multifunctional, being able to interact with nucleic acid and other proteins. For example, the transcription and replication of the measles virus, the RNA genome of which is encapsidated by the nucleoprotein (N), are initiated by the RNA-dependent RNA polymerase binding to the nucleocapsid via the phosphoprotein (P) (Longhi, 2009).

The packing of capsomers defines the shape of a viral capsid, which can be helical, icosahedral, or complex. Capsids of the *helical* or *filamentous* viruses are highly ordered helical structures consisting of a single type of capsomer stacked

around a central axis. The genetic material of these viruses, single-stranded RNA or, in some cases, single-stranded DNA is located inside a central cavity of the capsid, where it is bound to the capsid proteins via the electrostatic interactions between negative charges on nucleic acid and positive charges on the protein. The length of a helical capsid is dependent on the length of the viral nucleic acid, whereas its diameter is determined by the size and arrangement of capsomers. These rod-shaped or filamentous viruses can be short and highly rigid, or long and very flexible. Illustrative examples of the filamentous or helical viruses are tobacco mosaic virus (TMV), *Sulfolobus islandicus* filamentous virus (SIFV), *Acidianus* filamentous virus 1 (AFV1), filamentous bacteriophage fd, and others.

Capsids of the majority of viruses are *icosahedral* or near-spherical with icosahedral symmetry. A regular icosahedron is the optimal way to pack identical subunits to form a closed shell (Fig. 1.1a). Since there are 20 identical equilateral triangular faces in an icosahedron, the minimal number of identical subunits to form such a structure is 60 (Fig. 1.1b). Here, each triangular face is made up of three identical subunits. The capsomer of the icosahedral virus includes the five identical subunits that surround each vertex and are arranged in a fivefold symmetry (Fig. 1.1c).

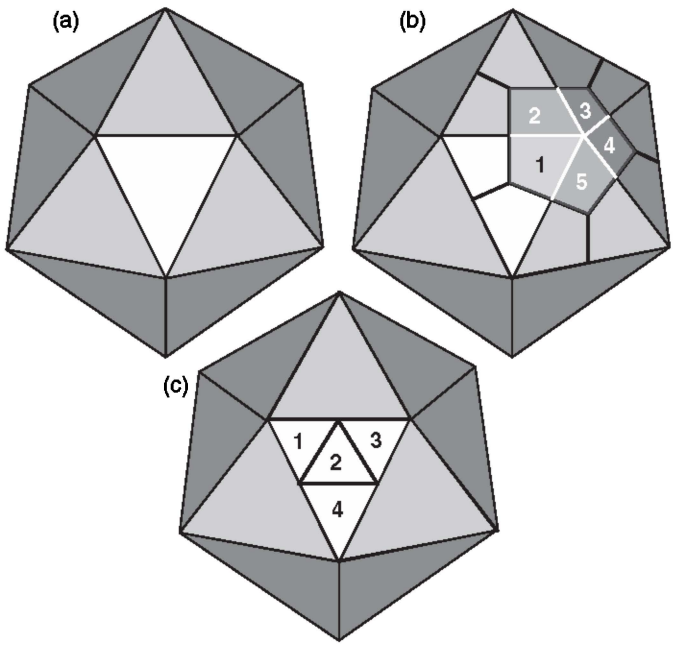


Figure 1.1 Icosahedron and virus capsid. (a) An icosahedron has 20 identical equilateral triangular faces. (b) In most icosahedral capsids of viruses, each triangular face is made up of three identical subunits. As a result, a typical viral capsid contains 60 subunits. The five subunits surrounding each vertex are arranged in a fivefold symmetry. (c) An icosahedral capsid of large viruses can consist of more than 60 subunits. Some of the triangular faces are made up of four subunits.

Therefore, there are typically 12 capsomers in the icosahedral virus. Many viruses have more than 60 subunits. In these viruses with large icosahedral capsids, the triangular faces are made up of four subunits (Fig. 1.1c).

Capsids of several viruses are neither purely helical nor purely icosahedral. These *complex* capsids may include extra structures, such as protein tails or complex outer walls. An illustrative example of such a complex virus is the well-known bacteriophage T4, which has an icosahedral head bound to a helical tail, which may have a hexagonal base plate with protruding protein tail fibers. This peculiar tail structure helps T4 to attach to the bacterial host and acts as a molecular syringe injecting the viral genome into the cell (Rossmann et al., 2004).

1.2.1.2 Viral Envelope In some viruses, the capsid is coated with a lipid membrane, known as the *viral envelope*, which is acquired by the capsid from an intracellular membrane of the virus host. Typically, in addition to the lipid membrane derived from a host, viral envelopes contain viral glycoproteins (e.g., hemagglutinin (HA), neuraminidase, and M2 protein, a proton-selective ion channel in influenza virus, or gp160 protein in human immunodeficiency virus (HIV), which consists of the structural subunit gp120, and the transmembrane subunit gp41). Some of these surface viral glycoproteins (HA, neuraminidase, and gp120) protrude from the viral lipid bilayer and play important roles in its attachment to and penetration into the target cells (Suzuki, 2005). Other viral envelope proteins are involved in various functions related to the virus life cycle. For example, a proton-selective ion channel M2 protein of influenza A virus enables hydrogen ions to enter the viral particle from the endosome, thus lowering pH of the inside of the virus. This decrease in pH triggers the dissociation of the viral matrix protein M1 from the ribonucleoprotein, therefore initiating the uncoating of the virus and exposing its content to the cytoplasm of the host cell (Cady et al., 2009).

1.2.1.3 Matrix In addition to membrane glycoproteins, enveloped viruses have matrix proteins, which link the viral envelope with the virus core. In general, viral matrix proteins are responsible for expelling the genetic material after a virus has entered a cell. However, they have several other biological functions. For example, in the influenza virus, one side of the matrix M1 protein possesses a specific affinity to the glycoproteins of the host cell membrane, whereas another side of this protein has nonspecific affinity for the viral RNA. As a result, a specific proteinaceous layer, or matrix, is formed under the membrane. The assembled complexes of viral ribonucleoprotein and viral RNA bind to the matrix and are enveloped and bud out of the cell as new mature viruses (Nayak et al., 2004, 2009). M1 protein also has multiple regulatory functions performed by interaction with the components of the host cell. These regulatory functions include a role in the export of the viral ribonucleoproteins from the host cell nucleus, inhibition of viral transcription, and a role in virus assembly and budding (Nayak et al., 2004, 2009).

1.2.2 Viral Nonstructural Proteins

Viral NS proteins are virus-encoded proteins that are not a part of the viral particle. Some of these proteins may play roles within the infected cell during virus replication, whereas others act in the regulation of virus replication or virus assembly. Specific functions of six NS HPV proteins were briefly introduced above. Below, three illustrative examples of the functions of viral NS proteins, namely, replicon formation, immunomodulation, and transactivation of genes encoding structural proteins are described.

1.2.2.1 Replicon Formation The hepatitis C virus's (HCV's) RNA replication complex formation requires interactions between the HCV NS proteins and a human cellular vesicle membrane transport protein hVAP-33 (Gao et al., 2004). The formation of this HCV replicon is initiated by the precursor of NS4B, which is able to anchor to the lipid raft membrane. Most of the other HCV NS proteins, including NS5A, NS5B, and NS3, are also localized to these lipid raft membranes, suggesting that protein–protein interactions among the various HCV NS proteins and hVAP-33 are important for the formation of HCV replication complex (Gao et al., 2004).

1.2.2.2 Immunomodulation The immunomodulatory function of West Nile virus NS protein NS1 was demonstrated by showing that the soluble and cell-surface-associated NS1 was able to bind to and recruit the complement regulatory protein factor H. This interaction led to decreased complement activation, minimizing immune system targeting of West Nile virus by decreasing complement recognition of infected cells (Chung et al., 2006). In rinderpest virus, the viral NS C protein was shown to block specifically the actions of type 1 and type 2 interferons, therefore suppressing the induction of the innate immune response (Boxer et al., 2009).

1.2.2.3 Transactivation of Genes Encoding Structural Proteins In the autonomous parvovirus minute virus of mice (MVM), whose genome contains two overlapping transcription units, the genes coding for the two NS proteins (NS1 and NS2) are transcribed from a promoter P04, whereas the promoter P39 controls the transcription of capsid protein genes. Intriguingly, the P39 promoter was shown to be activated by a viral NS protein NS1 (Doerig et al., 1988).

1.2.3 Viral Regulatory and Accessory Proteins

Viral regulatory and accessory proteins play a number of indirect roles in the viral function, for example, some of these proteins regulate the rate of transcription of viral structural genes. These proteins either regulate the expression of viral genes or are involved in modifying host cell functions. Many viral regulatory and accessory proteins serve multiple functions. For example, the active replication of HIV-1 is controlled by the production of several regulatory (Tat and Rev) and accessory

(Vpr, Vif, Vpu, and Nef) proteins (Seelamgari et al., 2004). Accessory proteins are important for the efficient *in vivo* infection. It is believed that Vif has evolved to overcome the antiviral defense mechanisms of the host, whereas accessory proteins such as Nef increase virus pathogenesis by targeting bystander cells. Therefore, these proteins control many aspects of the virus life cycle as well as host cell function, namely, gene regulation and apoptosis, mostly via interactions with other viral and cellular components (Seelamgari et al., 2004).

1.3 INTRINSIC DISORDER IN VIRAL PROTEINS

Most viral proteins (e.g., proteins involved in replication and morphogenesis of viruses and the major capsid proteins of icosahedral virions), being shared by many groups of RNA and DNA viruses, have no homologs in modern cells (Koonin et al., 2006). This clearly suggests that viruses are very antique and that viral genes primarily originated in the virosphere during replication of viral genomes and/or recruited from cellular lineages are now extinct (Forterre and Prangishvili, 2009). Viruses represent an interesting example of adaptation to extreme conditions, which include both environmental peculiarities and biological and genetic features of the hosts. Viruses have to survive outside and within the host cell (some viruses infect Archaea that are isolated from geothermally heated hot environments (Prangishvili et al., 2006)) and need to infect the host organism and replicate their genes while avoiding the host's countermeasures (Reanney, 1982). Genomes of many viruses are characterized by unusually high rates of mutation, which, being estimated as exchanges per nucleotide, per generation can be as high as 10^{-5}–10^{-3} for RNA viruses, 10^{-5} for single-stranded DNA viruses, and 10^{-8}–10^{-7} for double-stranded DNA viruses, compared to 10^{-10}–10^{-9} in bacteria and eukaryotes (Drake et al., 1998). Viral genomes are unusually compact and contain overlapping reading frames. Therefore, a single mutation might affect more than one viral protein (Reanney, 1982).

All these peculiarities raised an intriguing question on whether the viral proteins possess unique structural features. In an attempt to answer this question, a detailed analysis of viral proteins was undertaken (Tokuriki et al., 2009). First, 3D protein crystal structures of 123 representative single domain proteins of 70–250 amino acids that contain no covalent cofactors, and with a high resolution crystal structure, were analyzed. Of these 123 proteins, 26 were RNA viral proteins, 19 were DNA viral proteins (18 double-stranded and one single-stranded DNA virus), 26 were hypothermophilic, 26 were mesophilic eukaryotes, and 26 were mesophilic prokaryotes. The analysis revealed that viral proteins, especially RNA viral proteins, possessed systematically lower van der Waals contact densities than proteins from other groups. Furthermore, viral proteins were shown to have a larger fraction of residues that are not arranged in well-defined secondary structural elements such as helixes and strands. Finally, the effects of mutations on protein conformational stability ($\Delta\Delta G$ values) were compared for all these proteins. This analysis showed that viral proteins show lower average $\Delta\Delta G$ per residue

than proteins from other organisms. RNA viral proteins show particularly low average $\Delta\Delta G$ values, 0.20 kcal/mol lower than the mesophilic proteins of the same size and 0.26 kcal/mol lower than the thermophilic proteins (Tokuriki et al., 2009).

At the next stage, peculiarities of viral proteins were analyzed using approaches that are independent of structures, namely, amino acid composition profiling and disorder propensity calculations. These tools were applied to all available open reading frames (ORFs) in the relevant proteomes of 19 hyperthermophilic archaea, 35 mesophilic bacteria, 20 eukaryotes, and 30 single-stranded RNA, 30 single-stranded DNA, and 29 double-stranded DNA viruses (Tokuriki et al., 2009). In these analyses, viral proteomes were filtered to remove all annotated capsid/coat/envelope/structural proteins. Figure 1.2a represents the relative composition profiles calculated for various species as described by Vacic and colleagues (Vacic et al., 2007). Here, the fractional difference in composition between a given protein set and a set of completely ordered proteins was calculated for each amino acid residue. The fractional difference was evaluated as $(C_X - C_{order}/C_{order}$, where C_X is the content of a given amino acid in a given protein set and C_{order} is the corresponding content in the fully ordered data set (Xue et al., 2009a,b). In addition to the filtered data set of viral proteins from 89 proteomes, this figure also includes compositional profile calculated for the nonfiltered data set containing all viral proteins from ~2400 viral species. In general, viral proteins show a reduced fraction of hydrophobic and charged residues and a significantly increased proportion of polar resides. Figure 1.2b clearly shows that viral proteomes exhibit a very high propensity for an intrinsic disorder. In general, the amount of disorder in viruses was comparable with that in eukaryotes, which from previous studies were already known to possess the highest levels of disorder (Romero et al., 1998; Dunker et al., 2000, 2001; Ward et al., 2004; Oldfield et al., 2005a). Figure 1.2b illustrates that there was a fundamental difference between viral and eukaryotic proteomes since eukaryotes contained more proteins with long disordered regions, whereas viral proteomes were characterized by the dominance of short disordered segments (Tokuriki et al., 2009).

On the basis of these observations it has been concluded that in comparison with proteins from their hosts, viral proteins are less densely packed, possess a much weaker network of interresidue interactions (manifested by the lower contact density parameters, the increased fraction of residues not involved in secondary structure elements, and the abundance of short disordered regions), have an unusually high occurrence of polar residues, and are characterized by the lower destabilizing effects of mutations (Tokuriki et al., 2009). It has been concluded that the adaptive forces that shape viral proteins were different form those responsible for evolution of proteins of their hosts. In fact, as discussed, the abundance of polar residues, the lower van der Waals contact densities, high resistance to mutations, and the relatively high occurrence of flexible "coils" and numerous short disordered regions suggested that viral proteins are not likely to have evolved for higher thermodynamic stability but rather to be more adaptive for fast change in their biological and physical environments (Tokuriki et al., 2009).

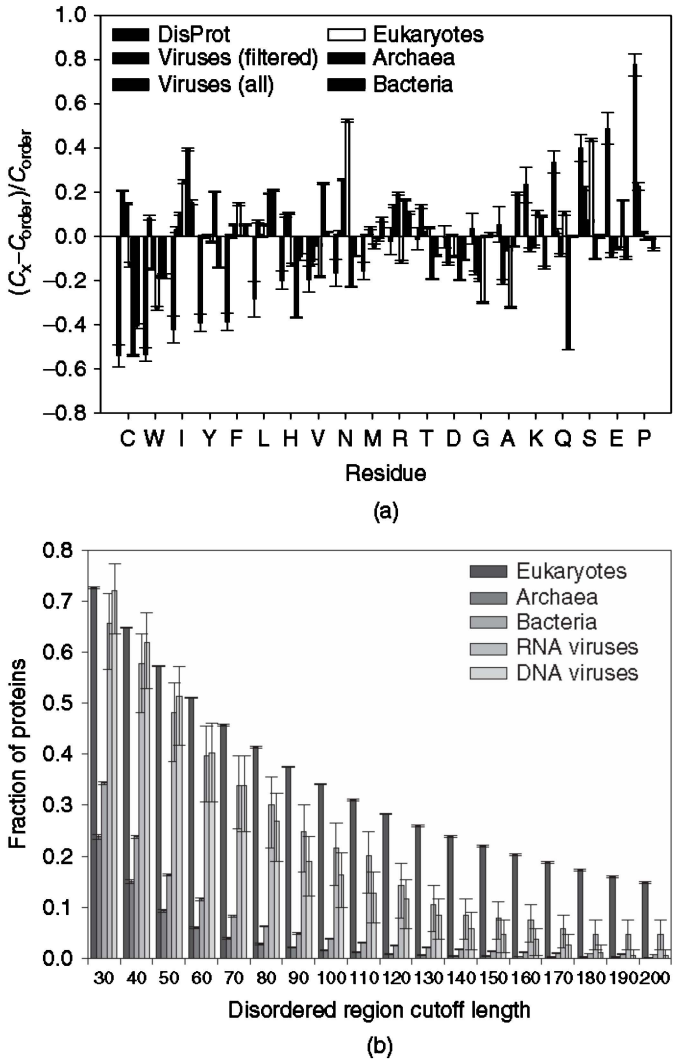


Figure 1.2 Evaluation of the uniqueness and abundance of intrinsic disorder in viral proteins. (a) Composition profile of amino acids for proteins from different organisms. Residues on the x-axis are arranged according to the increasing disorder tendency. The y-axis represents the relative compositional profile compared to a fully disordered data set. (b) The fraction of disordered regions within viral, mesophilic eukaryotic and prokaryotic, and thermophilic proteomes. Shown are the fractions of continuous disordered segments predicted for these proteomes for stretches of varying lengths, from $\geq$30 to $\geq$200 amino acids. (*See insert for color representation of the figure.*)

1.4 FUNCTIONALITY OF INTRINSIC DISORDER IN VIRAL PROTEINS

1.4.1 Intrinsic Disorder and Viral Pfam Domain Seeds

Proteins often contain one or more functional domains, different combinations of which give rise to the diverse range of proteins found in Nature. It has been recognized that the identification of domains that occur within proteins can therefore provide insights into their function. To find a correlation between intrinsic disorder and function in the viral proteins, we analyzed the abundance of intrinsic disorder in the Pfam database, which contains information on protein domains and families and uses hidden Markov models (HMMs) and multiple sequence alignments to identify members of its families emphasizing the evolutionary conservation of protein domains (Bateman et al., 2002, 2004; Finn et al., 2008). Each curated family in Pfam is represented by a seed and full alignment. The seed contains representative members of the family, while the full alignment contains all members of the family as detected with a profile HMM (Bateman et al., 2002). Since Pfam represents an important tool for understanding protein structure and function and since this database contains large amount of information on functional domains, the viral seed domains in the version 23.0 of the Pfam database were analyzed. There are 6360 Pfam domain seeds of viral origin. Figure 1.3 shows that intrinsic disorder is rather abundant among the viral Pfam seed domains. In fact, 535 Pfam domain seeds of viral origin were 50–98% disordered, and the length of disordered regions in the domains varied from 11 to 738 residues (Fig. 1.3a). Figure 1.3b shows that >100 domains ranging in length from 14 to 324 residues were almost completely disordered. Our analysis revealed that many Pfam domain

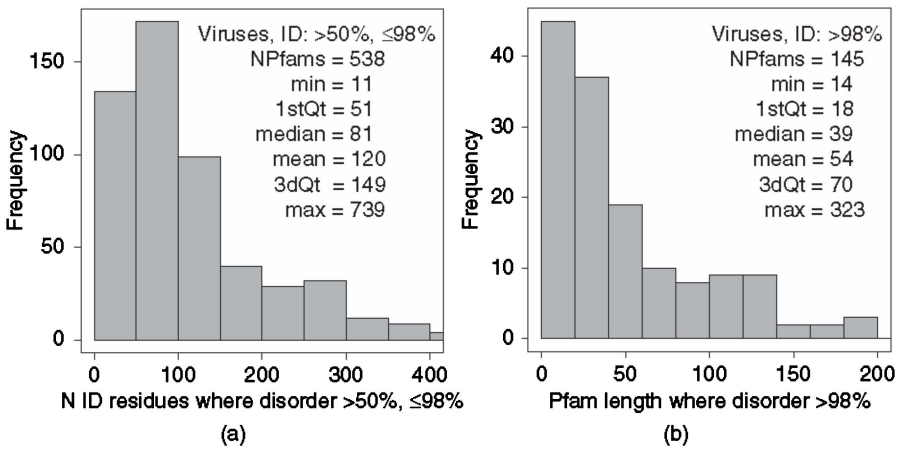


Figure 1.3 Intrinsic disorder distribution in Pfam domain seeds of the viral origin. (a) The length distribution of Pfam domain seeds of viral origin which are 50–98% disordered. (b) The length distribution of Pfam domains where disorder is observed for >98% residues.

seeds of the viral origin were completely disordered (they contain >98% disordered residues) but still possessed a number of crucial biological functions, mostly related to interaction with proteins, as well as recognition, regulation, and signal transduction (Xue et al., 2010). In other words, viral disordered domains possess functions similar to those of prokaryotic, archaeal, and eukaryotic proteins (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001, 2002a,b).

1.4.2 Intrinsic Disorder in Viral Structural Proteins

1.4.2.1 Capsids Capsids represent an economical use of multiple copies of a single or a few proteins to build a specific cage for genome transfer. In fact, this approach helps viruses to minimize the coding space for the capsid and also determines an easy and self-controlled mechanism of shell assembly, where only the fitting pieces can work. Since an icosahedral symmetry provides a low energy solution for the shell formation, it is commonly used by many isomeric (or icosahedral) viruses (Caspar and Klug, 1962). Sixty identical units can form an icosahedron. Although the majority of capsid proteins are relatively small, many viruses have very large capsids. These large capsids are built from a high number of building blocks, many times exceeding 60 units. The theory of quasi-equivalence, according to which the capsid is stabilized by the same type of interactions that are perturbed in slightly different ways in the non-symmetry-related environments, explained this apparent contradiction since multiples of 60 proteins can be arranged such that they will all be in nearly identical environments (Caspar and Klug, 1962).

Intriguingly, already in the first virus structures determined by X-ray crystallography (Harrison et al., 1978; Abad-Zapatero et al., 1980), the coat proteins appeared as globular parts (C-terminal domains) formed by two antiparallel four-stranded sheets with a jelly-roll or Swiss-roll topology and extended, partially invisible N-terminal segments (Liljas, 2004). In the polyoma virus and simian virus 40 (SV40), whose capsids are described by an icosahedral surface lattice with the triangulation number $T = 7d$, there are 360 units in the capsid (Baker et al., 1983; Liddington et al., 1991), which is noticeably less than the 420 units expected from the Caspar–Klug rules (Caspar and Klug, 1962). In these viruses, all 72 capsomers are pentamers of the structural protein VP1 (in polyoma virus) or of the coat protein (in SV40) rather than an expected mixture of pentamers and hexamers. Therefore, pentamers are found at the positions predicted to have a hexamer of subunits according to the Caspar–Klug hypothesis (Caspar and Klug, 1962). The apparent contradiction is resolved by intrinsically disordered arms of the capsid proteins: the intercapsomer contacts are established by the folded C-terminal domain, whereas the N-terminal domain of the capsid protein, the so-called arm, is extended and is present in six totally different conformations depending on its position in the lattice (Rayment et al., 1982; Liljas, 2004).

In agreement with the icosahedral symmetry, in the capsid of foot-and-mouth-disease virus there are 60 identical subunits, each of which is made up of four proteins: VP1, VP2, VP3, and VP4 (Fry et al., 2005). VP1, VP2, and VP3 are wedge-shaped, eight-stranded β-sandwiches. The loops connecting strands at the

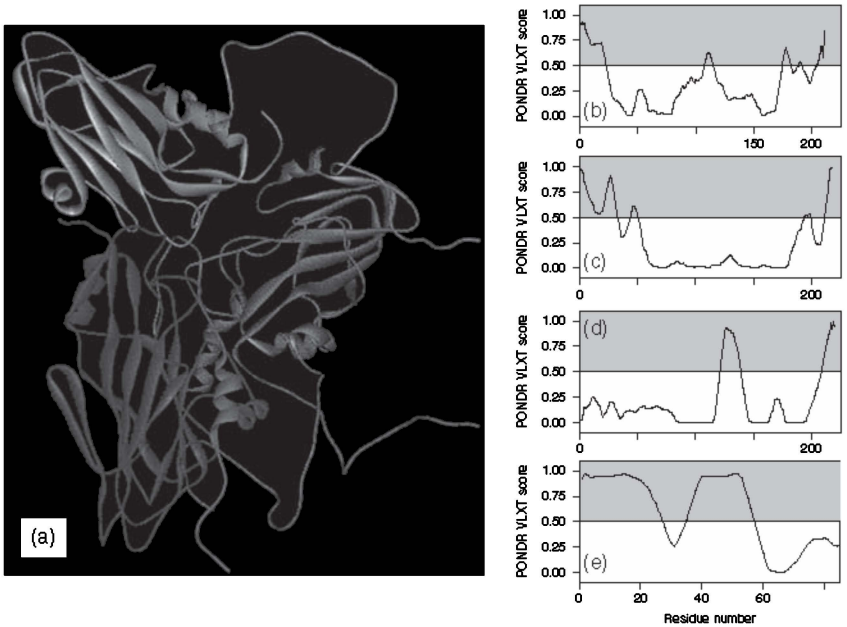


Figure 1.4 Intrinsic disorder in viral structural proteins. (a) Structure of the capsomer in the icosahedral capsid of the foot-and-mouth-disease virus. This capsid contains 60 capsomers, each of which is made up of four capsid proteins: VP1, VP2, VP3, and VP4. Distribution of predicted intrinsic disorder in capsid proteins of the foot-and-mouth-disease virus: VP1 (b), VP2 (c), VP3 (d), and VP4 (e).

narrow end of the wedge are less constrained by structural interactions and tend to mediate host interactions. VP4 and the N-termini of VP1 and VP3 are located at the capsid interior. It has been pointed out that the capsomer structure comprises residues 1–137 and 155–208 of VP1, 12–218 of VP2, 1–221 of VP3, and 15–39 and 62–85 of VP4, whereas residues 138–154 and 209–212 of VP1, 1–11 of VP2, and 1–14 and 40–61 of VP4 were too flexible to be modeled reliably (Fry et al., 2005). Therefore, the three major capsid proteins are mostly ordered and have a conserved iβ-barrel fold, whereas VP4 has a little regular secondary structure. This mostly disordered protein, VP4, is involved in the initial disassembly and final assembly stages of the virus.

Functional roles of disordered regions of capsid proteins extend far away from simple structural roles. In the excellent review by Liljas (2004), the functionality of various disordered arms of viral capsid proteins is systemized to show that these intrinsically disordered fragments can be used to stabilize the structure of a capsid, to control the capsid assembly and disassembly, and for the interaction with nucleic acids. To finish this part, Fig. 1.4 represents the results of disorder prediction for the capsid proteins discussed above together with the crystal structure of the foot-and-mouth-disease virus capsomer.

1.4.2.2 Viral Envelope Specific surface glycoproteins are used by the enveloped viruses, such as influenza, HIV-1, and Ebola, to enter target cells via fusion of the viral membrane with the target cellular membrane (Skehel and Wiley, 1998, 2000; Eckert and Kim, 2001). One of the most well-studied membrane fusion proteins is the influenza virus HA, which is a homotrimeric type I transmembrane surface glycoprotein responsible for virus binding to the host receptor, internalization of the virus, and subsequent membrane-fusion events within the endosomal pathway in the infected cell. HA is also the most abundant antigen on the viral surface and harbors the primary neutralizing epitopes for antibodies. Each 70-kDa HA subunit contains two disulfide-linked polypeptide chains, HA_1 and HA_2, created by proteolytic cleavage of the precursor protein HA_0 (Wiley and Skehel, 1987). Such a cleavage is absolutely crucial for membrane fusion (Wiley and Skehel, 1987). During membrane fusion, HA binds the virus to sialic acid receptors on the host cell surface, and following endocytosis, the acidic pH (pH 5–6) of endosomal compartments induces dramatic and irreversible reorganization of the HA structure (Skehel et al., 1982).

The HA trimer has a tightly intertwined "stem" domain at its membrane-proximal base, which is composed of HA_1 residues 11–51 and 276–329 and HA_2 residues 1–176. The dominant feature of this stalk region in the HA trimer is the three long, parallel α-helices (~50 amino acids in length each), one from each monomer, that associate to form a triple-stranded coiled coil. The membrane-distal domain consists of a globular "head," which is formed by HA_1 and which can be further subdivided into the R region (residues 108–261), containing the receptor-binding site and major epitopes for neutralizing antibodies, and the E region (residues 56–108 and 262–274), with close structural homology to the esterase domain of influenza C HA esterase fusion (HEF) protein (Stevens et al., 2004). The HA_2 chain contains two membrane-interacting hydrophobic peptide sequences: an N-terminal "fusion peptide" (residues 1–23), which interacts with the target membrane bilayer (Durrer et al., 1996), and a C-terminal transmembrane segment, which passes through the viral membrane.

Crystallographic studies suggested that the interaction with the host cell involves a dramatic structural reorganization of HA_2, which moves the fusion peptide from the interior approximately 100 AA toward the target membrane (Wilson et al., 1981; Bullough et al., 1994). In this process, the middle of the original long α-helix unfolds to form a reverse turn, jackknifing the C-terminal half of the long α-helix backward toward the N-terminus. These molecular rearrangements place the N-terminal fusion peptide and the C-terminal transmembrane anchor at the same end of the rod-shaped HA_2 molecule (Weber et al., 1994; Wharton et al., 1995), facilitating membrane fusion by bringing the viral and cellular membranes together.

Our recent analysis revealed that although many viral membrane glycoproteins are ordered, intrinsic disorder still is crucial for the biology of these proteins. For example, we have found some distinct differences in the disorder propensity between HA proteins of the virulent and nonvirulent strains of influenza A, especially in the region near residues 68–79 of the HA_2. This region represents the tip of the stalk that is in contact with the receptor chain, HA_1, and is therefore

likely to provide the greatest effect on the motions of the exposed portion of HA. Comparison of this region between virulent strains (1918 H1N1 and H5N1) and less virulent ones (H3N2 and 1930 H1N1) showed that this region is characterized by the increased level of intrinsic disorder in more virulent strains and subtypes of the virus but is predicted to be mostly ordered in less virulent strains (Goh et al., 2009).

1.4.2.3 Matrix We also analyzed the predisposition of several viral matrix proteins to intrinsic disorder (Goh et al., 2008a,b). These studies revealed that the matrix protein p17 from simian immunodeficiency virus (SIV_{mac}) and HIV-1 possesses high levels of predicted intrinsic disorder, whereas matrix proteins of the equine infectious anemia virus (EIAV) were characterized by noticeably lower levels of predicted disorder (Goh et al., 2008a).

1.4.3 Intrinsic Disorder in Viral Nonstructural, Regulatory, and Accessory Proteins

Since these proteins are responsible for the wide range of recognition- and regulation-based functions, including communication with the hosts and regulation of virus replication and assembly, they are frequently disordered. As illustrative examples, we are presenting below a brief overview of intrinsic disorder in several NS proteins from various viruses and the regulatory and accessory proteins from HIV-1.

1.4.3.1 Disorder in Viral Nonstructural Proteins As discussed earlier, an NS oncoprotein E7 of HPV is involved in regulation of cell growth and transformation, gene transcription, apoptosis, and DNA synthesis. It is known to interact with a number of cellular proteins, such as the Rb, p107 and p130, glycolytic enzymes, histone deacetylase, kinase p33CDK2, and cyclin A, and the cyclin-dependent kinase inhibitor $p21^{cip1}$ tumor suppressor. Importantly, E7 is involved in the pathogenesis and maintenance of human cervical cancers. The analysis of the E7 dimer from HPV45 by NMR revealed that each monomer contained an unfolded N-terminus and a well-structured C-terminal domain (Ohlenschlager et al., 2006). Later, a fragment of the oncoprotein E7 comprising the highly acidic N-terminal domain was confirmed to be intrinsically disordered by far-UV CD (circular dichroism) hydrodynamic analyses. Importantly, the N-terminal domain of this protein (residues 1–40) includes the retinoblastoma tumor suppressor binding and casein kinase II phosphorylation sites (Garcia-Alai et al., 2007).

There are more than 100 different types of HPVs, which are the causative agents of benign papillomas/warts and the cofactors in the development of carcinomas of the genital tract, head and neck, and epidermis. In respect to their association with cancer, HPVs are grouped into two classes, low risk (e.g., HPV-6 and HPV-11) and high risk (e.g., HPV-16 and HPV-18) types. Recently, in order to understand whether intrinsic disorder plays a role in the oncogenic potential of different HPV types, the bionformatics analysis of proteomes of high risk and low risk HPVs

with a major focus on E6 and E7 oncoproteins has been performed (Uversky et al., 2006). On the basis of the results of this analysis, it has been concluded that high risk HPVs are characterized by an increased amount of intrinsic disorder in transforming proteins E6 and E7 (Uversky et al., 2006).

Influenza virus NS protein 2 (NS2, or NEP) is known to interact with the nuclear export machinery during viral replication and serves as an adapter molecule between the nuclear export machinery and the viral ribonucleoprotein complex. Structural analysis of the recombinant NS2 by spectroscopy, differential scanning calorimetry, limited proteolysis, and hydrodynamic techniques revealed that this monomeric protein shows characteristics of the native molten globule under near physiological conditions being compact and highly flexible (Lommer and Luo, 2002).

1.4.3.2 Disorderedness of Viral Regulatory and Accessory Proteins Protein Tat is the HIV-1 transactivator of viral transcription and is an important factor in viral pathogenesis. Tat binds to a short nascent stem-bulge loop leader RNA, termed the *transactivation responsive region* (TAR), that is present at the 5′ extremity of all viral transcripts via its basic region and recruits the complex of cyclin T1 and cyclin-dependent kinase 9 (CDK9) forming the positive transcription elongation factor B complex. CDK9 hyperphosphorylates the carboxy terminus domain of RNA polymerase II, leading to the enhanced elongation of transcription from the viral promoter. However, Tat not only acts as the key transactivator of viral transcription but is also secreted by the infected cell and is taken up by neighboring cells where it has an effect both on infected and uninfected cells (Campbell and Loret, 2009).

The Tat amino acid sequence is characterized by a low overall hydrophobicity and a high net positive charge. This protein was predicted to be natively unfolded by several algorithms (Shojania and O'Neil, 2006). These predictions were in agreement with the lack of ordered secondary structure in this protein found by the CD analysis (Vendel and Lumb, 2003), and NMR chemical shifts and coupling constants suggested that Tat existed in a random coil conformation (Shojania and O'Neil, 2006).

Rev is another regulatory protein in HIV-1. This is a 116-residue basic protein that binds to multiple sites in the Rev-response element (RRE) of viral mRNA transcripts in nuclei of host cells, leading to transport of incompletely spliced and unspliced viral mRNA to the cytoplasm of host cells in the later phases of the HIV-1 life cycle. Therefore, Rev is absolutely required for viral replication (Blanco et al., 2001). On the basis of the detailed spectroscopic and hydrodynamic studies, it has been concluded that monomeric Rev is in a molten globule state (Surendran et al., 2004).

Vpr is a 96-residue HIV-1 accessory protein that shows multiple activities, including nuclear transport of the preintegration complex to the nucleus, activation of transcription, cell cycle arrest at the G2/M transition, and triggering of apoptosis. This protein controls many host cell functions through a variety of biological activities and by interaction with cellular biochemical pathways. For example, nuclear

import of Vpr may be due to its interaction with nuclear transport factors and components of the nuclear pore complex. Cell cycle arrest has been correlated with the binding to DCAF1, a cullin-4A-associated factor, and apoptosis may be facilitated by interaction with mitochondrial proteins in a caspase-dependent mechanism. Vpr also plays a critical role in long-term AIDS by inducing viral infection in nondividing cells such as monocytes and macrophages (Morellet et al., 2009). On the basis of the dynamic light scattering (DLS), CD, and ^{1}H NMR spectroscopy analyses, it has been concluded that Vpr was unstructured at neutral pH, whereas under acidic conditions or on addition of trifluorethanol it adopts α-helical structures (Henklein et al., 2000). On the basis of this pH-dependent folding switch, it has been suggested that the Vpr structure is dependent on the presence of specific binding factors (such as nucleic acids, proteins, or membrane components) or the environment of the cytosol, nucleus, mitochondrion, cellular membranes, and the extracellular space (Bruns et al., 2003).

Vif is another HIV-1 accessory protein that neutralizes the cellular defense mechanism against the virus. Many of the interactions of Vif are mediated via its C-terminal domain (residues 141–192). Detailed structural analysis has revealed that this fragment is mostly disordered, a conclusion based on the coil-like far-UV CD spectrum with some residual helical structure, unfolded features of the ^{15}N-HSQC NMR spectrum, and the extended size evaluated by size-exclusion chromatography. These findings have been further supported by the results of the computational analyses of the Vif C-terminal domain sequence. Importantly, CD analysis has revealed that this domain is able to fold upon interaction with membrane micelles, clearly showing that this natively unfolded domain may gain structure on binding its natural ligands (Reingewertz et al., 2009).

Vpu is an oligomeric type I integral membrane phosphoprotein that amplifies the release of virus particles from infected cells by mediation of degradation of the HIV receptor CD4 by the proteasome in the endoplasmic reticulum. Phosphorylation of Vpu at two sites, Ser52 and Ser56, on the motif DSGXXS is required for the interaction of Vpu with the ubiquitin ligase SCF-βTrCP, which triggers CD4 degradation by the proteasome. Vpu consists of a hydrophobic N-terminal membrane-anchoring domain and a polar C-terminal cytoplasmic domain (Gramberg et al., 2009). CD and NMR analyses of nine overlapping 15 amino acid fragments and 3 longer fragments in aqueous solutions have revealed that the C-terminal hydrophilic domain of Vpu is mostly disordered with some limited amounts of stable secondary structure. However, in the presence of trifluoroethanol, this domain protein is shown to fold into a helical conformation composed of two α-helices joined by a flexible region of six or seven residues, which contains the phosphorylation sites of Vpu at positions 52 and 55 (Wray et al., 1995).

Nef is an HIV-1 accessory protein that is known to interact with multiple cellular partners during the course of infection. The interactions of this viral protein with various cellular partners are mediated by the occurrence of ligand-induced conformational changes that direct the binding of Nef to subsequent partners. On the basis of the analysis of the available experimental data, it has been hypothesized that the binding-promoted conformational changes underwent by this protein

define a novel allosteric paradigm, namely, changes that involve conformations with large disordered regions (Leavitt et al., 2004). Importantly, these regions, being devoid of stable secondary or tertiary structure, contain the binding determinants for subsequent partners and only become functionally competent by ligand-induced folding and unfolding. This model of switching binding epitopes between structured and unstructured conformations provides a unique ability to modulate the binding affinity by several orders of magnitude (Leavitt et al., 2004).

1.5 INTRINSIC DISORDER, ALTERNATIVE SPLICING, AND OVERLAPPING READING FRAMES IN VIRAL GENOMES

Viruses have evolved a complex genetic organization for optimal use of their limited genomes and production of all necessary structural and regulatory proteins. The use of alternative splicing is essential for balanced expression of multiple viral regulators from one genomic polycistronic RNA. Furthermore, viruses use both sense and antisense transcriptions. For example, the genome of human T-cell lymphotropic virus type 1 (HTLV-1), which is a causative agent of adult T-cell leukemia (ATL), HTLV-1-associated myelopathy, and *Strongyloides stercoralis* hyperinfection, encodes common structural and enzymatic proteins found in many retroviruses (Gag, Pro, Pol, and Env). *Gag, Pro*, and *Pol* genes are translated as a series of polyproteins, Gag, Gag-Pro, and Gag-Pro-Pol, which are then cleaved posttranslationally to generate seven proteins. *Gag* gene encodes a polyprotein (Gag) whose cleavage products are the major structural proteins (matrix (MA), capsid (CA), and nucleocapsid (NC)) of the virus core. *Pro* encodes a middle part of a polyprotein (Gag-Pro or Gag-Pro-Pol) whose cleavage products include protease (PR). Finally, *Pol* encodes the last part of a polyprotein (Gag-Pro-Pol) whose cleavage products include reverse transcriptase (RT) and integrase (IN). *Env* encodes a polyprotein (Env) whose cleavage products SU (surface) and TM (transmembrane) are the structural proteins of the viral envelope. In addition to these common retroviral proteins, HTLV-1 encodes multiple regulatory and accessory proteins in four overlapping ORFs located in the pX region of the viral genome (Ciminale et al., 1992; Koralnik et al., 1992). The HTLV-1 basic leucine zipper factor (HBZ) is the product of the antisense transcription (Larocca et al., 1989). There are two transcripts that encode the *HBZ* gene, spliced (*sHBZ*) and unspliced (*usHBZ*). *sHBZ* has multiple transcriptional initiation sites in the U5 and R regions of the 3′ long terminal repeat (LTR), whereas the *usHBZ* gene initiates within the *tax* gene (Matsuoka and Green, 2009). Furthermore, expression of the various ORFs is controlled by differential splicing of the single genomic mRNA, producing unspliced, singly spliced, and multiply spliced mRNA (Kashanchi and Brady, 2005). Therefore, the replication of HTLV-1 is controlled by a group of nuclear and cytoplasmic processes, including transcription, splicing, alternative splicing, mRNA nuclear export, RNA stability, and translation (Baydoun et al., 2008).

Let us consider in more detail Tax and Rex, two HTLV-1 regulatory proteins needed for the viral genome expression. Tax is a transcriptional activator of the

viral promoter (Sodroski et al., 1984; Cann et al., 1985; Felber et al., 1985; Seiki et al., 1986; Derse, 1987; Boxus et al., 2008). Rex affects posttranscriptional regulatory steps by promoting transport of the unspliced and singly spliced mRNA from the nucleus to the cytoplasm and by promoting expression of the Gag, Pol, and Env proteins (Kiyokawa et al., 1985; Inoue et al., 1986, 1987; Derse, 1988; Seiki et al., 1988). Tax and Rex were shown to be expressed from two overlapping ORFs located in the distal part of the pX region of the virus by a bicistronic viral mRNA consisting of three exons (Seiki et al., 1983; Nagashima et al., 1986; Ciminale et al., 1992). There are two alternatively spliced isoforms of Rex, p27Rex and p21Rex. In p21Rex, residues 1–78 are missing. Furthermore, alternative splicing of the pX region in the ORF II generates two accessory proteins, p30 and p13.

Figure. 1.5 represents the HTLV-1 proteome map where each major product described above is present as a bar whose location corresponds to the location of the corresponding gene within the HTLV-1 genome. PONDR® VLXT disorder predictions for each (poly)protein are shown as solid lines inside the corresponding bars. A residue is considered to be disordered if its score is above 0.5. The top half of each bar shaded in gray corresponds to disorder scores >0.5. Therefore, inside each bar, pieces of the PONDR plots located in these shaded area correspond to protein fragments predicted to be disordered. Cleavage sites producing Gag, Pro, and Pol polyproteins are indicated by angled arrows and lettered. Cleavage sites, which are responsible for the posttranslational production of MA, CA, NC, RT, IN, SU, and TM proteins, are marked by short straight arrows and numbered. Gag, Pro, Pol, Env, p12, Tax, p27Rex, p21Rex, p30, and p13 are all the products of the genes produced by the sense transcription. Proteins usHBZ and sHBZ are produced from genes generated by antisense transcription. In Fig. 1.5, this fact is indicated by a long bold arrow marked with letters N and C to indicate the location of the beginnings and ends of the corresponding proteins, respectively. Obviously, the numbering of residues for the usHBZ and sHBZ presentation was inverted. p27Rex, p30, and sHBZ proteins are translated from the spliced genes. There are three alternatively spliced pairs of proteins in HTLV-1: p27Rex and p21Rex, p30 and p13, and usHBZ and sHBZ.

Analysis of this figure clearly shows that the economic usage of genetic material by HTLV-1 is translated into very important implementations of intrinsic disorder for the corresponding proteins.

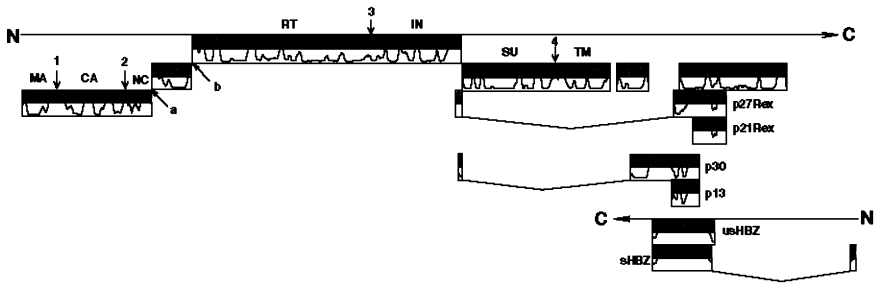


Figure 1.5 The proteome map of HTLV-1. See explanation in the text.

1. Prevailing intrinsic disorder is observed in posttranslational cleavage sites producing polyproteins Gag, Pro, and Pol from the grand-polyproteins Gag–Pro and Gar-Pro-Pol, as well as in the cleavage site producing MA, CA, NC, RT, IN, SU, and TM proteins from the corresponding polyproteins.
2. Proteins affected by alternative splicing are in general highly disordered. Furthermore, protein fragments removed by alternative splicing are mostly disordered.
3. Protein fragments corresponding to the overlapping genes are either disordered or possess complementary disorder distribution if these protein regions are not translated from the genes transcribed from the identical ORFs. For example, the N-terminal fragment of Tax, which overlaps with a significant portion of Rex, is mostly ordered, whereas the corresponding region in the Rex proteins is predominantly disordered. Similarly, the C-terminal region of p30, which overlaps with the ordered N-terminal fragment of Tax, is mostly disordered, as disordered is the p13 protein, which completely overlaps with the ordered N-terminal fragment of Tax.
4. Proteins translated from genes generated by antisense transcription are highly disordered.

The conclusion that the HTLV-1 proteins produced by the overlapping genes are intrinsically disordered in agreement with a recent study where the protein coded by overlapping genes from 43 genera of unspliced RNA viruses infecting eukaryotes has been analyzed (Rancurel et al., 2009). This study has revealed that overlapping proteins have a sequence composition globally biased toward disorder-promoting amino acids and are predicted to contain significantly more structural disorder than nonoverlapping proteins (Rancurel et al., 2009).

Importantly, many of the specific implementations of intrinsic disorder listed above are not unique to the viral proteins. In fact, sites of proteolytic cleavage of proteins from other organisms are frequently located in the disordered regions (Fontana et al., 1986, 1997a,b, 2004; Polverino de Laureto et al., 1995; Iakoucheva et al., 2001; de Laureto et al., 2006). It has been also shown that regions of mRNA that undergo alternative splicing code for disordered proteins much more often than they code for structured proteins (Romero et al., 2006). Finally, the so-called retro-proteins, that is, proteins whose sequence is read backward providing a new polypeptide that does not align with its parent sequence, were shown to lack an ordered 3D structure (Lacroix et al., 1998).

1.6 CONCLUDING REMARKS

The modern literature on protein intrinsic disorder in viral proteomes has been systematically analyzed. Published data clearly show that viral proteins are both different and similar to proteins from their hosts. On one hand, viral proteins are less densely packed, possess a much weaker network of interresidue interactions (manifested by the lower contact density parameters, the increased fraction of

residues not involved in secondary structure elements, and the abundance of short disordered regions), have an unusually high occurrence of polar residues, and are characterized by the lower destabilizing effects of mutations. On the basis of these peculiar features, it has been concluded that viral proteins are not likely to have evolved for higher thermodynamic stability but rather to be more adaptive for fast change in their biological and physical environments. On the other hand, recent studies clearly show that intrinsic disorder is widespread in viral proteomes and has a number of important functional implementations. In fact, almost all viral proteins, irrespective of their functions, have biologically important disordered regions. The list of functions attributed to these disorder regions of viral proteins overlaps with disorder-based activities of proteins from other organisms. In fact, many functional Pfam seed domains of the viral origin were shown to possess various levels of intrinsic disorder, with ~150 such seeds being completely disordered. Disordered Pfam domains were involved in various crucial functions, such as signaling, regulation, and interaction with nucleic acids and proteins, suggesting that similar to proteins from all domains of life, intrinsic disorder is heavily used by viral proteins in their functions. Therefore, although viral proteins possess a number of unique features, they still rely intensively on intrinsic disorder at almost all stages of their intriguing life cycle.

ACKNOWLEDGMENT

This work was supported in part by the grant EF 0849803 (to A. K. D and V. N. U.) from the National Science Foundation and the Program of the Russian Academy of Sciences for the "Molecular and Cellular Biology" (to V. N. U.). We gratefully acknowledge the support of the IUPUI Signature Centers Initiative. This chapter is based on a previous publication (Xue et al., 2010).

ABBREVIATIONS

AFV1	*Acidianus* filamentous virus 1
APMV	*Acanthamoeba polyphaga* mimivirus
ATL	adult T-cell leukemia
CA	capsid
CD	circular dichroism
CDK9	cyclin-dependent kinase 9
DLS	dynamic light scattering
EIAV	equine infectious anemia virus
HA	hemagglutinin
HBZ	basic leucine zipper factor
HCV	hepatitis C virus
HEF	HA esterase fusion
HIV-1	human immunodeficiency virus-1

HMM	hidden Markov model
HPV	human papilloma virus
HTLV-1	human T-cell lymphotropic virus type 1
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IN	integrase
MA	matrix
MVM	minute virus of mice
NC	nucleocapsid
NS	nonstructural
ORF	open reading frame
PKB	protein kinase P
PONDR	predictor of natural disordered regions
PP2A	protein phosphatase 2A
RRE	Rev-response element
RT	reverse transcriptase
SIFV	*Sulfolobus islandicus* filamentous virus
SIV_{mac}	simian immunodeficiency virus
SU	surface
TM	transmembrane
TMV	tobacco mosaic virus

REFERENCES

Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D, Tsukihara T. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 1980;286:33–39.

Baker TS, Caspar DL, Murakami WT. Polyoma virus hexamer tubes consist of paired pentamers. Nature 1983;303:446–448.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D141.

Baydoun HH, Bellon M, Nicot C. HTLV-1 Yin and Yang: Rex and p30 master regulators of viral mRNA trafficking. AIDS Rev 2008;10:195–204.

Bergh O, Borsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature 1989;340:467–468.

Blanco FJ, Hess S, Pannell LK, Rizzo NW, Tycko R. Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. J Mol Biol 2001;313:845–859.

Boxer EL, Nanda SK, Baron MD. The rinderpest virus non-structural C protein blocks the induction of type 1 interferon. Virology 2009;385:134–142.

Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L. The HTLV-1 Tax interactome. Retrovirology 2008;5:76.

Brazil DP, Hemmings BA. Ten years of protein kinase B signalling: a hard Akt to follow. Trends Biochem Sci 2001;26:657–664.

Brehm A, Nielsen SJ, Miska EA, McCance DJ, Reid JL, Bannister AJ, Kouzarides T. The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. EMBO J 1999;18:2449–2458.

Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 2005;13:278–284.

Brown DR, Kitchin D, Qadadri B, Neptune N, Batteiger T, Ermel A. The human papillomavirus type 11 E1–E4 protein is a transglutaminase 3 substrate and induces abnormalities of the cornified cell envelope. Virology 2006;345:290–298.

Bruns K, Fossen T, Wray V, Henklein P, Tessmer U, Schubert U. Structural characterization of the HIV-1 Vpr N terminus: evidence of cis/trans-proline isomerism. J Biol Chem 2003;278:43188–43201.

Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature 1994;371:37–43.

Cady SD, Luo W, Hu F, Hong M. Structure and function of the influenza A M2 proton channel. Biochemistry 2009;48:7356–7364.

Campbell GR, Loret EP. What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? Retrovirology 2009;6:50.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. Phage as agents of lateral gene transfer. Curr Opin Microbiol 2003;6:417–424.

Cann AJ, Rosenblatt JD, Wachsman W, Shah NP, Chen IS. Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. Nature 1985;318:571–574.

Caspar DL, Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 1962;27:1–24.

Chen JJ, Reid CE, Band V, Androphy EJ. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science 1995;269:529–531.

Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 2006;45:10448–10460.

Chung KM, Liszewski MK, Nybakken G, Davis AE, Townsend RR, Fremont DH, Atkinson JP, Diamond MS. West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. Proc Natl Acad Sci USA 2006;103:19111–19116.

Ciminale V, Pavlakis GN, Derse D, Cunningham CP, Felber BK. Complex splicing in the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and proteins produced by HTLV type I. J Virol 1992;66:1737–1745.

Cripe TP, Haugen TH, Turk JP, Tabatabai F, Schmid PG, 3rd, Durst M, Gissmann L, Roman A, Turek LP. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 transactivator and repressor gene products: implications for cervical carcinogenesis. Embo J 1987;6:3745–3753.

Davy CE, Ayub M, Jackson DJ, Das P, McIntosh P, Doorbar J. HPV16 E1;E4 protein is phosphorylated by Cdk2/cyclin A and relocalizes this complex to the cytoplasm. Virology 2006;349:230–244.

de Laureto PP, Tosatto L, Frare E, Marin O, Uversky VN, Fontana A. Conformational properties of the SDS-bound state of alpha-synuclein probed by limited proteolysis: unexpected rigidity of the acidic C-terminal tail. Biochemistry 2006;45:11523–11531.

Derse D. Bovine leukemia virus transcription is controlled by a virus-encoded trans-acting factor and by cis-acting response elements. J Virol 1987;61:2462–2471.

Derse D. trans-acting regulation of bovine leukemia virus mRNA processing. J Virol 1988;62:1115–1119.

Dey A, Atcha IA, Bagchi S. HPV16 E6 oncoprotein stimulates the transforming growth factor-beta 1 promoter in fibroblasts through a specific GC-rich sequence. Virology 1997;228:190–199.

Doerig C, Hirt B, Beard P, Antonietti JP. Minute virus of mice non-structural protein NS-1 is necessary and sufficient for trans-activation of the viral P39 promoter. J Gen Virol 1988;69(Pt 10):2563–2573.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics 1998;148:1667–1686.

Du M, Fan X, Hong E, Chen JJ. Interaction of oncogenic papillomavirus E6 proteins with fibulin-1. Biochem Biophys Res Comm 2002;296:962–969.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002a;41:6573–6582.

Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem 2002b;62:25–49.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 2005;272: 5129–5148.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.

Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 2000;11:161–171.

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 2008a;9(Suppl 2): S1.

Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Curr Opin Struct Biol 2008b;18:756–764.

Dunker AK, Uversky VN. Signal transduction via unstructured protein conduits. Nat Chem Biol 2008;4:229–230.

Durrer P, Galli C, Hoenke S, Corti C, Gluck R, Vorherr T, Brunner J. H^+-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. J Biol Chem 1996;271:13417–13421.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.

Dyson N, Howley PM, Munger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 1989;243:934–937.

Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem 2001;70:777–810.

Edwards RA, Rohwer F. Viral metagenomics. Nat Rev Microbiol 2005;3:504–510.

Fehrmann F, Klumpp DJ, Laimins LA. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. J Virol 2003;77:2819–2831.

Felber BK, Paskalis H, Kleinman-Ewing C, Wong-Staal F, Pavlakis GN. The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. Science 1985;229:675–679.

Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS. Abundance of intrinsically unstructured proteins in P. Falciparum and other apicomplexan parasite proteomes. Mol Biochem Parasitol 2006;150:256–267.

Filippova M, Song H, Connolly JL, Dermody TS, Duerksen-Hughes PJ. The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. J Biol Chem 2002;277:21730–21739.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. Nucleic Acids Res 2008;36:D281–D288.

Fontana A, de Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol 2004;51:299–321.

Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry 1986;25:1847–1851.

Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. Fold Des 1997a;2:R17–R26.

Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol 1997b;266:223–230.

Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 2006;117:5–16.

Forterre P, Prangishvili D. The origin of viruses. Res Microbiol 2009;160:466–472.

Frattini MG, Laimins LA. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. Proc Natl Acad Sci USA 1994;91:12398–12402.

Fry EE, Newman JW, Curry S, Najjam S, Jackson T, Blakemore W, Lea SM, Miller L, Burman A, King AM, Stuart DI. Structure of Foot-and-mouth disease virus serotype A10 61 alone and complexed with oligosaccharide receptor: receptor conservation in the face of antigenic variation. J Gen Virol 2005;86:1909–1920.

Gao L, Aizaki H, He JW, Lai MM. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. J Virol 2004;78:3480–3488.

Gao Q, Kumar A, Srinivasan S, Singh L, Mukai H, Ono Y, Wazer DE, Band V. PKN binds and phosphorylates human papillomavirus E6 oncoprotein. J Biol Chem 2000;275:14824–14830.

Gao Q, Srinivasan S, Boyer SN, Wazer DE, Band V. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. Mol Cell Biol 1999;19:733–744.

Garcia-Alai MM, Alonso LG, de Prat-Gay G. The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry 2007;46:10405–10412.

Genther SM, Sterling S, Duensing S, Munger K, Sattler C, Lambert PF. Quantitative role of the human papillomavirus type 16 E5 gene during the productive stage of the viral life cycle. J Virol 2003;77:2832–2842.

Gewin L, Galloway DA. E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. J Virol 2001;75:7198–7201.

Glaunsinger BA, Lee SS, Thomas M, Banks L, Javier R. Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. Oncogene 2000;19:5270–5280.

Gloss B, Bernard HU, Seedorf K, Klock G. The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones. Embo J 1987;6:3735–3743.

Goh GK, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. Virol J 2008a;5:126.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genomics 2008b;9(Suppl 2): S4.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. Virol J 2009;6:69.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci USA 2007;104:8685–8690.

Gramberg T, Sunseri N, Landau NR. Accessories to the crime: recent advances in HIV accessory protein biology. Curr HIV/AIDS Rep 2009;6:36–42.

Gross-Mesilaty S, Reinstein E, Bercovich B, Tobias KE, Schwartz AL, Kahana C, Ciechanover A. Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. Proc Natl Acad Sci USA 1998;95:8058–8063.

Halpern AL, Münger K. HPV-16 E7: Correspondence between primary structure and biological properties, HPV Sequence Database. Los Alamos: Los Alamos National Laboratory; 1995. pp. III–58–III–73.

Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Tomato bushy stunt virus at 2.9 Å resolution. Nature 1978;276:368–373.

Henklein P, Bruns K, Sherman MP, Tessmer U, Licha K, Kopp J, de Noronha CM, Greene WC, Wray V, Schubert U. Functional and structural characterization of synthetic HIV-1 Vpr that transduces cells, localizes to the nucleus, and induces G2 cell cycle arrest. J Biol Chem 2000;275:32016–32026.

Holmes EC. Viral evolution in the genomic age. PLoS Biol 2007;5:e278.

Hughes FJ, Romanos MA. E1 protein of human papillomavirus is a DNA helicase/ATPase. Nucleic Acids Res 1993;21:5817–5823.

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.

Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD, Ackerman EJ. Identification of intrinsic order and disorder in the DNA repair protein XPA. Protein Sci 2001;10:560–571.

Inoue J, Seiki M, Yoshida M. The second pX product p27 chi-III of HTLV-1 is required for gag gene expression. FEBS Lett 1986;209:187–190.

Inoue J, Yoshida M, Seiki M. Transcriptional (p40x) and post-transcriptional (p27x-III) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc Natl Acad Sci USA 1987;84:3653–3657.

Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 2006;117:156–184.

Jian Y, Schmidt-Grimminger DC, Chien WM, Wu X, Broker TR, Chow LT. Post-transcriptional induction of p21cip1 protein by human papillomavirus E7 inhibits unscheduled DNA synthesis reactivated in differentiated keratinocytes. Oncogene 1998;17:2027–2038.

Kashanchi F, Brady JN. Transcriptional and post-transcriptional gene regulation of HTLV-1. Oncogene 2005;24:5938–5951.

Kiyokawa T, Seiki M, Iwashita S, Imagawa K, Shimizu F, Yoshida M. p27x-III and p21x-III, proteins encoded by the pX sequence of human T-cell leukemia virus type I. Proc Natl Acad Sci USA 1985;82:8359–8363.

Kiyono T, Hiraiwa A, Fujita M, Hayashi Y, Akiyama T, Ishibashi M. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. Proc Natl Acad Sci USA 1997;94:11612–11616.

Koonin EV, Senkevich TG, Dolja VV. The ancient Virus World and evolution of cells. Biol Direct 2006;1:29.

Koralnik IJ, Gessain A, Klotman ME, Lo Monico A, Berneman ZN, Franchini G. Protein isoforms encoded by the pX region of human T-cell leukemia/lymphotropic virus type I. Proc Natl Acad Sci USA 1992;89:8813–8817.

Kukimoto I, Aihara S, Yoshiike K, Kanda T. Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. Biochem Biophys Res Comm 1998;249:258–262.

Kumar A, Zhao Y, Meng G, Zeng M, Srinivasan S, Delmolino LM, Gao Q, Dimri G, Weber GF, Wazer DE, Band H, Band V. Human papillomavirus oncoprotein E6 inactivates the transcriptional coactivator human ADA3. Mol Cell Biol 2002;22:5801–5812.

Lacroix E, Viguera AR, Serrano L. Reading protein sequences backwards. Fold Des 1998;3:79–85.

Larocca D, Chao LA, Seto MH, Brunck TK. Human T-cell leukemia virus minus strand transcription in infected T-cells. Biochem Biophys Res Comm 1989;163:1006–1013.

Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ. Structural and functional studies of archaeal viruses. J Biol Chem 2009;284:12599–12603.

Leavitt SA, SchOn A, Klein JC, Manjappara U, Chaiken IM, Freire E. Interactions of HIV-1 proteins gp120 and Nef with cellular partners define a novel allosteric paradigm. Curr Protein Pept Sci 2004;5:1–8.

Lee SS, Glaunsinger B, Mantovani F, Banks L, Javier RT. Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. J Virol 2000;74:9680–9693.

Leechanachai P, Banks L, Moreau F, Matlashewski G. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. Oncogene 1992;7:19–25.

Li S, Labrecque S, Gauzzi MC, Cuddihy AR, Wong AH, Pellegrini S, Matlashewski GJ, Koromilas AE. The human papilloma virus (HPV)-18 E6 oncoprotein physically associates with Tyk2 and impairs Jak-STAT activation by interferon-alpha. Oncogene 1999;18:5727–5737.

Li X, Coffino P. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. J Virol 1996;70:4509–4516.

Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. Structure of simian virus 40 at 3.8-Å resolution. Nature 1991;354:278–284.

Liljas, L. The role of disordered segments in viral coat proteins. In: Cheng, RH, Hammar, L, editors. Conformational proteomics of macromolecular architecture. New Jersy, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co.; 2004. pp.53–77.

Lommer BS, Luo M. Structural plasticity in influenza virus protein NS2 (NEP). J Biol Chem 2002;277:7108–7117.

Longhi S. Nucleocapsid structure and function. Curr Top Microbiol Immunol 2009;329:103–128.

Massimi P, Pim D, Banks L. Human papillomavirus type 16 E7 binds to the conserved carboxy-terminal region of the TATA box binding protein and this contributes to E7 transforming activity. J Gen Virol 1997;78(Pt 10):2607–2613.

Matsuoka M, Green PL. The HBZ gene, a key player in HTLV-1 pathogenesis. Retrovirology 2009;6:71.

Mazurek S, Zwerschke W, Jansen-Durr P, Eigenbrodt E. Effects of the human papilloma virus HPV-16 E7 oncoprotein on glycolysis and glutaminolysis: role of pyruvate kinase type M2 and the glycolytic-enzyme complex. Biochem J 2001;356:247–256.

McIntyre MC, Ruesch MN, Laimins LA. Human papillomavirus E7 oncoproteins bind a single form of cyclin E in a complex with cdk2 and p107. Virology 1996;215:73–82.

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Protein disorder in the human diseasome: Unfoldomics of human genetic diseases. BMC Genomics 2009;10(Suppl 1):S12.

Minezaki Y, Homma K, Kinjo AR, Nishikawa K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol 2006;359:1137–1149.

Mohan A, Sullivan WJ, Radivojac P, Dunker AK, Uversky VN, Jr. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. Mol Biosyst 2008;4:328–340.

Mohr IJ, Clark R, Sun S, Androphy EJ, MacPherson P, Botchan MR. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. Science 1990;250:1694–1699.

Morellet N, Roques BP, Bouaziz S. Structure-function relationship of Vpr: biological implications. Curr HIV Res 2009;7:184–210.

Morosov A, Phelps WC, Raychaudhuri P. Activation of the c-fos gene by the HPV16 oncoproteins depends upon the cAMP-response element at -60. J Biol Chem 1994;269:18434–18440.

Nagashima K, Yoshida M, Seiki M. A single species of pX mRNA of human T-cell leukemia virus type I encodes trans-activator p40x and two other phosphoproteins. J Virol 1986;60:394–399.

Nakagawa S, Huibregtse JM. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. Mol Cell Biol 2000;20:8244–8253.

Nayak DP, Balogun RA, Yamada H, Zhou ZH, Barman S. Influenza virus morphogenesis and budding. Virus Res 2009;143:147–161.

Nayak DP, Hui EK, Barman S. Assembly and budding of influenza virus. Virus Res 2004;106:147–165.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(Suppl 6):566–572.

Oh ST, Kyo S, Laimins LA. Telomerase activation by human papillomavirus type 16 E6 protein: induction of human telomerase reverse transcriptase expression through Myc and GC-rich Sp1 binding sites. J Virol 2001;75:5559–5566.

Ohlenschlager O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, Baum M, Korbas M, Meyer-Klaucke W, Durst M, Gorlach M. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene 2006;25:5953–5959.

Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry 2005a;44:1989–2000.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005b;44:12454–12470.

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 2008;9(Suppl 1): S1.

Patel D, Huang SM, Baglia LA, McCance DJ. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. Embo J 1999;18:5061–5072.

Pim D, Massimi P, Dilworth SM, Banks L. Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A. Oncogene 2005;24:7830–7838.

Polverino de Laureto P, De Filippis V, Di Bello M, Zambonin M, Fontana A. Probing the molten globule state of alpha-lactalbumin by limited proteolysis. Biochemistry 1995;34:12596–12604.

Prangishvili D, Forterre P, Garrett RA. Viruses of the Archaea: a unifying view. Nat Rev Microbiol 2006;4:837–848.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007;92:1439–1456.

Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J Virol 2009;83:10719–10736.

Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol 2008;6:315–319.

Rayment I, Baker TS, Caspar DL, Murakami WT. Polyoma virus capsid structure at 22.5 Å resolution. Nature 1982;295:110–115.

Reanney DC. The evolution of RNA viruses. Annu Rev Microbiol 1982;36:47–73.

Reingewertz TH, Benyamini H, Lebendiker M, Shalev DE, Friedler A. The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. Protein Eng Des Sel 2009;22:281–287.

Ristriani T, Masson M, Nomine Y, Laurent C, Lefevre JF, Weiss E, Trave G. HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. J Mol Biol 2000;296:1189–1203.

Ristriani T, Nomine Y, Masson M, Weiss E, Trave G. Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. J Mol Biol 2001;305:729–739.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Dunker AK. Identifying Disordered regions in proteins from amino acid sequences. IEEE Int Conf Neural Networks 1997;1:90–95.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK. Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 1998;437–448.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.

Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci USA 2006;103:8390–8395.

Ronco LV, Karpova AY, Vidal M, Howley PM. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. Genes Dev 1998;12:2061–2072.

Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol 2004;14:171–180.

Rybicki EP. The classification of organisms at the edge of life, or problems with virus systematics. S Afr J Sci 1990;86:182–186.

Sano E, Carlson S, Wegley L, Rohwer F. Movement of viruses between biomes. Appl Environ Microbiol 2004;70:5842–5846.

Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.

Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 1990;63:1129–1136.

Sedman SA, Barbosa MS, Vass WC, Hubbert NL, Haas JA, Lowy DR, Schiller JT. The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. J Virol 1991;65:4860–4866.

Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng L, Dadgar S, Bottazzi ME, Ghedin E, Pumfery A, Kashanchi F. Role of viral regulatory and accessory proteins in HIV-1 replication. Front Biosci 2004;9:2388–2413.

Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc Natl Acad Sci USA 1983;80:3618–3622.

Seiki M, Inoue J, Hidaka M, Yoshida M. Two cis-acting elements responsible for post-transcriptional trans-regulation of gene expression of human T-cell leukemia virus type I. Proc Natl Acad Sci USA 1988;85:7124–7128.

Seiki M, Inoue J, Takeda T, Yoshida M. Direct evidence that p40x of human T-cell leukemia virus type I is a trans-acting transcriptional activator. EMBO J 1986;5:561–565.

Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 2006;281:8347–8356.

Skehel JJ, Bayley PM, Brown EB, Martin SR, Waterfield MD, White JM, Wilson IA, Wiley DC. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc Natl Acad Sci USA 1982;79:968–972.

Skehel JJ, Wiley DC. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell 1998;95:871–874.

Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 2000;69:531–569.

Sodroski J, Trus M, Perkins D, Patarca R, Wong-Staal F, Gelmann E, Gallo R, Haseltine WA. Repetitive structure in the long-terminal-repeat element of a type II human T-cell leukemia virus. Proc Natl Acad Sci USA 1984;81:4617–4621.

Srivenugopal KS, Ali-Osman F. The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein. Oncogene 2002;21:5940–5945.

Stevens J, Corper AL, Basler CF, Taubenberger JK, Palese P, Wilson IA. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science 2004;303:1866–1870.

Straight SW, Hinkle PM, Jewers RJ, McCance DJ. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. J Virol 1993;67:4521–4532.

Surendran R, Herman P, Cheng Z, Daly TJ, Ching Lee J. HIV Rev self-assembly is linked to a molten-globule to compact structural transition. Biophys Chem 2004;108:101–119.

Suzuki Y. Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. Biol Pharm Bull 2005;28:399–408.

Thomas M, Banks L. Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. J Gen Virol 1999;80(Pt 6):1513–1517.

Thomas MC, Chiang CM. E6 oncoprotein represses p53-dependent gene activation via inhibition of protein acetylation independently of inducing p53 degradation. Mol Cell 2005;17:251–264.

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? Trends Biochem Sci 2009;34:53–59.

Tommasino M, Adamczewski JP, Carlotti F, Barth CF, Manetti R, Contorni M, Cavalieri F, Hunt T, Crawford L. HPV16 E7 protein associates with the protein kinase p33CDK2 and cyclin A. Oncogene 1993;8:195–202.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.

Tompa P. The functional benefits of protein disorder. J Mol Struct - Theochem 2003; 666–667:361–371.

Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays 2009;31:328–335.

Tong X, Boll W, Kirchhausen T, Howley PM. Interaction of the bovine papillomavirus E6 protein with the clathrin adaptor complex AP-1. J Virol 1998;72:476–482.

Tong X, Howley PM. The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. Proc Natl Acad Sci USA 1997;94:4412–4417.

Ustav M, Stenlund A. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. EMBO J 1991;10:449–457.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002a;11:739–756.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002b;269:2–12.

Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell Mol Life Sci 2003;60:1852–1871.

Uversky VN. Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res 2008a;5:260–287.

Uversky, VN. Intrinsic disorder in proteins associated with neurodegenerative diseases. In: Ovádi J, Orosz F, editors. Protein folding and misfolding: neurodegenerative diseases. New York, USA: Springer; 2008b.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18:343–384.

Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. Annu Rev Biophys Biomol Struct 2008;37:215–246.

Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res 2006;5:1829–1842.

Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics 2007;8:211.

Vendel AC, Lumb KJ. Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. Biochemistry 2003;42:910–916.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.

Weber T, Paesold G, Galli C, Mischler R, Semenza G, Brunner J. Evidence for H(+)-induced insertion of influenza hemagglutinin HA2N-terminal segment into viral membrane. J Biol Chem 1994;269:18353–18358.

Wharton SA, Calder LJ, Ruigrok RW, Skehel JJ, Steinhauer DA, Wiley DC. Electron microscopy of antibody complexes of influenza virus haemagglutinin in the fusion pH conformation. EMBO J 1995;14:240–246.

Wiley DC, Skehel JJ. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem 1987;56:365–394.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 1981;289:366–373.

Wilson R, Fehrmann F, Laimins LA. Role of the E1–E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. J Virol 2005;79:6732–6740.

Wray V, Federau T, Henklein P, Klabunde S, Kunert O, Schomburg D, Schubert U. Solution structure of the hydrophilic region of HIV-1 encoded virus protein U (Vpu) by CD and 1H NMR spectroscopy. Int J Pept Protein Res 1995;45:35–43.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.

Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol 2009;19:31–38.

Xue B, Li L, Meroueh SO, Uversky VN, Dunker AK. Analysis of structured and intrinsically disordered regions of transmembrane proteins. Mol Biosyst 2009a;5:1688–1702.

Xue B, Oldfield CJ, Dunker AK, Uversky VN. CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. FEBS Lett 2009b;583:1469–1474.

Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? Protein Pept Lett 2010;17:932–951.

Zhang B, Li P, Wang E, Brahmi Z, Dunn KW, Blum JS, Roman A. The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma. Virology 2003;310:100–108.

Zimmermann H, Degenkolbe R, Bernard HU, O'Connor MJ. The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. J Virol 1999;73:6209–6219.

Zwerschke W, Mazurek S, Massimi P, Banks L, Eigenbrodt E, Jansen-Durr P. Modulation of type M2 pyruvate kinase activity by the human papillomavirus type 16 E7 oncoprotein. Proc Natl Acad Sci USA 1999;96:1291–1296.

2

FUNCTIONAL ROLE OF STRUCTURAL DISORDER IN CAPSID PROTEINS

LARS LILJAS

2.1 INTRODUCTION

Capsid proteins form a protective coat around the nucleic acid of a virus. Sometimes called *coat proteins*, they are found in all virus particles. The coat always has some kind of symmetry, the most common being icosahedral. Icosahedral symmetry gives isometric structures, or variants thereof, that lead to elongated particles with rounded ends. The other type of symmetry found is helical symmetry, leading to rod-shaped or filamentous particles. In some viruses, there are no further components, but many of them have more than one layer of proteins. Enveloped viruses have a lipid bilayer with membrane-bound proteins. These outer layers may also be symmetric, but large viruses surrounded by membranes often lack a defined shape, even if the complex of nucleic acid and protein (the nucleocapsid) is symmetric.

It has been possible to determine the structure of many viral capsids using X-ray crystallography, and we therefore know much about the capsid proteins and their interaction with other proteins and the nucleic acid. This has been possible even for relatively large viruses, such as bluetongue virus cores with 900 protein subunits (Grimes *et al.*, 1998) and adenovirus with 780 subunits (Reddy *et al.*, 2010), but only in one case for a virus with a lipid membrane (Cockburn *et al.*, 2004). For enveloped viruses, our knowledge about capsid structure and function comes from cryo-electron microscopy, combined with structure determination of isolated capsid proteins that for some reason have not assembled into large structures. Recently, cryo-electron microscopy has reached a resolution where atomic structure of large

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

virus particles can be modeled (Liu *et al.*, 2010; Zhang *et al.*, 2010), but this has not yet been possible for any membrane virus.

The structures of capsid proteins and capsids have shown that many of these proteins consist of a globular part and extended arms (Chapman and Liljas, 2003; Liljas, 2004). The arms may be ordered in a capsid through interactions with other viral components, but it is apparent that they are flexible in the isolated protein. This chapter is about, among other things, how the disordered segments become ordered in the capsid. The crystallization process is insensitive to a possible internal asymmetry in a viral particle when the surface is icosahedrally symmetric. In structures of icosahedral capsids, those portions of the capsid protein that do not follow the icosahedral symmetry will have a random orientation in the crystal and appear disordered. It is not possible to decide whether these portions have a unique conformation, present in all particles and determined by interactions with the nucleic acid, or their conformation is random.

The extended arms in capsid proteins have a number of functions, and this chapter discusses the function of disordered or flexible arms for nucleic acid interactions, control of assembly, and stabilization of particles.

2.2 NUCLEIC ACID RECOGNITION AND BINDING

The symmetry of capsids is achieved by using many identical copies of the capsid protein. The capsid protein may interact with the nucleic acid both to recognize it, ensuring that the correct nucleic acid is included in the capsid, and to neutralize the negative charges of the phosphates. The nucleic acid is a single molecule (or a few molecules) and does not easily adopt the same type of symmetry, especially in the case of icosahedral symmetry. Many capsid proteins, therefore, use flexible arms to interact with the nucleic acid, and these arms are not visible in the structures because of their conformational flexibility.

The first crystal structures of viral particles to be determined were those of some icosahedral plant viruses with a single-stranded RNA genome. All these viruses had a capsid protein with a fold that dubbed the jelly-roll fold. They also shared the property of a disordered N-terminal extension. This arm had several positively charged and very few negatively charged residues and was therefore assumed to interact with the RNA. This observation has now been extended by many structures and holds for the plant viruses in the families Tombusviridae and Bromoviridae and the genus sobemoviruses, as well for the insect viruses in the genera Nodaviridae and Tetraviridae and for animal viruses in the family Polyomaviridae.

A typical example of that is the sobemoviruses, which are single-stranded RNA viruses. There are five known structures of sobemoviruses, all of which have 180 identical coat protein subunits of about 250 amino acid residues. A comparison of the sequences and structures of their ordered parts is found in Plevka *et al.* (2007). An N-terminal segment with a length of 26–38 amino acid residues is disordered in the capsid protein of all these viruses. This segment has 5–12 arginine/lysine residues, and all of them have at least one stretch of three or more positively

CfMV	VRKGAATKAPQQPKPRAQQQPGGRRRRRGRSMEP
RYMV	ARKGKKTNPNQGQQGKKKSRRPRGRS
SeMV	LSIQQLAKAIANTLETPPQPKAGRRRSAVQQLPPI
SCpMV	ATRLTKKQLAQAIQNTLPNPPRRKRRAKRRAAQVPKPT
RgMV	ARKKGKSASQVIVLKEKSRKKRQKSR

Figure 2.1 The disordered N-terminal segments of the coat proteins of five sobemoviruses for which the crystal structure of the capsid is known. The positively charged arginine and lysine residues are shown in bold. The sequences are from the cocksfoot mottle virus (CfMV), rice yellow mosaic virus (RYMV), sesbania mosaic virus (SeMV), southern cowpea mosaic virus (SCpMV), and ryegrass mosaic virus (RgMV).

charged residues (Fig. 2.1). There are at most a single negatively charged residue, no aromatic residues, and only a few nonpolar residues. The sequences are, therefore, of a type that suggests that they are disordered (Uversky *et al.*, 2000). These sequences are not conserved and cannot be aligned reliably, but all of them share the characteristic of being positively charged.

In the insect viruses of the Nodaviridae family, the capsid is built up in a similar way as in the sobemoviruses with 180 identical subunits encapsulating two single-stranded RNA molecules. The capsid protein has the same jelly-roll topology and an N-terminal segment that is not part of the globular fold of the subunit. The N-terminal segments in these viruses have a length of 47–55 amino acid residues and have about 30% positively charged residues, with a strong (or complete) preference for arginine residues (Fig. 2.2). The properties of these segments are similar to those of the sobemoviruses. In contrast to the sobemoviruses, where the crystal structure shows essentially no trace of the RNA molecule, parts of the nucleic acid are visible in the crystal structures of nodaviruses. This means that some of the RNA molecule has adapted to the symmetrical shell and is bound similarly to the protein. The RNA is bound as a double-stranded segment. In one of the nodaviruses, Pariacoto virus (Tang *et al.*, 2001), the N-terminal segment is mostly ordered in 60 of the 180 identical subunits (Fig. 2.3). This segment interacts directly with one of the double-stranded RNA segments with several of the arginine and lysine residues close to the phosphate groups in RNA. The N-terminal segments of the other 120 subunits probably interact with segments of the RNA that are further away from

```
BBV  VRNNNRRRQRTQRIVTTTTQTAPVPQQN-----VPKQPRRRRNRARRNRRQGRAMNMGAL
FHV  VNNNRPRRQRAQRVVVTTTQTAPVPQQN-----VPRNGRRRRNRTRRNRRRVRGMNMAAL
PaV  VSRTKNRRNKARKVVSRSTALVPMAPASQRTGPAPRKPRKRNQALVRN-----------
NoV  VSKAARRR----------RAAPRQQQRQQSNRASNQPRRR--RARRTRRQQRMAATNNM
```

Figure 2.2 Aligned sequences of the disordered N-terminal segments of the coat proteins of four nodaviruses. The positively charged arginine and lysine residues are shown in bold. The sequences are from the black beetle virus (BBV), Flock House virus (FHV), Pariacoto virus (PaV), and Nodamura virus (NoV).

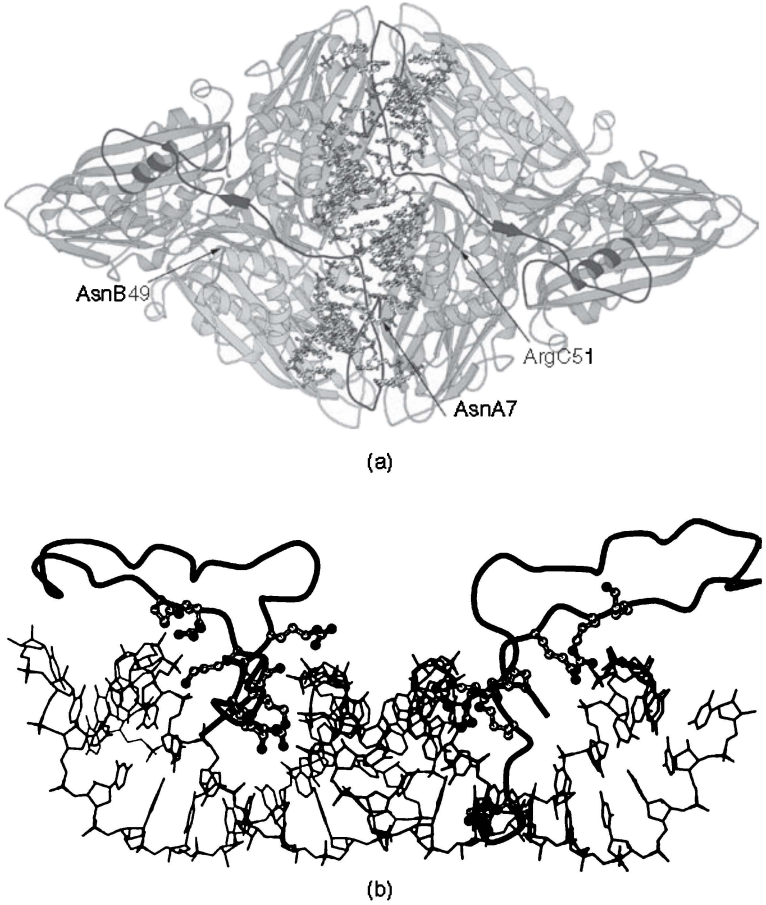


Figure 2.3 (a). Two trimers of Pariacoto virus showing the interaction with segments of double-stranded RNA (PDB code 1f8v). The three types of subunits are colored blue, green, and red (subunits A, B, and C, respectively). The partially ordered N-terminal arms of the A subunit are shown in dark blue. These arms interact with the RNA segments. The corresponding arms of the other subunits are disordered. (b). Details of the interaction showing the interaction between positively charged residues in the ordered arm and the RNA (gray lines with colored phosphate groups). Only the ordered arms of the two A subunits (residues 7–46) are shown. The view is parallel to the inner surface along the major groove of the double helix. (*See insert for color representation of the figure.*)

the symmetrical capsid and no longer adapt to the icosahedral symmetry of the protein layer.

Although the capsid proteins of many simple viruses have positively charged N-terminal segments, there are several examples of viruses that have proteins with a similar jelly-roll fold but no extended arms interacting with the RNA. The plant viruses in the Tymoviridae family (Canady *et al.*, 1996) are one example. In this

family of viruses, the single-stranded RNA binds spermidine molecules, which interact with the phosphate groups in the nucleic acid.

Positively charged N-terminal segments that bind to nucleic acid are not restricted to the simple viruses previously described. In alphaviruses, which are enveloped animal viruses, a capsid with 240 identical protein subunits encloses a single-stranded RNA molecule. The capsid protein has an N-terminal segment of about 110 amino acid residues and a C-terminal globular domain of about 100 residues that has the fold of a serine protease (Choi *et al.*, 1991). The N-terminal segment has about 30 positively charged residues that form patches interrupted by proline-rich and glutamine-rich regions. In cryo-electron microscopy, fitting of the protease part in the density of the capsid leaves density corresponding to a segment connecting the protease with the RNA region of the map (Mukhopadhyay *et al.*, 2006).

Likewise, in hepatitis B viruses, a segment of the capsid protein is positively charged and interacts with the nucleic acid molecule, in this case double-stranded DNA. Hepatitis B virus is an enveloped virus. The conformation of the capsid is known from recombinant expression of a fragment of the capsid protein where the nucleic-acid-binding C-terminal extension is removed (Wynne *et al.*, 1999). The protein has a completely helical fold, another example illustrating that enveloped viruses mostly do not have capsid proteins with the jelly-roll fold, in contrast to the nonenveloped viruses. Among the 36 residues of the C-terminal extension, 17 are arginine residues.

In addition to the nonspecific interactions that occur between the flexible segments of the protein and the nucleic acid, there is also specific recognition of the nucleic acid by capsid proteins. These interactions control that the correct nucleic acid molecules are encapsidated. They have been difficult to study structurally, and the only case of a binding between a capsid protein and a recognition signal in a nucleic acid where structural data are available is that of the small bacterial virus MS2 (Valegård *et al.*, 1994). In this case, there are, however, no extended or disordered segments involved in the interaction. In other viruses, especially in viruses with capsid proteins with the jelly-roll fold, such as the nodaviruses, both the general neutralization and the specific recognition of the viral nucleic acid may use the extended disordered segments (Marshall and Schneemann, 2001; Schneemann and Marshall, 1998).

2.3 CONTROL OF ASSEMBLY

Icosahedral viruses are built up of 60 identical units related by the two-, three-, and fivefold symmetry axes of such objects. Many viruses have capsids built up of multiples of 60 identical subunits, and this means that chemically identical subunits have distinct environments after capsid formation. To explain how the capsid proteins assemble into stable structures with these differences in interactions, Caspar and Klug (1962) proposed the quasi-equivalence hypothesis. This states that the subunits form similar (quasi-equivalent) interactions according to a scheme that

leads to a series of allowed triangulation numbers ($T = 3$, 4, 7, and so on), which indicate the multiples of 60 subunits in the capsid. The basis of the hypothesis is that the differences between hexameric and pentameric arrangements of proteins are small enough to allow similar interactions to form. This hypothesis has been shown to hold for the great majority of viruses, but it does not explain in what way the quasi-equivalent interactions differ in detail and how the correct interaction forms at assembly.

The first crystal structures of icosahedral viruses were of plant viruses with $T = 3$ quasi-symmetry and showed that the correct assembly was controlled by N-terminal arms that were ordered only in one set of subunits and disordered in the other (Abad-Zapatero *et al.*, 1980; Harrison *et al.*, 1978). The ordered arms were bound in one of the sets of quasi-equivalent subunit–subunit interfaces, making it distinct (flat) compared to the other (bent) contacts (Fig. 2.4). This order–disorder switching was found in all sobemoviruses and tombusviruses, which have similar capsids. Removal of the arm in these viruses results in formation of capsids with only 60 subunits (Erickson *et al.*, 1985; Hsu *et al.*, 2006; Sangita *et al.*, 2004). A similar arrangement of subunits is also found in structures of viruses in the Caliciviridae, Nodaviridae, and Tetraviridae families. In nodavirus capsids, the switching is controlled both through order/disorder in an N-terminal arm and through binding of RNA elements (Fig. 2.3). In the tetravirus Nudaurelia capensis ω virus Nudaurelia Capensis ω Virus (NwV), an insect virus with $T = 4$ quasi-symmetry, a C-terminal region of the protein shows an order–disorder switching, where ordered segments in 120 subunits create flat contacts, and bent contacts form where the corresponding segments are disordered (Helgstrand *et al.*, 2004).

Another way of controlling the assembly of capsids with quasi-equivalent symmetry using order/disorder is found in the $T = 3$ capsids of viruses in the Bromoviridae and Tymoviridae families. Here, extended N-terminal arms from two of the three types of subunits form a sixfold arrangement stabilizing a hexamer of proteins, while the corresponding arms in the third type of subunit, forming pentamers, are disordered (Canady *et al.*, 1996; Speir *et al.*, 1995).

An even more complex way of controlling quasi-equivalence is found in polyomaviruses. These viruses have subunits arranged in a $T = 7$ quasi-symmetric lattice, but at the sixfold positions, only pentamers are found. The central part of all the 72 pentamers have the same conformation, but extended C-terminal arms from each subunit are used to control the packing (Liddington *et al.*, 1991). There are three distinct types of pentamer–pentamer interactions, and in all cases, part of the C-terminal segment interacts intimately, and in the same way, with one of the capsid proteins in the neighboring pentamer (Fig. 2.5). In this family of viruses, removal of a suitable segment of the C-terminal region also leads to formation of $T = 1$ particles (Yokoyama *et al.*, 2007).

Extended arms and order–disorder switching are used to control quasi-symmetry in many nonenveloped icosahedral virus capsids. In these viruses, the capsid proteins with extended arms have the same jelly-roll fold. In capsids formed by proteins with other types of folds, the control of quasi-equivalence seems to use other mechanisms. The $T = 3$ leviviruses (Valegård *et al.*, 1990) are examples for capsids that

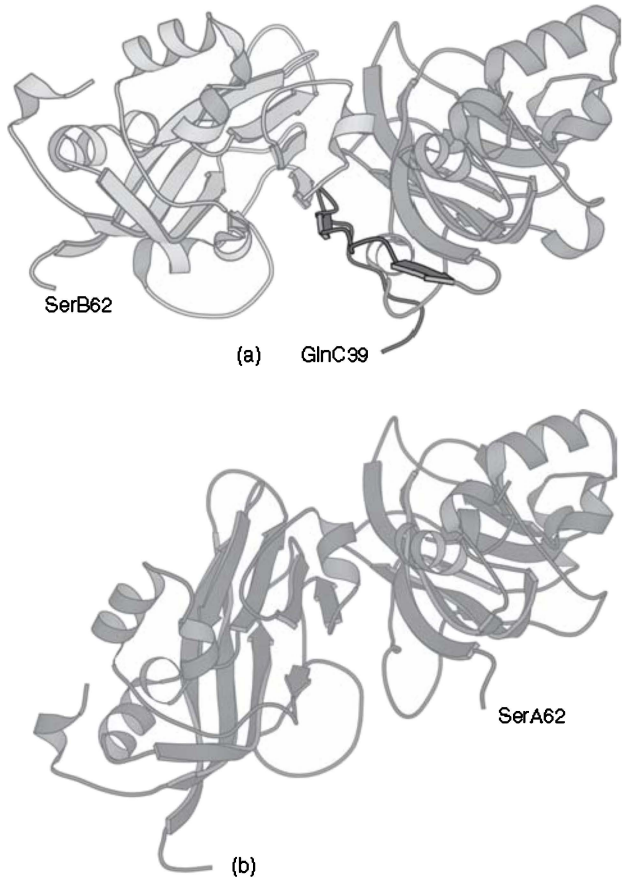


Figure 2.4 The bent and flat contacts in a sobemovirus (PDB code 4sbv). (a) The flat contact between the B (left) and C (right) subunits (the quasi-sixfold contact). The partly ordered arm in the C subunits is shown in dark. It is inserted between the B and C subunits, and the visible N-terminus (marked) forms a trimeric structure called the β-*annulus* with the corresponding arms from two other C subunits (not illustrated here). (b) The bent contact between two A subunits (the fivefold contact). The visible N-terminus is indicated. To simplify the comparison of (a) and (b), the molecules to the right have the same orientation.

assemble without order–disorder mechanisms. Many large viruses use scaffolding proteins that are removed after assembly.

2.4 STABILIZATION OF THE CAPSID AND CONTROL OF NUCLEIC ACID RELEASE

A complex use of extended arms that has been studied in structural detail is found in the picornaviruses, which are small RNA viruses. In picornaviruses, such as

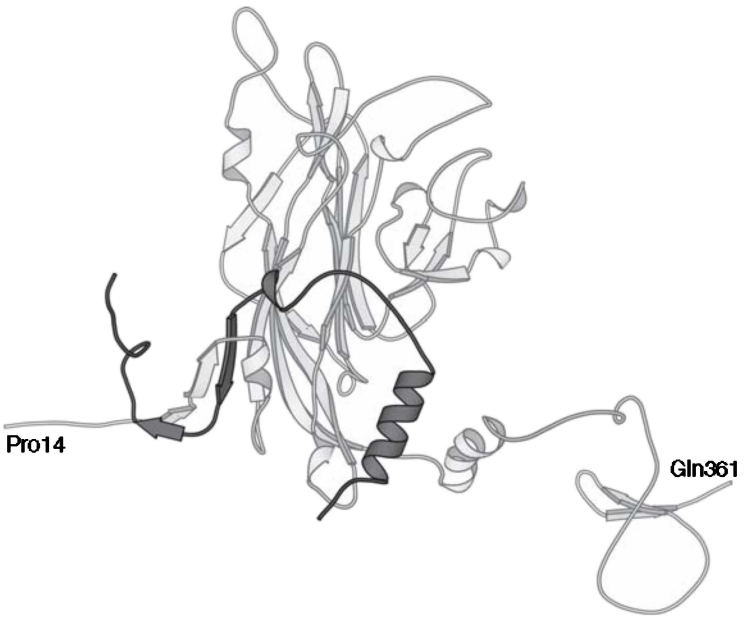


Figure 2.5 The capsid protein of SV40, illustrating the use of arms to link pentamers in the capsid (PDB code 1sva). The visible N- and C-termini are marked. In dark is the invading C-terminal arm (residues 296–355) from a subunit in another pentamer. The arm forms a β-strand that is clamped by the N-terminal segment of the invaded subunit.

rhinovirus, poliovirus, and foot-and-mouth disease virus, the capsid shell is assembled from 180 subunits, but these are of three distinct types, and the capsid does not show quasi-symmetry (Acharya *et al.*, 1989; Hogle *et al.*, 1985; Rossmann *et al.*, 1985). The three different subunits are called VP0, VP1, and VP3. At a late stage during assembly, part of the N-terminal of VP0 is cleaved to create the VP4 peptide of about 70 amino acid residues and the major VP2 protein. Such a maturation cleavage also is found in many other types of viral capsids, for example, nodaviruses, tetraviruses, and the influenza virus, and it is mostly a necessary step to allow infection of the host cell. Since the capsid should release the nucleic acid in some way upon infection, a maturation cleavage of a peptide creates a metastable capsid that is sensitive to some signal created during the infection process. The mechanisms leading to release of the nucleic acid are widely different, but the signal may be exposure to slightly acidic pH, such as found in endosomes, or interaction with receptors.

VP1, VP2, and VP3 all have a jelly-roll fold and extended N-terminal arms with lengths of 50–80, about 60, and about 45 residues, respectively. The arrangement in the capsid of the jelly-roll part of the subunits resembles that of small plant $T = 3$ viruses such as the sobemoviruses. The N-terminal arms of the three subunits, together with VP4, form an extended network on the inside of the capsid

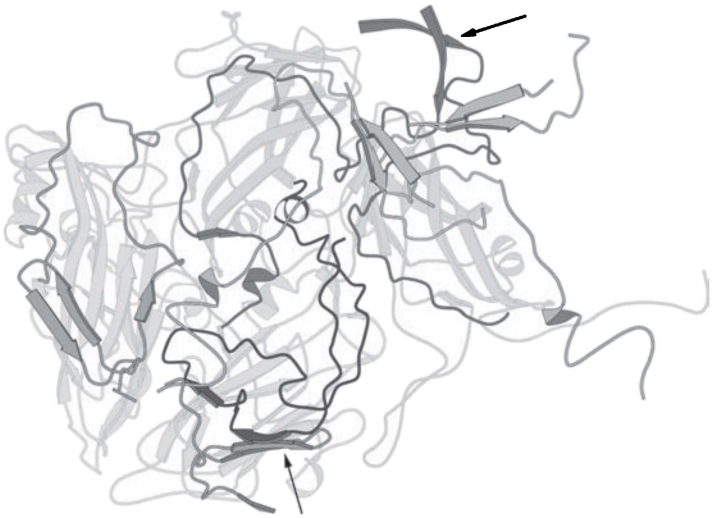


Figure 2.6 The arms on the inside of the protein shell in a picornavirus (PDB code 1pvc). One protomer of VP1 (blue), VP2 (green), VP3 (red), and VP4 (turquoise) is shown. The jelly-roll domain of VP1 and the N-terminal segments of VP2, VP3, and VP4 from neighboring protomers are included to illustrate some of the contacts. The N-terminal arms are in dark colors. Conserved interactions between arms in picornaviruses are indicated by arrows. (*See insert for color representation of the figure.*)

(Fig. 2.6). Only short segments of the N-terminal parts of VP1 and VP2 are disordered in the intact virions. This network stabilizes the different types of contacts between the subunits in the capsid. It is formed during assembly and is completed after the cleavage of VP0 into VP4 and VP2 (Basavappa *et al.*, 1994). The conformation of these arms is partially conserved among the picornaviruses, and a similar network is also found in the cricket paralysis virus, a member of the Dicistroviridae family (Liljas *et al.*, 2002).

The amino acid sequences of the extended arms are not of the same kind as the arms interacting with the nucleic acid. The sequences do not show preferences for a few kinds of amino acids as is the case for the arms interacting with nucleic acid, and there are aromatic residues, which are lacking in the other types of arms. This probably reflects that these arms have a unique way of interacting with the compact part of the subunits. The contact surfaces have, therefore, evolved to be of the same kind as contact surfaces within a globular domain, but the contacts are formed only after the assembly and cleavage of VP0.

In the case of polioviruses, the signal leading to conformational changes and release of the RNA is the binding to the receptor. Binding to the host cell or even to soluble receptor molecules leads to the release of VP4 and exposure of the N-terminal portion of VP1 (Fricks and Hogle, 1990). The released VP4 and N-terminal of VP1 are thought to form a membrane pore through which the RNA can enter the host cell (Bubeck *et al.*, 2005).

ACKNOWLEDGMENT

Terese Bergfors has provided valuable comments to this chapter.

REFERENCES

Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AGW, Rayment I, Rossmann MG, Suck D, Tsukihara T. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 1980;286:33–39.

Acharya KR, Fry E, Stuart D, Fox G, Rowlands D, Brown F. The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. Nature 1989;337:709–716.

Basavappa R, Syed R, Flore O, Icenogle JP, Filman DJ, Hogle JM. Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: Structure of the empty capsid assembly intermediate at 2.9 Å resolution. Protein Sci 1994;3:1651–1669.

Bubeck D, Filman DJ, Cheng N, Steven AC, Hogle JM, Belnap DM. The structure of the poliovirus 135S cell entry intermediate at 10-angstrom resolution reveals the location of an externalized polypeptide that binds to membranes. J Virol 2005;79:7745–7755.

Canady MA, Larson SB, Day J, McPherson A. Crystal structure of turnip yellow mosaic virus. Nature Struct Biol 1996;3:771–781.

Caspar DLD, Klug A. Physical principles in the construction of regular viruses. Cold Spring Harbor Symp Quant Biol 1962;27:1–24.

Chapman, MS, Liljas, L. Structural folds of viral proteins. Volume 64. In: Chiu W, Johnson JE, editors. Advances in protein chemistry. San Diego: Academic Press; 2003. pp. 125–196.

Choi H-K, Tong L, Minor W, Dumas P, Boege U, Rossmann MG. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature 1991;354:37–43.

Cockburn JJ, Abrescia NG, Grimes JM, Sutton GC, Diprose JM, Benevides JM, Thomas GJ, Bamford JK, Bamford DH, Stuart DI, Jr. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature 2004;432:122–125.

Erickson JW, Silva AM, Murthy MRN, Fita I, Rossmann MG. The structure of a $T = 1$ icosashedral empty particle from southern bean mosaic virus. Science 1985;229:625–629.

Fricks CE, Hogle JM. Cell-induced conformational change in poliovirus: Externalization of the amino terminus of VP1 is responsible for liposome binding. J Virol 1990;64:1934–1945.

Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PP, Stuart DI. The atomic structure of the bluetongue virus core. Nature 1998;395:470–478.

Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Tomato bushy stunt virus at 2.9 Å resolution. Nature 1978;276:368–373.

Helgstrand C, Munshi S, Johnson JE, Liljas L. The refined structure of Nudaurelia capensis w virus reveals control elements for a T = 4 capsid maturation. Virology 2004;318:192–204.

Hogle JM, Chow M, Filman DJ. Three-dimensional structure of poliovirus at 2.9 A resolution. Science 1985;229:1358–1365.

Hsu C, Singh P, Ochoa W, Manayani DJ, Manchester M, Schneemann A, Reddy VS. Characterization of polymorphism displayed by the coat protein mutants of tomato bushy stunt virus. Virology 2006;349:222–229.

Liddington R, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. Structure of simian virus 40 at 3.8 Å resolution. Nature 1991;354:278–284.

Liljas, L. The role of disordered segments in viral coat proteins. In: Cheng RH, Hammar L, editors. Conformational proteomics of macromolecular architecture. Singapore: World Scientific Publishing; 2004. pp. 53–77.

Liljas L, Lin T, Tate J, Christian P, Johnson JE. Evolution of the picornavirus superfamily: implications of conserved structural motifs between picornaviruses and insect picorna-like viruses. Arch Virol 2002;147:59–84.

Liu H, Jin L, Koh SBS, Atanasov I, Schein S, Wu L, Zhou ZH. Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science 2010;329:1038–1043.

Marshall D, Schneemann A. Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein. Virology 2001;285(1):165–175.

Mukhopadhyay S, Zhang W, Gabler S, Chipman PR, Strauss EG, Strauss JH, Baker TS, Kuhn RJ, Rossmann MG. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. Structure 2006;14:63–73.

Plevka P, Tars K, Zeltins A, Balke I, Truve E, Liljas L. The three-dimensional structure of ryegrass mottle virus at 2.9 Å resolution. Virology 2007;369:364–374.

Reddy VS, Natchiar SK, Stewart PL, Nemerow GR. Crystal structure of human adenovirus at 3.5 Å resolution. Science 2010;329:1071–1075.

Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, Johnson JE, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 1985;317:145–153.

Sangita V, Lokesh GL, Satheshkumar PS, Vijay CS, Saravanan V, Savithri HS, Murthy MR. $T = 1$ capsid structures of Sesbania mosaic virus coat protein mutants: determinants of $T = 3$ and $T = 1$ capsid assembly. J Mol Biol 2004;342:987–999.

Schneemann A, Marshall D. Specific encapsidation of nodavirus RNAs is mediated through the C terminus of capsid precursor protein alpha. J Virol 1998;72:8738–8746.

Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. Structure 1995;3:63–78.

Tang L, Johnson KN, Ball LA, Lin T, Yeager M, Johnson JE. The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. Nature Struct Biol 2001;8:77–83.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–427.

Valegård K, Liljas L, Fridborg K, Unge T. The three-dimensional structure of the bacterial virus MS2. Nature 1990;345:36–41.

Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature 1994;371:623–626.

Wynne SA, Crowther RA, Leslie AG. The crystal structure of the human hepatitis B virus capsid. Mol Cell 1999;3:771–780.

Yokoyama N, Kawano MA, Tsukamoto H, Enomoto T, Inoue T, Takahashi RU, Nakanishi A, Imai T, Wada T, Handa H. Mutational analysis of the carboxyl-terminal region of the SV40 major capsid protein VP1. J Biochem 2007;141:279–286.

Zhang X, Jin L, Fang Q, Hui WH, Zhou ZH. 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell 2010;141:472–482.

3

STRUCTURAL DISORDER WITHIN THE NUCLEOPROTEIN AND PHOSPHOPROTEIN FROM MEASLES, NIPAH, AND HENDRA VIRUSES

JOHNNY HABCHI, LAURENT MAMELLI, AND SONIA LONGHI

3.1 THE REPLICATIVE COMPLEX OF MEASLES, NIPAH, AND HENDRA VIRUSES

Measles virus (MeV) is a Morbillivirus member within the Paramyxoviridae family of the Mononegavirales order. This order includes several human pathogens with a strong socioeconomical impact and comprises both well-characterized viruses (e.g., MeV, mumps virus (MuV) and parainfluenza, rabies, and Ebola viruses) and emerging viruses, such as the Nipah virus (NiV) and Hendra virus (HeV). Although these latter viruses share the same overall genome organization of members of the Paramyxovirinae subfamily (Lamb and Kolakofsky, 2001; Lamb and Parks, 2007), a few distinctive properties, including the much larger size of their genome and their broad host range, led to the creation of the Henipavirus genus to accommodate these newly emerged zoonotic viruses (Wang et al., 2000; Eaton et al., 2007).

Mononegavirales are enveloped viruses. Their envelope, which is composed of a lipid bilayer derived from the plasma membrane of the host cell, contains the attachment (H) and fusion (F) glycoproteins. Beneath the envelope, the viral matrix protein associates with the cytoplasmic tails of the H and F proteins, as well as with the viral core particle or nucleocapsid. The genome of Mononegavirales consists of

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

a nonsegmented, single-stranded RNA molecule of negative polarity (e.g., whose sequence is complementary to that of the encoded open reading frames (ORFs)). The genome of Paramyxoviridae encodes six proteins, namely, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the F and H glycoproteins, and the polymerase "large" L protein. With 18,234 (HeV) or 18,246 (NiV) nucleotides, the genome of henipaviruses is larger than that of MeV, which is 15,892 nucleotides in length. The extra length of the Henipavirus genome mainly arises from additional, unique long untranslated sequences at the 3′ end of the N, P, M, F, and G genes.

As in all Mononegavirales members, the genome is encapsidated by N to form a helical nucleocapsid. In all Paramyxovirinae members, the genome length is divisible by six, this property being related to the fact that each N monomer in the viral nucleocapsid binds to six nucleotides (Lamb and Parks, 2007). This is also true for HeV and NiV despite their much larger genome sizes. The viral RNA is tightly bound within the nucleocapsid and does not dissociate during RNA synthesis, as well illustrated by the resistance of the MeV genome to silencing by small interfering RNA (Bitko and Barik, 2001). Hence, this N–RNA complex, rather than naked RNA, is the template for both transcription and replication. These latter activities are carried out by the RNA-dependent RNA polymerase that is composed of the large (L) protein and the phosphoprotein (P). P is an essential polymerase cofactor in that it tethers the L protein onto the nucleocapsid template. This ribonucleoprotein complex made of RNA, N, P, and L constitutes the basic elements of the viral transcriptase and replicase (i.e., the viral replicative unit) (Fig. 3.1a). Although minigenome replicon studies showed that Henipavirus N, P, and L proteins are necessary and sufficient to sustain replication of viral RNA (Halpin et al., 2004), only few functional studies that focused on the replicative machinery of henipaviruses have been published so far. Hence, in this chapter, and unless differently specified, the functions of the Henipavirus N, P, and L proteins are taken to be similar to those of their counterparts from other Paramyxoviridae members based on analogy.

Once the viral ribonucleoprotein complexes are released into the cytoplasm of infected cells, transcription of viral genes occurs using endogenous NTPs as substrate. Following primary transcription, the polymerase switches to a processive mode and ignores the gene junctions to synthesize a full, complementary strand of genome length. This positive-stranded RNA (antigenome) does not serve as a template for transcription, and its unique role is to provide an intermediate in genome replication. In Mononegavirales, the intracellular concentration of the N protein is thought to be the main element controlling the relative level of transcription versus replication. When N is limiting, the polymerase functions preferentially as a transcriptase, thus leading to an increase in the intracellular concentration of viral proteins, including N. When N levels are high enough to allow encapsidation of the nascent RNA chain, the polymerase functions preferentially as a replicase (Plumet et al., 2005) (see Longhi and Canard (1999), Lamb and Kolakofsky (2001), Albertini et al., (2005), Roux (2005), and Lamb and Parks (2007) for reviews on transcription and replication). Studies on Sendai virus (SeV, a paramyxovirus)

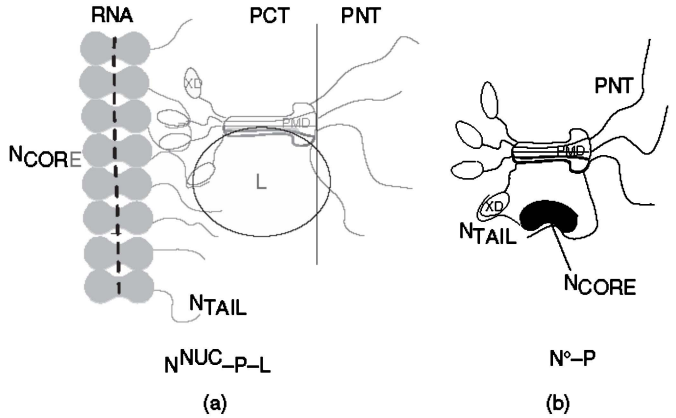


Figure 3.1 Schematic representation of the (a) N^{NUC}-P-L and (b) N^0–P complexes of Paramyxovirinae members. The disordered N_{TAIL} and PNT regions are represented by lines. The encapsidated RNA is shown as a dotted line embedded in the middle of N by analogy with RV, VSV, and RSV N–RNA complexes (Albertini et al., 2006; Green et al., 2006; Tawar et al., 2009). The multimerization domain of P (PMD) is represented with a dumbbell shape according to Tarbouriech et al. (2000b). P is depicted as a tetramer by analogy with SeV P (Tarbouriech et al., 2000b). The polymerase complex is formed by L and a tetramer of P. The tetrameric P is shown bound to N^{NUC} through three of its four C-terminal XD "arms," as in the model of Curran and Kolakofsky (Curran and Kolakofsky, 1999). The segment connecting PMD and XD is represented as disordered by analogy to MeV (Karlin et al., 2003; Longhi et al., 2003). The L protein is shown as an oval contacting P through both PMD, by analogy with SeV (Smallwood et al., 1994), and PNT, by analogy with rinderpest virus (Sweetman et al., 2001). *Source*: Modified from Bourhis et al. (2005, 2006).

(Horikami et al., 1992) and on the vesicular stomatitis virus (VSV, a rhabdovirus) (Qanungo et al., 2004) have shown that the polymerase in replication mode consists of an L-P-N complex, whereas in transcription mode, it is a complex of L, P, and cellular proteins.

By virtue of their role in encapsidating the genome, the Ns from MeV, NiV, and HeV are the most abundant viral structural proteins. They all consist of a globular N-terminal domain referred to as N_{CORE} (roughly spanning the first 400 residues) and of a flexible, C-terminal domain referred to as N_{TAIL} (aa 401–525 for MeV, and aa 400–532 for both NiV and HeV). As we will see, the combination of these two domains supports a dynamic range of N protein functions that go far beyond those of a static structural component of the viral core particle. Within MeV-infected cells, N is found in a soluble, monomeric form (referred to as N^0) and in a nucleocapsid assembled form (referred to as N^{NUC}). The soluble form of MeV N is localized in both the cytosol and the nucleus (Gombart et al., 1993; Horikami and Moyer, 1995). Sato and coworkers have recently identified the determinants of the cytoplasmic to nuclear trafficking within the N sequence of MeV and canine distemper virus (CDV), a closely related Morbillivirus. They both possess a novel

nuclear localization signal (NLS) at positions 70–77 and a nuclear export signal (NES). The NLS has a novel leucine/isoleucine-rich motif (TGILISIL), whereas the NES is composed of a leucine-rich motif (LLRSLTLF). While in CDV the NES occurs at positions 4–11, in MeV it is located in the C-terminus (Sato et al., 2006). In both viruses, the nuclear export of N is CRM1-independent. At present, the intranuclear function(s) of MeV N protein is unknown.

Following the synthesis of the N protein, a chaperone is required to maintain this protein in a soluble and monomeric form in the cytoplasm. This role is played by the P protein, whose association simultaneously prevents illegitimate self-assembly of N and retains N in the cytoplasm (Huber et al., 1991; Spehner et al., 1997). This soluble N^0–P complex (Fig. 3.1b) is used as the substrate for the encapsidation of the nascent genomic RNA chain during replication. The assembled form of N also forms complexes with P, either isolated (N^{NUC}–P) or bound to L (N^{NUC}-P-L), which are essential to RNA synthesis by the viral polymerase (Ryan and Portner, 1990; Buchholz et al., 1994).

The viral polymerase, which is responsible for both transcription and replication, is poorly characterized. It is thought to carry out most (if not all) enzymatic activities required for transcription and replication, including nucleotide polymerization, mRNA capping, and polyadenylation. So far, no functional Paramyxoviridae polymerase has been purified to homogeneity, with the only exception of the rinderpest virus L–P complex, which has been partially purified (Gopinath and Shaila, 2008). Consequently, most of our present knowledge arises from bioinformatics studies. Notably, using bioinformatics approaches, we identified a ribose-2′-*O*-methyltransferase domain possibly involved in capping of viral mRNAs, within the C-terminal region of Mononegavirales polymerases (with the exception of Bornaviridae and nucleorhabdoviruses) (Ferron et al., 2002). Consistent with this prediction, the methyltransferase activity has been demonstrated biochemically within the C-terminal region of the closely related SeV polymerase (aa 1756–2228) (Ogino et al., 2005). Interestingly, both ribose-2′-*O* and guanosine-N-7 methyltransferase acivities have been mapped to a conserved C-terminal motif of VSV L (Rahmeh et al., 2009) using a purified recombinant form of the protein (Li et al., 2008).

For all Mononegavirales members, the viral genomic RNA is always encapsidated by the N protein, and genomic replication does not occur in the absence of N^0 and without concurrent encapsidation of the nascent genomic RNA chain. Therefore, during RNA synthesis, the viral polymerase has to interact with the N–RNA complex and use the N^0–P complex as substrate for encapsidation of nascent genomic RNA. Hence, the components of the viral replication machinery, namely, P, N, and L proteins, engage in a complex macromolecular ballet.

Although the understanding of the precise role(s) of N, P, and L within the replicative complex of MeV has benefitted significant breakthroughs in recent years (see Bourhis et al., (2006), Bourhis and Longhi (2007), Longhi (2009), and Longhi and Oglesbee (2010) for reviews), rather limited three-dimensional information on the replicative machinery is available. In the case of henipaviruses, even less structural data are available, with only one paper focused on the structural characterization of their N and P proteins having been published so far (Habchi et al.,

2010). The scarcity of high resolution structural data for Paramyxovirinae members stems from several facts: (i) the difficulty of obtaining homogenous polymers of N suitable for X-ray analysis (Schoehn et al., 2001; Karlin et al., 2002); (ii) the low abundance of L in virions and its very large size that is a challenge to heterologous expression; and (iii) the structural flexibility of N and P (Karlin et al., 2002, 2003; Longhi et al., 2003; Bourhis et al., 2004, 2005, 2006; Longhi, 2009; Habchi et al., 2010; Longhi and Oglesbee, 2010).

Indeed, in the course of the structural and functional characterization of MeV, NiV, and HeV replicative complex proteins, we discovered that the N and P proteins contain disordered regions of up to 400 residues in length that possess the sequence and biochemical features that typify intrinsically disordered proteins (IDPs) (Karlin et al., 2002, 2003; Longhi et al., 2003; Bourhis et al., 2004, 2005, 2006; Longhi, 2009; Habchi et al., 2010). IDPs are functional proteins that lack highly populated secondary and tertiary structures under physiological conditions of pH and salinity in the absence of a partner, and they exist as dynamic ensembles of conformers (for a review see Radivojac et al. (2007)).

Using bioinformatics approaches (as described in Ferron et al. (2006) and Bourhis et al. (2007), we further extended these results to the N and P proteins of viruses of the Paramyxovirinae subfamily (Karlin et al., 2003). By combining computational and experimental approaches (as described in Receveur-Bréchot et al. (2006)), we showed that large disordered regions also occur within the P protein of rabies virus (RV, a Rhabdoviridae member) (Gerard et al., 2009) and of respiratory syncytial virus (RSV, a Pneumovirinae member within the Paramyxoviridae family) (Llorente et al., 2006). Altogether, these data pointed out that structural disorder is a conserved and widespread property within these two viral families, thus implying functional relevance.

3.2 STRUCTURAL ORGANIZATION OF THE PHOSPHOPROTEIN

Like the N, the P protein provides several functions in transcription and replication. Beyond serving as a chaperone for N, P binds to the nucleocapsid, thus tethering the polymerase onto the nucleocapsid template. The actual oligomeric state of P of MeV and Henipavirus is unknown. However, by analogy with the closely related SeV (Tarbouriech et al., 2000a, 2000b), it is thought to be tetrameric (Fig. 3.1).

The P genes of MeV and of henipaviruses encode multiple proteins, including P, V, and C (for reviews see Longhi and Canard (1999), Lamb and Kolakofsky (2001), Eaton et al. (2007), and Lamb and Parks (2007)). While the C protein is encoded by an alternate ORF within the P gene through ribosome initiation at an alternative translation codon, the V protein is translated from a P messenger obtained on cotranscriptional insertion of a G at the editing site of the P mRNA. The V proteins thus share with the P proteins the N-terminal module (MeV PNT (P N-terminal), aa 1–230; NiV PNT, aa 1–406; and HeV PNT, aa 1–404) and possess a unique C-terminal, zinc-binding domain (ZnBD). The organization of the P gene suggests that P is a modular protein, consisting of at least two domains: an N-terminal domain

(PNT) common to both P and V and a C-terminal domain (PCT) unique to the P protein (Figs. 3.2a and 3.3a). Transcription requires only the PCT domain, whereas genome replication requires PNT also. Within Paramyxovirinae, PNT plays the role of a chaperone for newly synthesized N (N^0), and it is this interaction that leads to the formation of the encapsidation complex (N^0–P) that is used as a substrate by the polymerase during RNA replication (for review Longhi and Canard (1999) Lamb and Kolakofsky (2001), and Lamb and Parks (2007)). Within the MeV N^0–P complex, P to N binding is mediated by the dual PNT–N_{CORE} and PCT–N_{TAIL} interaction (Chen et al., 2003) (Fig. 3.1b). It has recently been suggested that the MeV N^0–P complex has a very short half-life (Plumet et al., 2005).

3.2.1 The Intrinsically Disordered PNT Domains of Measles, Nipah, and Hendra Viruses

After purification from the soluble fraction of *Escherichia coli*, the PNT domains from MeV, NiV, and HeV display an abnormally slow migration in SDS-PAGE, with an apparent molecular mass of 32 kDa for MeV PNT (expected mass 25 kDa) (Karlin et al., 2002) and 60 kDa for both NiV and HeV PNT domains (expected mass 45 kDa) (Habchi et al., 2010). Notably, in all cases, mass spectrometry analyses confirmed that the recombinant products possess the expected molecular mass. This anomalous behavior, which constitutes a hallmark of structural disorder, is related to a rather high content in acidic residues. This characteristic has been previously described in the literature for other Paramyxovirinae P proteins (Lamb and Kolakofsky, 2001; Lamb and Parks, 2007) and, more generally, in other IDPs (Tompa, 2002).

The hydrodynamic properties of PNT domains inferred from size exclusion chromatography (SEC) are consistent with these protein domains possessing extended conformations in solution (Table 3.1). Indeed, they display rather large values of hydrodynamic radius (Stokes radius, R_S) (41 Å and 44 Å for MeV and Henipavirus PNT, respectively) compared to the theoretical values of R_S expected for globular proteins having the same molecular mass as that of either Henipavirus PNT (29 Å) or MeV PNT (23 Å) (Table 3.1). In addition, the theoretical expected R_S for fully unfolded proteins of the same molecular mass as that of either MeV or Henipavirus PNT domains are 46 Å and 60 Å, respectively (Table 3.1). Thus, the experimentally observed R_S are not compatible with the values expected for globular proteins, but rather with those expected for (at least partly) unfolded polypeptide chains.

The absence of a globular core has been further demonstrated by limited proteolysis experiments, which showed that PNT domains are fully exposed to the solvent (Karlin et al., 2002; Habchi et al., 2010). In addition, the very negative ellipticity at 200 nm observed in the far-UV CD (circular dichroism) spectra of MeV (Fig. 3.2b) and henipaviruses PNT (Fig. 3.3b), together with the small spread of the resonance frequencies (between 7.8 and 8.7 ppm) of the Nuclear Overhauser Effect Spectroscopy (NOESY) spectra (Fig. 3.3c and data not shown), support the absence of stable secondary structures (Karlin et al., 2002; Habchi et al., 2010).

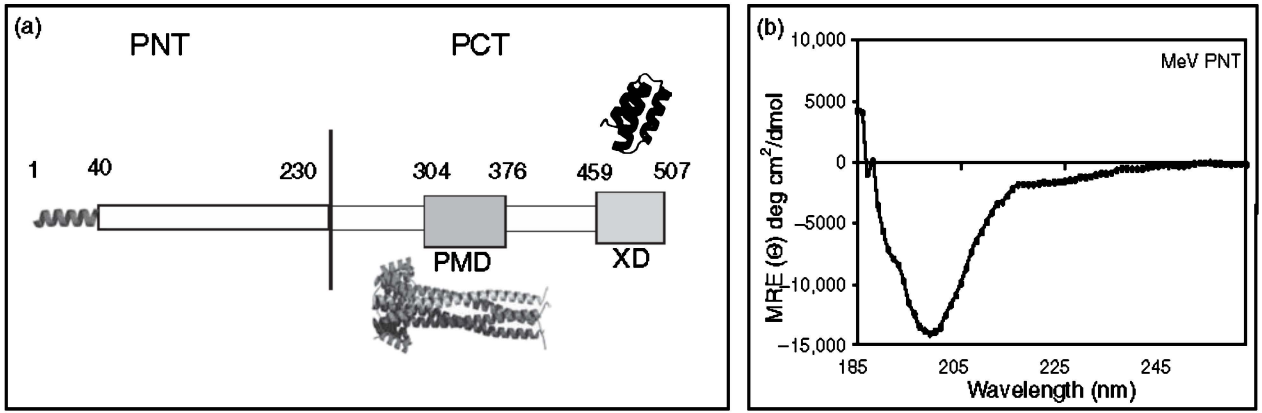


Figure 3.2 (a) Schematic representation of the modular organization of MeV P, where globular and disordered regions are represented by large and narrow boxes, respectively. The vertical line separating PNT and PCT is located at the border between the region shared by P and V and the region unique to P (see text). The hydrophobic region with α-helical folding potential at the N-terminus of P is highlighted. The crystal structures of MeV XD (PDB code 1OKS) (Johansson et al., 2003) and SeV PMD (PDB code 1EZJ) (Tarbouriech et al., 2000b) are also shown. Structures were drawn using Pymol (DeLano, 2002). (b) Far-UV CD spectrum of MeV PNT (0.1 mg/mL) in 10 mM sodium phosphate buffer pH 7 at 20°C. (c) MeDor ouput of MeV PNT (accession number CAA91364). The sequence is represented as a single, continuous horizontal line below the predicted secondary structure elements. Below the sequence is shown the hierarchical cluster analysis (HCA) plot and the predicted regions of disorder that are represented by bidirectional arrows. The region highlighted corresponds to a putative α-MoRE possibly involved in N^0 binding. MRE, mean residue ellipticity. *Source*: Modified from Longhi and Oglesbee (2010).

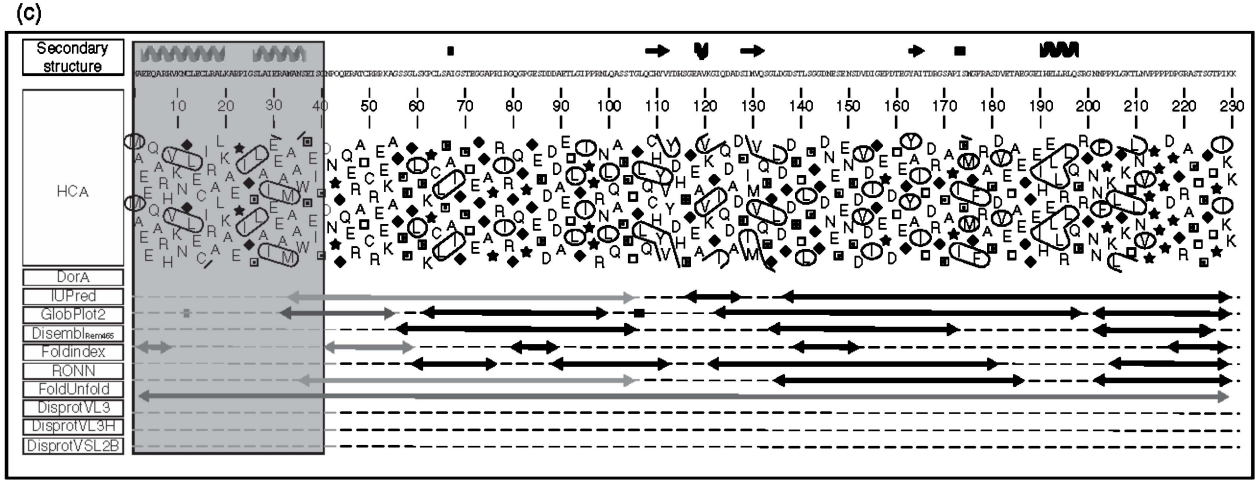


Figure 3.2 (*Continued*)

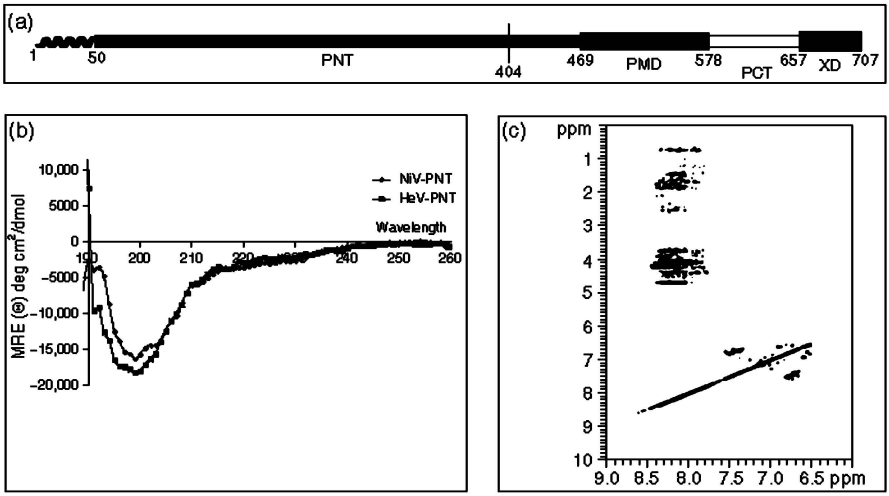


Figure 3.3 (a) Schematic representation of the modular organization of HeV P, where globular and disordered regions are represented by large and narrow boxes, respectively. The vertical line separating PNT and PCT is located at the border between the region shared by P and V and the region unique to P. The hydrophobic α-helical region at the N-terminus is highlighted. (b) Far-UV circular dichroism spectra of Henipavirus PNT (0.1 mg/mL) in 10 mM sodium phosphate buffer pH 7 at 20°C. (c) Two-dimensional ^{1}H NMR NOESY spectrum of NiV PNT (0.1 mM) in 10 mM sodium phosphate pH 7, 150 mM NaCl, and 10% D_2O recorded at 100K. ppm: values for resonance shifts in parts per million of the spectrophotometer frequency. The frame shows the small spread of the resonance frequencies for amide protons. MRE, mean residue ellipticity.

TABLE 3.1 Hydrodynamic Properties of MeV and Henipavirus PNT

	MM^{app} (kDa)	MM^{theo} (kDa)	R_S^{SEC} (Å)	R_S^{DLS} (Å)	R_S^{DLS} + UREA (Å)	R_S^{NF} (Å)	R_S^{PMG} (Å)	R_S^{U} (Å)
NiV	137	45	44 ± 2	44 ± 3	55 ± 2	29	43	60
HeV	138	45	44 ± 2	50 ± 3	57 ± 2	29	43	60
MeV	115	24	41 ± 3	47 ± 4	ND	23	34	46

[a]The Stokes radii of MeV and Henipavirus PNT domains as obtained by either SEC (R_S^{SEC}) or DLS (R_S^{DLS}) analyses in the presence or absence of urea are shown. The average values as obtained from three independent measurements are shown. The apparent molecular masses either inferred from the SEC calibration column (MM^{app}) or expected based on the amino acid sequence (MM^{theo}) are shown. The values of the Stokes radii expected for the various conformational states are indicated. Abbreviations: NF, natively unfolded; PMG, premolten globule; U, fully unfolded; ND, not determined. *Source*: Data were taken from Habchi et al. (2010) and Karlin et al. (2002).

Moreover, PNT domains are consistently predicted to be disordered by several disorder predictors (Karlin et al., 2002; Habchi et al., 2010), including the method of the hydrophobicity/mean charge ratio (Uversky et al., 2000), as well as most disorder predictors implemented within the MeDor metaserver for the prediction of disorder (Lieutaud et al., 2008) (Fig. 3.2c and data not shown). Using computational approaches, we have extended these results to the W protein of SeV (the PNT counterpart), as well as to other PNT domains from other Morbillivirus members (Karlin et al., 2002). The consistent disorder prediction for PNT domains indicates that the lack of stable secondary structure does not arise from a purification artifact, being rather an intrinsic property encoded in their primary structure.

Interestingly, a short (40–50 residues) ordered region is consistently predicted at the N-terminus of P of MeV, NiV, and HeV by all the predictors implemented in MeDor (Fig. 3.2c and data not shown). This N-terminal module with α-helical folding propensities corresponds to a conserved region amongst Avulavirus, Henipavirus, and Rubulavirus members (Karlin et al., 2003), with that from rubulaviruses having been shown to be involved in N^0-binding (Watanabe et al., 1996). Therefore, this region likely corresponds to an α-helical Molecular Recognition Element (α-MoRE), where MoREs are short, order-prone regions within IDPs, which have a certain propensity to bind to a partner and thereby to undergo induced folding (i.e., a disorder-to-order transition) (Garner et al., 1999; Oldfield et al., 2005; Mohan et al., 2006; Vacic et al., 2007).

Consistent with the predicted occurrence of short regions with some folding potential, hydrodynamic studies showed that Henipavirus PNT domains are not fully extended in solution and rather adopt a premolten globule (PMG) conformation. PMGs are characterized by a conformational state intermediate between a random coil (RC) and a molten globule and possess a certain degree of residual compactness because of the presence of residual and fluctuating secondary and/or tertiary structures (Dunker et al., 2001; Uversky, 2002). As shown in Table 3.1, the measured hydrodynamic radii of Henipavirus PNT, as inferred from both SEC and dynamic light scattering (DLS) studies, are close to the values expected for native PMG conformations (Table 3.1). Unexpectedly (i.e., contrary to predictions), the measured hydrodynamic radius of MeV PNT is close to the value expected for a fully unfolded form (Table 3.1). In further support of the occurrence of some residual structure within Henipavirus PNT, DLS studies pointed out a significant increase in the R_S of the PNT domains on addition of urea (Table 3.1). Furthermore, the HeV PNT domain has a notable behavior in that its R_S, as measured by DLS, is slightly, but significantly, larger than that inferred from SEC studies (Table 3.1). This observation is consistent with HeV PNT, adopting a slightly more extended conformation with respect to NiV PNT. The more extended conformation of MeV and HeV PNT domains with respect to NiV PNT is also revealed by their far-UV CD parameters. Indeed, Uversky noticed that IDPs can be subdivided into PMG-like and RC-like as a function of their ellipticity values at 200 and 222 nm (Uversky, 2002). Strikingly, both MeV and HeV PNT domains are closer to the random coil-like region of the plot than NiV PNT (Fig. 3.4).

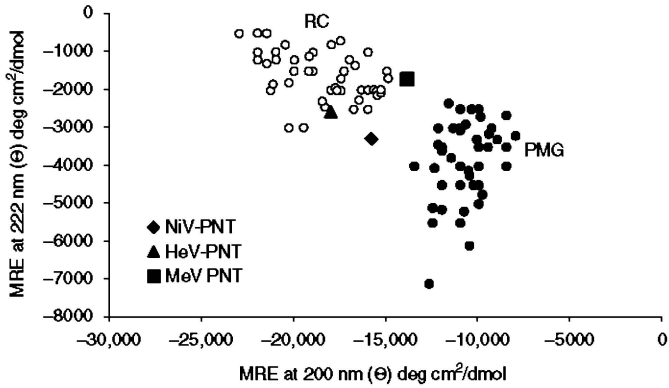


Figure 3.4 A 222-200 nm ellipticity plot. The mean residue ellipticity (MRE) values at 222 nm of a set of well-characterized unfolded or premolten globule proteins (Uversky, 2002) have been plotted against the MRE values at 200 nm. The position in the plot of MeV and Henipavirus PNT is highlighted. RC, random coil; PMG, premolten globule. *Source*: Modified from Habchi et al. (2010).

Thus the MeV and henipaviruses PNT domains are not fully unfolded, but they conserve some transiently populated secondary structure content typical of the PMG subfamily of IDPs, with the extent of residual compactness following the order NiV PNT > HeV PNT > MeV PNT. The residual ordered secondary structure present within NiV PNT, and to a lower extent within HeV PNT, likely arises from a transiently populated α-MoRE. That the conformational space sampled by these interaction-prone short segments within intrinsically disordered regions can be restricted even in the absence of the partner has already been reported (Fuxreiter et al., 2004). It has been proposed that the residual intramolecular interactions that typify the PMG state may enable a more efficient start of the folding process induced by a partner by lowering the entropic cost of the folding-coupled-to-binding process (Tompa, 2002; Fuxreiter et al., 2004; Lacy et al., 2004).

What are the functional implications of the intrinsic disorder of PNT domains? PNT domains are reminiscent of transcriptional acidic activator domains (AADs). AADs play a role in recruiting the transcriptional machinery via protein–protein interactions, and their function does not rely on a precise tertiary structure. Recently, several studies have shown that AADs act through bulky hydrophobic residues scattered within acidic residues. The numerous, charged residues would help to keep these hydrophobic residues in an aqueous environment, allowing them to establish weak, short-distance contacts with hydrophobic patches in their targets. According to this model, specificity of AADs for their physiological partners is determined by other factors dictating their colocalization with those partners (on DNA), and affinity is ensured by the intervention of multiple activation domains that strengthen the interaction (Melcher, 2000). AADs undergo induced folding in the presence of their physiological partner (Melcher, 2000). In the case of SeV, PNT was shown to be required for the synthesis of genomic RNA (Curran et al., 1995).

This activity is related to the involvement of PNT domain in the formation of the N^0–P complex that is the substrate for the encapsidation of the nascent RNA chain, but it could also be attributed to a weak PNT–L interaction, as shown in the case of rinderpest virus (Sweetman et al., 2001). In contrast, a stable L–P interaction site has been mapped within the PCT domain (Smallwood et al., 1994). Given the similarity of AADs to PNT, the colocalization of PNT and L on the N–RNA complex would be ensured by the presence of a stable, independent L–P interaction site. The colocalization, together with the presence of multiple PNT domains within the P tetramer, would strengthen the PNT–L interaction.

Interestingly, the incubation of MeV and henipaviruses PNT domains in the presence of increasing concentrations of 2,2,2-trifluoroethanol (TFE) induces a pronounced gain of α-helicity (Karlin et al., 2002; Habchi et al., 2010, p. 1536). TFE is an organic solvent that mimics the hydrophobic environment experienced by proteins during protein–protein interactions and is thus widely used as an empirical probe of hidden structural propensities of proteins and to unveil regions that have a propensity to undergo induced folding (Dahlman-Wright and McEwan, 1996; Hua et al., 1998). In the case of MeV PNT, limited proteolysis experiments in the presence of TFE led to the identification of a thermolysin-resistant fragment. This fragment, spanning residues 27–99 for MeV PNT, contains a protein region (aa 27–38) with a strong propensity to fold as an α-helix. This α-helix may well represent an α-MoRE involved in a disorder-to-order transition of PNT domain on binding to a partner (Karlin et al., 2002).

Does a PNT domain actually undergo induced folding on interaction with its physiological partner(s)? It is conceivable that the N-terminus of P folds on binding to N^0. This gain of structure may favor recognition of the N^0–P complex by the polymerase and the proper positioning of N monomers on the nascent RNA chain. However, direct answers to this question await the availability of the two potential physiological partners of PNT, namely, N^0 and the L protein.

3.2.2 The Partly Disordered PCT Domain

Beyond PNT domain, other disordered regions have been identified within the P of MeV and Henipavirus. Indeed, the PCT domain has a modular organization, being composed of alternating disordered and structured regions (Karlin et al., 2003, Habchi, 2010 p. 1536) (Figs. 3.2a and 3.3a). Within MeV PCT, the region spanning residues 304–376 of P (referred to as PMD for *P*multimerization *d*omain) is responsible for the oligomerization of P (Chen et al., 2005), while the XD is responsible for binding to both N_{TAIL} (Johansson et al., 2003) and the cellular ubiquitin E3 ligase Pirh2 (Chen et al., 2005). In morbilliviruses and henipaviruses, sequence analysis predicts a coiled-coil region within the PMD. The coiled-coil organization has been experimentally confirmed in the case of PMDs of SeV (Tarbouriech et al., 2000b) and rinderpest virus (Rahaman et al., 2004). Although no experimental data are available on the oligomeric state of Henipavirus P, the protein is thought to oligomerize via PMD by analogy to Morbillivirus members.

We have previously reported the crystal structure of MeV XD and shown that it consists of a triple α-helical bundle (Johansson et al., 2003) (Fig. 3.2a). High resolution structural data are also available for the XDs of the closely related MuV and SeV, the structure of which have been solved by X-ray crystallography and nuclear magnetic resonance (NMR), respectively (Blanchard et al., 2004; Kingston et al., 2008). The MuV XD (aa 343–391 of P) has a few notable distinguishing properties with respect to MeV and SeV XDs. Indeed, MuV XD has been shown to exist as a molten globule in solution, being loosely packed and devoid of a stable tertiary structure (Kingston et al., 2004, 2008). In addition, contrary to the MeV and SeV XDs, MuV XD does not interact with N_{TAIL} and rather establishes contacts with the structured N_{CORE} region of N (Kingston et al., 2004). Structural data are also available for the C-terminal domain of Rhabdoviridae members, namely, RV (Mavrakis et al., 2004), VSV (Ribeiro et al., 2008), and Mokola virus (Assenberg et al., 2010). Interestingly, comparison of the P nucleocapsid-binding domains solved so far suggests that the N–RNA binding domains are structurally conserved among Paramyxoviridae and Rhabdoviridae P in spite of low sequence conservation (Delmas et al., 2010). They, therefore, provide a fascinating example of convergent evolution.

In all Paramyxovirinae members, PMD and XD are separated by a flexible linker region predicted to be poorly ordered (Karlin et al., 2003). Indeed, in the case of SeV, NMR studies carried out on the 474–568 region of P (referred to as PX) showed that the region upstream XD (aa 474–515 of P) is disordered (Bernado et al., 2005; Houben et al., 2007a). The flexibility and solvent exposure of this linker region has been also experimentally determined in the case of MeV, where recombinant PCT was shown to undergo spontaneous cleavage at position 436 (Longhi et al., 2003). Finally, in morbilliviruses, an additional flexible region (referred to as "spacer") occurs upstream PMD (Karlin et al., 2003).

The disordered nature of PNT domains and of the "spacer" region connecting PNT to PMD likely reflects a way of alleviating evolutionary constraints within overlapping ORFs, in agreement with previous reports that pointed out a relationship between overlapping genes and structural disorder (Jordan et al., 2000; Kovacs et al., 2010; Narechania et al., 2005; Rancurel et al., 2009). Indeed, PNT partially overlaps the C protein (being encoded by the same RNA region), and the spacer region partially overlaps the ORF encoding the ZnBD of the V protein (Fig. 3.2a) (Lamb and Kolakofsky, 2001; Lamb and Parks, 2007). Disorder, which is encoded by a much wider portion of sequence space as compared to order, can indeed represent a strategy by which genes encoding overlapping reading frames can lessen evolutionary constraints imposed on their sequence by the overlap, allowing the encoded overlapping protein products to sample a wider sequence space without losing function.

In the same vein, by comparing the modular organization of the P proteins within the Paramyxovirinae subfamily, we noticed that a larger PNT domain in henipaviruses accounts for the extra length of their P protein (Figs. 3.2a and 3.3a). This finding is consistent with the higher tolerance of disordered regions to insertions or major rearrangements as compared to ordered ones.

3.3 STRUCTURAL ORGANIZATION OF THE NUCLEOPROTEIN

Bioinformatics, deletion, and electron microscopy studies have shown that Paramyxoviridae Ns are divided into two regions: a structured N-terminal moiety, N_{CORE} (aa 1–400 in MeV and aa 1–399 in henipaviruses), which contains all the regions necessary for self-assembly and RNA binding (Buchholz et al., 1993; Curran et al., 1993; Bankamp et al., 1996; Liston et al., 1997; Myers et al., 1997; Myers et al., 1999; Karlin et al., 2002; Kingston et al., 2004), and a C-terminal domain, N_{TAIL} (aa 401–525 in MeV and aa 400–532 in Henipavirus), which is intrinsically disordered (Longhi et al., 2003; Habchi et al., 2010) (Figs. 3.5a and 3.6a). N_{TAIL} protrudes from the globular body of N_{CORE} and is exposed at the surface of the viral nucleocapsid (Heggeness et al., 1980, 1981; Karlin et al.,

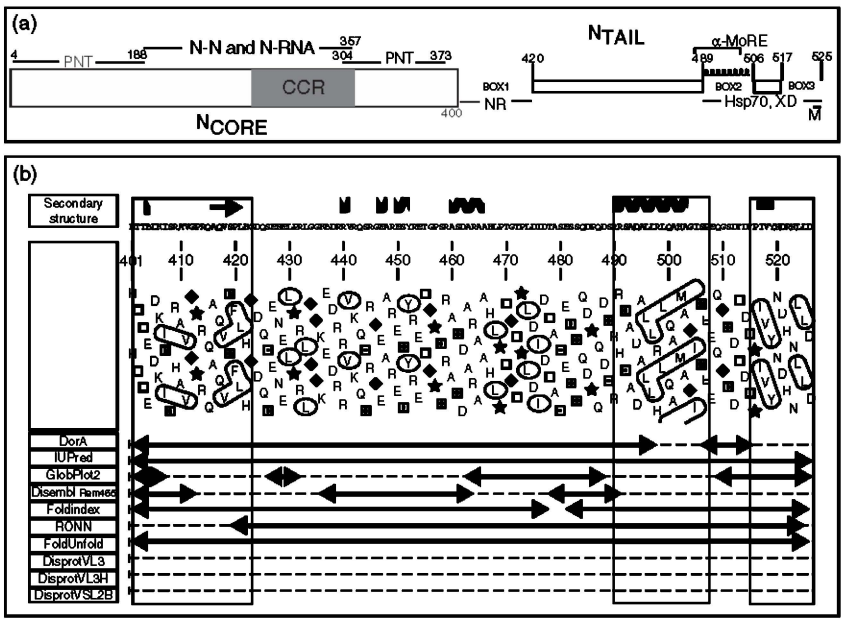


Figure 3.5 (a) Schematic representation of the modular organization of MeV N, where the disordered N_{TAIL} domain is represented by a narrow box. The location of N-N, N-P, and RNA-binding sites is indicated. The central conserved region (CCR, aa 258–357) involved in oligomerization and in RNA binding is depicted in dark gray. The three N_{TAIL} Boxes conserved among Morbillivirus members (Diallo et al., 1994) are shown, as well as the location of the binding sites of NR, hsp70, XD, and M. The experimentally characterized α-MoRE involved in binding to XD is shown (Kingston et al., 2004). (b) MeDor output of MeV N_{TAIL} (accession number P04851). The first and the last regions highlighted in light gray correspond to a putative α-MoRE involved in binding to NR (Laine et al., 2005) and to an I-MoRE contributing to the interaction with XD (Bourhis et al., 2005) (see text). The second region highlighted in light gray corresponds to the experimentally characterized α-MoRE (Kingston et al., 2004). Conventions are the same as in Fig. 3.2c.

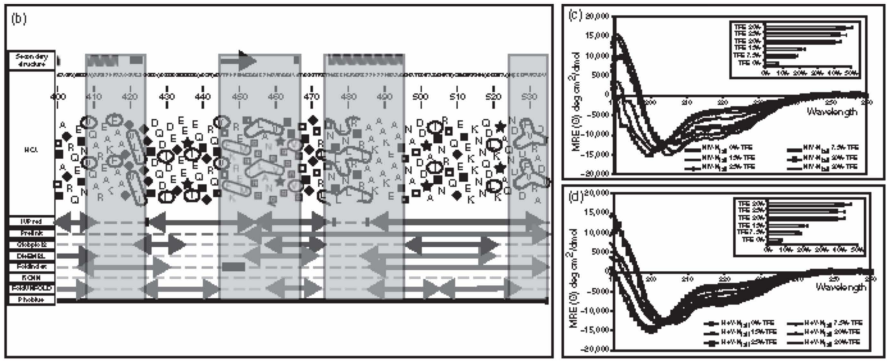


Figure 3.6 (a) Schematic representation of the modular organization of Henipavirus N. The disordered N_{TAIL} domain is represented by a narrow box. The two putative α-MoREs (aa 408–422 and aa 473–493), as well as a putative irregular MoRE (I-MoRE) (aa 523–532) and a putative MoRE of dubious state (aa 444–464, see region contoured by a dashed line) are shown. (b) MeDor output of NiV N_{TAIL} where the four putative MoREs are highlighted in dark gray. Conventions are the same as in Fig. 3.2c. Far-UV CD spectra of (c) NiV and (d) HeV N_{TAIL}. The spectra were recorded in 10 mM sodium phosphate pH 7 at 20°C in the presence of increasing concentrations of TFE (0%, 7.5%, 15%, 20%, and 30%). Protein concentrations ranged from 0.06 mg/mL (30% TFE) to 0.1 mg/mL (0% TFE). The insets show the α-helical content at the various TFE concentrations as estimated by CDSSTR. *Source*: Modified from Habchi et al. (2010).

2002). N_{TAIL} contains the regions responsible for binding to P in both N^0–P and N^{NUC}–P complexes (Fig. 3.1) (Bankamp et al., 1996; Liston et al., 1997; Longhi et al., 2003; Kingston et al., 2004).

When expressed in heterologous systems, MeV N and NiV N self-assemble to form large helical nucleocapsidlike particles with a broad size distribution (Spehner et al., 1991; Warnes et al., 1995; Bhella et al., 2002; Tan et al., 2004; Kerdiles et al., 2006). MeV nucleocapsids, as visualized by negative stain transmission electron microscopy, have a typical herringbonelike appearance (Bhella et al., 2002, 2004; Karlin et al., 2002; Longhi et al., 2003; Schoehn et al., 2004). The nucleocapsid of all Paramyxoviridae members has a considerable conformational flexibility and can adopt different helical pitches (the axial rise per turn) and twists (the number of subunits per turn) resulting in conformations differing in their extent of compactness (Heggeness et al., 1980, 1981; Egelman et al., 1989; Bhella et al., 2002, 2004; Schoehn et al., 2004). Because of this property, Paramyxoviridae Ns are poorly amenable to high resolution structural characterizations. Because of variable helical parameters, the recombinant or viral nucleocapsids are also difficult to analyze using electron microscopy coupled to image analysis. Despite these technical drawbacks, elegant electron microscopy studies by two independent groups led to real-space helical reconstruction of MeV nucleocapsids (Bhella et al., 2004; Schoehn et al.,

2004). These studies showed that the most extended nucleocapsid conformation has a helical pitch of 66 Å, while twist varies from 13.04 to 13.44 (Bhella et al., 2004). Notably, these studies also highlighted a cross talk between N_{CORE} and N_{TAIL}, which was judged based on the observation that removal of the disordered N_{TAIL} domain leads to increased nucleocapsid rigidity, with significant changes in both pitch and twist (see Longhi et al., 2003; Bhella et al., 2004; Schoehn et al., 2004). Distinct nucleocapsid morphologies have been associated with either high or low levels of viral transcriptional activity, making variations in N_{TAIL}–N_{CORE} interaction a possible determinant of viral gene expression (Robbins et al., 1980).

Conversely, high resolution structural data are available for two Rhabdoviridae members, namely, RV and VSV (Albertini et al., 2006; Green et al., 2006; Luo et al., 2007), and for RSV (Tawar et al., 2009), for which the crystal structures of N–RNA rings have been solved. The N of these viruses consists of two lobes and possesses an extended terminal arm that makes contacts with a neighboring N monomer. The RNA is tightly packed between the two N lobes and, in the case of RV and VSV, points toward the inner cavity of the N–RNA rings. Conversely, in the case of RSV, the RNA is located on the external face of the N–RNA rings (Tawar et al., 2009). Thus, in all these Ns, the RNA is not accessible to the solvent and has to be partially released from N to become accessible to the polymerase. In addition, the crystal structure of the N of the Borna disease virus has also been solved. Although it also shares a bilobal morphology with the N proteins from the above-mentioned viruses, its structure was not solved from N–RNA rings but from the N protein alone and was shown to consist of a tightly packed homotetramer centered at the fourfold crystallographic axis of the crystal (Rudolph et al., 2003).

Functional and structural similarities between the Ns of Rhabdoviridae and of Paramyxovirinae members are well established. In particular, they share the same organization in two well-defined regions, N_{CORE} and N_{TAIL}, and in both families a central conserved region (CCR) is involved in RNA binding and self-assembly of Ns (Kouznetzoff et al., 1998; Myers et al., 1999). Incubation of the RV nucleocapsid with trypsin results in the removal of the C-terminal region (aa 377–450) (Kouznetzoff et al., 1998). This N_{TAIL}-free nucleocapsid is no longer able to bind to P, thus suggesting that in the family Rhabdoviridae, N_{TAIL} plays a role in the recruitment of P as in the case of Paramyxoviridae (Schoehn et al., 2001). However, in contrast to NiV, HeV, and MeV, the RV N_{TAIL} domain is structured, as is that of VSV (Albertini et al., 2006; Green et al., 2006). In the case of RSV, the last 12–20 residues of N are disordered in the crystal (Tawar et al., 2009). Although VSV, RV, and RSV Ns share the same bilobed morphology, presently it is not known whether this morphology is also conserved in MeV and henipaviruses. Notably, bioinformatics analyses predict that MeV N_{CORE} is organized into two subdomains (aa 1–130 and aa 145–400) (Ferron et al., 2006; Bourhis et al., 2007) separated by a variable, antigenic loop (aa 131–149) (Giraudon et al., 1988), which might fold cooperatively into a bilobed morphology.

As indicated previously, the N region responsible for binding to P within the MeV N^{NUC}–P complex is located within N_{TAIL}, whereas the N^0–P complex

involves an additional interaction between N_{CORE} and the disordered N-terminal domain of P (PNT) (Fig. 3.1). In other words, the MeV N^0–P complex reflects dual PNT–N_{CORE} and PCT–N_{TAIL} interactions (Chen et al., 2003) (Fig. 3.1). Studies on SeV suggested that an N^0–P complex is absolutely necessary for the polymerase to initiate encapsidation (Baker and Moyer, 1988). Therefore, formation of the N^0–P complex would have at least two separate functions: (i) prevent illegitimate self-assembly of N and (ii) allow the polymerase to deliver N to the nascent RNA to initiate replication.

The regions within N_{CORE} responsible for binding to PNT within the N^0–P complex have been mapped to residues 4–188 and 304–373, with the latter region being not strictly required for binding but rather favors it (Bankamp et al., 1996) (Fig. 3.5a). However, precise mapping of such regions is hard because N_{CORE} does not have a modular structure, and consequently, it is difficult to distinguish between gross structural defects and specific effects of deletions.

3.3.1 The Intrinsically Disordered N_{TAIL} Domains of Measles, Nipah, and Hendra Viruses

In morbilliviruses, N_{TAIL} is responsible for binding to P in both N^0–P and N^{NUC}–P complexes (Bankamp et al., 1996; Liston et al., 1997; Longhi et al., 2003; Kingston et al., 2004). Within the MeV N^{NUC}–P complex, N_{TAIL} is also responsible for the interaction with the polymerase (L–P) complex (Bankamp et al., 1996; Liston et al., 1997; Longhi et al., 2003; Kingston et al., 2004) (Fig. 3.1).

N_{TAIL} domains of NiV, HeV, and MeV possess features that are hallmarks of intrinsic disorder: (i) they are hypersensitive to proteolysis (Karlin et al., 2002; Habchi et al., 2010); (ii) they cannot be visualized in cryo-electron microscopy reconstructions of nucleocapsids (Bhella et al., 2004); and (iii) they have an amino acid sequence that is highly variable amongst phylogenetically related members (see Fig. S1 in Habchi et al. (2010)).

In agreement, the sequence properties of N_{TAIL} domains conform to those of IDPs. IDPs have been shown to be enriched in charged and polar residues (R, Q, S, and E), which are generally found in solvent-exposed loops of structured proteins (and hence referred to as *disorder-promoting residues*), and to be depleted in hydrophobic, bulky residues (W, C, F, Y, I, L), which are generally buried in structured proteins (and hence referred to as *order-promoting residues*) (Dunker et al., 2001). While the amino acid composition of N_{CORE} does not deviate significantly from the average composition of proteins found in the Protein Data Bank (PDB), the N_{TAIL} regions are depleted in order promoting residues and enriched in disorder promoting residues (Longhi et al., 2003; Habchi et al., 2010). Moreover, MeV, NiV, and HeV N_{TAIL} domains are predicted to be mainly (if not fully) disordered by the secondary structure and disorder predictors implemented within the MeDor metaserver (Lieutaud et al., 2008) (Figs. 3.5b and 3.6b). Notably, however, short order-prone regions, possibly corresponding to MoREs, can be readily detected within NiV, HeV, and MeV N_{TAIL} domains (Figs. 3.5b and 3.6b).

TABLE 3.2 Hydrodynamic Properties of MeV and Henipavirus N_{TAIL}

	MM^{app} (kDa)	MM^{theo} (kDa)	R_S^{SEC} (Å)	R_S^{DLS} (Å)	R_S^{DLS} + UREA (Å)	R_S^{NF} (Å)	R_S^{PMG} (Å)	R_S^{U} (Å)
NiV	43	15	28 ± 2	28 ± 3	37 ± 2	19	28	34
HeV	43	15	28 ± 2	26 ± 1	37 ± 2	19	28	34
MeV	36	15	27 ± 3	30 ± 2	ND	19	28	34

The Stokes radii of MeV and Henipavirus N_{TAIL} domains as obtained by either SEC (R_S^{SEC}) or DLS (R_S^{DLS}) analyses in the presence or absence of urea are shown. The average values as obtained from three independent measurements are shown. The apparent molecular masses either inferred from the SEC calibration column (MM^{app}) or expected based on the amino acid sequence (MM^{theo}) are shown. The values of the Stokes radii expected for the various conformational states are indicated. Abbreviations: NF, natively folded; PMG, premolten globule; U, fully unfolded; ND, not determined. *Source*: Data were taken from Habchi et al. (2010) and from Longhi et al. (2003).

Biochemical, hydrodynamic, and spectroscopic analyses, similar to those carried out during the characterization of PNT domains, confirmed that MeV and Henipavirus N_{TAIL} domains belong to the family of IDPs (Longhi et al., 2003; Habchi et al., 2010). Although these N_{TAIL} domains are mostly unfolded in solution, they have been shown to retain a certain degree of compactness based on their R_S (Table 3.2) and ellipticity values at 200 and 222 nm (Longhi et al., 2003; Habchi et al., 2010). Furthermore, DLS studies in the presence of urea showed an increase in the R_S for all these domains, thereby confirming their native PMG state (Table 3.2). Altogether, these characteristics indicate that MeV, NiV, and HeV N_{TAIL} domains can be classified within the PMG subfamily (Longhi et al., 2003; Bourhis et al., 2004; Habchi et al., 2010).

As already mentioned, the functional relevance of PMGs is related to a facilitated induced folding process. That N_{TAIL} does indeed undergo induced folding was pointed out by CD studies. Those studies showed that MeV N_{TAIL} undergoes an α-helical transition in the presence of PCT (Longhi et al., 2003), while both NiV and HeV N_{TAIL} were shown to gain α-helicity in the presence of TFE (Habchi et al., 2010) (Fig. 3.6c,d). In the case of MeV, an α-MoRE (aa 488–499 of N) has been identified within one (namely, Box 2) out of three N_{TAIL} regions (referred to as Box 1, Box 2, and Box 3) that are conserved within Morbillivirus members (Diallo et al., 1994) (Fig. 3.5). The role of the MeV α-MoRE in binding to P and in the α-helical induced folding of N_{TAIL} was further confirmed by spectroscopic and biochemical experiments carried out on a truncated N_{TAIL} form devoid of the 489–525 region (Bourhis et al., 2004). Apart from the α-MoRE, two additional short regions with some folding potential can be detected within MeV N_{TAIL}: one located in the 400–420 region and another located at the C-terminus (aa 516–525) (Fig. 3.5b). Notably, the former region was shown to represent the binding site to an yet uncharacterized N Receptor, referred to as NR, which is expressed at the surface of dendritic cells of lymphoid origin (both normal and tumoral) (Laine et al., 2003) and of T and B lymphocytes (Laine et al., 2005), while the latter region has been shown to play a role in the binding to XD (Section 3.3.2).

In the case of Henipavirus N_{TAIL}, four putative MoREs were identified: (i) two α-MoREs (aa 408–422 and aa 473–493), roughly corresponding to the NR-binding site and to the experimentally characterized α-MoRE of MeV N_{TAIL}, respectively; (ii) an irregular MoRE (I-MoRE) (aa 523–532) corresponding to the XD-binding region located at the C-terminus of MeV N_{TAIL}; and (iii) an additional putative α-MoRE (aa 444–464) (Fig. 3.6b). However, based on the experimentally observed behavior of both the Henipavirus N_{TAIL}, the latter MoRE is more probably of α-helical type: indeed far-UV CD studies pointed out an increase in the α-helical content of HeV and NiV N_{TAIL} domains on addition of increasing concentrations of TFE, while no concomitant dose-dependant increase in the content of β-content was detected (see insets in Fig. 3.6c,d).

3.3.2 Molecular Mechanisms of the XD-Induced Folding of Measles Virus N_{TAIL}

In the case of MeV, the P region responsible for the interaction with and induced folding of N_{TAIL} has been mapped to the C-terminal XD (aa 459–507) of P (Fig. 3.1), and its crystal structure has been solved at 1.8 Å resolution (Johansson et al., 2003). A model of the interaction between XD and the α-MoRE of MeV N_{TAIL} was then built, in which N_{TAIL} is embedded in a large XD hydrophobic cleft delimited by helices α2 and α3 of XD. According to this model, burying hydrophobic residues of the α-MoRE would provide the driving force to induce the folding of the α-MoRE, thus leading to a pseudo-four-helix arrangement that occurs frequently in Nature (Johansson et al., 2003). This model has been subsequently validated by Kingston and coworkers who solved the crystal structure of a chimeric form mimicking this complex (Fig. 3.7a,b) (Kingston et al., 2004).

The C-terminal region of both the Henipavirus P proteins (also referred to as XD, see Fig. 3.3a) is predicted to be structured and to adopt a predominant α-helical conformation (data not shown). By analogy with other Paramyxovirinae members, we can speculate that this globular region could be the counterpart of the MeV XD and thus be involved in binding to N_{TAIL}. Definite answers to the possible conservation amongst these viruses of a similar N_{TAIL}–XD interaction profile await additional experimental work that is presently in progress in our laboratory.

Small angle X-ray scattering (SAXS) studies of the MeV N_{TAIL}–XD complex provided a low resolution model of the N_{TAIL} bound form, which showed that most of N_{TAIL} (aa 401–488) remains disordered within the complex (Fig. 3.7c,d) (Bourhis et al., 2005). The lack of a protruding appendage corresponding to the extreme C-terminus of N_{TAIL} suggests that, besides Box 2, Box 3 could also be involved in binding to XD (Bourhis et al., 2005). That MeV XD does indeed affect the conformation of Box 3 has been confirmed by heteronuclear NMR (HN-NMR) studies, where analysis of the HSQC (heteronuclear single quantum coherence) spectra of ^{15}N-labeled N_{TAIL} (or of a truncated form devoid of Box 3) showed that the addition of XD triggered both α-helical folding of Box 2 and a minor, but significant, magnetic perturbation within Box 3 (Bourhis et al., 2005; Gely et al., 2010).

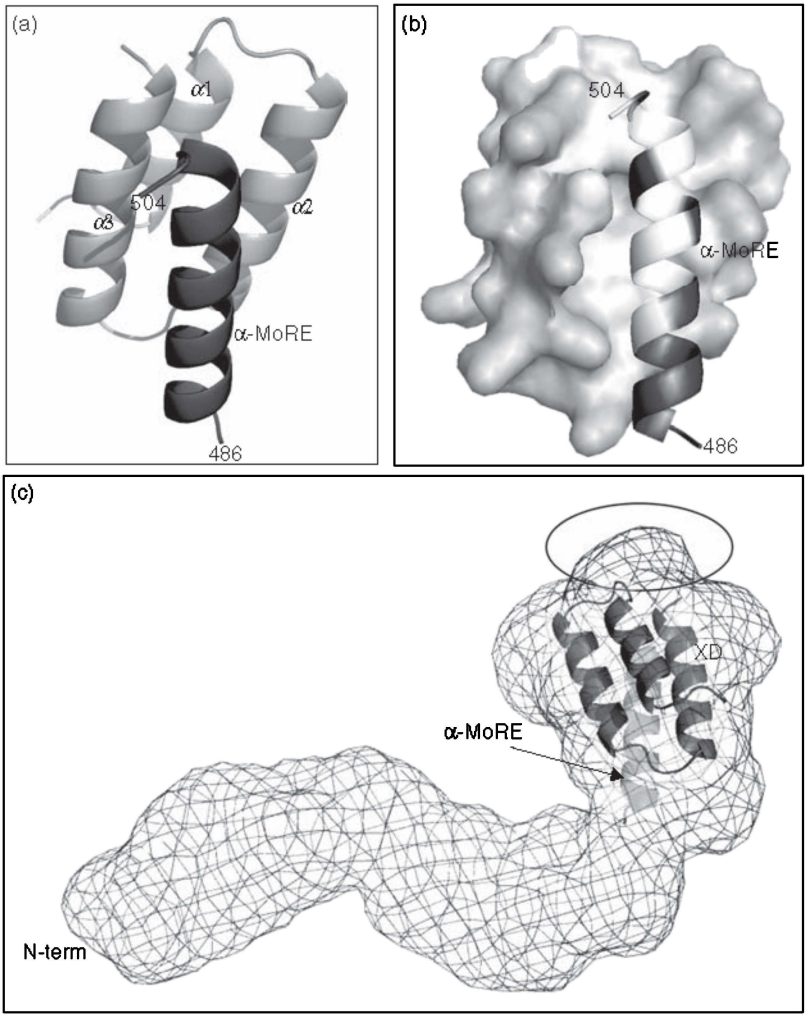


Figure 3.7 (a,b) Crystal structure of the chimeric construct between XD and the α-MoRE of MeV N_{TAIL} (aa 486–504 of N, dark gray, ribbon representation) (PDB code 1T6O) (Kingston et al., 2004). XD is shown by a ribbon (a) or surface (b) representation. In b, the hydrophobic residues of XD and of the α-MoRE are shown in white. (c) Model of the N_{TAIL}–XD complex as derived by small angle X-ray scattering studies (Bourhis et al., 2005). The circle points to the lack of a protruding shape from the globular body of the model. The crystal structure of the chimera between XD and the N_{TAIL} region encompassing residues 486–504 (PDB code 1T6O) (Kingston et al., 2004) is shown. The picture was drawn using Pymol (DeLano, 2002). (d) Low resolution model of the N_{TAIL}–XD complex showing that (i) the 401–488 region of N_{TAIL} is disordered and exposed to the solvent; (ii) the α-MoRE is packed against XD; and (iii) the C-terminus of N_{TAIL} (Box 3) does not protrude from the globular part of the model. *Source*: Modified from Bourhis et al. (2005).

Figure 3.7 (*Continued*)

In further support of the role of Box 3 in binding to XD, surface plasmon resonance (SPR) studies showed that removal of either Box 3 alone or Box 2 *plus* Box 3 results in a strong increase (three orders of magnitude) in the equilibrium dissociation constant, with the K_D increasing from 80 nM to 12 μM (Bourhis et al., 2005) (Table 3.3). Conversely, and as expected from the low resolution model inferred from SAXS data, Box 1 does not contribute to binding to XD (Table 3.3). When synthetic peptides mimicking Box 1, Box 2, and Box 3 were used, Box 2 peptide (aa 487–507) was found to display an affinity for XD (K_D of 20 nM) that was similar to that between XD and N_{TAIL} (K_D of 80 nM), consistent with the role of Box 2 as the primary binding site (Table 3.3). Surprisingly, however, Box 3 peptide exhibits an insignificant affinity for XD (K_D of approximately 1 mM) (Table 3.3). In the same vein, HN-NMR experiments using ^{15}N-labeled XD pointed out lack of magnetic perturbation within the domain on addition of unlabeled Box 3 peptide, consistent with the lack of stable contacts between XD and Box 3 (Bernard et al., 2009). The discrepancy between the data obtained with N_{TAIL}-truncated proteins and with peptides could be accounted for by assuming that Box 3 would act only in the context of N_{TAIL} and not in isolation. Consistent with this hypothesis, both N_{TAIL} and $N_{TAIL\Delta3}$ (i.e., N_{TAIL} devoid of Box 3 and thus spanning residues 401–516 of N) trigger slightly more pronounced chemical shift variations within ^{15}N-labeled XD than Box 2 peptide alone, indicating that the region downstream Box 2 (either Box 3 itself or the region connecting Box 2 to Box 3) contributes to binding to XD, although only when acting in concert with Box 2 (Bernard et al., 2009). Thus, according to this model, Box 3 and Box 2 would be functionally coupled in the binding of N_{TAIL} to XD. One can speculate that the burying of the hydrophobic side of α-MoRE in the hydrophobic cleft formed by helices α2 and α3 of XD would provide the primary driving force in the N_{TAIL}–XD interaction and that Box 3 would act to further stabilize the bound conformation.

In striking contrast with these data, in a recent study by the group of Kingston, where synthetic peptides corresponding to the 477–505 or 477–525 region of N_{TAIL} were used, no evidence was obtained in support of a role of the 505–525 region in binding to XD (Yegambaram and Kingston, (2010)). In addition, in that study, the experimentally determined K_D of the XD-binding reaction was found to be either

TABLE 3.3 Calculated Equilibrium Dissociation Constants (K_D) between MeV XD and MeV N_{TAIL} Proteins and Peptides Using Surface Plasmon Resonance

Analyte	Quality of Fit		
	Residuals	Chi^2	K_D
Box1 peptide (aa 401–421) (TTEDKISRAVGPRQAQVSFLH)	−0.6 to 0.6	0.048	9 mM
Box 2 peptide (aa 487–507) (DSRRSADALLRLQAMAGISEE)	−1.8 to 1.8	0.209	20 nM
Box 3 peptide (aa 505–525) (SEEQGSDTDTPIVYNDRNLLD)	−0.6 to 0.6	0.038	1 mM
N_{TAILHN} (aa 401–525 with an N-term 6-His tag)	−1.2 to 1.6	0.150	92 nM
$N_{TAILHNFC}$ (aa 401–525 with an N-term 6-His tag and a C-term DYKDDDDK flag sequence)	−1.5 to 1.5	0.676	80 nM
$N_{TAIL\Delta3}$ (aa 401–516 with an N-term 6-His tag and a C-term DYKDDDDK flag sequence)	−0.6 to 0.4	0.043	12 μM
$N_{TAIL\Delta2,3}$ (aa 401–488 with an N-term 6-His tag and a C-term DYKDDDDK flag sequence)	−2.4 to 2.4	0.919	41 μM
$N_{TAIL\Delta1}$ (aa 421–525 with an N-term 6-His tag and a C-term DYKDDDDK flag sequence)	−1.2 to 1.2	0.239	49 nM

Source: Data were taken from Bourhis et al. (2005) and Longhi and Oglesbee (2010).

7.4 or 15 μM depending on whether the 477–525 or the 477–505 peptide was used, respectively (Yegambaram and Kingston, 2010).

A role of SeV XD in binding and α-helical folding of N_{TAIL} has also been shown (Houben et al., 2007b). A few features, however, distinguish the SeV N_{TAIL}–XD complex from that of MeV. In particular, in SeV N_{TAIL}, the XD-binding region is restricted to a region with α-helical propensities (aa 472–493 of N), while the C-terminus is not involved in the interaction (Houben et al., 2007b). Also, while binding between MeV XD and MeV N_{TAIL} occurs with a rather strong affinity (K_D of 80 nM) (Bourhis et al., 2005), the experimentally determined K_D for the SeV couple is much higher (60 μM) (Houben et al., 2007b). In this regard, however, it should be pointed out that since the dissociation constant for the SeV N_{TAIL}-XD binding reaction has been determined using NMR titration, it may have been overestimated as a result of the generally high protein concentrations used in these studies, leading to nonideal experimental conditions. Finally, contrary to the MeV N_{TAIL}–XD complex that is mediated by hydrophobic interactions (Johansson et al., 2003; Kingston et al., 2004), the SeV-binding interface is dominated by charged residues (Houben et al., 2007b).

In view of unraveling the precise MeV N_{TAIL} region undergoing α-helical folding, as well as the impact of XD binding on Box 3, the N_{TAIL}–XD interaction has been also investigated by using site-directed spin-labeling (SDSL) electron paramagnetic resonance (EPR) spectroscopy. The basic strategy of SDSL involves the introduction of a paramagnetic nitroxide side chain through covalent modification of a selected protein site. This is usually accomplished by cysteine-substitution

mutagenesis, followed by covalent modification of the unique sulfydryl group with a selective nitroxide reagent, such as the methanethiosulfonate derivative (see Feix and Klug (1998), Hubbell et al. (1998), Biswas et al. (2001), and Belle et al. (2010) for reviews). From the EPR spectral shape of a spin-labeled protein, one can extract information in terms of radical mobility, which reflects the local mobility of residues in the proximity of the radical. Variations in the radical mobility can, therefore, be monitored in the presence of partners, ligands, or organic solvents.

Fourteen single-site MeV N_{TAIL} cysteine variants were designed, purified, and labeled, thus enabling grafting of a nitroxide paramagnetic probe on 12 sites scattered in the 488–525 region and on 2 sites located outside the reported region of interaction with XD (Morin et al., 2006; Belle et al., 2008). EPR spectra were then recorded in the presence of either the secondary structure stabilizer TFE or XD.

Different regions of N_{TAIL} were shown to contribute to a different extent to the binding to XD: while the mobility of the spin labels grafted at positions 407 and 460 was unaffected on addition of XD, the mobility of spin labels grafted within the 488–502 and the 505–522 regions was severely and moderately reduced, respectively (Belle et al., 2008). Furthermore, EPR experiments in the presence of 30% sucrose (i.e., under conditions in which the intrinsic motion of the protein becomes negligible with respect to the intrinsic motion of the spin label), allowed precise mapping of the N_{TAIL} region undergoing α-helical folding to residues 488–502. The drop in the mobility of the 505–522 region on binding to XD was shown to be comparable to that observed in the presence of 20% TFE (Belle et al., 2008), a condition where Box 2 undergoes α-helical folding, whereas Box 3 does not (Bourhis et al., 2005). This observation suggests that the restrained mobility experienced by the Box 3 region on binding to XD is due to neither a steric hindrance exerted by XD nor a direct interaction with XD, but arises from the α-helical folding of the neighboring Box 2.

The mobility of the 488–502 region was found to be restrained even in the absence of the partner (Belle et al., 2008; Kavalenka et al., 2010), a behavior that could be accounted for by the existence of a transiently populated folded state. That the region spanning residues 491–499 adopts an α-helical conformation in about 50% of the conformers sampled by the free form of MeV N_{TAIL} has also been confirmed by HN-NMR studies (Gely et al., 2010). These findings are in agreement with previous reports that showed that the conformational space of MoREs (Oldfield et al., 2005) in the unbound state is often restricted by their inherent conformational propensities, thereby reducing the entropic cost of binding (Tompa, 2002; Fuxreiter et al., 2004; Lacy et al., 2004; Sivakolundu et al., 2005).

Using a novel approach that relies on a combination of SDSL EPR spectroscopy and modeling of local rotation conformational spaces, we could describe the structure of the partly disordered MeV N_{TAIL}–XD complex and showed that in spite of the local gain of rigidity induced by XD binding, the 505–525 region of N_{TAIL} conserves a significant degree of freedom even in the bound form (Fig. 3.8) (Kavalenka et al., 2010).

Finally, EPR equilibrium displacement experiments showed that the XD-induced folding of MeV N_{TAIL} is a reversible phenomenon (Morin et al., 2006; Belle et al.,

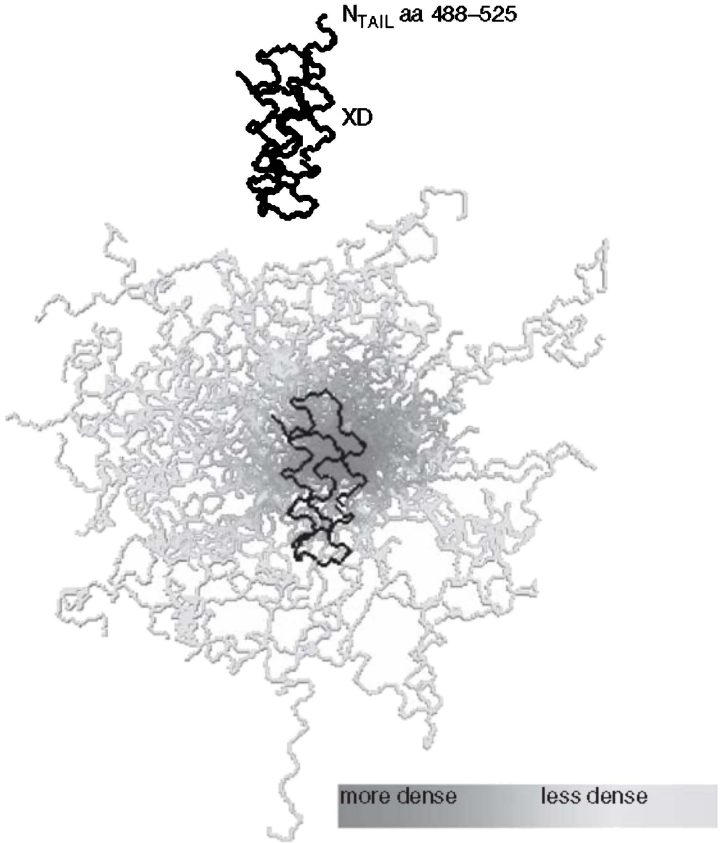


Figure 3.8 Model of the partly disordered MeV N_{TAIL}–XD complex as a conformational ensemble. Fifty best-fit structures of the 488–525 region of N_{TAIL} in complex with XD. The N_{TAIL} conformers are depicted with a color gradient ranging from dark grey to light grey with decreasing structural density, while XD is shown in black. *Source*: Modified from Kavalenka et al. (2010).

2008). These results represent the first experimental evidence indicating that N_{TAIL} adopts its original PMG conformation after dissociation of XD. This point is particularly relevant, taking into consideration that the contact between XD and N_{TAIL} within the replicative complex has to be dynamically made and broken to allow the polymerase to progress along the nucleocapsid template during both transcription and replication. Hence, the complex cannot be excessively stable for this transition to occur efficiently at a high rate (Section 3.4).

In conclusion, using a panel of various physicochemical approaches, the interaction between MeV N_{TAIL} and XD was shown to imply the stabilization of the helical conformation of α-MoRE, which is otherwise only transiently populated in the unbound form. The occurrence of a transiently populated α-helix even in the

absence of the partner suggests that the molecular mechanism governing the folding of N_{TAIL} induced by XD could rely on conformer selection (i.e., selection by the partner of a preexisting conformation) (Tsai et al., 2001a,b). Recent data based on a quantitative analysis of NMR titration studies (Gely et al., 2010) suggest, however, that the binding reaction may also imply a binding intermediate in the form of a weak, nonspecific encounter complex and hence may also occur through a "fly casting" mechanism (Shoemaker et al., 2000) (Fig. 3.9). That binding-coupled-to-folding events rather may rely on a mixed mechanism implying both fly casting and conformational selection has already been reported (Espinoza-Fonseca, 2009).

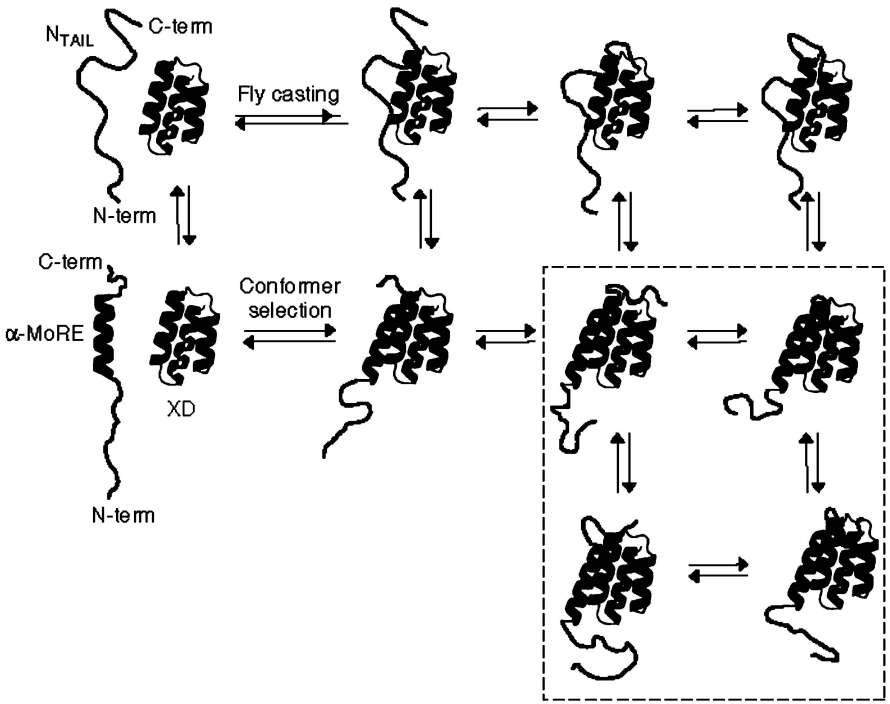


Figure 3.9 Model for MeV N_{TAIL}–XD complex formation that utilizes both conformer selection and nonspecific encounter complex formation. The helix, corresponding to the primary binding site for XD, is partly preformed in the absence of XD. In the nonspecific binding model, a weak encounter complex may also form between N_{TAIL} and XD. This encounter complex is converted to a tightly bound complex by the folding of the α-MoRE. It is unclear if this folding occurs during the lifetime of the encounter complex. In the conformer selection model, the preformed helix interacts with XD to form a tightly bound complex. In both cases, following α-helical folding of Box 2, Box 3 becomes more rigid. The four conformers that are contoured schematically represent the final stage in complex formation, which consists of an ensemble of conformers in which Box 3 has a reduced conformational freedom that may favor the establishment of weak, nonspecific contacts with XD. *Source*: Modified from Gely et al. (2010).

XD-induced stabilization of the helical conformation of the α-MoRE is also accompanied by a reduction in the mobility of the downstream region. The lower flexibility of the region downstream Box 2 is not due to the gain of α-helicity and it cannot be ascribed to a restrained motion due to the presence of XD or to the establishment of stable contacts with XD. Rather, it likely arises from the gain of rigidity brought by the α-helical folding of the neighboring Box 2 region, which results in a reduced Box 3 conformational sampling. The reduced conformational freedom of Box 3 may favor the establishment of weak, nonspecific contacts with XD (Fig. 3.9). At present, the exact role that Box 3 plays in the stabilization of the MeV N_{TAIL}–XD complex remains to be unraveled. Indeed, although recent data clearly indicate that transient long-range tertiary contacts between Box 2 and Box 3 are unlikely (Gely et al., 2010), it is still unclear whether Box 3 contributes to binding to XD through weak (transient) nonspecific contacts or through another unknown mechanism. The possibility that Box 3 may act by reducing the entropic penalty of the binding-coupled-to-folding process (Tompa and Fuxreiter, 2008) seems to be unlikely: indeed, replacement of the Box 3 region by an irrelevant, eight-residues-long Flag sequence (DYKDDDDK) led to a dramatic drop (two orders of magnitude) in the affinity toward XD, which is contrary to what would be expected for a nonbinding region contributing to the overall entropy of the system (Table 3.3).

3.4 FUNCTIONAL ROLE OF STRUCTURAL DISORDER WITHIN N AND P IN TERMS OF TRANSCRIPTION AND REPLICATION

The K_D value of the MeV N_{TAIL}-XD binding reaction, as experimentally determined by SPR studies, is in the 100 nM range (Bourhis et al., 2005). This affinity is considerably higher than that derived from both isothermal titration calorimetry and SPR studies (K_D of 7.4–13 μM) (Kingston et al., 2004; Yegambaram and Kingston, 2010). However, it should be pointed out that these studies were carried out using N_{TAIL} peptides encompassing residues 477–505 or 477–525 rather than full-length N_{TAIL}. A weak binding affinity, implying fast association and dissociation rates, would ideally fulfill the requirements of a polymerase complex that has to cartwheel on the nucleocapsid template during both transcription and replication. However, a K_D in the μM range would not seem to be physiologically relevant considering the low intracellular concentrations of P in the early phases of infection, and the relatively long half-life of active P–L transcriptase complex tethered on the nucleocapsid template, which has been determined to be well over 6 h (Plumet et al., 2005). Moreover, such a weak affinity is not consistent with the ability to readily purify MeV nucleocapsid–P complexes using rather stringent techniques such as CsCl isopycnic density centrifugation (Robbins and Bussell, 1979; Stallcup et al., 1979; Robbins et al., 1980; Oglesbee et al., 1989). On the other hand, a more stable XD–N_{TAIL} complex would be expected to hinder the processive movement of P along the nucleocapsid template. In agreement with this model, the elongation rate of MeV polymerase was found to be rather slow (three nucleotides per second) (Plumet et al., 2005), and MeV N_{TAIL} amino acid substitutions that lower the

affinity toward XD result in enhanced transcription and replication levels, as well as in increased polymerase rate (Oglesbee, Gerlier and Longhi, unpublished data). In addition, the C-terminus of MeV N_{TAIL} has been shown to have an inhibitory role on transcription and genome replication, as indicated by minireplicon experiments, where deletion of the C-terminus of N enhances basal reporter gene expression (Zhang et al., 2002). Deletion of the C-terminus of MeV N also reduces the affinity of XD for N_{TAIL}, providing further support for modulation of XD-N_{TAIL} binding affinity as a basis for polymerase elongation rate. Thus, Box 3 would dynamically control the strength of the N_{TAIL}–XD interaction by stabilizing the complex probably through several weak, nonspecific contacts with XD. Removal of Box 3 or interaction of Box 3 with other partners (Section 3.5), would reduce the affinity of MeV N_{TAIL} for XD, thus stimulating transcription and replication. Modulation of MeV XD-N_{TAIL} binding affinity could be dictated by interactions between N_{TAIL} and cellular and/or viral cofactors. Indeed, the requirement for cellular or viral cofactors in both transcription and replication has been already documented in the case of MeV (Vincent et al., 2002) and other Mononegavirales members (Fearns and Collins, 1999; Hartlieb et al., 2003). Furthermore, in both CDV and MeV, viral transcription and replication are enhanced by the major inducible heat shock protein (hsp70), and this stimulation relies on interaction with N_{TAIL} (Oglesbee et al., 1993, 1996; Vasconcelos et al., 1998a,b; Zhang et al., 2002, 2005; Carsillo et al., 2006; Oglesbee, 2007). These cofactors may serve as elongation factors and could act by modulating the strength of the interaction between the polymerase complex and the nucleocapsid template (Section 3.5).

In MeV, N_{TAIL} also influences the physical properties of the nucleocapsid helix that is formed by N_{CORE} (Longhi et al., 2003; Schoehn et al., 2004). Electron microscopy analyses of nucleocapsids formed by either N or N_{CORE} indicate that the presence of N_{TAIL} was associated with a greater degree of fragility as evidenced by the tendency of helices to break into individual ring structures. This fragility is associated with the evidence of increased nucleocapsid flexibility, with helices formed by N_{CORE} alone forming rods (Longhi et al., 2003; Schoehn et al., 2004). It is, therefore, conceivable that the induced folding of N_{TAIL} resulting from the interaction with P (and/or other physiological partners) could also exert an impact on the nucleocapsid conformation in such a way as to affect the structure of the replication promoter. Indeed, the replication promoter, located at the 3′ end of the viral genome, is composed of two discontinuous elements that form a functional unit when juxtaposed on two successive helical turns (Tapparel et al., 1998) (Fig. 3.10). The switch between transcription and replication could be dictated by variations in the helical conformation of the nucleocapsid, which would result in a modification in the number of N monomers (and thus of nucleotides) per turn, thereby disrupting the replication promoter in favor of the transcription promoter (or vice versa). Morphological analyses, showing the occurrence of a large conformational flexibility within Paramyxoviridae nucleocapsids (Oglesbee et al., 1989, 1990; Bhella et al., 2002, 2004), tend to corroborate this hypothesis. In further support for a possible role of cellular cofactors in affecting the nucleocapsid conformation thereby favoring transcription and replication, hsp70–nucleocapsid complexes of CDV exhibit

Figure 3.10 (a) Schematic representation and (b) cryo-electron microscopy reconstruction of the MeV nucleocapsid highlighting the bipartite structure of the replication promoter composed of two discontinuous units juxtaposed on successive helical turns (see regions wrapped by the N monomers highlighted by an asterisk). *Source*: Courtesy of D. Bhella, MRC, Glasgow, UK; modified from Longhi (2009).

an expanded helical diameter, an increased fragility (as judged by tendency to fragment into rings), and an enhanced exposure of the genomic RNA to nuclease degradation (Oglesbee et al., 1989).

In the same vein, preliminary data indicate that incubation of MeV nucleocapsids in the presence of XD triggers unwinding of the nucleocapsid, thus possibly enhancing the accessibility of genomic RNA to the polymerase complex (Bhella and Longhi, unpublished data). Hence, it is tempting to propose that the XD-induced α-helical folding of N_{TAIL} could trigger the opening of the two lobes of N_{CORE}, thus rendering the genomic RNA accessible to the solvent. Altogether, these data establish a relationship between N_{TAIL}-binding partners, nucleocapsid conformation, and exposure of the genomic RNA.

The presence of unstructured domains in both N and P would allow for coordinated interactions between the polymerase complex and a large surface area of the nucleocapsid template, including successive turns of the helix. Unstructured regions are in fact considerably more extended in solution than globular ones. For instance, MeV PNT has a Stokes radius of 4 nm (Karlin et al., 2002). However, the Stokes radius only reflects a mean dimension. Indeed, the maximal extension of PNT, as measured by SAXS (Longhi and Receveur-Bréchot, unpublished data), is

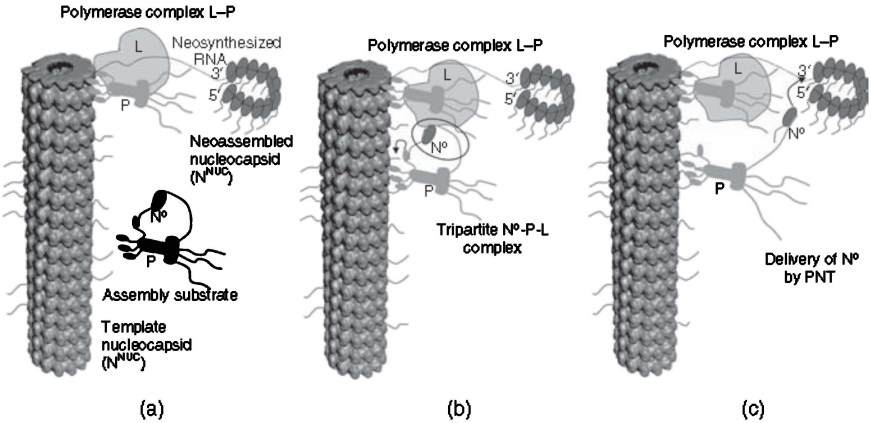


Figure 3.11 Model of the MeV polymerase complex actively replicating genomic RNA. Disordered regions are represented by lines. The location of the viral RNA (dotted line) within the N^{NUC}–P complex is schematically represented at the interior of the nucleocapsid only for clarity. Within the N^{NUC}–P complex, P is represented as bound to N^{NUC} through three of its four terminal XD arms according to the model of Curran and Kolakofsky (1999). For the sake of clarity, only a few N_{TAIL} regions are drawn. PNT regions within the L–P complex have not been represented in panels 2 and 3. The numbering of the different panels indicates the chronology of events. (a) L is bound to a P tetramer. The newly synthesized RNA is shown as already partially encapsidated. (b) The encapsidation complex N^0–P binds to the nucleocapsid template through three of its four XD arms. The extended conformation of N_{TAIL} and PNT would allow the formation of a tripartite complex between N^0, P, and the polymerase (circled). It is tempting to imagine that the proximity of the polymerase (or an unknown signal from the latter) may promote the release of N^0 by XD, thus leading to N^0 incorporation within the assembling nucleocapsid. The N^0 release would also lead to cartwheeling of the L–P complex through binding of the free XD arm onto the nucleocapsid template (arrow) as in the model of Curran and Kolakofsky (1999). (c) PNT delivers N^0 to the newly assembled nucleocapsid (arrow). *Source*: Modified from Bourhis et al. (2005, 2006).

considerably larger (> 40 nm). In comparison, one turn of the MeV nucleocapsid is 18 nm in diameter and 6 nm high (Bhella et al., 2002). Thus, PNT could easily stretch over several turns of the nucleocapsid, and since P is multimeric, N^0–P might have a considerable extension (Fig. 3.11). In the same vein, it is striking that SeV and MeV PCT domains, which interact with the intrinsically disordered N_{TAIL} domain, comprise a flexible linker (Marion et al., 2001; Longhi et al., 2003; Blanchard et al., 2004; Bernado et al., 2005). This certainly suggests the need for a great structural flexibility. This flexibility could be necessary for the tetrameric P to bind several turns of the helical nucleocapsid. Indeed, the promoter signals for the polymerase are located on the first and the second turn of the SeV nucleocapsid (Tapparel et al., 1998).

Likewise, the maximal extension of MeV N_{TAIL} in solution is 13 nm (Longhi et al., 2003). The very long reach of disordered regions could enable them to act as

linkers and to tether partners on large macromolecular assemblies, thereby acting as scaffolding engines as already described for intrinsically disordered scaffold proteins (Cortese et al., 2008; Balazs et al., 2009). Accordingly, one role of the tentaclelike N_{TAIL} projections in actively replicating nucleocapsids could be to put into contact several proteins within the replicative complex, such as the N^0–P and the P–L complexes (Fig. 3.11).

In respiroviruses and morbilliviruses, PNT contains binding sites for N^0 (Curran et al., 1994; Harty and Palese, 1995; Sweetman et al., 2001) and L (Curran and Kolakofsky, 1991; Curran et al., 1995; Sweetman et al., 2001). This pattern of interactions among N^0, P, and L, mediated by unstructured regions of either P or N, suggests that N^0, P, and L might interact simultaneously at some point during replication. Notably, the existence of a N-P-L tripartite complex has been recently proved by coimmunoprecipitation studies in the case of VSV, where this tripartite complex constitutes the replicase complex, as opposed to the transcriptase complex that consists of L, P, and host cellular proteins, such as hsp60 (Gupta et al., 2003; Qanungo et al., 2004).

A model can be proposed, where during the replication of MeV, the extended conformation of PNT and N_{TAIL} would be key to allowing contact between the assembly substrate (N^0–P) and the polymerase complex (L–P), thus leading to a tripartite N^0-P-L complex (Fig. 3.11). This model emphasizes the plasticity of intrinsically disordered regions, which might give a considerable reach to the elements of the replicative machinery.

Interestingly, there is a striking parallelism between the MeV N_{TAIL}–XD interaction and the MeV PNT–N_{CORE} interaction (Fig. 3.1b). Both interactions are not stable by themselves and must be strengthened by the combination of other interactions. This might ensure easy breaking and reforming of interactions and could result in transient, easily modulated interactions. One can speculate that the gain of structure of MeV N_{TAIL} on binding to XD could result in stabilization of the N–P complex. At the same time, folding of N_{TAIL} would result in a modification in the pattern of solvent-accessible regions, resulting in the shielding of specific regions of interaction. As a result, N_{TAIL} would no longer be available for binding to its other partners. Although induced folding likely enhances the affinity between interacting proteins, the dynamic nature of these interactions could rely on (i) the intervention of viral and/or cellular cofactors modulating the strength of such interactions and (ii) the ability of the IDP to establish weak affinity interactions through residual disordered regions.

Finally, as binding of MeV N_{TAIL} to XD allows tethering of the L protein on the nucleocapsid template, the N_{TAIL}–XD interaction is crucial for both viral transcription and genomic replication. As neither N_{TAIL} nor XD has cellular homologs, this interaction is an ideal target for antiviral inhibitors. *In silico* screening of small compounds for their ability to bind to the hydrophobic cleft of XD is in progress. Using HN-NMR, a few candidate molecules are being tested for their ability to bind to XD and to prevent interaction with N_{TAIL} (Morelli, Guerlesquin, and Longhi, unpublished data).

3.5 STRUCTURAL DISORDER AND MOLECULAR PARTNERSHIP

One of the functional advantages of disorder is related to an increased plasticity that enables disordered regions to bind to numerous structurally distinct targets (Dunker et al., 2005; Uversky et al., 2005; Haynes et al., 2006). In agreement, intrinsic disorder is a distinctive and common feature of "hub" proteins, with disorder serving as a determinant of protein interactivity (Dunker et al., 2005; Uversky et al., 2005; Haynes et al., 2006). The lack of a rigid 3D structure allows IDPs to establish interactions that are characterized by a high specificity and a low affinity; while high specificity is ensured by the very large surface area that is generally buried in complexes involving IDPs (Tompa, 2003), low affinity arises from an unfavorable entropic contribution associated to the disorder-to-order transition (Dunker et al., 1998, 2001, 2002; Wright and Dyson, 1999; Dunker and Obradovic, 2001; Gunasekaran et al., 2003; Uversky et al., 2002; Dyson and Wright, 2005; Fink, 2005). The extent of the entropic penalty is, however, tightly related to the extent of conformational sampling of the prerecognition state, that is, on the degree to which MoREs are preconfigured in solution before binding; the occurrence of a partly preconfigured MoRE in the unbound state in fact reduces the entropic cost of binding, thereby enhancing affinity (Tompa, 2002; Fuxreiter et al., 2004; Lacy et al., 2004; Sivakolundu et al., 2005; Mohan et al., 2006; Vacic et al., 2007). As such, IDPs exhibit a wide binding diversity, with some of them binding their partners with strong affinities, as is the case of the MeV N_{TAIL}–XD and MeV N_{TAIL}–hsp70 (see below) couples.

The disordered nature of N_{TAIL} confers to this N domain a large malleability, enabling it to adapt to various partners and to form complexes that are critical for both transcription and replication. Hence, thanks to its exposure at the surface of the viral nucleocapsid, MeV N_{TAIL} establishes numerous interactions with various viral partners, including P, the P–L complex, and the matrix protein (Iwasaki et al., 2009) (Fig. 3.12). Beyond viral partners, MeV N_{TAIL} also interacts with cellular proteins, including hsp70 (Zhang et al., 2002, 2005), the interferon regulatory factor 3 (IRF3) (Colombo et al., 2009; tenOever et al., 2002), the cell protein responsible for the nuclear export of N (Sato et al., 2006), and, possibly, components of the cell cytoskeleton (Moyer et al., 1990; De and Banerjee, 1999) (Fig. 3.12). Moreover, MeV N_{TAIL} within viral nucleocapsids released from infected cells also binds to cell receptors involved in MeV-induced immunosuppression (Marie et al., 2001; Laine et al., 2003, 2005) (Fig. 3.12). Likewise, both MeV and SeV PNT domains have been reported to interact with multiple partners, with the former interacting with N (Chen et al., 2003) and cellular proteins (Liston et al., 1995) and the latter interacting with the unassembled form of N (N^0) and the L protein (Curran et al., 1994, 1995).

In both CDV and MeV, the interaction between N_{TAIL} and hsp70 stimulates both transcription and genome replication (Oglesbee et al., 1993, 1996; Vasconcelos et al., 1998a,b; Zhang et al., 2002, 2005; Carsillo et al., 2006; Oglesbee, 2007). In MeV, two binding sites for hsp70 have been identified (Zhang et al., 2002, 2005). High affinity binding (K_D of 10 nM) is supported by α-MORE, and

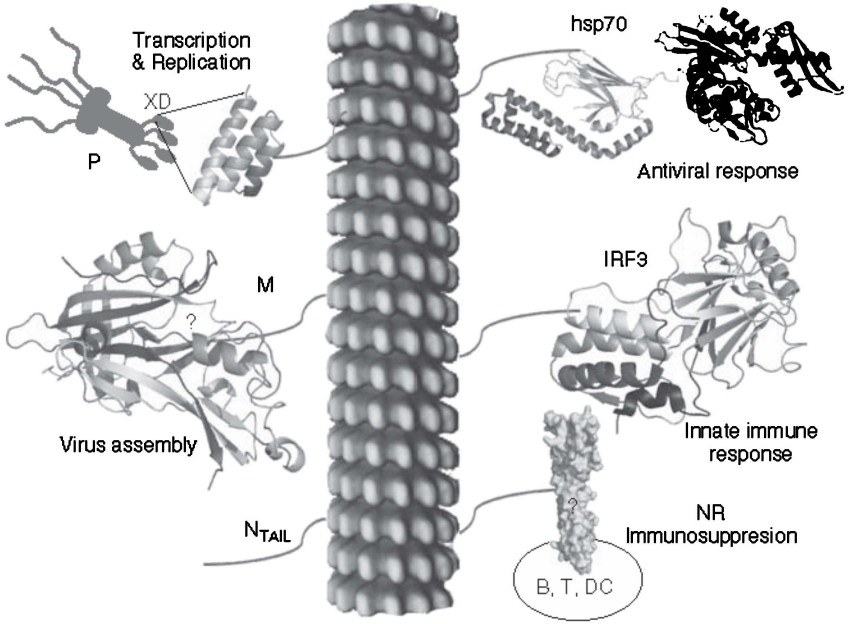


Figure 3.12 Schematic representation of the complex molecular partnership established by MeV N_{TAIL}. Cryo-electron microscopy reconstruction of MeV nucleocapsid according to Bhella et al. (2004) and Bhella (2007), onto which only a few N_{TAIL} regions have been depicted for the sake of clarity. The ribbon representations of MeV XD (PDB code 1OKS) (Johansson et al., 2003), the regulatory domain of IRF3 (PDB code 1QWT) (Qin et al., 2003), and the *Escherichia coli* DnaK chaperone (the prokaryotic counterpart of hsp70) (PDB code 2KHO) (Bertelsen et al., 2009) are shown. Since no structural data are available for MeV M, the ribbon representation of the matrix protein from RSV (PDB code 2VQP) (Money et al., 2009) is shown. The question mark symbolizes the lack of structural data for both MeV M and NR. The latter is represented inserted in the membrane of B/T lymphocytes and dendritic cells (DCs). The pictures were drawn using Pymol (DeLano, 2002).

hsp70 was shown to competitively inhibit binding of XD to N_{TAIL} (Zhang et al., 2005). A second low affinity binding site is present within Box 3 (Zhang et al., 2002; Carsillo et al., 2006). Variability in Box 3 sequence does not influence the binding affinity of N_{TAIL} for hsp70 since this reaction is primarily determined by α-MORE, yet the ability of hsp70 to directly bind Box 3 does have functional consequences (Zhang et al., 2005). Analysis of an infectious virus containing the N522D substitution within Box 3 shows loss of hsp70-dependent stimulation of transcription but not genome replication (Zhang et al., 2005). These findings suggest that hsp70 could enhance transcription and genome replication by reducing the stability of P–N_{TAIL} complexes, thereby promoting successive cycles of binding and release that are essential to polymerase movement along the nucleocapsid template (Bourhis et al., 2005; Zhang et al., 2005). hsp70-dependent reduction of the

stability of P–N_{TAIL} complexes would rely on competition between hsp70 and XD for binding to N_{TAIL} through (i) competition for binding to α-MoRE (and this would occur at low hsp70 concentrations) and (ii) neutralization of the contribution of the C-terminus of the N_{TAIL} to the formation of a stable P–N_{TAIL} complex (and this would occur in the context of elevated cellular levels of hsp70 and only for MeV strains that support hsp70 binding in this region) (Zhang et al., 2005). Another possible mechanism for the hsp70-dependent enhancement of viral transcription and replication is related to the ability of hsp70 to maintain the nucleocapsid in a transcription-competent form (Oglesbee and Gerlier, unpublished data).

The basis for the separable effects of hsp70 on genome replication versus transcription remains to be shown, with template changes unique to a replicase versus transcriptase being a primary candidate. The latter could involve unique nucleocapsid ultrastructural morphologies, with hsp70-dependent morphologies being well documented for CDV (Oglesbee et al., 1989, 1990). Another possibility is that the transcriptase and replicase complexes may be different, with the latter possibly including hsp70 by analogy to the replicase complex of VSV that was shown to incorporate hsp60 (Qanungo et al., 2004).

Notably, many heat shock proteins function as chaperone complexes, and we have recently shown that high affinity MeV N_{TAIL}-hsp70 binding requires an hsp40 co-chaperone that interacts primarily with the hsp70 nucleotide-binding domain and displays no significant affinity for N_{TAIL}. hsp40 directly enhances hsp70 ATPase activity in an N_{TAIL}-dependent manner, and formation of hsp40-hsp70-N_{TAIL} intracellular complexes requires the presence of N_{TAIL} Box 2 and Box 3. These results are consistent with the functional interplay between the hsp70 nucleotide and substrate-binding domains, where ATP hydrolysis is rate limiting to high affinity binding to client proteins and is enhanced by hsp40 (Couturier et al., 2009). Further investigation is required to determine if limitations in intracellular levels of hsp40 may influence permissiveness to viral gene expression or development of antiviral immune responses.

Other cellular partners for MeV N_{TAIL} binding may similarly require cofactors not identified in the final complex. We have recently reported that the MeV N_{TAIL}–IRF3 interaction requires a specific cellular environment, such as that provided by 293T human cells (Colombo et al., 2009). Indeed, using a large panel of spectroscopic approaches, we failed to detect any direct interaction between purified IRF3 regulatory domain and either full-length N or N_{TAIL}, nor could we detect such an interaction in *E coli* and in a yeast two-hybrid assay. This dependence from a specific cellular context likely reflects the requirement for a human or mammalian cellular cofactor that remains to be identified (Colombo et al., 2009).

MeV N_{TAIL} binding to cellular partners may also occur in the extracellular environment. After apoptosis of infected cells, the viral nucleocapsid is released in the extracellular compartment where it becomes available to cell surface receptors. While N_{CORE} specifically interacts with FcγRII (Laine et al., 2005), N_{TAIL} interacts with the yet uncharacterized NR (Laine et al., 2003, 2005). The MeV N_{TAIL}–NR interaction was shown to trigger an arrest in the G_0/G_1 phase of cell cycle, whereas

the N_{CORE}–FcγRII interaction triggers apoptosis (Laine et al., 2005). Both mechanisms have the potential to contribute to immunosuppression that is a hallmark of MeV infections (Laine et al., 2005). Flow cytofluorimetry studies carried out on truncated forms of N_{TAIL} allowed the identification of the N_{TAIL} region responsible for the interaction with NR (Box1, aa 401–420) (Laine et al., 2005). Whether MeV N_{TAIL} directly interacts with NR and whether this interaction requires a cofactor, as in the case of hsp70 and IRF3, remain to be elucidated.

3.6 CONCLUSIONS

As thoroughly discussed in this chapter, N and P are multifunctional proteins exerting multiple biological functions. The molecular bases for this property reside in the ability of their disordered regions to establish interactions with various partners. Since many, if not all, of these interactions are critical for transcription and genome replication, they provide excellent targets for antiviral agents. In this context, the discovery that the N and P domains supporting these multiple protein interactions are intrinsically disordered is particularly relevant: indeed protein–protein interactions mediated by disordered regions provide interesting drug discovery targets with the potential to increase significantly the discovery rate for new compounds (see Cheng et al. (2006) and Uversky (2010) and references therein cited).

Although we cannot formally rule out the possibility that MeV and Henipavirus PNT and N_{TAIL} domains may undergo some degree of folding in the context of the full-length P and N proteins, a few studies directly support the conclusion that disorder is retained in the context of the full-length proteins. Indeed, in SeV, the PNT region has been shown to be disordered not only in isolation but also in the context of the full-length protein (Chinchar and Portner, 1981; Deshpande and Portner, 1985), as were the linker region between PMD and XD in the context of PCT (Tarbouriech et al., 2000a) and the region upstream XD in both SeV and MeV (Longhi et al., 2003; Bernado et al., 2005; Houben et al., 2007a). Likewise, the MeV N_{TAIL} region is also disordered within the entire N, being equally accessible to monoclonal antibodies in isolation and within recombinant nucleocapsids and being not visible by electron microscopy (Longhi et al., 2003). Experiments aimed at assessing the disordered state of N_{TAIL} in the context of full-length N proteins from both MeV and henipaviruses are in progress. As for the possibility that PNT and N_{TAIL} are effectively disordered *in vivo*, further experimental work is required to address this question. In this regard, it is noteworthy that available data addressing the disordered state of IDPs *in vivo* are mixed. On one hand, in-cell NMR experiments on the natively unfolded FlgM protein suggest a more folded conformation in the cellular environment of live bacteria (Dedmon et al., 2002). On the other hand, a few reports support a disordered state for other IDPs either in living cells or in the presence of crowding agents that mimic the crowded environment of cells (Longhi et al., 2003; McNulty et al., 2006; Bodart et al., 2008).

Recent studies showed that viruses and eukaryota have 10 times more conserved disorder (roughly 1%) than archaea and bacteria (0.1%) (Chen et al., 2006) and

also pointed out that viral proteins, and, in particular, proteins from RNA viruses, are enriched in short disordered regions (Tokuriki et al., 2009). In this latter study, the authors propose that beyond affording a broad partnership, the wide occurrence of disordered regions in viral proteins could also be related to the typical high mutation rates of RNA viruses, representing a strategy for buffering the deleterious effects of mutations (Tokuriki et al., 2009).

Taking into account these considerations, as well as the correlation between overlapping genes and disorder (Jordan et al., 2000; Narechania et al., 2005; Rancurel et al., 2009; Kovacs et al., 2010), we propose that the main advantage of the abundance of disorder within viruses would reside in pleiotropy and genetic compaction. Indeed, disorder provides a solution to reduce both genome size and molecular crowding, where a single gene would (i) encode a single (regulatory) protein product that can establish multiple interactions via its disordered regions and hence exert multiple concomitant biological effects and/or (ii) encode more than one product by means of overlapping reading frames. In fact, since disordered regions are less sensitive to structural constraints than ordered ones, the occurrence of disorder within one or both protein products encoded by an overlapping reading frame can represent a strategy to alleviate evolutionary constraints imposed by the overlap. As such, disorder would confer to viruses the ability to "handle" overlaps, thus further expanding the coding potential of viral genomes.

ACKNOWLEDGMENTS

We wish to thank all the persons who contributed to the works herein described. In particular, within the AFMB laboratory, we would like to thank Jean-Marie Bourhis, Benjamin Morin, Stéphanie Costanzo, Matteo Colombo, Marie Couturier, Sabrina Rouger, Elodie Liquière, Bruno Canard, Kenth Johansson, David Karlin, François Ferron, Véronique Receveur-Brechot, Hervé Darbon, Cédric Bernard, and Valérie Campanacci. We also thank our coworkers Michael Oglesbee (Ohio State University), Keith Dunker (Indiana University), David Bhella (MRC, Glasgow, UK), Denis Gerlier (LabVirPAth, Lyon, France), Hélène Valentin and Chantal Rabourdin-Combe (INSERM, Lyon, France), Gary Daughdrill (University of South Florida), Martin Blackledge and Malene Ringkjobing-Jensen (IBS, Grenoble, France), and André Fournel, Valérie Belle, and Bruno Guigliarelli (BIP, CNRS, Marseille, France). We are also grateful to David Bhella who is the author of Fig. 3.10. We also want to thank Vladimir Uversky and Denis Gerlier for stimulating discussions. Finally, S. L. wishes to express her gratitude to Frédéric Carrière (EIPL, CNRS, Marseille, France) for having introduced her to EPR spectroscopy and for the constant support. The studies mentioned in this review were carried out with the financial support of the European Commission, program RTD, QLK2-CT2001-01225, of the Agence Nationale de la Recherche, specific programs "Microbiologie et Immunologie" (ANR-05-MIIM-035-02) and "Physico-chimie du vivant" (ANR-08-PCVI-0020-01), and of the National Institute of Neurological Disorders and Stroke (R01 NS031693-11A2).

ABBREVIATIONS

MeV	measles virus
NiV	Nipah virus
HeV	Hendra virus
H	attachment glycoprotein
F	fusion glycoprotein
ORF	open reading frame
N	nucleoprotein
P	phosphoprotein
M	matrix protein
L	large protein
NTP	nucleotide triphosphate
SeV	Sendai virus
VSV	vesicular stomatitis virus
N_{CORE}	N-terminal, structured domain of N
N_{TAIL}	C-terminal disordered domain of N
N^0	monomeric, soluble form of N
N^{NUC}	assembled form of N
CDV	canine distemper virus
NLS	nuclear localization signal
NES	nuclear export signal
IDP	intrinsically disordered protein
RV	rabies virus
RSV	respiratory syncytial virus
PNT	P N-terminal domain
ZnBD	zinc-binding domain
PCT	P C-terminal domain
SEC	size exclusion chromatography
DLS	dynamic light scattering
R_S	Stokes radius
MoRE	molecular recognition element
PMG	premolten globule
RC	random coil
AAD	acidic activator domain
TFE	2,2,2-trifluoro ethanol
PMD	P multimerization domain
XD	X domain
MuV	mumps virus
NMR	nuclear magnetic resonance
PX	protein X
CCR	central conserved region
PDB	protein data bank
NR	nucleoprotein receptor
HN-NMR	heteronuclear NMR

SPR	surface plasmon resonance
SDSL-EPR	site-directed spin-labeling electron paramagnetic resonance
hsp70	major inducible heat shock protein 70 kDa
hsp40	heat shock protein 40 kDa
IRF3	interferon regulator factor 3.

REFERENCES

Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW. Crystal structure of the rabies virus nucleoprotein-RNA complex. Science 2006;313(5785):360–363.

Albertini AAV, Schoehn G, Ruigrok RW. Structures impliquées dans la réplication et la transcription des virus à ARN non segmentés de sens négatif. Virologie 2005;9(2):83–92.

Assenberg R, Delmas O, Ren J, Vidalain PO, Verma A, Larrous F, Graham SC, Tangy F, Grimes JM, Bourhy H. Structure of the nucleoprotein binding domain of Mokola virus phosphoprotein. J Virol 2010;84(2):1089–1096.

Baker SC, Moyer SA. Encapsidation of Sendai virus genome RNAs by purified NP protein during *in vitro* replication. J Virol 1988;62(3):834–838.

Balazs A, Csizmok V, Buday L, Rakacs M, Kiss R, Bokor M, Udupa R, Tompa K, Tompa P. High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1. FEBS J 2009;276(14):3744–3756.

Bankamp B, Horikami SM, Thompson PD, Huber M, Billeter M, Moyer SA. Domains of the measles virus N protein required for binding to P protein and self-assembly. Virology 1996;216(1):272–277.

Belle V, Rouger S, Costanzo S, Liquiere E, Strancar J, Guigliarelli B, Fournel A, Longhi S. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. Protein Struct Funct Bioinform 2008;73(4):973–988.

Belle V, Rouger S, Costanzo S, Longhi S, Fournel A. Site-directed spin labeling EPR spectroscopy. In: Uversky VN, Longhi S, editors. Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation. Hoboken, New Jersey: John Wiley and Sons; 2010.

Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 2005;102(47):17002–17007.

Bernard C, Gely S, Bourhis JM, Morelli X, Longhi S, Darbon H. Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. FEBS Lett 2009;583(7):1084–1089.

Bertelsen EB, Chang L, Gestwicki JE, Zuiderweg ER. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proc Natl Acad Sci USA 2009;106(21):8471–8476.

Bhella D. Measles virus nucleocapsid structure, conformational flexibility and the rule of six. In: Longhi S, editor. Measles virus nucleoprotein. Hauppage, NY: Nova Publishers Inc.; 2007.

Bhella D, Ralph A, Murphy LB, Yeo RP. Significant differences in nucleocapsid morphology within the Paramyxoviridae. J Gen Virol 2002;83(Pt 8):1831–1839.

Bhella D, Ralph A, Yeo RP. Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. J Mol Biol 2004;340(2):319–331.

Biswas R, Kuhne H, Brudvig GW, Gopalan V. Use of EPR spectroscopy to study macromolecular structure and function. Sci Prog 2001;84(Pt 1):45–67.

Bitko V, Barik S. Phenotypic silencing of cytoplasmic genes using sequence-specific double-stranded short interfering RNA and its application in the reverse genetics of wild type negative-strand RNA viruses. BMC Microbiol 2001;1:34.

Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP, Ruigrok RW, Marion D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. Virology 2004;319(2):201–211.

Bodart JF, Wieruszeski JM, Amniai L, Leroy A, Landrieu I, Rousseau-Lescuyer A, Vilain JP, Lippens G. NMR observation of Tau in Xenopus oocytes. J Magn Reson 2008;192(2):252–257.

Bourhis J, Johansson K, Receveur-Bréchot V, Oldfield CJ, Dunker AK, Canard B, Longhi S. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. Virus Res 2004;99:157–167.

Bourhis JM, Canard B, Longhi S. Désordre structural au sein du complexe réplicatif du virus de la rougeole: implications fonctionnelles. Virologie 2005;9:367–383.

Bourhis JM, Canard B, Longhi S. Structural disorder within the replicative complex of measles virus: functional implications. Virology 2006;344:94–110.

Bourhis JM, Canard B, Longhi S. Predicting protein disorder and induced folding: from theoretical principles to practical applications. Curr Protein Pept Sci 2007;8(2): 135–149.

Bourhis JM, Longhi S. Measles virus nucleoprotein: structural organization and functional role of the intrinsically disordered C-terminal domain. In: Longhi S, editor. Measles virus nucleoprotein. Hauppage, NY: Nova Publishers Inc.; 2007. pp. 1–35.

Bourhis JM, Receveur-Bréchot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, Canard B, Finet S, Longhi S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci 2005;14:1975–1992.

Buchholz CJ, Retzler C, Homann HE, Neubert WJ. The carboxy-terminal domain of Sendai virus nucleocapsid protein is involved in complex formation between phosphoprotein and nucleocapsid-like particles. Virology 1994;204(2):770–776.

Buchholz CJ, Spehner D, Drillien R, Neubert WJ, Homann HE. The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. J Virol 1993;67(10):5803–5812.

Carsillo T, Zhang X, Vasconcelos D, Niewiesk S, Oglesbee M. A single codon in the nucleocapsid protein C terminus contributes to *in vitro* and *in vivo* fitness of Edmonston measles virus. J Virol 2006;80(6):2904–2912.

Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. J Proteome Res 2006;5(4):879–887.

Chen M, Cortay JC, Gerlier D. Measles virus protein interactions in yeast: new findings and caveats. Virus Res 2003;98(2):123–129.

Chen M, Cortay JC, Logan IR, Sapountzi V, Robson CN, Gerlier D. Inhibition of ubiquitination and stabilization of human ubiquitin E3 ligase PIRH2 by measles virus phosphoprotein. J Virol 2005;79(18):11824–11836.

Cheng Y, Legall T, Oldfield CJ, Mueller JP, Van YY, Romero P, Cortese MS, Uversky VN, Dunker AK. Rational drug design via intrinsically disordered protein. Trends Biotechnol 2006;24(10):435–442.

Chinchar VG, Portner A. Functions of Sendai virus nucleocapsid polypeptides: enzymatic activities in nucleocapsids following cleavage of polypeptide P by Staphylococcus aureus protease V8. Virology 1981;109(1):59–71.

Colombo M, Bourhis JM, Chamontin C, Soriano C, Villet S, Costanzo S, Couturier M, Belle V, Fournel A, Darbon H, Gerlier D, Longhi S. The interaction between the measles virus nucleoprotein and the interferon regulator factor 3 relies on a specific cellular environment. Virol J 2009;6(1):59.

Cortese MS, Uversky VN, Dunker AK. Intrinsic disorder in scaffold proteins: getting more from less. Prog Biophys Mol Biol 2008;98(1):85–106.

Couturier M, Buccellato M, Costanzo S, Bourhis JM, Shu Y, Nicaise M, Desmadril M, Flaudrops C, Longhi S, Oglesbee M. High affinity binding between Hsp70 and the C-terminal domain of the Measles virus nucleoprotein requires an Hsp40 co-chaperone. J Mol Recogn 2009;23(3):301–315.

Curran J, Homann H, Buchholz C, Rochat S, Neubert W, Kolakofsky D. The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. J Virol 1993;67(7):4358–4364.

Curran J, Kolakofsky D. Replication of paramyxoviruses. Adv Virus Res 1999;54: 403–422.

Curran J, Marq JB, Kolakofsky D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. J Virol 1995;69(2):849–855.

Curran J, Pelet T, Kolakofsky D. An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. Virology 1994; 202(2):875–884.

Curran JA, Kolakofsky D. Rescue of a Sendai virus DI genome by other parainfluenza viruses: implications for genome replication. Virology 1991;182(1):168–176.

Dahlman-Wright K, McEwan IJ. Structural studies of mutant glucocorticoid receptor transactivation domains establish a link between transactivation activity *in vivo* and alpha-helix-forming potential *in vitro*. Biochemistry 1996;35(4):1323–1327.

De BP, Banerjee AK. Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses. Microsc Res Tech 1999;47(2):114–123.

Dedmon MM, Patel CN, Young GB, Pielak GJ. FlgM gains structure in living cells. Proc Natl Acad Sci USA 2002;99(20):12681–12684.

DeLano WL. The PyMOL molecular graphics system. Protein Struct Funct Bioinform 2002;30:442–454.

Delmas O, Assenberg R, Grimes JM, Bourhy H. The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae. RNA Biol 2010;7(3).

Deshpande KL, Portner A. Monoclonal antibodies to the P protein of Sendai virus define its structure and role in transcription. Virology 1985;140(1):125–134.

Diallo A, Barrett T, Barbron M, Meyer G, Lefevre PC. Cloning of the nucleocapsid protein gene of peste-des-petits-ruminants virus: relationship to other morbilliviruses. J Gen Virol 1994;75(Pt 1):233–237.

Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem 2002;62:25–49.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. FEBS J 2005;272(20):5129–5148.

Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998;3:473–484.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19(1):26–59.

Dunker AK, Obradovic Z. The protein trinity-linking function and disorder. Nat Biotechnol 2001;19(9):805–806.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6(3):197–208.

Eaton BT, Mackenzie JS, Wang LF. Henipaviruses. In: Fields BN, Knipe DM, Howley PM, editors. Fields virology. 5th ed. Philadelphia: Lippincott-Raven; 2007. pp. 1587–1600.

Egelman EH, Wu SS, Amrein M, Portner A, Murti G. The Sendai virus nucleocapsid exists in at least four different helical states. J Virol 1989;63(5):2233–2243.

Espinoza-Fonseca LM. Reconciling binding mechanisms of intrinsically disordered proteins. Biochem Biophys Res Comm 2009;382(3):479–482.

Fearns R, Collins PL. Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. J Virol 1999;73(7):5852–5864.

Feix JB, Klug CS. Site-directed spin-labeling of membrane proteins and peptide-membrane interactions. Volume 14, In: Berliner L, editor. Biological magnetic resonance. Volume spin labeling: the next millenium. New York: Plenum Press; 1998. pp. 251–281.

Ferron F, Longhi S, Canard B, Karlin D. A practical overview of protein disorder prediction methods. Proteins 2006;65(1):1–14.

Ferron F, Longhi S, Henrissat B, Canard B. Viral RNA-polymerases- a predicted 2'-O-ribose methyltransferase domain shared by all Mononegavirales. Trends Biochem Sci 2002;27(5):222–224.

Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15(1):35–41.

Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 2004;338(5): 1015–1026.

Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting binding regions within disordered proteins. Genome Inform Ser Workshop Genome Inform 1999;10:41–50.

Gely S, Lowry DF, Bernard C, Ringkjobing-Jensen M, Blackledge M, Costanzo S, Darbon H, Daughdrill GW, Longhi S. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. J Mol Recogn 2010;23:435–447.

Gerard FC, Ribeiro Ede A, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RW, Jamin M Jr. Modular organization of rabies virus phosphoprotein. J Mol Biol 2009;388(5):978–996.

Giraudon P, Jacquier MF, Wild TF. Antigenic analysis of African measles virus field isolates: identification and localisation of one conserved and two variable epitope sites on the NP protein. Virus Res 1988;10(2–3):137–152.

Gombart AF, Hirano A, Wong TC. Conformational maturation of measles virus nucleocapsid protein. J Virol 1993;67(7):4133–4141.

Gopinath M, Shaila MS. Recombinant L and P protein complex of Rinderpest virus catalyses mRNA synthesis *in vitro*. Virus Res 2008;135(1):150–154.

Green TJ, Zhang X, Wertz GW, Luo M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 2006;313(5785):357–360.

Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 2003;28(2):81–85.

Gupta AK, Shaji D, Banerjee AK. Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA. J Virol 2003;77(1):732–738.

Habchi J, Mamelli L, Darbon H, Longhi S. Structural disorder within henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. PLoS ONE 2010;5(7):e11684.

Halpin K, Bankamp B, Harcourt BH, Bellini WJ, Rota PA. Nipah virus conforms to the rule of six in a minigenome replication assay. J Gen Virol 2004;85(Pt 3):701–707.

Hartlieb B, Modrof J, Muhlberger E, Klenk HD, Becker S. Oligomerization of Ebola virus VP30 is essential for viral transcription and can be inhibited by a synthetic peptide. J Biol Chem 2003;278(43):41830–41836.

Harty RN, Palese P. Measles virus phosphoprotein (P) requires the NH2- and COOH-terminal domains for interactions with the nucleoprotein (N) but only the COOH terminus for interactions with itself. J Gen Virol 1995;76(Pt 11):2863–2867.

Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2006;2(8):e100.

Heggeness MH, Scheid A, Choppin PW. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. Proc Natl Acad Sci USA 1980;77(5):2631–2635.

Heggeness MH, Scheid A, Choppin PW. The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. Virology 1981;114(2):555–562.

Horikami SM, Curran J, Kolakofsky D, Moyer SA. Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication *in vitro*. J Virol 1992;66(8):4901–4908.

Horikami SM, Moyer SA. Structure, transcription, and replication of measles virus. Curr Top Microbiol Immunol 1995;191:35–50.

Houben K, Blanchard L, Blackledge M, Marion D. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. Biophys J 2007a;93(8):2830–2844.

Houben K, Marion D, Tarbouriech N, Ruigrok RW, Blanchard L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J Virol 2007b;81(13): 6807–6816.

Hua QX, Jia WH, Bullock BP, Habener JF, Weiss MA. Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. Biochemistry 1998;37(17):5858–5866.

Hubbell WL, Gross A, Langen R, Lietzow MA. Recent advances in site-directed spin labeling of proteins. Curr Opin Struct Biol 1998;8(5):649–656.

Huber M, Cattaneo R, Spielhofer P, Orvell C, Norrby E, Messerli M, Perriard JC, Billeter MA. Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. Virology 1991;185(1):299–308.

Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein. J Virol 2009;83(20):10374–10383.

Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. J Biol Chem 2003;278(45):44567–44573.

Jordan IK, Sutter BA, McClure MA. Molecular evolution of the Paramyxoviridae and Rhabdoviridae multiple-protein-encoding P gene. Mol Biol Evol 2000;17(1):75–86.

Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84(Pt 12):3239–3252.

Karlin D, Longhi S, Canard B. Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. Virology 2002;302(2):420–432.

Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. Virology 2002;296(2):251–262.

Kavalenka A, Urbancic I, Belle V, Rouger S, Costanzo S, Kure S, Fournel A, Longhi S, Guigliarelli B, Strancar J. Conformational analysis of the partially disordered measles virus NTAIL-XD complex by SDSL EPR spectroscopy. Biophys J 2010;98(6):1055–1064.

Kerdiles YM, Cherif B, Marie JC, Tremillon N, Blanquier B, Libeau G, Diallo A, Wild TF, Villiers MB, Horvat B. Immunomodulatory properties of morbillivirus nucleoproteins. Viral Immunol 2006;19(2):324–334.

Kingston RL, Gay LS, Baase WS, Matthews BW. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. J Mol Biol 2008;379(4):719–731.

Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci USA 2004;101(22):8301–8306.

Kingston RL, Walter AB, Gay LS. Characterization of nucleocapsid binding by the measles and the mumps virus phosphoprotein. J Virol 2004;78(16):8615–8629.

Kouznetzoff A, Buckle M, Tordo N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. J Gen Virol 1998;79(Pt 5): 1005–1013.

Kovacs E, Tompa P, Liliom K, Kalmar L. Dual coding in alternative reading frames correlates with intrinsic protein disorder. Proc Natl Acad Sci USA 2010;107(12): 5429–5434.

Lacy ER, Filippov I, Lewis WS, Otieno S, Xiao L, Weiss S, Hengst L, Kriwacki RW. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. Nat Struct Mol Biol 2004;11(4):358–364.

Laine D, Bourhis J, Longhi S, Flacher M, Cassard L, Canard B, Sautés-Fridman C, Rabourdin-Combe C, Valentin H. Measles virus nucleoprotein induces cell proliferation arrest and apoptosis through NTAIL/NR and NCORE/FcgRIIB1 interactions, respectively. J Gen Virol 2005;86(6):1771–1784.

Laine D, Trescol-Biémont M, Longhi S, Libeau G, Marie J, Vidalain P, Azocar O, Diallo A, Canard B, Rabourdin-Combe C, Valentin H. Measles virus nucleoprotein binds to a novel cell surface receptor distinct from FcgRII via its C-terminal domain: role in MV-induced immunosuppression. J Virol 2003;77(21):11332–11346.

Lamb RA, Kolakofsky D. Paramyxoviridae: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM, editors. Fields virology. 4th ed. Philadelphia, PA: Lippincott-Raven; 2001. pp. 1305–1340.

Lamb RA, Parks GD. Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, editors. Fields virology. 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2007. pp. 1450–1497.

Li J, Rahmeh A, Morelli M, Whelan SP. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. J Virol 2008;82(2):775–784.

Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. BMC Genomics 2008;9(Suppl 2):S25.

Liston P, Batal R, DiFlumeri C, Briedis DJ. Protein interaction domains of the measles virus nucleocapsid protein (NP). Arch Virol 1997;142(2):305–321.

Liston P, DiFlumeri C, Briedis DJ. Protein interactions entered into by the measles virus P, V, and C proteins. Virus Res 1995;38(2–3):241–259.

Llorente MT, Barreno-Garcia B, Calero M, Camafeita E, Lopez JA, Longhi S, Ferron F, Varela PF, Melero JA. Structural analysis of the human respiratory syncitial virus phosphoprotein: characterization of an a-helical domain involved in oligomerization. J Gen Virol 2006;87:159–169.

Longhi S. Nucleocapsid structure and function. Curr Top Microbiol Immunol 2009;329: 103–128.

Longhi S, Canard B. Mécanismes de transcription et de réplication des paramyxoviridae. Virologie 1999;3(3):227–240.

Longhi S, Oglesbee M. Structural disorder within the measles virus nucleoprotein and phosphoprotein. Protein Pept Lett 2010;17(8):961–978.

Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 2003;278(20):18638–18648.

Luo M, Green TJ, Zhang X, Tsao J, Qiu S. Conserved characteristics of the rhabdovirus nucleoprotein. Virus Res 2007;129(2):246–251.

Marie JC, Kehren J, Trescol-Biemont MC, Evlashev A, Valentin H, Walzer T, Tedone R, Loveland B, Nicolas JF, Rabourdin-Combe C, Horvat B. Mechanism of measles virus-induced suppression of inflammatory immune responses. Immunity 2001;14(1):69–79.

Marion D, Tarbouriech N, Ruigrok RW, Burmeister WP, Blanchard L. Assignment of the 1H, 15N and 13C resonances of the nucleocapsid- binding domain of the Sendai virus phosphoprotein. J Biomol NMR 2001;21(1):75–76.

Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J Mol Biol 2004;343(4):819–831.

McNulty BC, Young GB, Pielak GJ. Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. J Mol Biol 2006;355(5):893–897.

Melcher K. The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. J Mol Biol 2000;301(5):1097–1112.

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362(5): 1043–1059.

Money VA, McPhee HK, Mosely JA, Sanderson JM, Yeo RP. Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. Proc Natl Acad Sci USA 2009;106(11):4441–4446.

Morin B, Bourhis JM, Belle V, Woudstra M, Carriére F, BGuigliarelli B, Fournel A, Longhi S. Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling EPR spectroscopy. J Phys Chem B 2006;110(41):20596–20608.

Moyer SA, Baker SC, Horikami SM. Host cell proteins required for measles virus reproduction. J Gen Virol 1990;71(Pt4):775–783.

Myers TM, Pieters A, Moyer SA. A highly conserved region of the Sendai virus nucleocapsid protein contributes to the NP-NP binding domain. Virology 1997;229(2): 322–335.

Myers TM, Smallwood S, Moyer SA. Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. J Gen Virol 1999;80(Pt6):1383–1391.

Narechania A, Terai M, Burk RD. Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. J Gen Virol 2005;86(Pt5):1307–1313.

Ogino T, Kobayashi M, Iwama M, Mizumoto K. Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. J Biol Chem 2005;280(6):4429–4435.

Oglesbee M. Nucleocapsid protein interactions with the major inducible 70kDa heat shock protein. In: Longhi S, editor. Measles virus nucleoprotein. Hauppage, NY: Nova Publishers Inc.; 2007. pp. 53–98.

Oglesbee M, Ringler S, Krakowka S. Interaction of canine distemper virus nucleocapsid variants with 70K heat-shock proteins. J Gen Virol 1990;71(Pt7):1585–1590.

Oglesbee M, Tatalick L, Rice J, Krakowka S. Isolation and characterization of canine distemper virus nucleocapsid variants. J Gen Virol 1989;70(Pt9):2409–2419.

Oglesbee MJ, Kenney H, Kenney T, Krakowka S. Enhanced production of morbillivirus gene-specific RNAs following induction of the cellular stress response in stable persistent infection. Virology 1993;192(2):556–567.

Oglesbee MJ, Liu Z, Kenney H, Brooks CL. The highly inducible member of the 70kDa family of heat shock proteins increases canine distemper virus polymerase activity. J Gen Virol 1996;77(Pt9):2125–2135.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled Folding and Binding with alpha-Helix-Forming Molecular Recognition Elements. Biochemistry 2005;44(37):12454–12470.

Plumet S, Duprex WP, Gerlier D. Dynamics of viral RNA synthesis during measles virus infection. J Virol 2005;79(11):6900–6908.

Qanungo KR, Shaji D, Mathur M, Banerjee AK. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. Proc Natl Acad Sci USA 2004;101(16):5952–5957.

Qin BY, Liu C, Lam SS, Srinath H, Delston R, Correia JJ, Derynck R, Lin K. Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. Nat Struct Biol 2003;10(11):913–921.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007;92(5):1439–1456.

Rahaman A, Srinivasan N, Shamala N, Shaila MS. Phosphoprotein of the rinderpest virus forms a tetramer through a coiled coil region important for biological function. A structural insight. J Biol Chem 2004;279(22):23606–23614.

Rahmeh AA, Li J, Kranzusch PJ, Whelan SP. Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. J Virol 2009;83(21):11043–11050.

Rancurel C, Khosravi M, Dunker KA, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J Virol 2009;83(20):10719–10736.

Receveur-Bréchot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Protein Struct Funct Bioinform 2006;62:24–45.

Ribeiro EA, Favier A, Gerard FC, Leyrat C, Brutscher B, Blondel D, Ruigrok RW, Blackledge M, Jamin M Jr. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. J Mol Biol 2008;382(2):525–538.

Robbins SJ, Bussell RH. Structural phosphoproteins associated with purified measles virions and cytoplasmic nucleocapsids. Intervirology 1979;12(2):96–102.

Robbins SJ, Bussell RH, Rapp F. Isolation and partial characterization of two forms of cytoplasmic nucleocapsids from measles virus-infected cells. J Gen Virol 1980;47(2):301–310.

Roux L. Dans le génome des Paramyxovirinae, les promoteurs et leurs activités sont façonnés par la régle de six. Virologie 2005;9(1):19–34.

Rudolph MG, Kraus I, Dickmanns A, Eickmann M, Garten W, Ficner R. Crystal structure of the borna disease virus nucleoprotein. Structure (Camb) 2003;11(10):1219–1226.

Ryan KW, Portner A. Separate domains of Sendai virus P protein are required for binding to viral nucleocapsids. Virology 1990;174(2):515–521.

Sato H, Masuda M, Miura R, Yoneda M, Kai C. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. Virology 2006;352(1):121–130.

Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RW. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. J Virol 2001;75(1):490–498.

Schoehn G, Mavrakis M, Albertini A, Wade R, Hoenger A, Ruigrok RW. The 12 A structure of trypsin-treated measles virus N-RNA. J Mol Biol 2004;339(2):301–312.

Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 2000;97(16): 8868–8873.

Sivakolundu SG, Bashford D, Kriwacki RW. Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. J Mol Biol 2005;353(5): 1118–1128.

Smallwood S, Ryan KW, Moyer SA. Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. Virology 1994;202(1):154–163.

Spehner D, Drillien R, Howley PM. The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. Virology 1997;232(2):260–268.

Spehner D, Kirn A, Drillien R. Assembly of nucleocapsidlike structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein. J Virol 1991;65(11):6296–6300.

Stallcup KC, Wechsler SL, Fields BN. Purification of measles virus and characterization of subviral components. J Virol 1979;30(1):166–176.

Sweetman DA, Miskin J, Baron MD. Rinderpest virus C and V proteins interact with the major (L) component of the viral polymerase. Virology 2001;281(2):193–204.

Tan WS, Ong ST, Eshaghi M, Foo SS, Yusoff K. Solubility, immunogenicity and physical properties of the nucleocapsid protein of Nipah virus produced in Escherichia coli. J Med Virol 2004;73(1):105–112.

Tapparel C, Maurice D, Roux L. The activity of Sendai virus genomic and antigenomic promoters requires a second element past the leader template regions: a motif (GNNNNN)3 is essential for replication. J Virol 1998;72(4):3117–3128.

Tarbouriech N, Curran J, Ebel C, Ruigrok RW, Burmeister WP. On the domain structure and the polymerization state of the sendai virus P protein. Virology 2000a;266(1): 99–109.

Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nat Struct Biol 2000b;7(9):777–781.

Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagné N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eleouet JF, Rey FA. 3D structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. Science 2009;326:1279–1283.

tenOever BR, Servant MJ, Grandvaux N, Lin R, Hiscott J. Recognition of the measles virus nucleocapsid as a mechanism of IRF-3 activation. J Virol 2002;76(8):3659–3669.

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? Trends Biochem Sci 2009;34(2):53–59.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27(10):527–533.

Tompa P. The functional benefits of disorder. J Mol Struct (Theochem) 2003;666–667; 361–371.

Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci 2008;33(1):2–8.

Tsai CD, Ma B, Kumar S, Wolfson H, Nussinov R. Protein folding: binding of conformationally fluctuating building blocks via population selection. Crit Rev Biochem Mol Biol 2001a;36(5):399–433.

Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R. Structured disorder and conformational selection. Protein Struct Funct Bioinform 2001b;44(4):418–427.

Uversky VN. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. Expert Rev Proteomics 2010;7(4):543–564.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11(4):739–756.

Uversky VN, Gillespie JR, Fink AL. Why are natively unfolded proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–427.

Uversky VN, Li J, Souillac P, Jakes R, Goedert M, Fink AL. Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma- synucleins. J Biol Chem 2002;25:25.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 2005;18(5):343–384.

Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK. Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 2007;6(6):2351–2366.

Vasconcelos D, Norrby E, Oglesbee M. The cellular stress response increases measles virus-induced cytopathic effect. J Gen Virol 1998a;79(Pt7):1769–1773.

Vasconcelos DY, Cai XH, Oglesbee MJ. Constitutive overexpression of the major inducible 70kDa heat shock protein mediates large plaque formation by measles virus. J Gen Virol 1998b;79(Pt9):2239–2247.

Vincent S, Tigaud I, Schneider H, Buchholz CJ, Yanagi Y, Gerlier D. Restriction of measles virus RNA synthesis by a mouse host cell line: trans-complementation by polymerase components or a human cellular factor(s). J Virol 2002;76(12):6121–6130.

Wang LF, Yu M, Hansson E, Pritchard LI, Shiell B, Michalski WP, Eaton BT. The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. J Virol 2000;74(21):9972–9979.

Warnes A, Fooks AR, Dowsett AB, Wilkinson GW, Stephenson JR. Expression of the measles virus nucleoprotein gene in Escherichia coli and assembly of nucleocapsid-like structures. Gene 1995;160(2):173–178.

Watanabe N, Kawano M, Tsurudome M, Kusagawa S, Nishio M, Komada H, Shima T, Ito Y. Identification of the sequences responsible for nuclear targeting of the V protein of human parainfluenza virus type 2. J Gen Virol 1996;77(Pt2):327–338.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293(2):321–331.

Yegambaram K, and Kingston RL. The feet of the measles virus polymerase bind the viral nucleocapsid protein at a single site. Protein Sci 2010;19(4):893–899.

Zhang X, Bourhis JM, Longhi S, Carsillo T, Buccellato M, Morin B, Canard B, Oglesbee M. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. Virology 2005;337(1):162–174.

Zhang X, Glendening C, Linke H, Parks CL, Brooks C, Udem SA, Oglesbee M. Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. J Virol 2002;76(17):8737–8746.

4

STRUCTURAL DISORDER WITHIN SENDAI VIRUS NUCLEOPROTEIN AND PHOSPHOPROTEIN

MALENE RINGKJØBING JENSEN, PAU BERNADÓ, ROB W. H. RUIGROK, AND MARTIN BLACKLEDGE

4.1 INTRODUCTION

The anticipation that protein structure determination can provide the key to molecular function has been challenged in recent years by the realization that up to 40% of the proteins encoded by the human genome are intrinsically disordered or contain disordered regions of significant length (>50 residues) (Tompa, 2002; Uversky, 2002; Fink, 2005; Dyson and Wright, 2005b; Dunker et al., 2008). Intrinsically disordered proteins (IDPs) are functional despite the lack of a well-defined three-dimensional structure and, therefore, escape the attention of classical structure biology (Wright and Dyson, 1999).

One of the more intriguing characteristics of IDPs is their capacity to undergo structural transitions to partially folded forms on binding to physiological partners (Demchenko, 2001; Dyson and Wright, 2002; Sugase et al., 2007). Such disorder-to-order transitions are a key feature of the complex relationship between the structural dynamics and function of IDPs. The thermodynamics and kinetics of the coupled folding and binding reaction are certainly influenced by the intrinsic propensity of the chain to form its bound conformation, although the molecular mechanisms controlling this behavior remain largely unknown. Therefore, in order to obtain a complete picture of how these proteins carry out their biological function, atomic resolution models of the proteins in their free, pre-recognition states and in complex with their partners are necessary, as well as a determination of

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

the kinetics and the exact nature of the associated disorder-to-order transitions. In the absence of appropriate technology to describe this information, it has become essential to develop new experimental methods that allow an atomic resolution insight into the relationship between primary sequence and molecular function of IDPs.

Over the last few years, it has become increasingly clear that intrinsic disorder is highly abundant within viruses (Xue et al., 2010). Even though the viral genome often codes for a restricted number of proteins, they engage in a multitude of interactions in order for the transcription and replication machinery to function. IDPs play a crucial role in this context, ensuring highly flexible and dynamic interactions. Here, we describe structural disorder within one of the paramoxyviruses, namely, Sendai virus (SeV) (Skiadopoulos et al., 2002). The family of paramyxoviruses consists of enveloped viruses with nonsegmented, negative-stranded RNA genomes and includes, in addition to SeV, measles, Hendra, Nipah, and Mumps viruses.

SeV is responsible for a highly transmissible respiratory tract infection in rats, mice, and hamsters (Skiadopoulos et al., 2002). In this virus, the viral RNA is protected by multiple copies of the nucleoprotein N, forming a helical nucleocapsid (Arnheiter et al., 1985). Protein N is composed of a highly conserved core region, N_{CORE} (amino acids 1–400) necessary for interaction with the RNA, and a C-terminal domain, N_{TAIL} (amino acids 401–524) (Fig. 4.1). The atomic structure of the core region of N is not known, while N_{TAIL} is predicted to be an IDP (Houben et al., 2007).

The polymerase for viral transcription and replication consists of a complex of two proteins, the phosphoprotein P and the large protein L. The L protein holds the polymerase activity, while P serves as a cofactor to stabilize L and to place the polymerase complex on the N–RNA template. L alone is unable to interact with N–RNA. No atomic resolution structural information has so far been obtained for the L protein; however, recent mutational analyses have identified residues within L that are important for virus replication (Murphy et al., 2010).

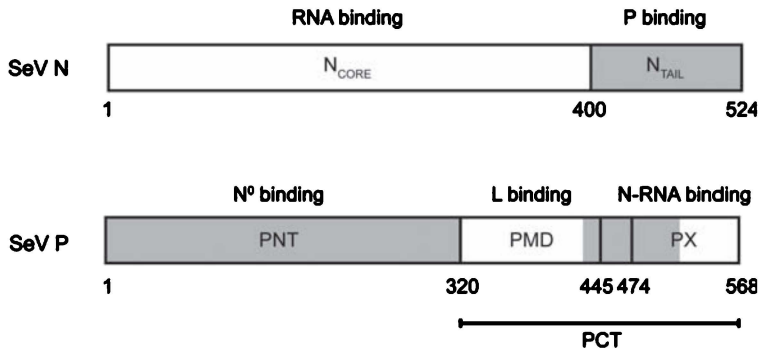


Figure 4.1 Domain organization of SeV nucleoprotein N and phosphoprotein P. Regions predicted to possess high levels of intrinsic disorder are indicated in gray.

The SeV P is a 568-amino-acid protein made up of two main domains: the N-terminal domain (PNT, amino acids 1–319) and the C-terminal domain (PCT, amino acids 320–568) (Fig. 4.1). PNT is believed to possess high levels of intrinsic disorder (Karlin et al., 2002, 2003; Gerard et al., 2009). So far, no further structural information has been obtained in this region; however, it has been shown that a small part of this domain (amino acids 33–41) plays an essential role in the formation of the N^0–P complex, which keeps N in a soluble and RNA-free form during the initial stages of genome replication (Curran et al., 1995). PCT also possesses intrinsic disorder and consists of two subdomains, the oligomerization domain (P multimerization domain (PMD), amino acids 320–445) and the X domain (PX, amino acids 474–568) (Fig. 4.1).

Positioning of the polymerase complex on the N–RNA template requires an apparently dynamic interaction between N_{TAIL} and PX (Curran et al., 1993; Tuckis et al., 2002). N_{TAIL} is believed to undergo α-helical folding of the interaction site on binding to PX. This chapter describes the structural characterization of N_{TAIL} and PCT using a powerful combination of X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and small angle X-ray scattering (SAXS). NMR is one of the most powerful tools for obtaining an atomic resolution insight into IDPs and their structural transitions and physiological interactions, while SAXS provides information describing the overall dimensions of proteins and, as such, is a highly complementary technique to NMR.

4.2 CHARACTERIZING DISORDERED PROTEINS BY NMR SPECTROSCOPY

The study of IDPs is complicated by the fact that these proteins are highly dynamic and therefore do not adopt a single conformation in solution but rather a continuum of rapidly interconverting structures. Classical methods of structural biology are inadequate for characterizing the conformational properties of IDPs because of the inherent conformational heterogeneity of the chain. Thus, crystallization of IDPs, even if successful, could only capture at most one conformation from a dynamic continuum and thereby provide only limited information. Any representation of the disordered chain as a single structure would be inappropriate, and an ensemble of interconverting conformers provides a more representative description.

NMR spectroscopy has recently evolved into one of the most powerful techniques for studying IDPs (Dyson and Wright, 2004, 2005a; Mittag and Forman-Kay, 2007; Eliezer, 2009; Jensen et al., 2009). Although the protein adopts many different conformers in solution, NMR allows site-specific characterization of average properties over all conformers of the disordered chain. Even the simplest measurement, such as the chemical shifts of the backbone nuclei, depends on a population-weighted average over all rapidly exchanging local conformations sampled by all molecules in the ensemble, on timescales up to the millisecond (Jensen et al., 2010). Not surprisingly then, the average chemical shifts measured from a broad conformational equilibrium can be interpreted in terms of local conformational

propensity of the ensemble through the so-called secondary chemical shift (SCS) (Wishart et al., 1992; Wishart and Sykes, 1994a, 1994b):

$$\delta_{SCS} = \delta_{EXP} - \delta_{RC} \tag{4.1}$$

Here, δ_{EXP} is the experimental chemical shift and δ_{RC} is the tabulated amino-acid-specific random coil (RC) values, that is, the expected chemical shifts if no secondary structure is present (Wishart et al., 1995; Zhang et al., 2003). In particular, Cα, Cβ, and C′ chemical shifts are highly sensitive to the presence of secondary structures. For example, α-helical propensity is characterized by positive Cα SCSs, while β-sheet propensity is characterized by negative Cα SCSs. In addition, by comparing the experimental chemical shifts with those expected for a fully formed α-helix or β-sheet, an estimate of the population of secondary structures along the disordered chain can be obtained. This is indeed the basis of the secondary structure propensity (SSP) score proposed recently, which combines the chemical shifts of multiple nuclei (Cα, Cβ, and C′) in one population score, while at the same time avoiding the common chemical shift referencing errors (Marsh et al., 2006).

Dipolar couplings are also very sensitive probes of time and ensemble-averaged conformational equilibria exchanging on timescales up to the millisecond and can therefore be used to characterize both the structure and dynamics of disordered proteins (Jensen et al., 2009). The dipolar coupling D_{ij} between two spins i and j is given by (Emsley and Lindon, 1975)

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3} \left\langle \frac{3\cos^2 \Omega(t) - 1}{2} \right\rangle \tag{4.2}$$

where Ω is the angle of the internuclear vector relative to the direction of the static magnetic field; γ_i and γ_j are the gyromagnetic ratios of spin i and j, respectively; and r is the internuclear distance. The parentheses indicate an average over all conformation exchanging on timescales faster than the millisecond. In solution NMR, the dipolar coupling between two nuclei is effectively averaged to zero because all orientations of the protein molecule are equally probable (isotropic solution). However, the dipolar coupling can be reintroduced, at least partially, by weakly aligning the protein molecules in the magnetic field, for example, using an anisotropic solution (Tjandra and Bax, 1997), or exploiting the magnetic anisotropy of paramagnetic metal ions (Tolman et al., 1995). These so-called residual dipolar couplings (RDCs) add to the experimentally measured scalar coupling (J-coupling), and RDCs can therefore be obtained by comparing the coupling splitting in the NMR spectra recorded in isotropic and anisotropic solutions. A variety of anisotropic solutions exist for weakly aligning proteins in the magnetic field, for example, lipid bicelles (Tjandra and Bax, 1997), filamentous phages (Torbet and Maret, 1979; Clore et al., 1998; Hansen et al., 1998), lyotropic ethylene glycol/alcohol phases (Rückert and Otting, 2000), and polyacrylamide gels, that have been strained either laterally or longitudinally to produce anisotropic cavities (Sass et al., 2000; Tycko et al., 2000). Alignment results from either a steric repulsion between the protein and the medium or a combination of electrostatic and steric interactions.

Experimentally measured RDCs report on orientations of internuclear bond vectors (e.g., N-HN, Cα-Hα, Cα-C′, and HN-C′) that can be expressed with respect to a common external reference frame in terms of a second-rank tensor describing the overall alignment of the protein molecule in the magnetic field (Blackledge, 2005):

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3}\left[A_a(3\cos^2\theta - 1) + \frac{3}{2}(A_r \sin^2\theta \cos 2\varphi)\right] \quad (4.3)$$

Here, A_a and A_r are the axial and rhombic components of the alignment tensor, respectively, and θ and φ represent the orientation of the internuclear vector with respect to the alignment tensor. RDCs have been used extensively in the determination of the structure of folded proteins (Tjandra et al., 1997; Bax, 2003; Bax and Grishaev, 2005) and for determining the relative orientation of proteins involved in protein–protein complexes (Clore, 2000; Ortega-Roldan et al., 2009). In addition, RDCs from several different alignment media with independent alignment tensors have been used to report on local dynamics occurring in proteins on timescales up to the millisecond (Meiler et al., 2001; Tolman, 2002; Clore and Schwieters, 2004; Bouvignies et al., 2005; Lange et al., 2008; Salmon et al., 2009). For a steric alignment medium, the alignment tensor can be accurately predicted from the overall shape of the protein molecules, for example, using the program PALES (Zweckstetter and Bax, 2000; Zweckstetter, 2008), allowing one to obtain information about average bond vector orientations in IDPs. Below, we demonstrate the remarkable ability of chemical shifts and RDCs in combination with SAXS to describe the conformational behavior of the intrinsically disordered domains of SeV P and N.

4.3 STRUCTURAL CHARACTERIZATION OF SENDAI VIRUS PHOSPHOPROTEIN

The structure of PMD (see Fig. 4.1 for the domain organization of P) was solved by X-ray crystallography (Tarbouriech et al., 2000). It is a tetramer in which each monomer is composed of a long α-helix that forms a tetrameric coiled coil and a helical bundle of three short α-helices (Fig. 4.2). The central domain contains the binding site for protein L, in which the residues K408, R409, E412, K415, and E416 have been found to be essential for the formation of the P–L complex (Bowman et al., 1999).

The structure of PX was studied by NMR spectroscopy. According to prediction, this domain contains both structured and disordered parts (Fig. 4.1). The ^{1}H-^{15}N HSQC spectrum of PX contains both sharp resonances in a reduced ^{1}H chemical shift range (8.0–8.5 ppm) indicative of a disordered chain, as well as slightly broader resonances in a more dispersed ^{1}H chemical shift range (6.8–9.3 ppm) indicative of a globular domain (Fig. 4.3) (Marion et al., 2001). Using standard short-range (<6 Å) distance constraints between protons (nuclear Overhauser enhancements (NOEs)), it was found that PX contains a folded, three-helix bundle structure (amino acids 516–568) and a long disordered strand (amino acids

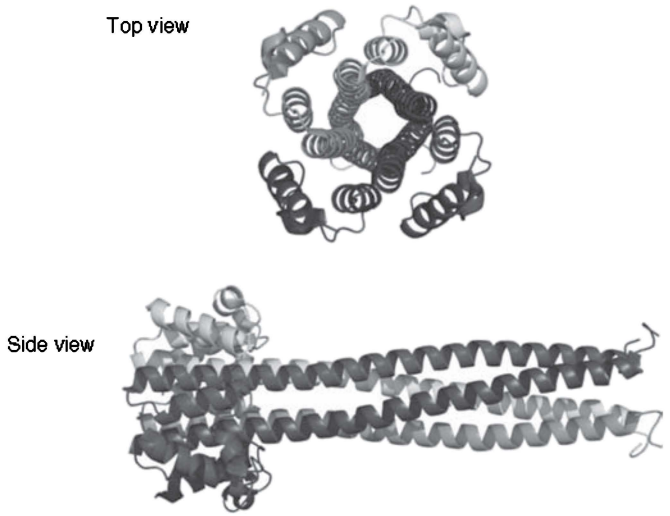


Figure 4.2 Two views of the crystal structure of the tetramerization domain (PMD) of P (PDB accession number 1EZJ).

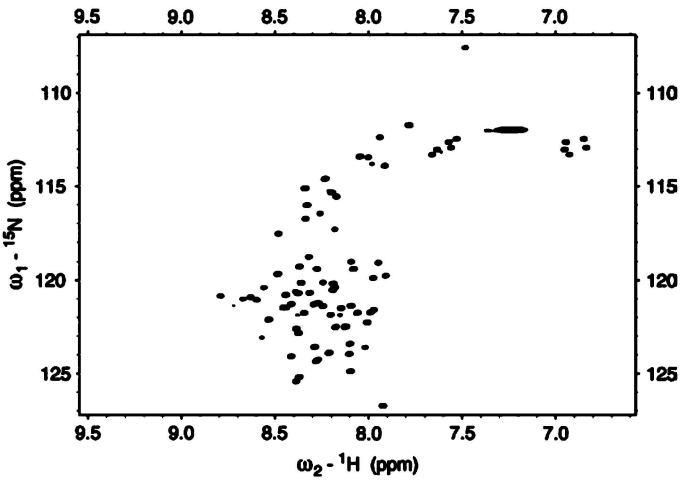


Figure 4.3 ^{1}H-^{15}N HSQC spectrum of PX. The sample contained 50 mM phosphate buffer and 0.5 M NaCl at pH 6.0. The spectrum was recorded at a ^{1}H frequency of 800 MHz at 25°C.

474–515) via which it is attached to the tetramerization domain of P (Fig. 4.4) (Blanchard et al., 2004).

One of the major challenges in the study of IDPs is to develop explicit ensemble descriptions on the basis of experimental data that in turn can be used to understand the conformational behavior of the unfolded chains. A structural model of the disordered domain of PX was obtained on the basis of experimental RDCs and SAXS (Bernadó et al., 2005). The experimental N-HN and C′-HN RDCs could be

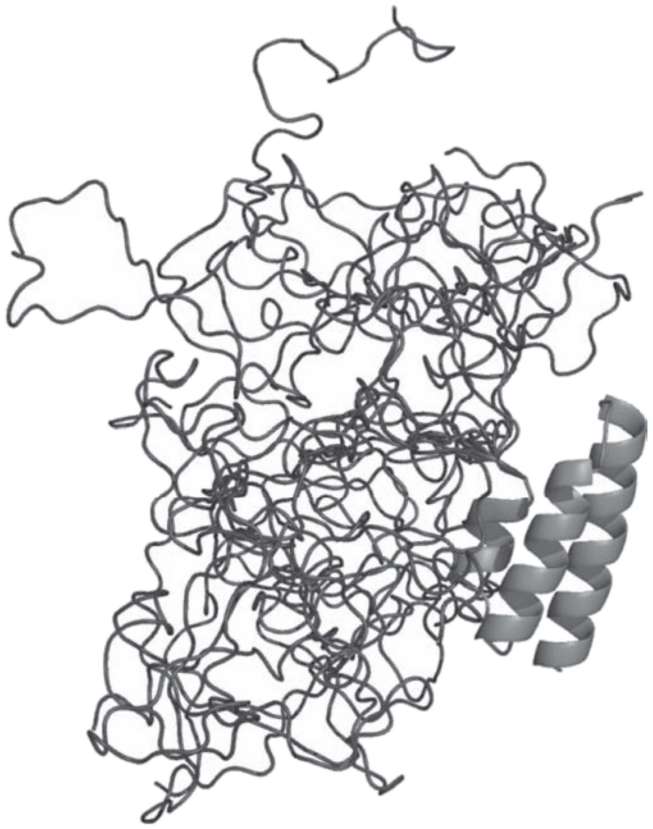


Figure 4.4 Representation of the structure of PX consisting of a folded three-helix bundle domain and a long disordered strand via which it is attached to the tetramerization domain of P (PDB accession number 1R4G). Fifteen copies of the unfolded domain are shown.

obtained in the disordered domain of PX despite the heterogeneous conformational behavior of the chain, although the couplings are smaller than those measured in the folded domain (Fig. 4.5). In order to model the diverse conformational sampling in the disordered domain of PX, an amino-acid-specific random-coil ϕ/ψ database was created from loop regions of high resolution X-ray structures. A structural ensemble of PX was generated using Monte-Carlo-based sampling of the coil database, while at the same time avoiding clashes between different parts of the chain using a simple steric exclusion model (Bernadó et al., 2005). This algorithm, *Flexible-Meccano*, allows for a very efficient, restraint-free sampling to fill the conformational space available to the disordered sequence. The alignment tensor was estimated on the basis of the shape of each member of the ensemble using PALES, and the RDCs were calculated using Equation 4.3 and finally averaged over the whole ensemble. To ensure convergence of the RDCs in the disordered domain, 50,000 conformers were included in the PX ensemble. A relatively good

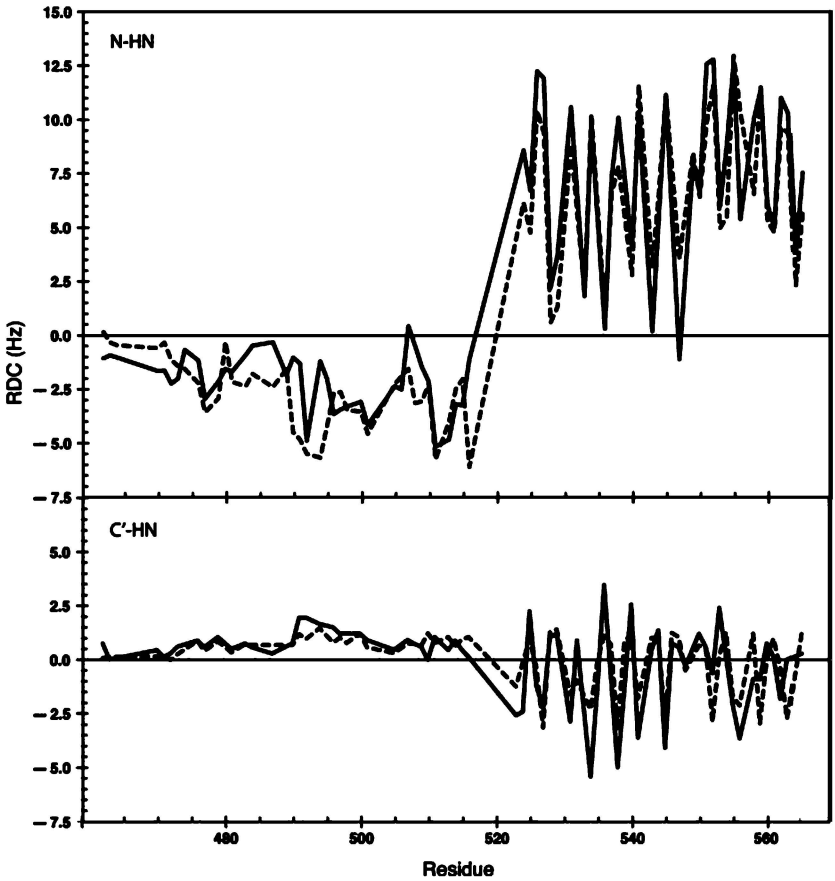


Figure 4.5 Comparison of experimental (solid line) and predicted (dashed line) RDCs in PX. Two types of RDCs are shown: N-HN (top) and C′-HN (bottom). The experimental RDCs were obtained by aligning the protein in a liquid crystal composed of polyethylene glycol and 1-hexanol. The predicted RDCs were obtained from a structural ensemble of PX created using *Flexible-Meccano*. The protein construct contained an N-terminal his-tag of 14 amino acids such that residues 474–568 correspond to intact PX.

reproduction of experimental RDCs is obtained throughout the chain (Fig. 4.5) where the amplitude and distribution of RDCs are reproduced. This, and related observations in other proteins, indicate that loop regions of high resolution X-ray structures constitute a reasonable model of the disordered state. Note that in this particular system, the RDCs from each copy of the protein, in both folded and disordered domains, depend on the relative alignment of the two domains, constituting a quantitative test of the validity of the approach.

The model of PX was further validated using experimental SAXS data obtained under the same conditions as the NMR experiments. Side chains were added to each *Flexible-Meccano* conformer in the unfolded domain using the SCCOMP program (Eyal et al., 2004), and SAXS intensity curves were predicted using the program

CRYSOL (Svergun et al., 1995). The intensities from each of the conformers were averaged, and 2000 conformers were sufficient to achieve convergence of the relevant scattering properties simulated from the ensemble. The effective radius of gyration extracted from the PX ensemble is 28.7 Å, which is in reasonable agreement with the experimental value of 29.7 Å. In addition, the overall shape of the experimental scattering profile agrees well with the profile predicted from the PX ensemble, indicating that the obtained structural model of the disordered domain is reliable (Fig. 4.6).

In order to obtain further insight into the overall dimensions of P, additional SAXS measurements were carried out on PCT. The dimensions of the tetramerization domain were calculated from the crystal structure, resulting in a length of 102 Å, a radius of gyration of 33 Å, and a cross-sectional radius of gyration (measure for the width of the domain) of 13 Å. From the SAXS measurements of PCT, a radius of gyration of 48 Å and a cross-sectional radius of gyration of 14 Å were obtained (Blanchard et al., 2004). In particular, the value of the cross-sectional radius, that is, only 1 Å greater than the corresponding value calculated for PMD, indicates that PCT can be modeled as a cylinder with a width of 40 Å and a length of 160 Å. Although these dimensions may suggest that the four PX domains are positioned close to the axis of the tetramerization domain, significant dynamic averaging due to the flexibility of the unfolded domain may also produce

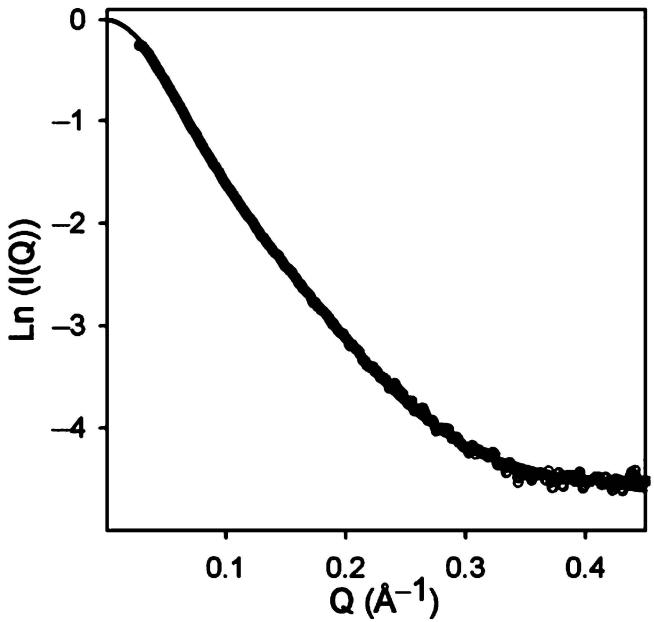


Figure 4.6 Defining the size and shape of PX from SAXS. Comparison of background-corrected experimental SAXS curve (intensity I(Q) vs. scattering vector Q, empty gray circles) and predicted curve (solid line) from the structural ensemble of PX created using *Flexible-Meccano*.

only a small increase in lateral dimensions, even in the presence of large amplitude excursions. A slight overlap between the lengths of PMD (102 Å) and PX (70 Å) was interpreted to give a total length of 160 Å (Blanchard et al., 2004).

To obtain an insight into the P–N interaction, mapping of the interaction site of N_{TAIL} on PX was performed by following the changes in the ^{1}H-^{15}N HSQC spectrum of PX on addition of unlabeled N_{TAIL} (Houben et al., 2007). It was found that N_{TAIL} binds to the folded domain of PX in the cleft between the helices $\alpha 2$ and $\alpha 3$, with an estimated dissociation constant of 57 μM. A surface representation of PX in terms of hydrophobic and electrostatic residues shows that N_{TAIL} binds to a patch on PX that contains predominately negatively charged residues suggesting that the SeV N_{TAIL}–PX interaction is mainly governed by electrostatic forces. This is in contrast to the corresponding N_{TAIL}–PX interaction in measles virus, which is dominated by hydrophobic forces (Kingston et al., 2004).

4.4 STRUCTURAL CHARACTERIZATION OF THE SENDAI VIRUS NUCLEOPROTEIN

Structural characterization of the C-terminal domain, N_{TAIL} of N, was performed using NMR spectroscopy. The ^{1}H-^{15}N HSQC spectrum of N_{TAIL} (residues 443–524) displays poor dispersion of the signals in the ^{1}H dimension, indicating that N_{TAIL} lacks a tertiary structure and therefore most likely belongs to the family of IDPs (Fig. 4.7). The amount of residual structure in N_{TAIL} was initially probed by combining the experimental $C\alpha$ and $C\beta$ chemical shifts in the SSP score, showing that N_{TAIL}, although intrinsically disordered, possesses up to 80% helix in the 476–495 region (Fig. 4.8).

Mapping (in the opposite sense to that described above) of the interaction site of PX on N_{TAIL} was performed by following the changes in the ^{1}H-^{15}N HSQC spectrum of N_{TAIL} on addition of unlabeled PX (Houben et al, 2007). The largest chemical shift changes were observed for the residues 476–495. Interestingly, this region possesses helical propensity already in the prerecognition state, and it was shown that it becomes more rigid on interaction with PX (Houben et al., 2007).

To obtain more detailed information about the α-helical sampling in the interaction site of N_{TAIL}, different types of RDCs (N-HN, Cα-Hα, Cα-C′, and HN-C′) were measured in the protein (Jensen et al., 2008). N-HN RDCs have been shown to be particularly sensitive to α-helical elements because they change sign from negative in a completely disordered protein to positive within helices (Jensen and Blackledge, 2008). The observed change in sign can be explained as follows: the average orientation of the N-HN bond vector in a disordered chain where the protein is preferentially aligned in a direction parallel to the magnetic field, for example, in an elongated cavity, would be expected to be approximately orthogonal to the field and therefore gives rise to negative couplings (Eq. 4.2). Within a helical element, the bond vector would be aligned more or less parallel to the average chain direction and therefore gives rise to positive dipolar couplings (Fig. 4.9). The N-HN RDCs obtained in N_{TAIL} are positive with a wavelike fine structure within the molecular

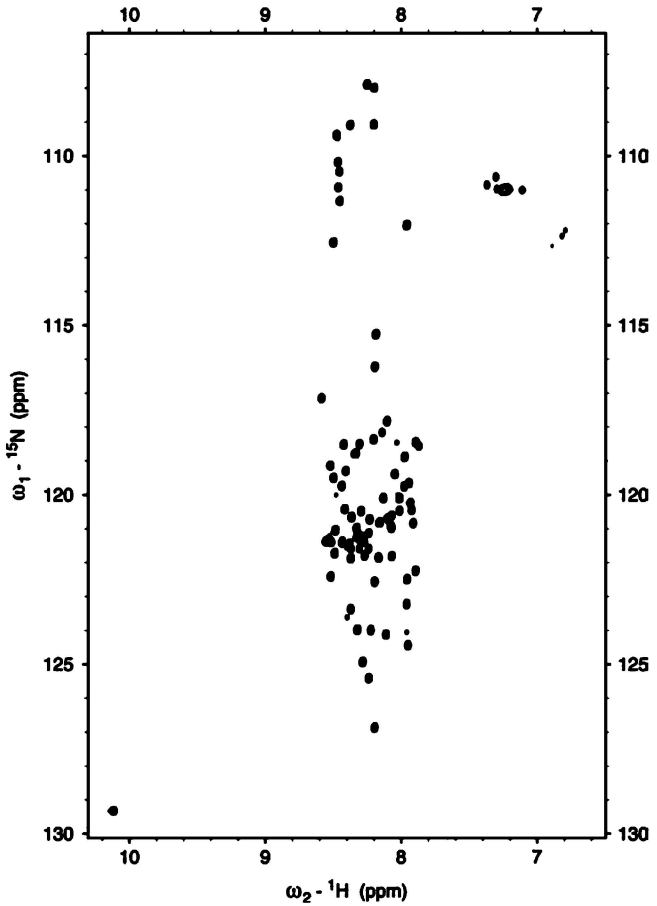


Figure 4.7 ^{1}H-^{15}N HSQC spectrum of N_{TAIL} comprising residues 443–524. The sample contained 50 mM phosphate buffer and 500 ml NaCl at pH 6.0. The spectrum was recorded at a ^{1}H frequency of 600 MHz at 25°C.

recognition element (MoRE), while the RDCs in the disordered chains outside the MoRE are predominately negative (Jensen et al., 2008). The dipolar waves in the MoRE are due to the aligned helical elements are slightly tilted with respect to the direction of the magnetic field. It has been shown that the tilt angle of the helical element with respect to the magnetic field depends in a predictable manner on the direction of the disordered chains projected from the helix caps (Jensen and Blackledge, 2008). Thus, if the disordered chains are projected in approximately the same direction, the tilt angle is close to zero, while a larger tilt angle is obtained if the chains are projected in opposite directions. For these reasons, the size of the RDCs and the observed dipolar waves are sensitive to helix length, population, and capping.

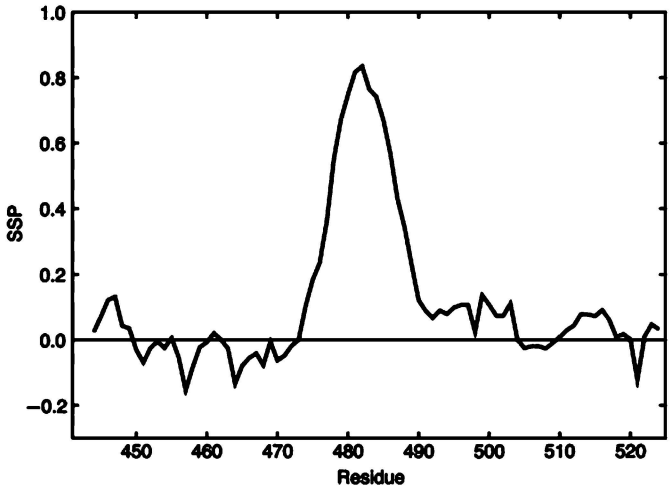


Figure 4.8 Secondary structure propensity (SSP) score of N_{TAIL} (construct comprising residues 443–524). Positive SSP score indicates the formation of α-helical elements, while a negative score indicates the formation of β-sheetlike structure. The SSP score was obtained by combining experimental Cα and Cβ chemical shifts (see text).

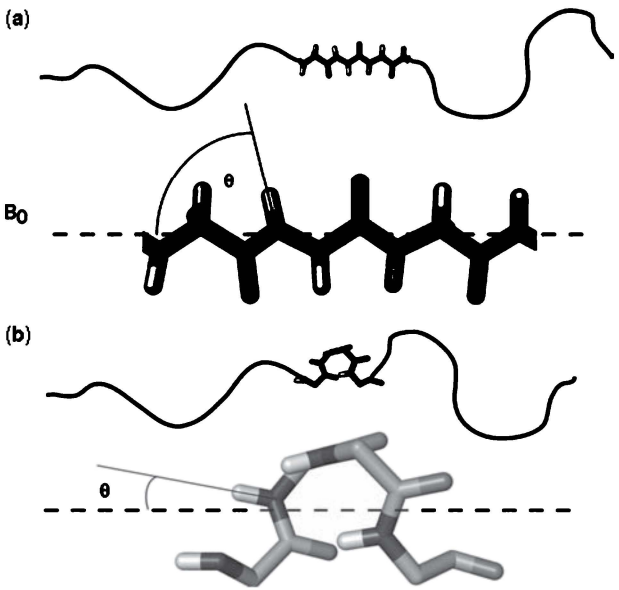


Figure 4.9 Figurative representation of effective angular averaging properties of ^{1}H-^{15}N bond vectors in a disordered protein dissolved in a weakly aligning medium with the director along the magnetic field. (a) RDCs measured for ^{1}H-^{15}N bond vectors in more extended conformations, as commonly found in disordered proteins, will have negative values (Ω close to 90°). (b) RDCs measured for ^{1}H-^{15}N bond vectors in turns or helical conformations will have positive values (Ω close to 0°).

In order to exploit the available information in the experimental RDCs, we developed a minimal ensemble approach that allowed us to quantitatively characterize at the atomic level the α-helical sampling in the MoRE (residues 476–495) of N_{TAIL} (Jensen et al., 2008). We used *Flexible-Meccano* to generate conformational ensembles of N_{TAIL} with helices of different lengths and positions within the MoRE. Each ensemble consisted of 10,000 structures in order to ensure convergence of the RDCs when averaged over the ensemble. A total of 153 different ensembles were necessary to include all possible helices with a minimum length of 4 amino acids to a maximum length of 20 amino acids. This approach serves to find the minimal ensemble that best fits the experimental RDCs by combining the predicted RDCs from the different ensembles and optimizing the populations of the helical elements:

$$D_{ij,\text{eff}} = \sum_{k=1}^{n} p_k D_{ij}^k + \left(1 - \sum_{k=1}^{n} p_k\right) D_{ij}^U \tag{4.4}$$

Here, p_k represents the populations of the n helical conformers for which D_{ij}^k is the individual predicted coupling between nuclei i and j and D_{ij}^U is the predicted coupling from the completely disordered state. These effective couplings are compared to the experimental data through

$$\chi^2 = \sum \frac{\left(D_{ij,\text{eff}} - D_{ij,\text{exp}}\right)^2}{\sigma_{ij}^2} \tag{4.5}$$

where σ_{ij} is the estimated uncertainty on the experimental couplings. All four measured types of couplings were included in the fitting procedure. The number of helical conformers was gradually increased (starting from $n = 1$), and F-tests were applied to verify the statistical significance of the improvement of the fitting of the experimental data. It was found that $n = 3$ is the minimum number of helices that is needed to describe the experimental RDCs (Fig. 4.10). Thus, the MoRE of N_{TAIL} is adopting a conformational equilibrium between three specific helical conformers and a completely unfolded form (Fig. 4.11). A maximum helical propensity in the MoRE of 75% was obtained from the RDCs, which is in good agreement with the value obtained on the basis of the experimental chemical shifts (Fig. 4.8).

The model of N_{TAIL} obtained using the minimal ensemble approach shows several features important for molecular recognition in the N_{TAIL}/PX system. Thus, rather than fraying randomly, the MoRE of N_{TAIL} preferentially populates specific α-helical conformers. As mentioned above, the interaction between N_{TAIL} and PX is most likely mediated by electrostatic forces, where positively charged residues of N_{TAIL} interact with a patch of negative residues located on the surface of PX. Three positively charged side chains, R482, R486, and R490, are found in the MoRE of N_{TAIL}. These arginines are separated by four amino acids in the primary sequence, ensuring that they are located on the same side of helices formed in the prerecognition state of N_{TAIL} (Fig. 4.11).

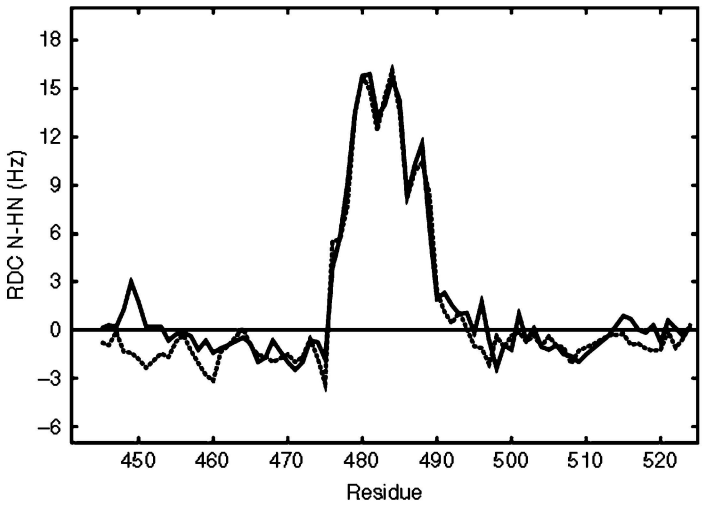


Figure 4.10 Comparison of experimental N-HN RDCs (solid line) and predicted RDCs (dashed line) in N_{TAIL}. The experimental RDCs were obtained by aligning the protein in a liquid crystal composed of polyethylene glycol and 1-hexanol. The predicted RDCs were calculated from the model of N_{TAIL} shown in Fig. 4.11 obtained using the minimal ensemble approach.

Interestingly, the helices selected by the minimal ensemble approach are all preceded by aspartic acids or serines—two amino acids capable of forming stabilizing N-capping interactions (Serrano and Fersht, 1989; Serrano et al., 1992). Thus, these residues act as helix promoters in an otherwise unfolded protein showing how the favored conformations are encoded in the primary sequence of the MoRE. Possibly, equally important is the fact that the direction in which the disordered strands are projected from the helix caps is selectively controlled by the presence of these stabilizing interactions. This suggests a mechanism by which the partially folded form of N_{TAIL} projects the disordered strands in the most functionally useful direction to achieve efficient fly-casting interactions. The fly-casting effect, first hypothesized as a mechanism for the folding of IDPs on binding, promotes interactions between the IDP and its partner by first binding weakly and nonspecifically to its partner, then folding as it approaches its binding conformation (Shoemaker et al., 2000).

In conclusion, the SeV phosphoprotein and nucleoprotein contains large regions of intrinsic disorder, and the combination of NMR spectroscopy and SAXS has proven extremely valuable for obtaining an insight into both local conformational sampling and overall dimensions of such large, highly dynamic proteins. The P protein is envisaged to cartwheel on the N–RNA template via the simultaneous breaking and reforming of contacts with N subunits such that at least two arms of the tetramer continuously engage the N–RNA. It is clear that the intrinsic dynamics of both P and N plays a major role in this context and that the atomic resolution models of P and N provide valuable insight into the molecular basis of transcription and replication of paramyxoviruses.

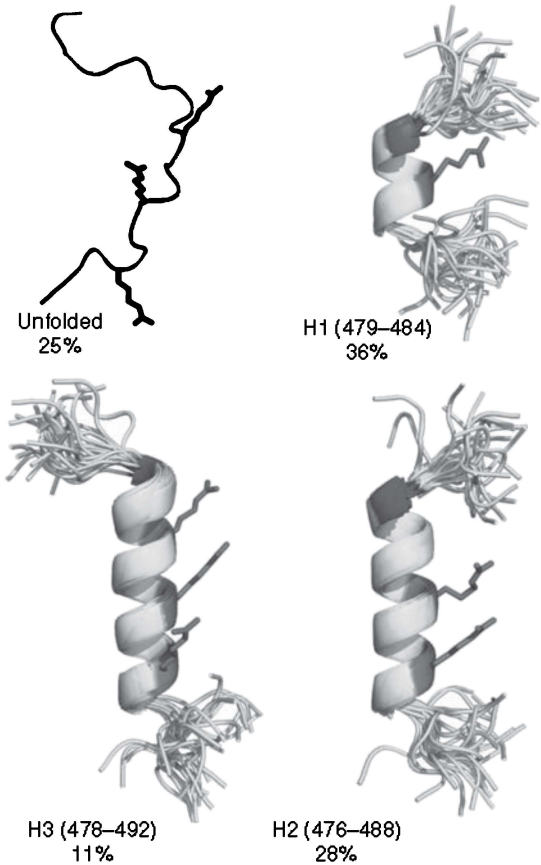


Figure 4.11 The conformational equilibrium adopted by the MoRE of N_{TAIL} in solution. The four most populated conformations are shown as a single structure representing the 25% disordered conformers, a helix comprising the six amino acids 479–484 populated at a level of 36%, a helix 476–488 populated at a level of 28%, and a helix 478–492 populated at a level of 11%. The side chains of the interaction site of arginines are shown in red in each case. Twenty randomly selected conformers are shown for each of the helical segments to illustrate the directionality of the adjacent chains projected from the helix caps. The location of the helices selected using the minimal ensemble approach is shown in the primary sequence. All helices are preceded by aspartic acids or serines—two residues capable of forming N-capping interactions. (*See insert for color representation of the figure.*)

ACKNOWLEDGMENTS

This work was supported by the Commisariat à l'energy atomique et au energies alternatives; the Centre National de la Recherche Scientifique; the Université Joseph Fourier, Grenoble; the Finovi Foundation of Lyon, France; and the ANR PCV 2007 Protein Motion. M. R. J. acknowledges support from Lundbeckfonden, Denmark and EMBO. P. B. holds a Ramón y Cajal contract partly funded by the Spanish Ministerio de Educación y Ciencia. The authors thank the Partnership for Structural Biology for the excellent structural biology environment.

ABBREVIATIONS

HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
MoRE	molecular recognition element
NaCl	sodium chloride
NMR	nuclear magnetic resonance
N	Sendai virus nucleoprotein
N_{CORE}	amino acids 1–400 of Sendai virus nucleoprotein
N_{TAIL}	amino acids 401–524 of Sendai virus nucleoprotein
PEG	polyethylene glycol
P	Sendai virus phosphoprotein
PCT	amino acids 320–568 of Sendai virus phosphoprotein
PMD	amino acids 320–445 of Sendai virus phosphoprotein
PNT	amino acids 1–320 of Sendai virus phosphoprotein
PX	amino acids 474–568 of Sendai virus phosphoprotein
RDC	residual dipolar coupling
RNA	ribonucleic acid
SCS	secondary chemical shift
SSP	secondary structure propensity

REFERENCES

Arnheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. Cell 1985;41:259–267.

Bax A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 2003;12:1–16.

Bax A, Grishaev A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 2005;15:563–570.

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 2005;102:17002–17007.

Blackledge M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog Nucl Magn Reson Spectros 2005;46:23–61.

Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP, Ruigrok RWH, Marion D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. Virology 2004;319:201–211.

Bouvignies G, Bernadó P, Meier S, Cho K, Grzesiek S, Brüschweiler R, Blackledge M. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. Proc Natl Acad Sci USA 2005;102:13885–13890.

Bowman MC, Smallwood S, Moyer SA. Dissection of individual functions of the Sendai virus phosphoprotein in transcription. J Virol 1999;73:6474–6483.

Clore GM. Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. Proc Natl Acad Sci USA 2000;97:9021–9025.

Clore GM, Schwieters CD. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation?. J Am Chem Soc 2004;126:2923–2938.

Clore GM, Starich MR, Gronenborn AM. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. J Am Chem Soc 1998;120:10571–10572.

Curran J. Reexamination of the Sendai virus P protein domains required for RNA synthesis: a possible supplemental role for the P protein. Virology 1996;221:130–140.

Curran J, Homann H, Buchholz C, Rochat S, Neubert W, Kolakofsky D. The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. J Virol 1993;67:4358–4364.

Curran J, Marq JB, Kolakofsky D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. J Virol 1995;69:849–855.

Demchenko AP. Recognition between flexible protein molecules: induced and assisted folding. J Mol Recognit 2001;14:42–61.

Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Curr Opin Struct Biol 2008;18:756–764.

Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.

Dyson HJ, Wright PE. Elucidation of the protein folding landscape by NMR. Meth Enzymol 2005a;394:299–321.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005b;6:197–208.

Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev 2004;104:3607–3622.

Eliezer D. Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 2009;19:23–30.

Emsley JW, Lindon JC. NMR spectroscopy using liquid crystal solvents, Oxford (NY): Pergamon Press; 1975.

Eyal E, Najmanovich R, McConkey BJ, Edelman M, Sobolev V. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. J Comput Chem 2004;25:712–724.

Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15:35–41.

Gerard FCA, Ribeiro EDA, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RWH, Jamin M. Modular organization of rabies virus phosphoprotein. J Mol Biol 2009;388:978–996.

Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–1074.

Houben K, Marion D, Tarbouriech N, Ruigrok RWH, Blanchard L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J Virol 2007;81: 6807–6816.

Jensen MR, Blackledge M. On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. J Am Chem Soc 2008;130:11266–11267.

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, Blackledge M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc 2008;130:8055–8061.

Jensen MR, Markwick PRL, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P, Blackledge M. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 2009;17:1169–1185.

Jensen MR, Salmon L, Nodet G, Blackledge M. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. J Am Chem Soc 2010;132:1270–1272.

Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84:3239–3252.

Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins. Virology 2002;296:251–262.

Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci USA 2004;101:8301–8306.

Lange OF, et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 2008;320:1471–1475.

Marion D, Tarbouriech N, Ruigrok RW, Burmeister WP, Blanchard L. Assignment of the 1H, 15N and 13C resonances of the nucleocapsid-binding domain of the Sendai virus phosphoprotein. J Biomol NMR 2001;21:75–76.

Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Protein Sci 2006;15:2795–2804.

Meiler J, Prompers JJ, Peti W, Griesinger C, Brüschweiler R. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. J Am Chem Soc 2001;123:6098–6107.

Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 2007;17:3–14.

Murphy AM, Moerdyk-Schauwecker M, Mushegian A, Grdzelishvili VZ. Sequence-function analysis of the Sendai virus L protein domain VI. Virology 2010;405: 370–382.

Ortega-Roldan JL, Jensen MR, Brutscher B, Azuaga AI, Blackledge M, van Nuland NAJ. Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar couplings: application to the CD2AP SH3-C:ubiquitin complex. Nucleic Acids Res 2009;37:e70.

Rückert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc 2000;122:7793–7797.

Salmon L, et al. Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. Angew Chem Int Ed Engl 2009;48:4154–4157.

Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J Biomol NMR 2000;18:303–309.

Serrano L, Fersht AR. Capping and alpha-helix stability. Nature 1989;342:296–299.

Serrano L, Sancho J, Hirshberg M, Fersht AR. Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. J Mol Biol 1992;227:544–559.

Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 2000;97: 8868–8873.

Skiadopoulos MH, et al. Sendai virus, a murine parainfluenza virus type 1, replicates to a level similar to human PIV1 in the upper and lower respiratory tract of African green monkeys and chimpanzees. Virology 2002;297:153–160.

Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007;447:1021–1025.

Svergun D, Barberato C, Koch MHJ. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 1995;28:768–773.

Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nat. Struct. Biol 2000;7:777–781.

Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 1997;278:1111–1114.

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar 1H{-}15N and 1H{-}13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 1997;4:732–738.

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci USA 1995;92:9279–9283.

Tolman JR. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. J Am Chem Soc 2002;124:12020–12030.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.

Torbet J, Maret G. Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field. J Mol Biol 1979;134:843–845.

Tuckis J, Smallwood S, Feller JA, Moyer SA. The C-terminal 88 amino acids of the Sendai virus P protein have multiple functions separable by mutation. J Virol 2002;76:68–77.

Tycko R, Blanco FJ, Ishii Y. Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. J Am Chem Soc 2000;122:9340–9341.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 1995;5:67–81.

Wishart DS, Sykes BD. Chemical shifts as a tool for structure determination. Meth Enzymol 1994a;239:363–392.

Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994b;4:171–180.

Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 1992;31:1647–1651.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.

Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features?. Protein Pept. Lett 2010;17:932–951.

Zhang H, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 2003;25:173–195.

Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 2000;122:3791–3792.

Zweckstetter M. NMR: prediction of molecular alignment from structure using the PALES software. Nat Protocol 2008;3:679–690.

5

STRUCTURAL DISORDER IN PROTEINS OF THE RHABDOVIRIDAE REPLICATION COMPLEX

CÉDRIC LEYRAT, FRANCINE C.A. GÉRARD, EURIPEDES A. RIBEIRO, JR, IVAN IVANOV, AND MARC JAMIN

5.1 INTRODUCTION

5.1.1 The Rhabdoviruses

The Rhabdoviridae is a large family of enveloped viruses that have a nonsegmented genome made of a single strand of negative-sense RNA. The Rhabdoviridae is grouped in the order Mononegavirales (MNV) (the name is a composition of three elements: mono, single; nega, negative sense; virales, virus) with three other families, the Filoviridae (Ebola and Marburg viruses), the Paramyxoviridae (measles, mumps, respiratory syncytial viruses), and the Bornaviridae (Borna disease virus), because all these viruses, in addition to having a nonsegmented negative-sense RNA genome, share similar genome and structural organizations as well as similar strategies of RNA replication and transcription (Pringle, 1997). Viruses are classified as Rhabdoviridae member and currently grouped into six genera on the basis of structural properties, antigenicity, and, more recently, phylogenetic analyses (Tordo et al., 2004). Viruses from three genera (*Lyssavirus, Vesiculovirus*, and *Ephemerovirus*) infect a variety of animal hosts, including mammals, fish, and arthropods; viruses from the fourth genus infect numerous fishes (*Novirhabdovirus*); and viruses from the last two genera are arthropod-borne and infect plants (*Cytorhabdovirus* and *Nucleorhabdovirus*), while numerous other rhabdoviruses remain unclassified.

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

The most prominent member of the Rhabdoviridae is the rabies virus (RV) in the genus *Lyssavirus*. RV is the main causative agent of rabies, a multihost zoonotic disease, whose transmission by dog bite and fatal nature have been known since the twenty-third century BC (Théodoridés, 1986). RV is a neurotropic virus that causes encephalitis with a complex pathogenesis leading to different forms of the disease ("furious" or encephalitic, "dumb" or paralytic, nonclassic rabies) (Hemachudha et al., 2002; Dietzschold et al., 2008). The encephalitis is incurable and fatal in 100% of the cases, but the infection can be prevented. Since the first successful use of rabies vaccine by Louis Pasteur in 1885, the disease is under control in Western Europe and North America by the application of preexposure and postexposure treatments (Rupprecht et al., 2006). However, rabies remains an important, albeit neglected, human health problem, killing more than 50,000 people each year worldwide (one person every 10 min) mainly in Asia and Africa (see WHO Fact Sheet N° 99, 2008) and imposing important economic burden on the affected populations (Knobel et al., 2005). Moreover, recent reports suggest that the number of cases is substantially underestimated in poor rural areas (Cleaveland et al., 2002). The main natural reservoirs of RV are bats or carnivores such as dog, fox, raccoon, and skunk. The mechanism of cross-species transmission remains poorly understood (Streicker et al., 2010), but new strains, potentially unsusceptible to vaccination against classical rabies (Hanlon et al., 2005; Weyer et al., 2008), could emerge and spread again in the human population (Fooks, 2005; Leslie et al., 2006). Recently, RV has been developed as a neuronal tracer to identify neuronal networks within the brain and for studying the innervations of motor muscles (Ugolini, 2010).

The vesicular stomatitis virus (VSV) in the genus *Vesiculovirus* is another well-known member of the Rhabdoviridae. It is an arthropod-transmitted virus that causes a mild disease in horses, cattle, and pigs and rarely transmits to human. It has economical impacts on the farm industry. For many years, VSV has been used as a model system for studying the replication cycle of the Rhadoviridae and, more generally, the interactions between viruses and their host cells. More recently, VSV has been used as a vaccine vector and as an oncolytic virus (Lichty et al., 2004; Rose et al., 2008). Other vesiculoviruses such as Chandipura virus or Piry virus also infect humans. In recent outbreaks in India, Chandipura virus produced acute encephalitis with high fatality ratio. Finally, numerous Rhabdoviridae members infect fishes or plants and are of major economic importance to fish farming and agricultural industries.

5.1.2 Genome Organization and the Virion Structure

The viral genome of the Rhabdoviridae members is made of a single molecule of negative-sense RNA of 9–18 kb flanked on both sides by untranslated leader (le) and trailer (tr) RNA regions (Fig. 5.1a). It comprises up to 10 genes, but RV and VSV genomes contain only five genes that are common to all members of the MNV and encode successively from the 3′ terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G), and the large subunit of the RNA-dependent RNA polymerase (L) (Fig. 5.1a).

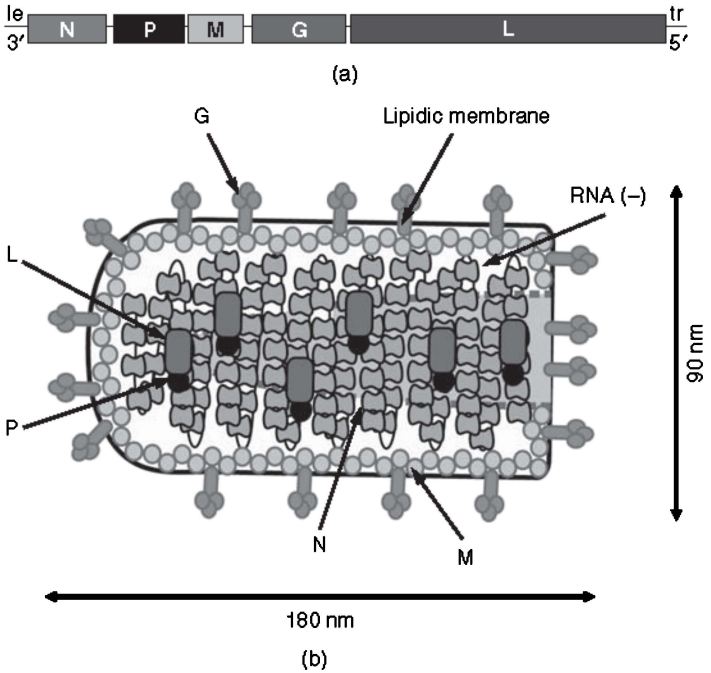


Figure 5.1 The rabies virus. (a) RV genome. The genome encodes five proteins and contains untranslated 3′ leader (le) and 5′ trailer (tr) sequences. (b) Structure of the virion. Schematic representation of the bullet-shaped RV showing the dimensions of the particle and the relative organization of the glycoprotein (G), matrix protein (M), nucleoprotein–RNA complex (N–RNA), phosphoprotein (P), and large subunit of the RNA-dependent RNA polymerase (L). The L and P proteins are localized inside the N–RNA complex but are shown here outside for presentation purpose.

Animal rhabdoviruses are enveloped viruses with an overall bullet shape with one round end and the other flat, while plant rhabdoviruses have a bacillus shape. RV and VSV particles are 180–170 nm in length and 90–70 nm in diameter (Schoehn et al., 2001; Ge et al., 2010) (Fig. 5.1b). The cryo-electron microscopy (EM) reconstruction of the virion of VSV revealed the molecular organization of the viral proteins in three concentric layers (Ge et al., 2010). The outer shell is a host-cell-derived lipidic membrane decorated with trimeric G spikes. The middle layer is formed by M proteins that are organized in a regular helix onto which the trimeric G spikes attach. The inner layer is the nucleocapsid (NC), the actual infectious part of the virus, composed of the genomic RNA molecule associated with the N, P, and L proteins (Szilagyi and Uryvayev, 1973). N has two domains forming a cavity into which nine ribonucleotides bind through interactions with the sugar–phosphate backbone (Iseni et al., 2000; Schoehn et al., 2001; Green et al., 2006; Albertini et al., 2006b). N completely packs the genomic RNA into a long N–RNA complex organized in a regular helix (Schoehn et al., 2001) that hides the RNA molecule away from host cell nucleases and innate immune receptors and

is likely involved in the presentation of the genome to the L polymerase. P and L proteins attach to this N–RNA complex but cannot be visualized in the cryo-EM reconstruction because they are not organized in a highly symmetrical fashion.

5.1.3 The Viral Replication Cycle

All intricacies of the life cycle of rhabdoviruses have not yet been unraveled. Most of the studies of intracellular replication were carried out with VSV and RV, which replicate in the host cell cytoplasm. The current knowledge about the RV replication cycle was recently reviewed (Schnell et al., 2010). Although numerous features of this process are common to both RV and VSV, RV replicates more slowly and is less cytopathic than VSV, indicating different regulatory mechanisms and different modes of interference with the host cell.

The replication cycle of rhabdoviruses has three main phases, which are summarized in Fig. 5.2. The first phase includes the attachment of the viral particle

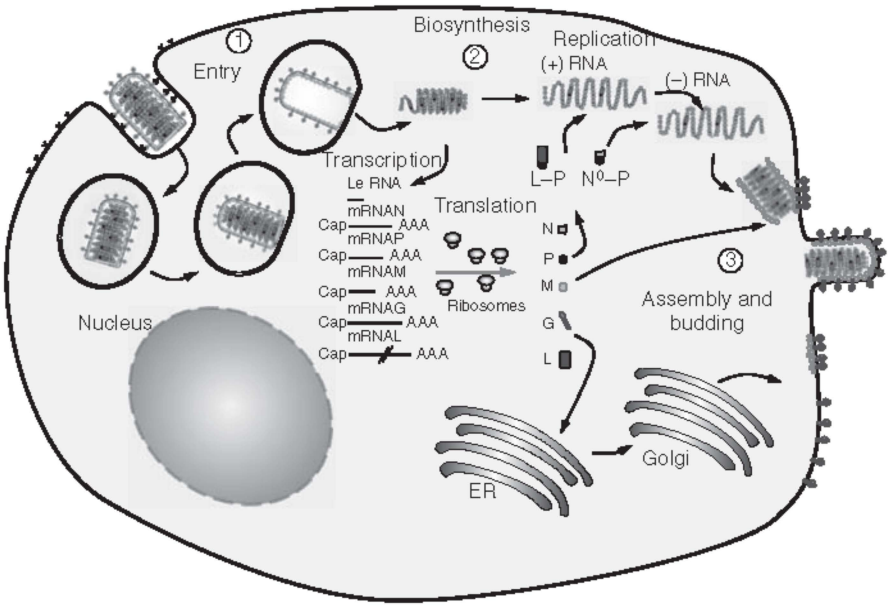


Figure 5.2 RV replication cycle. The replication of rabies virus occurs in three main phases. The first phase (1) includes the attachment of the viral particle onto the surface of the host cell, followed by its internalization into an endosome and the release of the NC in the cytoplasm through fusion of the viral membrane with the endosomal membrane. In the second phase (2), the viral genome is transcribed and replicated within the viral transcription/replication complexes, and the viral proteins are translated by the host cell translation machineries. In the third phase (3), the different viral components migrate to the host cell plasma membrane where they assemble, and new viral particles bud out from the host cell.

onto the surface of the host cell, followed by its internalization into an endosome and, in the case of RV, its intracellular axonal transport toward the cell body. Next, the NC is released in the cytoplasm through fusion of the viral membrane with the endosomal membrane. In the second phase, the viral genome is transcribed and replicated. These processes occur in cytoplasmic inclusion bodies that are named Negri bodies in the case of RV (Lahaye et al., 2009; Heinrich et al., 2010). Initially, the incoming viral genome is transcribed in a sequential process, producing a positive-stranded leRNA and five capped and polyadenylated mRNAs (Banerjee, 1987), and the viral proteins are produced by the host cell translation machineries. Later, the viral polymerase switches from transcription to replication of the RNA genome and starts producing a full-length positive-sense antigenome that subsequently serves as a template for the synthesis of new viral RNA genomes. The mRNAs are not encapsidated by N (Patton et al., 1984), but both the antigenomes and newly synthesized genomes are encapsidated by N, and therefore, replication requires the continuous production of soluble N (Patton et al., 1984). The mechanism of the switch from transcription to replication is unknown. It is regulated by the accumulation of soluble N proteins (Howard and Wertz, 1989), but in VSV, it also depends on the phosphorylation of P, the interaction of P with leRNA (Blumberg et al., 1981), and the M protein (Finke and Conzelmann, 2003). In the third phase, the viral components migrate to the host cell plasma membrane where they assemble. New viral particles bud out of the cell, taking away a piece of host cell membrane, which contains the G protein. The M protein interacts with both the N protein and host membrane (Dancho et al., 2009) and could therefore serve as a scaffold for assembling the viral components at the inner face of the cellular membrane.

5.2 THE RHABDOVIRUS REPLICATION COMPLEX

Only three of the viral proteins, N, P, and L, are required to synthesize viral RNA in an efficient and regulated manner. For VSV, transcription could be reconstituted *in vitro* using a mixture of L, P, and N–RNA complex (Emerson and Yu, 1975; De and Banerjee, 1984), and both transcription and replication could be reconstituted *in vivo* using a three-plasmid (L, P, and N) genetic system (Pattnaik et al., 1997). In VSV, two multimolecular complexes that contain the viral L, P, and N proteins, but differ in their content in host cell proteins, were postulated to separately catalyze transcription and replication (Gupta et al., 2003; Qanungo et al., 2004). The N packs the RNA genome in a single-stranded form, and this N–RNA complex, rather than the naked RNA, serves as the active template for both transcription and replication by the viral RNA-dependent RNA polymerase (Arnheiter et al., 1985). P is an essential, noncatalytic cofactor that plays multiple roles in the viral life cycle (Emerson and Yu, 1975). First, it assembles with the L protein to form the active RNA polymerase complexes (Emerson and Yu, 1975; De and Banerjee, 1984). In these complexes, L carries out all enzymatic activities involved in the transcription and replication of the viral genome, including RNA

synthesis, mRNA cap synthesis and methylation, and mRNA polyadenylation. P is an essential noncatalytic cofactor, but its precise role remains unknown (Barr et al., 2002). In the absence of P, L initiates RNA synthesis and produces short oligonucleotides of two to four ribonucleotides, but P is required for RNA elongation (De and Banerjee, 1985). L is a processive enzyme, which must remain attached as it moves along its template (Stillman and Whitt, 1999), and L alone has a weak binding affinity for the N–RNA complex and is unable to reassociate with its template (Mellon and Emerson, 1978). P has binding sites for both L and the N–RNA complex and is presumed to correctly position L onto the N–RNA template and maintain its attachment during the displacements of the polymerase complex along the template. The binding of P to L involves the N-terminal part of P and the last 500 residues of L (Emerson and Schubert, 1987; Castel et al., 2009), while the binding of P to the N–RNA complex involves the C-terminal domain of P and the C-terminal domain of N (Chenik et al., 1994; Fu et al., 1994; Green et al., 2000; Jacob et al., 2001; Schoehn et al., 2001).

During viral replication, the newly synthesized RNA genomes and antigenomes must be encapsidated by N. N has a strong affinity for RNA and binds nonspecifically to cellular RNA. A second role of P is to act as a chaperone of nascent N molecules. P forms with the N a complex named N^0–P complex (the superscript ^{0} denotes the absence of RNA) that maintains N in a soluble and RNA-free form until N is transferred, by an unknown mechanism, from P to the newly synthesized viral RNA (Masters and Banerjee, 1988; Peluso and Moyer, 1988; Howard and Wertz, 1989; Mavrakis et al., 2006; Chen et al., 2007). The binding site for N^0 was localized in the N-terminal part of P (Mavrakis et al., 2006; Chen et al., 2007). The RNA-binding cavity of N has a highly positive electrostatic surface potential (Green et al., 2006; Albertini et al., 2006b), whereas the N-terminal part of P is rich in negatively charged residues, suggesting that P binds in the same cavity of N as the RNA (Hudson et al., 1986; Mavrakis et al., 2006).

In addition to its roles in the transcription and replication of the viral genome, P also interferes with the host cell innate immune response by blocking the interferon response pathway and thereby modulating the pathogenicity of the virus (Ito et al., 2010). Several N-terminally truncated forms of P are produced from alternative in-frame AUG initiation codons by a leaky scanning mechanism (Herman, 1986; Spiropoulou and Nichol, 1993; Chenik et al., 1995). In RV, depending on the presence of the N-terminal nuclear export signal (NES) sequence, the truncated forms of P are located either in the cytoplasm or in the nucleus and therefore can interact with various components of the interferon response in both subcellular compartments, blocking the interferon production (Brzozka et al., 2005), the interferon-induced signaling cascade (Vidy et al., 2005), and some products of interferon-stimulated genes (Blondel et al., 2002). Finally, the VSV P and RV P interact with the host cell cytoskeleton (Poisson et al., 2001; Das et al., 2006). The interaction of RV P with the light chain 8 of the dynein motor could allow retroaxonal transport or intracellular transport.

In the last five years, the structural characterization of P, the N–RNA complex, and the N-RNA-P_{CTD} complex of both RV and VSV brought a new light on the

structural organization of the transcription/replication complex of the rhabdoviruses and highlighted the roles played by disordered regions in different molecular recognition events occurring within this complex (Ding et al., 2006; Green et al., 2006; Albertini et al., 2006b; Gérard et al., 2007, 2009; Ribeiro et al., 2008; Green and Luo, 2009; Ribeiro et al., 2009). These results are discussed here in the context of a scenario where P plays a central role in the assembly of the transcription/replication complex and in its progression along the N–RNA template during transcription and replication.

5.3 A META-PREDICTION OF PROTEIN DISORDERED REGIONS

Recently, the structure/function paradigm was challenged with the discovery that many proteins are intrinsically disordered proteins (IDPs) or contain intrinsically disordered regions (IDRs) under physiological conditions (Wright and Dyson, 1999; Uversky et al., 2000; Tompa, 2002; Uversky, 2002). Since the early days of protein folding studies, biological activities have been associated with the adoption by the polypeptide chain of a well-defined three-dimensional structure, but recent findings showed that disorder is required for the assembly of multimolecular complexes (Spolar and Record, 1994; Wright and Dyson, 2009) or for the biological functions of numerous proteins involved in molecular recognition, cell signaling, and regulation control (Wright and Dyson, 1999; Dyson and Wright, 2002; Tompa, 2002; Dunker et al., 2002a,b; Iakoucheva et al., 2004; Romero et al., 2004; Tompa and Csermely, 2004; Uversky et al., 2005; Liu et al., 2006). In the last decade, new approaches were developed for identifying and characterizing these flexible proteins, and a special section for disorder prediction has been added to the biannual Critical Assessment of Techniques for Protein Structure Prediction (CASP) exercise since 2002 (Bordoli et al., 2007). NMR spectroscopy (Dyson and Wright, 2004; Eliezer, 2009; Jensen et al., 2009) and small angle scattering experiments (Bernado et al., 2007; Bernado and Blackledge, 2009) are particularly suitable for characterizing their structural properties, and various algorithms were developed for identifying IDP or proteins containing IDRs (He et al., 2009). Complete genome surveys revealed that as much as one-third of the proteins from eukaryotes contain long IDRs (Tokuriki et al., 2009). However, these different predictors yield different results, with large discrepancies in the number, position, and length of predicted IDRs, and the use of multiple predictions was recommended (Ferron et al., 2006; Bourhis et al., 2007). These discrepancies argue for the existence of different types ("flavors") of disorder (Vucetic et al., 2003; Ferron et al., 2006) but raise the problem of choosing the best algorithm and/or strategy for locating the boundaries of the disordered regions in a given protein (Ferron et al., 2006; Bourhis et al., 2007). Recently, meta-servers and meta-predictors were developed, which automatically present the results from different predictors (Lieutaud et al., 2008) or calculate a consensus prediction (Ishida and Kinoshita, 2008; Deng et al., 2009; Schlessinger et al., 2009; Xue et al., 2010; Mizianty et al., 2010).

When we initiated our work on RV and VSV P, we devised our own meta-predictor using a simple binary voting procedure that integrates predictions from

16 different algorithms (Ribeiro et al., 2008; Gérard et al., 2009). The amino acid sequence was submitted to different algorithms available through web servers, and for each prediction, each residue was ranked in a binary manner as ordered or disordered using the default thresholds set in the web servers (Gérard et al., 2009). A score of zero was attributed to each residue predicted to be disordered, whereas a score of one was attributed to each residue predicted to be in a structured region. For each residue, the values obtained from the different predictors were added and divided by the number of used algorithms in order to normalize the score. The calculated D-score varied from zero to one, and consensus disordered regions were arbitrarily defined as regions with a normalized D-score of ≤ 0.50. This meta-approach was found to be very successful for localizing known structured domains within the Sendai virus P (Tarbouriech et al., 2000; Blanchard et al., 2004; Bernado et al., 2005; Ding et al., 2006; Ribeiro et al., 2008) and RV P (Mavrakis et al., 2004) and for identifying unknown structured domains in VSV P (Ribeiro et al., 2008) and RV P (Ivanov et al., 2010). This analysis, backed by biochemical and biophysical data, led to a model of modular organization of the Rhabdoviridae P protein (Gérard et al., 2007, 2009) in agreement with that proposed for the Paramyxoviridae P protein (Karlin et al., 2003; Habchi and Longhi, 2010).

5.4 THE MODULAR ORGANIZATION OF RHABDOVIRUS PHOSPHOPROTEIN

5.4.1 Functional and Structural Modules

Early studies based on deletions and mutations revealed the existence of several independent and conserved functional regions in VSV P and RV P (Gill et al., 1986; Emerson and Schubert, 1987; Gigant et al., 2000; Mavrakis et al., 2004; Das and Pattnaik, 2005; Mavrakis et al., 2006) that suggested a modular organization. The N-terminal part of P is highly acidic and contains binding sites for L (Emerson and Schubert, 1987; Chenik et al., 1998; Castel et al., 2009) and N^0 (Takacs et al., 1993; Mavrakis et al., 2006; Chen et al., 2007). It also contains phosphorylation sites, which in VSV are involved in transcriptional regulation (Barik and Banerjee, 1992a,b). The C-terminal part of P contains the binding site for the N–RNA template (Paul et al., 1988; Das et al., 1997; Green et al., 2000; Jacob et al., 2001; Schoehn et al., 2001; Mavrakis et al., 2004) and phosphorylation sites, which control replication in VSV (Hwang et al., 1999). The intermediate region is less conserved and includes the self-association domain (Jacob et al., 2001; Ding et al., 2006).

The application of our meta-prediction to the location of IDRs in VSV P and RV P hinted at a modular structural organization in broad agreement with the functional map, although the boundaries between structural and functional modules were slightly different. For RV P, the consensus analysis predicted three structured domains, P_{NTD}, P_{CED}, and P_{CTD}, separated by two long IDRs (Fig. 5.3a,c). P_{NTD} contained the identified N^0-binding site (Mavrakis et al., 2006), P_{CED} was part of the region required for the oligomerization of P (Jacob et al., 2001), and P_{CTD}

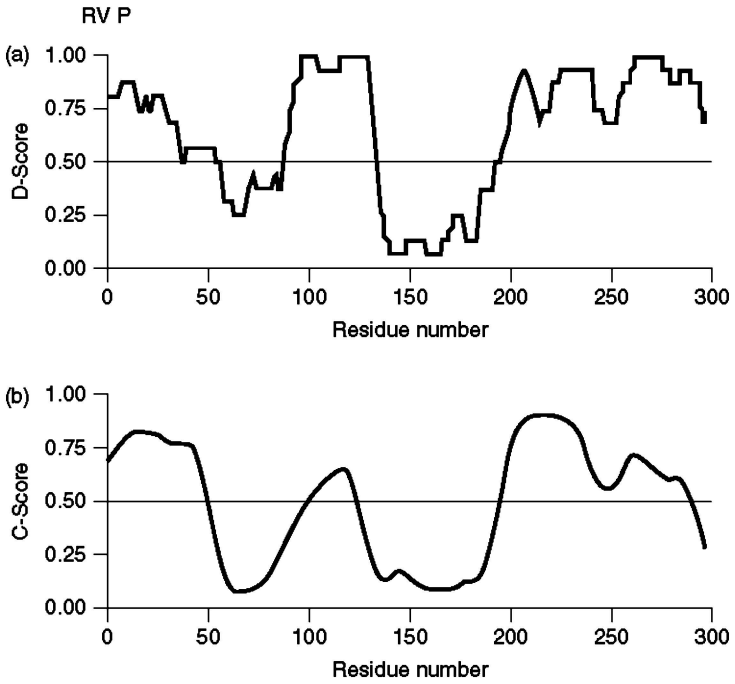


Figure 5.3 Modular organization of RV P. (a) Meta-prediction of IDRs. A consensus prediction for the localization of IDRs was obtained by combining results from 16 independent predictors available through WEB servers (Gérard et al., 2009). For each disorder prediction, residues were simply ranked in a binary manner as ordered or disordered using the default threshold set for each algorithm. For each predicting algorithm, a score of zero was attributed to a residue when it was predicted to be disordered, whereas a score of one was attributed when it was predicted to be in a structured region. The D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined a threshold at 0.50, above which residues are assigned as structured, and which residues are assigned as disordered. (b) Sequence conservation. The amino acid sequences of the P protein from 13 *Lyssavirus* were aligned, and the conservation rate for each amino acid was scored by using the AL2CO software (Pei and Grishin, 2001). A C-score above 0.5 indicates a conserved position. This analysis revealed the presence of the sequence of three conserved regions (aa 1–50, 101–126, and 194–292) that are located within or correspond to the otherwise predicted structured regions. (c) Consensus localization of structured and disordered regions. The consensus disordered regions determined from the D-score suggest the presence of two disordered regions (aa 53–89 and 131–194) and three structured domains (aa 1–52, 90–130, and 195–297). (d) Structured domains. The high resolution structures of two domains corresponding closely to the predicted P_{CED} and P_{CTD} have been determined by X-ray crystallography (Mavrakis et al., 2004; Ivanov et al., 2010). The known three-dimensional structures of P_{CED} and P_{CTD} are shown as cartoon models and the disordered regions are shown as dotted lines. The dotted circles represent the localization of different binding sites, which could fold on binding to their partner.

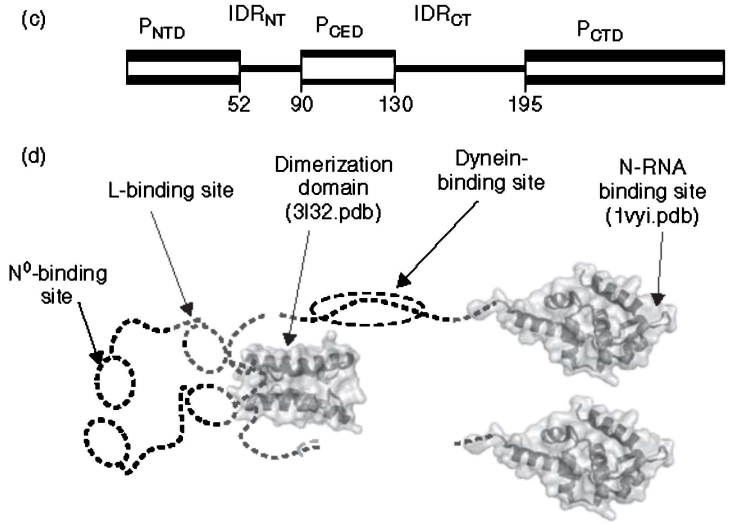


Figure 5.3 (*Continued*)

corresponded to the N–RNA binding domain (Mavrakis et al., 2004) that also interacts with several components of the cellular interferon response (Chelbi-Alix et al., 2006) (Fig. 5.3d). IDR_{NT} contains a CRM1-dependent NES (Pasdeloup et al., 2005), phosphorylation sites (Gupta et al., 2000), and the binding site for L (Castel et al., 2009), whereas the IDR_{CT} contains a binding site for the light chain 8 of the dynein motor machinery (Poisson et al., 2001). A multiple alignment of the amino acid sequence of P proteins from 13 members of the genus *Lyssavirus* revealed that the three predicted structured domains were more conserved than the predicted IDRs (Gérard et al., 2009) (Fig. 5.3b).

The consensus prediction obtained for VSV P suggested the presence of four structured domains separated by three IDRs (Fig. 5.4a,c). As in RV P, the predicted P_{NTD} contains the binding site for N^0. The dimerization domain was previously identified in the central part of P by limited proteolysis, and its structure was solved by X-ray crystallography (Ding et al., 2004, 2006) (Fig. 5.4d). The meta-analysis of disorder, however, predicted the existence of two short structured domains (aa 98–105 and aa 136–186) separated by a short IDR_{CE} and failed to identify the crystallized dimerization domain (Ding et al., 2006) (Fig. 5.4a). The predicted C-terminal domain corresponded to the functional domains required for transcription and replication and for binding to the N–RNA template. The IDR_{NT} contains phosphorylation sites that control viral transcription, and a multiple alignment of the amino acid sequence of P proteins from 12 members of the genus *Vesiculovirus* showed that the three predicted structured domains as well as IDR_{NT} are conserved (Gérard et al., 2009) (Fig. 5.4b). The C-terminal IDR_{CT} was short, but its presence between the dimerization domain and P_{CTD} was confirmed by NMR spectroscopy (Ribeiro et al., 2008).

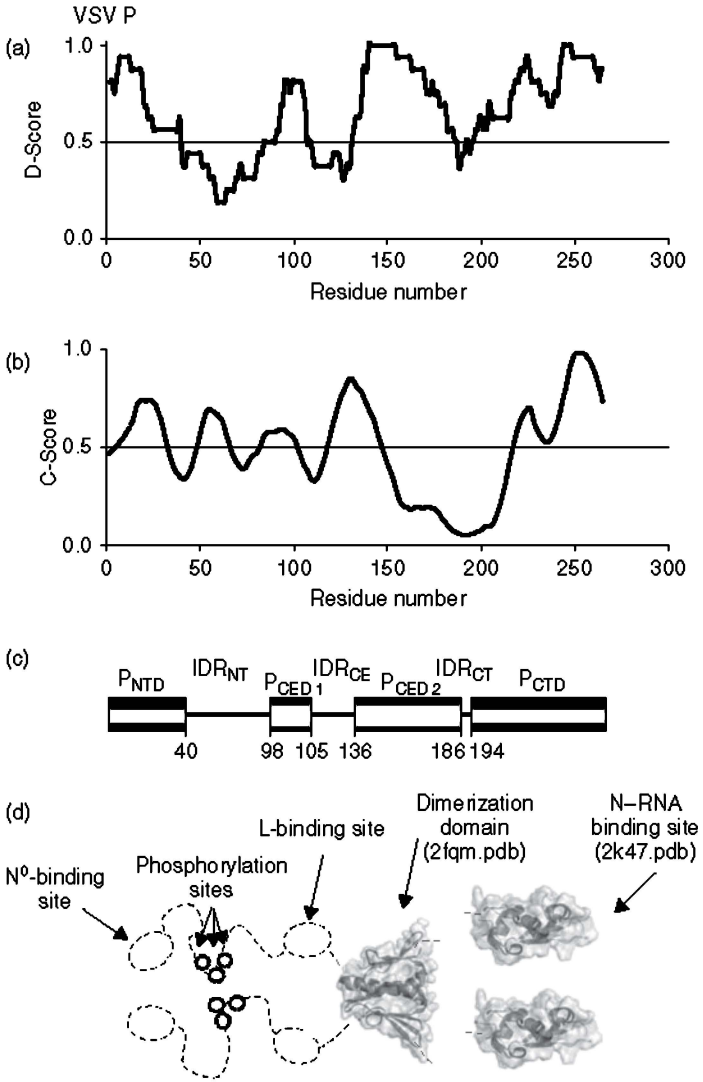


Figure 5.4 Modular organization of VSV P. (a) Meta-prediction of IDRs. (b) Sequence conservation. The amino acid sequences of the P protein from 12 *Vesiculovirus* were aligned. This analysis revealed that most of the N-terminal regions (aa 1–150) and the C-terminal regions (aa 216–265) are conserved. (c) Consensus localization of structured and disordered regions. The consensus disordered regions determined from the D-score suggest the presence of three disordered regions (aa 40–98, 105–136, and 186–194) and four structured domains (aa 1–39, 99–104, 137–185, and 195–265). (d) Structured domains. The high resolution structures of two domains corresponding closely to the predicted P_{CED} and P_{CTD} have been determined by X-ray crystallography and NMR, respectively (Ding et al., 2006; Ribeiro et al., 2008).

5.4.2 Two Ordered Domains of P are Autonomous Folding Units

In RV P and VSV P, P_{CED} and P_{CTD} are autonomous folding units, that is, domains that fold by themselves when expressed as recombinant proteins. VSV P_{CTD} was also shown to fold and unfold reversibly *in vitro* (Ribeiro et al., 2008). Nonphosphorylated RV P and VSV P produced in bacteria form dimers (Gérard et al., 2007). The predicted RV P_{CED} forms dimers in solution, whereas the protein deleted from the central region ($P_{\Delta 91-131}$) forms monomers (Gérard et al., 2009). In the crystal structure, P_{CED} also forms dimers (Fig. 5.5a) (Ivanov et al., 2010). Similarly, VSV P_{CED} identified previously by limited proteolysis (VSV P_{CED}, aa 107–170) is dimeric in the crystal structure (Ding et al., 2004, 2006) (Fig. 5.5a). The predicted RV P_{CTD} (aa 195–297) corresponded closely to the previously identified autonomous folding N–RNA binding domain whose structure was solved by X-ray crystallography (aa 186–296) (Mavrakis et al., 2004) (Fig. 5.5b). The equivalent domain was correctly localized in VSV P from the meta-prediction (aa 197–265), and its structure was solved by NMR spectroscopy (Ribeiro et al., 2008) (Fig. 5.5b).

Although multiple sequence alignments indicate almost no sequence similarity between P proteins from VSV and RV, the structure of VSV P_{CTD} is homologous to that of RV P (Fig. 5.5b). The topology of the backbone is conserved, although

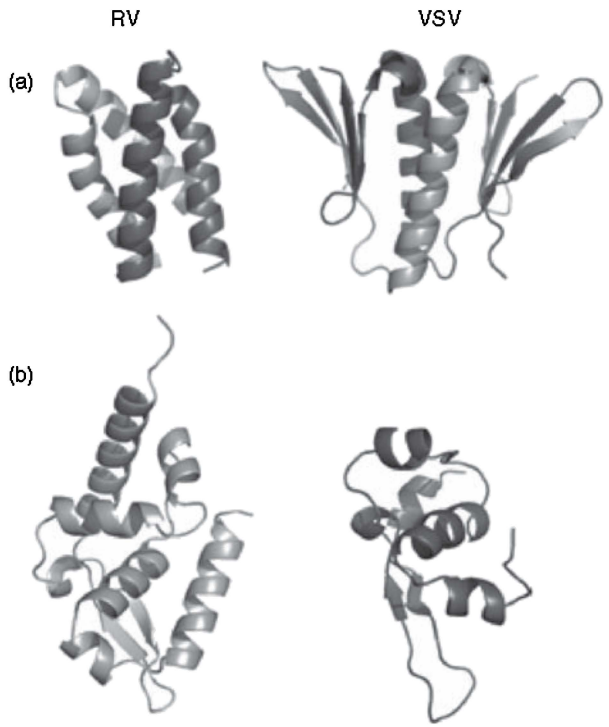


Figure 5.5 Comparison of the oligomerization (a) and N–RNA binding domains (b) of the phosphoproteins of RV and VSV.

VSV P_{CTD} lacks helices α_3 and α_6 of RV P_{CTD} (Ribeiro et al., 2008). This structure suggested that the C-terminal domain and its associated biological functions are conserved among the Rhabdoviridae members, a proposal recently confirmed by the structure of Mokola virus P_{CTD} (Assenberg et al., 2009). Conversely, a comparison of the dimeric structures of RV P_{CED} and VSV P_{CED} showed that they are so different from each other that it is impossible to determine whether these domains have evolved from a common ancestor (Ivanov et al., 2010) (Fig. 5.5a). The monomer of VSV P_{CED} is made of a central α-helix flanked by two two-stranded β-sheets. In the dimer, the helices from the two monomers pack against each other in a parallel orientation, and on each side of the helical bundle, the N-terminal β-sheet of one monomer forms a four-stranded β-sheet with the C-terminal β-sheet of the other monomer (Ding et al., 2006). The monomer of RV P_{CED} is made of two α-helices forming a hairpin, and the dimer is assembled by packing two monomers in a parallel orientation with an angle of about 50° (Ivanov et al., 2010).

5.4.3 The N-Terminal Region Contains a MoRE

Although our meta-prediction of IDRs suggested that the N-terminal part of RV P and VSV P contains a structured domain (Figs. 5.3a and 5.4a) that is highly conserved within their respective genera (Figs. 5.3b and 5.4b), peptides corresponding to the first 60 residues of RV P and VSV P appeared globally disordered (Gérard et al., 2009). Their Stokes radii and their radii of gyration were similar to those expected for a fully unfolded protein of this size, and the NMR spectra exhibited poor chemical shift dispersion and narrow resonances typical of disordered proteins. However, circular dichroism spectroscopy revealed low amounts of secondary structure, and NMR spectroscopy revealed the presence of two transient α-helical elements (aa 2–8 and aa 25–38) in VSV P_{NTD}, that is, α-helices that are not stable enough to be present all the time in the protein, but which fold and unfold continuously (Leyrat et al., unpublished data). This local order in the N-terminal part of P could indicate that P_{NTD} contains a molecular recognition element (MoRE), that is, a short motif within a disordered protein that promotes association with a partner by folding on binding (Mohan et al., 2006; Jensen et al., 2008).

5.4.4 The Overall Structure of P

P dimer has overall dimensions larger than those expected for a globular protein of the same molecular mass (Gérard et al., 2007). On a plot of the hydrodynamic radius as a function of molecular mass, full-length P was intermediate between native and unfolded proteins, suggesting that this protein is not fully structured (Gérard et al., 2009). Small angle neutron scattering data indicated that the molecule has an overall cylindrical shape with length about three times longer than the diameter (Gérard et al., 2007). The modular organization of P with a folded central dimeric domain, a folded C-terminal domain, and two disordered regions conveys a very dynamic picture of P with four flexible "arms" emerging from the central core (Figs. 5.3d and 5.4d).

IDRs have the ability to interact with multiple structurally diverse partners, and their binding energetics may favor exchange between different partners. The N-terminal "arm" (P_{NTD} + IDR_{NT}) is highly flexible, contains multiple binding sites, and can serve to recruit different viral and cellular partners. In RV P, it (P_{NTD} + IDR_{NT}, aa 1–92) contains the binding sites for N^0, L, host cell kinase and exportin. The whole disordered N-terminal part of VSV P is highly conserved among vesiculoviruses and contains binding sites for a cellular kinase and a binding site for L at the junction between IDR_{CT} and the dimerization domain (aa 79–123) (Emerson and Schubert, 1987). Our meta-analysis predicted a short structured domain (aa 91–106) within this L-binding region, suggesting that it could also fold on binding to its partner. The N-terminal part of RV P and VSV P, including the P_{NTD} and the IDR_{NT}, is rich in acidic residues and exhibits a net negative charge. The requirement of a globally disordered acidic region for the correct assembly of the transcription/replication complex is reminiscent of eukaryotic transcription regulator proteins, such as GAL4 and GCN4. These proteins comprise disordered acidic regions that activate transcription by recruiting transcription factors to the promoter and by helping in the assembly of the transcription complex (Ptashne and Gann, 1997). The acidic character of these regions, and not their specific amino acid sequence, is the major determinant of their activity, and transcriptional activation could be achieved with random polypeptides of identical composition (Ma and Ptashne, 1987). In VSV, a chimeric P protein, in which the N-terminal part was replaced by β-tubulin, a protein with an acidic C-terminal region, was active in transcription (Chattopadhyay and Banerjee, 1988). The N-terminal disordered region of P (P_{NTD} + IDR_{NT}) likely recruits the L polymerase to the site of RNA synthesis and allows the assembly of the multimolecular transcription/replication complexes.

The C-terminal "arm" contains a terminal folded domain (P_{CTD}) tethered to the dimeric core by a flexible linker (IDR_{CT}). The C-terminal domain is involved in the interaction with the N–RNA template and with various cellular proteins (importin, exportin, kinase, STATs, PML, and microtubules). In RV P, the C-terminal flexible linker (IDR_{CT}) also contains a binding site for a cellular partner. P dimer can thus be envisioned as an ensemble of interconverting conformers, in which the molecule continuously expands and contracts and adapts its shape to bind to different partners.

5.5 FLEXIBLE LOOPS IN N PARTICIPATE IN BINDING P

A disordered region in RV N also plays a major role in the assembly of the N–RNA complex with the C-terminal domain of P, a critical step for the attachment of the polymerase complex onto its template (Fig. 5.6a). The viral NCs are long and flexible particles and so far resisted to high resolution structure determination (Schoehn et al., 2001). However, expression of recombinant N in bacteria or insect cells leads to the formation of circular N–RNA complexes (Iseni et al., 1998; Green et al., 2000; Schoehn et al., 2001), and homogenous preparations of circular complexes were obtained and crystallized (Green et al., 2000; Albertini et al.,

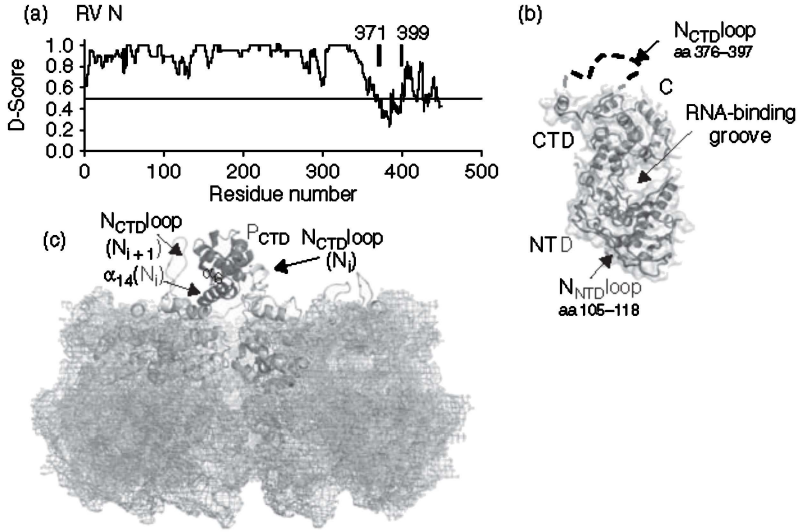


Figure 5.6 The flexible N_{CTD} loop participates in binding P_{CTD}. (a) D-score calculated for RV N. (b) Structure of N. Ribbon diagram within a space-filled model of the N protomer in a lateral view showing that the RNA-binding groove is closed. The N_{NTD} and N_{CTD} loops are shown as dotted lines. (c) Structure of the N-RNA-PCTD complex. Representation of the complex formed between RV P_{CTD} and the N_{11}–RNA complex obtained by molecular docking and modeling. The ribbon representation shows the location of P_{CTD} at the surface of the circular N_{11}–RNA ring. The meshed structure shows the circular N_{11}–RNA complex. The light gray ribbon representations show the three N_{CTD} domains used in the molecular modeling, and the dark gray ribbon representation shows the P_{CTD}. The N_{CTD} loop of protomers N_i and N_{i+1}, as well as the helix α_6 of P_{CTD} and the helix α_{14} of the N_i protomer are shown.

2006a). The crystal structures revealed that the N protomer consists of an N-terminal core domain (N_{NTD}), composed of helices connected by large loops, and a C-terminal core domain (N_{CTD}), composed of helices joined by short loops (Green et al., 2006; Albertini et al., 2006b) (Fig. 5.6b). Both the N_{NTD} and N_{CTD} act as jaws that close around the single-stranded RNA. In addition to the core domains, two smaller subdomains participate in domain exchange between protomers and stabilize polymeric forms of N–RNA complexes (Green et al., 2006; Albertini et al., 2006b).

In the RV structure, two regions, the N_{NTD} loop (aa 105–118) and the N_{CTD} loop (aa 376–397), are flexible and invisible in the crystal structure. Limited trypsin digestion cleaves the protein at Lys376 (Kouznetzoff et al., 1998), and trypsinated N–RNA complexes lose their ability to bind P_{CTD}, although they retain their overall structure (Kouznetzoff et al., 1998; Schoehn et al., 2001). The application of our meta-prediction to the analysis of RV P predicted the presence of a disordered region between residues 371 and 399 (Fig. 5.4c), in good agreement with the location of the flexible N_{CTD} loop.

Titration experiments revealed that two monomers of RV P_{CTD} bind independently to one N–RNA ring, with a dissociation constant near 100 nM (Ribeiro et al., 2009). A model of the complex was generated by molecular docking and molecular dynamics simulations, using small angle X-ray scattering (SAXS) profiles and mutational data as constraints (Jacob et al., 2001), that revealed the role of the flexible N_{CTD} loop in the recognition process. In this structural model, P_{CTD} lies on the top of an N protomer (N_i), and the N_{CTD} loops of this protomer (N_i) and of the adjacent one (N_{i+1}) close around the P_{CTD} by an induced-fit mechanism, making extensive protein–protein contacts with the two faces of the P_{CTD} (Ribeiro et al., 2009) (Fig. 5.6c). In the meantime, the structure of the homologous complex from VSV was determined by X-ray crystallography (Green and Luo, 2009). In both VSV and RV, P_{CTD} exhibits the same orientation on the top side of the N–RNA ring, its N-terminal extremity pointing toward the center of the ring, thus correctly positioning the N-terminal part of P near the RNA molecule (Fig. 5.6c). The involvement of two adjacent protomers indicates that the complete binding site for P_{CTD} only exists in the multimeric N–RNA complexes that serve as template for the polymerase. Also, the structure of both complexes revealed no conformational change in N that would significantly modify the accessibility to the RNA.

A main difference between the RV and VSV complexes resides in the length of the N_{CTD} loop, which is longer and more flexible in RV than in VSV. The N_{CTD} loop is not visible in the crystal structure of the RV N–RNA complex, and loops from two adjacent N protomers in RV hide a larger part of P_{CTD} in RV than in VSV. The higher flexibility of the N_{CTD} loop might explain that the RV N-RNA-P_{CTD} complex could not be crystallized. Another difference resides in the occupancy of the potential P_{CTD}-binding sites on the N–RNA ring. In the VSV crystal structure, one P_{CTD} is bound to each N protomer, although reduced occupancy was observed in some positions (Green and Luo, 2009). In the RV complex, the involvement of N_{CTD} loops from adjacent protomers in binding one P_{CTD} makes unlikely to reach a stoichiometry larger than one P_{CTD} for two N protomers, and titration experiments showed that only two P_{CTD} bind to one N–RNA ring (Ribeiro et al., 2009).

Also, in RV, phosphorylation at S389 of N enhances the binding affinity for P_{CTD} (Toriumi and Kawai, 2004). Our model of RV N-RNA-P_{CTD} complex suggested that phosphorylation of this serine residue creates an additional network of stabilizing electrostatic interactions (Ribeiro et al., 2009), and recent data show that phosphorylation of N achieves a 1000-fold increase in binding affinity (Ribeiro et al., unpublished data).

5.6 ROLES OF DISORDER IN THE RHABDOVIRUS TRANSCRIPTION/REPLICATION COMPLEX

IDPs can confer different advantages over structured proteins in molecular recognition and multimolecular assembly processes (Dyson and Wright, 2005; Uversky et al., 2005; Mittag et al., 2010), and several could be at play in the Rhabdovirus transcription/replication complex. The conformational plasticity offered by IDRs

can allow a protein to adapt its shape to different partners, and numerous hub proteins are highly flexible (Haynes et al., 2006). The plasticity of RV P and VSV P might be important for connecting multiple viral and host cell partners at different stages of the replication cycle or for interfering with various cellular components. A flexible binding site can provide a better access to the active site of an enzyme or to the binding site of a partner. Phosphorylation and other posttranslational modifications often involve flexible or disordered regions (Iakoucheva et al., 2004). The phosphorylation sites in VSV P that control transcriptional activation are localized in the IDR_{NT} (Chattopadhyay and Banerjee, 1987), and the binding site for dynein LC8 in RV P binds in an extended conformation (Poisson et al., 2001). IDRs might also affect the kinetics and thermodynamics of a binding reaction. By comparison with the mechanism of transcription/replication of Sendai virus (Paramyxoviridae), it was speculated that oligomerization of P allows the protein to move along the N–RNA template by a cartwheeling mechanism (Kolakofsky et al., 2004). In Sendai virus, P forms tetramers through a central coiled-coil domain that is required for transcription. In RV, the self-assembly domain of P is dispensable for transcription (Jacob et al., 2001). The high flexibility of *Rhabdovirus* P and the high binding affinity of P_{CTD} for the N–RNA template suggest an alternative mechanism for the movement of the polymerase along its N–RNA template, in which dimerization of P is dispensable. Highly flexible P molecules tethered to the N–RNA template through their P_{CTD} will have a large capture radius that could increase the on-rate for binding its partner as proposed in the "fly-casting" mechanism (Pontius, 1993; Shoemaker et al., 2000). Also, by sampling a large conformational space around its attachment point on the N–RNA template, P dimers act like brushes and thereby regulate the spacing between P molecules along the NC. If P molecules are distributed at regular intervals along the template, exposing their N^0- and L-binding sites, the polymerase could move along the template by jumping from one P to the next (Fig. 5.7). The long N-terminal arms of P could fetch the polymerase upstream, maintain its attachment to the template during the time it replicates a stretch of RNA, and then deliver it downstream to the next P molecule. Similarly, the N-terminal arms could catch the N^0 molecules and position them correctly for encapsidation of the newly synthesized RNA molecule.

Depending on the extent of the disorder-to-order transition that accompanies their binding, IDRs can differently affect the binding affinity. During the viral replication cycle, the N^0–P complex is not an end product. It forms only transiently as an intermediate during the synthesis of new NCs. The N^0-bound P must be outcompeted by the newly synthesized genomic RNA, and therefore, the binding affinity of P for N^0 should not be too high. The mechanism of folding on binding provides a means of specific recognition without the corollary of high affinity (Dyson and Wright, 2002; Wright and Dyson, 2009). Indeed, the folding or the adoption of a rigid structure by the ligand when it binds to its receptor will generally lead to the formation of multiple specific intermolecular interactions that confer a great specificity and consequently, a high affinity. If the ligand is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. We might thus speculate that by being disordered in its free

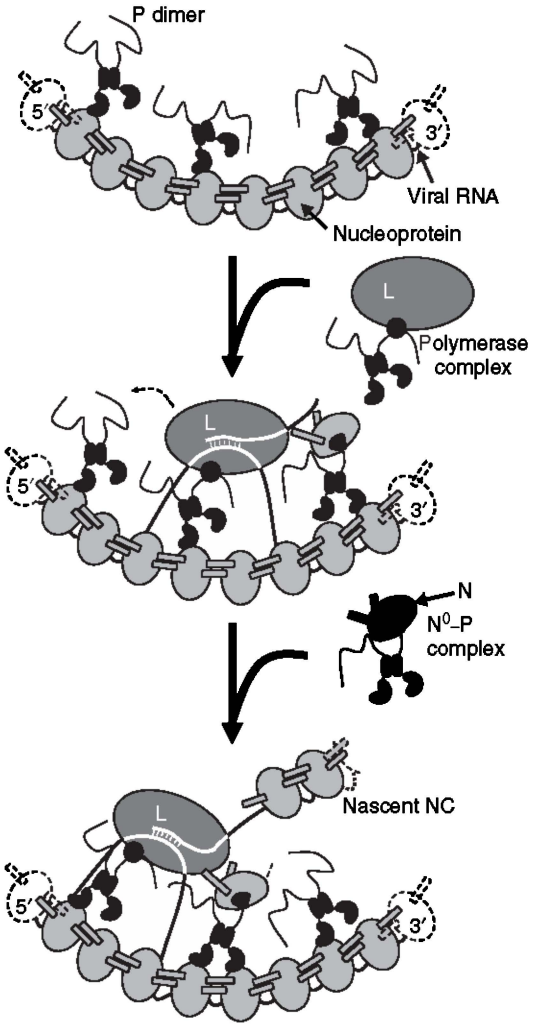


Figure 5.7 Schematic representation of the mechanisms of transcription and replication of *Rhabdovirus*. During RNA synthesis, the polymerase must move along its RNA template and must displace the RNA from the nucleoproteins. P provides the physical link by which the polymerase attaches to its template. In RV, the dimerization domain of P is dispensable for transcription and cannot therefore move by a cartwheeling mechanism. Because of the strong binding affinity of P_{CTD} for the N–RNA template, we suppose that P dimers are bound at regular intervals along the N–RNA complex. During transcription, the polymerase attaches to the N-terminal region of P and transcribes a stretch of RNA. Because P is elongated and flexible, the N-terminal arm of the adjacent P molecule can extend and catch the polymerase, which is then transferred downstream. The polymerase can thus move along the N–RNA complex by jumping between P molecules. During replication, the polymerase can move by a similar mechanism, but in addition, the newly synthesized genomic RNA is encapsidated by nascent N molecules. Nascent N molecules form soluble N^0–P complexes, which can also attach to the N–RNA complex through their P_{CTD}. N^0 can then be delivered to the site of RNA synthesis, yielding new N–RNA complexes.

form, P_{NTD} recognizes N^0 with high specificity but moderate affinity, thus allowing its displacement not only by the newly synthesized viral RNA but also by host cell RNAs.

Finally, IDRs have also been proposed to confer evolutionary advantages by allowing the accumulation of mutations, by facilitating the creation of binding sites for new partners, or by limiting the size of the protein required to achieve a given specificity (Gunasekaran et al., 2003; Tokuriki et al., 2009). The modular organization of the P protein from the Rhabdoviridae members is similar to that of P from the Paramyxoviridae members (Karlin et al., 2003; Gérard et al., 2009). In the Paramyxoviridae family, as in the Rhabdoviridae family, structured domains alternate with IDRs and modules playing the same role are organized in the same order along the sequence. The N-terminal region of P is involved in binding to N^0, the central part of P constitutes the self-assembly domain, and the C-terminal domain of P binds to the N–RNA template. This similar organization suggests that beyond the large differences in length, sequence, and even structure of the ordered domains, the P proteins have evolved from a common ancestor by only conserving their modular organization and the presence of IDRs (Habchi and Longhi, 2010; Ruigrok and Blackledge, 2010).

ACKNOWLEDGMENTS

This work was supported by grants from the French ANR (ANR-07-001-01 (ANRAGE)), the FINOVI foundation, and Lyonbiopôle. The work of E. A. R. was supported by postdoctoral fellowships from both the ANR and the FINOVI programs. Works of C. L. and F. G. were supported by MENRT fellowships from the French government and I. I. by a fellowship from the Institut Laue Langevin. We thank the Partnership for Structural Biology for the excellent structural biology environment.

ABBREVIATIONS

MNV	Mononegavirales order
VSV	vesicular stomatitis virus
RV	rabies virus
RNA	ribonucleic acid
N	nucleoprotein
N–RNA	nucleoprotein–ribonucleic acid complex
P	phosphoprotein
L	large subunit of the RNA-dependent RNA polymerase
NC	nucleocapsid
N^0–P	complex between the RNA-free nucleoprotein and the phosphoprotein
P_{CTD}	C-terminal domain of P
P_{CED}	central domain of P

P_{NTD}	N-terminal domain of P
N_{NTD}	N-terminal domain of N
N_{CTD}	C-terminal domain of N
IDR	intrinsically disordered region
IDP	intrinsically disordered protein
SEC	size exclusion chromatography
MALLS	multiangle laser light scattering
SANS	small angle neutron scattering
SAXS	small angle X-ray scattering
NMR	nuclear magnetic resonance

REFERENCES

Albertini AA, Clapier CR, Wernimont AK, Schoehn G, Weissenhorn W, Ruigrok RW. Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. J Struct Biol 2006a;158:129–133.

Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW. Crystal structure of the rabies virus nucleoprotein-RNA complex. Science 2006b;313:360–363.

Arnheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. Cell 1985;41: 259–267.

Assenberg R, Delmas O, Ren J, Vidalain PO, Verma A, Larrous F, Graham SC, Tangy F, Grimes JM, Bourhy H. The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein. J Virol 2009;84:1089–1096.

Banerjee AK. Transcription and replication of rhabdoviruses. Microbiol Rev 1987;51: 66–87.

Barik S, Banerjee AK. Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. Proc Natl Acad Sci USA 1992a;89:6570–6574.

Barik S, Banerjee AK. Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation. J Virol 1992b;66:1109–1118.

Barr JN, Whelan SP, Wertz GW. Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. Biochim Biophys Acta 2002;1577:337–353.

Bernado P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J 2009;97:2839–2845.

Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 2005;102:17002–17007.

Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. J. Am. Chem. Soc. 2007;129:5656–5664.

Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP, Ruigrok RW, Marion D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. Virology 2004;319:201–211.

Blondel D, Regad T, Poisson N, Pavie B, Harper F, Pandolfi PP, De The H, Chelbi-Alix MK. Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. Oncogene 2002;21:7957–7970.

Blumberg BM, Leppert M, Kolakofsky D. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. Cell 1981;23:837–845.

Bordoli L, Kiefer F, Schwede T. Assessment of disorder predictions in CASP7. Proteins 2007;69(Suppl 8):129–136.

Bourhis JM, Canard B, Longhi S. Predicting protein disorder and induced folding: from theoretical principles to practical applications. Curr Protein Pept Sci 2007;8:135–149.

Brzozka K, Finke S, Conzelmann KK. Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3. J Virol 2005;79:7673–7681.

Castel G, Chteoui M, Caignard G, Prehaud C, Mehouas S, Real E, Jallet C, Jacob Y, Ruigrok RW, Tordo N. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. J Virol 2009;83:10808–10820.

Chattopadhyay D, Banerjee AK. Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*. Cell 1987;49:407–414.

Chattopadhyay D, Banerjee AK. NH2-terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin. Proc Natl Acad Sci USA 1988;85:7977–7981.

Chelbi-Alix MK, Vidy A, El Bougrini J, Blondel D. Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, and PML nuclear bodies. J Interferon Cytokine Res 2006;26:271–280.

Chen M, Ogino T, Banerjee AK. Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. J Virol 2007;81:13478–13485.

Chenik M, Chebli K, Blondel D. Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. J Virol 1995;69:707–712.

Chenik M, Chebli K, Gaudin Y, Blondel D. *In vivo* interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. J Gen Virol 1994;75:2889–2896.

Chenik M, Schnell M, Conzelmann KK, Blondel D. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. J Virol 1998;72:1925–1930.

Cleaveland S, Fevre EM, Kaare M, Coleman PG. Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries. Bull World Health Organ 2002;80:304–310.

Dancho B, McKenzie MO, Connor JH, Lyles DS. Vesicular stomatitis virus matrix protein mutations that affect association with host membranes and viral nucleocapsids. J Biol Chem 2009;284:4500–4509.

Das SC, Nayak D, Zhou Y, Pattnaik AK. Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells. J Virol 2006;80:6368–6377.

Das SC, Pattnaik AK. Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. J Virol 2005;79:8101–8112.

Das T, Pattnaik AK, Takacs AM, Li T, Hwang LN, Banerjee AK. Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. Virology 1997;238:103–114.

De BP, Banerjee AK. Specific interactions of vesicular stomatitis virus L and NS proteins with heterologous genome ribonucleoprotein template lead to mRNA synthesis *in vitro*. J Virol 1984;51:628–634.

De BP, Banerjee AK. Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process *in vitro*. Biochem Biophys Res Comm 1985;126:40–49.

Deng X, Eickholt J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions. BMC Bioinformatics 2009;10: 436.

Dietzschold B, Li J, Faber M, Schnell M. Concepts in the pathogenesis of rabies. Future Virol 2008;3:481–490.

Ding H, Green TJ, Lu S, Luo M. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J Virol 2006;80:2808–2814.

Ding H, Green TJ, Luo M. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). Acta Crystallogr D Biol Crystallogr 2004;60:2087–2090.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002a;41:6573–6582.

Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem 2002b;62:25–49.

Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.

Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev 2004;104:3607–3622.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.

Eliezer D. Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 2009;19:23–30.

Emerson SU, Schubert M. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. Proc Natl Acad Sci USA 1987;84:5655–5659.

Emerson SU, Yu Y. Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. J Virol 1975;15:1348–1356.

Ferron F, Longhi S, Canard B, Karlin D. A practical overview of protein disorder prediction methods. Proteins 2006;65:1–14.

Finke S, Conzelmann KK. Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. J Virol 2003;77:12074–12082.

Fooks AR. Rabies remains a "neglected disease". Euro Surveill 2005;10:211–212.

Fu ZF, Zheng Y, Wunner WH, Koprowski H, Dietzschold B. Both the N- and the C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein. Virology 1994;200:590–597.

Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. Science 2010;327:689–693.

Gérard F, Ribeiro E, Albertini A, Zaccai G, Ebel C, Ruigrok R, Jamin M. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. Biochemistry 2007;46:10328–10338.

Gérard FCA, Ribeiro EA, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RWH, Jamin M. Modular organization of rabies virus phosphoprotein. J Mol Biol 2009;388:978–996.

Gigant B, Iseni F, Gaudin Y, Knossow M, Blondel D. Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. J Gen Virol 2000;81:1757–1761.

Gill DS, Chattopadhyay D, Banerjee AK. Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription *in vitro*. Proc Natl Acad Sci USA 1986;83:8873–8877.

Green TJ, Luo M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. Proc Natl Acad Sci USA 2009;106:11721–11726.

Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, Luo M. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. J Virol 2000;74:9515–9524.

Green TJ, Zhang X, Wertz GW, Luo M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 2006;313:357–360.

Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 2003;28:81–85.

Gupta AK, Blondel D, Choudhary S, Banerjee AK. The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. J Virol 2000;74:91–98.

Gupta AK, Shaji D, Banerjee AK. Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA. J Virol 2003;77:732–738.

Habchi, J, Longhi, S. Structural disorder within the nucleocapsids and phosphoproteins of measles, Nipah and Hendra viruses. In: Uversky, V, Longhi, S, editors. Flexible viruses: structural disorder in viral proteins. Hoboken, NJ: John Wiley & Sons; 2010. Chapter 3.

Hanlon CA, Kuzmin IV, Blanton JD, Weldon WC, Manangan JS, Rupprecht CE. Efficacy of rabies biologics against new lyssaviruses from Eurasia. Virus Res 2005;111:44–54.

Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2006;2: e100.

He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. Cell Res 2009;19:929–949.

Heinrich BS, Cureton DK, Rahmeh AA, Whelan SP. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. PLoS Pathog 2010;6: e1000958.

Hemachudha T, Laothamatas J, Rupprecht CE. Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges. Lancet Neurol 2002;1: 101–109.

Herman RC. Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein. J Virol 1986;58:797–804.

Howard M, Wertz G. Vesicular stomatitis virus RNA replication: a role for the NS protein. J Gen Virol 1989;70(Pt 10):2683–2694.

Hudson LD, Condra C, Lazzarini RA. Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template. J Gen Virol 1986;67(Pt 8):1571–1579.

Hwang LN, Englund N, Das T, Banerjee AK, Pattnaik AK. Optimal replication activity of vesicular stomatitis virus RNA polymerase requires phosphorylation of a residue(s) at carboxy-terminal domain II of its accessory subunit, phosphoprotein P. J Virol 1999;73:5613–5620.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004;32:1037–1049.

Iseni F, Barge A, Baudin F, Blondel D, Ruigrok RW. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. J Gen Virol 1998;79(Pt 12):2909–2919.

Iseni F, Baudin F, Blondel D, Ruigrok RW. Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. RNA 2000;6:270–281.

Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 2008;24:1344–1348.

Ito N, Moseley GW, Blondel D, Shimizu K, Rowe CL, Ito Y, Masatani T, Nakagawa K, Jans DA, Sugiyama M. Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity. J Virol 2010;84:6699–6710.

Ivanov I, Crepin T, Jamin M, Ruigrok R. Structure of the dimerisation domain of the rabies virus phosphoprotein. J Virol 2010;84:3707–3710.

Jacob Y, Real E, Tordo N. Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. J Virol 2001;75:9613–9622.

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc 2008;130:8055–8061.

Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernado P, Blackledge M. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 2009;17:1169–1185.

Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84:3239–3252.

Knobel DL, Cleaveland S, Coleman PG, Fevre EM, Meltzer MI, Miranda ME, Shaw A, Zinsstag J, Meslin FX. Re-evaluating the burden of rabies in Africa and Asia. Bull World Health Organ 2005;83:360–368.

Kolakofsky D, Le Mercier P, Iseni F, Garcin D. Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. Virology 2004;318:463–473.

Kouznetzoff A, Buckle M, Tordo N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. J Gen Virol 1998;79(Pt 5):1005–1013.

Lahaye X, Vidy A, Pomier C, Obiang L, Harper F, Gaudin Y, Blondel D. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. J Virol 2009;83:7948–7958.

Leslie MJ, Messenger S, Rohde RE, Smith J, Cheshier R, Hanlon C, Rupprecht CE. Bat-associated rabies virus in Skunks. Emerg Infect Dis 2006;12:1274–1277.

Lichty BD, Power AT, Stojdl DF, Bell JC. Vesicular stomatitis virus: re-inventing the bullet. Trends Mol Med 2004;10:210–216.

Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. BMC Genomics 2008;9(Suppl 2): S25.

Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry 2006;45:6873–6888.

Ma J, Ptashne M. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 1987;48:847–853.

Masters PS, Banerjee AK. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. J Virol 1988;62:2658–2664.

Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J Mol Biol 2004;343:819–831.

Mavrakis M, Mehouas S, Real E, Iseni F, Blondel D, Tordo N, Ruigrok RW. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. Virology 2006;349:422–429.

Mellon MG, Emerson SU. Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. J Virol 1978;27:560–567.

Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recogn 2010;23:105–116.

Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM, Kurgan L. Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics 2010;26: i489–i496.

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362:1043–1059.

Pasdeloup D, Poisson N, Raux H, Gaudin Y, Ruigrok RW, Blondel D. Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal. Virology 2005;334:284–293.

Pattnaik AK, Hwang L, Li T, Englund N, Mathur M, Das T, Banerjee AK. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. J Virol 1997;71:8167–8175.

Patton JT, Davis NL, Wertz GW. N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. J Virol 1984;49:303–309.

Paul PR, Chattopadhyay D, Banerjee AK. The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype). Virology 1988;166:350–357.

Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics 2001;17:700–712.

Peluso RW, Moyer SA. Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. Virology 1988;162: 369–376.

Poisson N, Real E, Gaudin Y, Vaney MC, King S, Jacob Y, Tordo N, Blondel D. Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P. J Gen Virol 2001;82:2691–2696.

Pontius BW. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends Biochem Sci 1993;18:181–186.

Pringle CR. The order Mononegavirales—current status. Arch Virol 1997;142: 2321–2326.

Ptashne M, Gann A. Transcriptional activation by recruitment. Nature 1997;386:569–577.

Qanungo KR, Shaji D, Mathur M, Banerjee AK. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. Proc Natl Acad Sci USA 2004;101:5952–5957.

Ribeiro EA, Favier A, Gerard FC, Leyrat C, Brutscher B, Blondel D, Ruigrok RW, Blackledge M, Jamin M Jr. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. J Mol Biol 2008;382:525–538.

Ribeiro EA, Leyrat C, Gérard FC, Albertini AA, Falk C, Ruigrok RW, Jamin M. Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. J Mol Biol 2009;394:558–575.

Romero P, Obradovic Z, Dunker AK. Natively disordered proteins: functions and predictions. Appl Bioinformatics 2004;3:105–113.

Rose NF, Publicover J, Chattopadhyay A, Rose JK. Hybrid alphavirus-rhabdovirus propagating replicon particles are versatile and potent vaccine vectors. Proc Natl Acad Sci USA 2008;105:5839–5843.

Ruigrok, R, Blackledge, M. Structural disorder within the Sendai virus nucleoprotein and phosphoprotein. In: Uversky, V, Longhi, S, editors. Flexible viruses: structural disorder in viral proteins. Hoboken, NJ: John Wiley & Sons; 2010. Chapter 4.

Rupprecht CE, Willoughby R, Slate D. Current and future trends in the prevention, treatment and control of rabies. Expert Rev Anti Infect Ther 2006;4:1021–1038.

Schlessinger A, Punta M, Yachdav G, Kajan L, Rost B. Improved disorder prediction by combination of orthogonal approaches. PLoS One 2009;4: e4433.

Schnell MJ, McGettigan JP, Wirblich C, Papaneri A. The cell biology of rabies virus: using stealth to reach the brain. Nat Rev Microbiol 2010;8:51–61.

Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RW. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. J Virol 2001;75:490–498.

Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 2000;97: 8868–8873.

Spiropoulou CF, Nichol ST. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. J Virol 1993;67:3103–3110.

Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. Science 1994;263:777–784.

Stillman EA, Whitt M. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. J Virol 1999;73:7199–7209.

Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science 2010;329:676–679.

Szilagyi JF, Uryvayev L. Isolation of an infectious ribonucleoprotein from vesicular stomatitis virus containing an active RNA transcriptase. J Virol 1973;11:279–286.

Takacs AM, Das T, Banerjee AK. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. Proc Natl Acad Sci USA 1993;90:10375–10379.

Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nat Struct Biol 2000;7:777–781.

Théodoridés J. Histoire de la rage. Paris: Masson; 1986.

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? Trends Biochem Sci 2009;34:53–59.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.

Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. Faseb J 2004;18:1169–1175.

Tordo, N, Benmansour, A, Calisher, C, Dietzgen, RG, Fang, R-X, Jackson, AO, Kurath, G, Nadin-Davis, S, Tesh, RB, Walker, PJ. Rhabdoviridae. In: Fauquet, M, Mayo, MA, Maniloff, J, Desselberger, U, Ball, LA, editors. Virus taxonomy, VIIIth Report of the ICTV. London: Elsevier/Academic Press; 2004. pp.623–644.

Toriumi H, Kawai A. Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N). Microbiol Immunol 2004;48:399–409.

Ugolini G. Advances in viral transneuronal tracing. J Neurosci Methods 2010;194:2–20.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18:343–384.

Vidy A, Chelbi-Alix M, Blondel D. Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. J Virol 2005;79:14411–14420.

Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. Proteins 2003;52:573–584.

Weyer J, Kuzmin IV, Rupprecht CE, Nel LH. Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length lyssavirus glycoprotein genes. Epidemiol Infect 2008;136:670–678.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.

Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol 2009;19:31–38.

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010;1804: 996–1010.

6

STRUCTURAL DISORDER IN MATRIX PROTEINS OF HIV-RELATED VIRUSES

GERARD K.-M. GOH, BIN XUE, A. KEITH DUNKER, AND VLADIMIR N. UVERSKY

6.1 INTRODUCTION

The viral matrix protein underlies the envelope of a virion, representing essentially a link between the envelope and the nucleocapsid (Cannon et al., 1997; Turner and Summers, 1999). The functions of matrix proteins are usually multifaceted and not completely understood (Dorfman et al., 1994; Harris et al., 1999; Hearps and Jans, 2007). They are, however, known to be involved in the viral assembly and stabilization of the lipid envelope (Lyles et al., 1992). Matrix proteins of different viral types are often structurally, functionally, and evolutionarily related (Harris et al., 1999). For instance, the influenza M1 and HIV p17 proteins are known to be related, and both have similar RNA- and membrane-binding domains (Harris et al., 1999).

Lentivirinae is a subfamily of slow-acting viruses in the Retroviridae family. Members of the Lentivirinae subfamily are among the genera of viruses that possess a matrix layer (Clements and Zink, 1996; Goudsmit, 1997). Viruses that belong to this genus include human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), bovine immunodeficiency virus (BIV), equine infectious anemia virus (EIAV), maedi-visna virus (MVV), and caprine arthritis-encephalitis virus (CAEV). The viruses in this family have

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

different characteristics especially with respect to the onset of diseases such as AIDS, the viral loads, and the success or failure in finding vaccines (Marx et al., 1991; Clements and Zink, 1996; Leroux et al., 2004). These viruses affect their hosts differently, possessing variable virulence and different modes of interaction with their host's immune systems. A question then arises whether some of the mentioned variability in the behavior can be reflected in some peculiar features of the corresponding viral proteins. This chapter examines matrix proteins of several viruses of the Lentivirinae subfamily using computational tools such as intrinsic disorder predictors to search for the crucial differences in the levels and distributions of intrinsic disorder in these matrix proteins.

The concept of protein intrinsic disorder is used in this chapter to investigate characteristics pertaining to the various viral matrix proteins. Intrinsically disordered proteins (IDPs) have been described by other names such as "intrinsically unstructured" (Wright and Dyson, 1999; Tompa, 2002), "natively unfolded" (Weinreb et al., 1996; Uversky et al., 2000), and "natively disordered" (Daughdrill et al., 2005). Historically, the investigation of intrinsic disorder began with finding and characterizing several proteins—exceptions from the paradigm stating that unique rigid protein structure is an unavoidable prerequisite for the specific protein function. Although such counterexamples were periodically observed, it was not till the end of the last century that researchers started to pay significant attention to this phenomenon (Dunker et al., 2008). As a result, the last decade witnessed the real rise of unfoldomics, a new field of protein science dealing with the various aspects of IDPs. It is recognized now that many crucial biological functions are performed by proteins that lack ordered tertiary and/or secondary structures, that is, by IDPs (Wright and Dyson, 1999; Uversky et al., 2000, 2005; Dunker et al., 2001, 2002, 2008; Tompa, 2002; Daughdrill et al., 2005; Dyson and Wright, 2005; Vucetic et al., 2007; Xie et al., 2007ab).

The fact that amino acid sequences/compositions of IDPs and ordered proteins are rather different was utilized to develop numerous disorder predictors, which became instrumental in the pursuit of a greater understanding of intrinsic disorder. Access to important information on many of these predictors is provided via the DisProt database (Sickmeier et al., 2007). In this chapter, we utilized two members of the PONDR® family of disorder predictors, VLXT and VL3 (Dunker et al., 1998; Romero et al., 1998, 2001; Garner et al., 1999; Obradovic et al., 2003; Vucetic et al., 2003), to examine the matrix proteins of the various viruses especially those related to HIV. PONDR VL3 was chosen because of its high accuracy in the prediction of long disordered regions (Obradovic et al., 2003), whereas PONDR VLXT was shown to be extremely sensitive for finding function-related disordered regions (Garner et al., 1999; Oldfield et al., 2005; Cheng et al., 2007). Uniqueness of this study is in the fact that we applied disorder predictors to proteins with known 3D structure. This approach revealed some peculiar patterns of predicted intrinsic disorder (PID) that can be used to better understand the behavior of the HIV matrix proteins.

6.2 INTRODUCING LENTIVIRINAE SUBFAMILY

Figure 6.1 represents a model of the HIV virion as an illustration of the Lentivirinae virion structure. The surface of the HIV virion is the viral envelope that is made of the cellular membrane, which is acquired when the virus leaves the host cell. Protruding from the envelope is the viral glycoprotein gp160, which is made up of two component parts, the structural unit (SU) gp120 and the transmembrane (TM) gp41. These two surface proteins play important roles in attachment and penetration of HIV into target cells. Inside the lipid envelope, there is a matrix formed by Gag protein p17, which holds the RNA-containing core in place. This cylindrical core is a proteinaceous capsid made of p24 protein. The capsid contains two copies of the single-stranded RNA genome and three key enzymes, namely, protease (PR, p11), integrase (IN, p32), and reverse transcriptase (RT, p66), as well as some other proteins.

There are three known HIV viruses in the world today, HIV-0, HIV-1, and HIV-2 (Marx et al., 1991; Jurriaans et al., 1994; Goudsmit, 1997). The latter two are of the most interest to our study. The HIV-1 is the predominant virus spreading around the globe. HIV-2, by contrast, is predominantly spread in certain parts of Africa, is being found in about 10% of HIV cases in West Africa, and has recently been found to be spreading in some parts of India (Jurriaans et al., 1994; Goudsmit, 1997). While the onset of AIDS usually occurs within an average of six years of

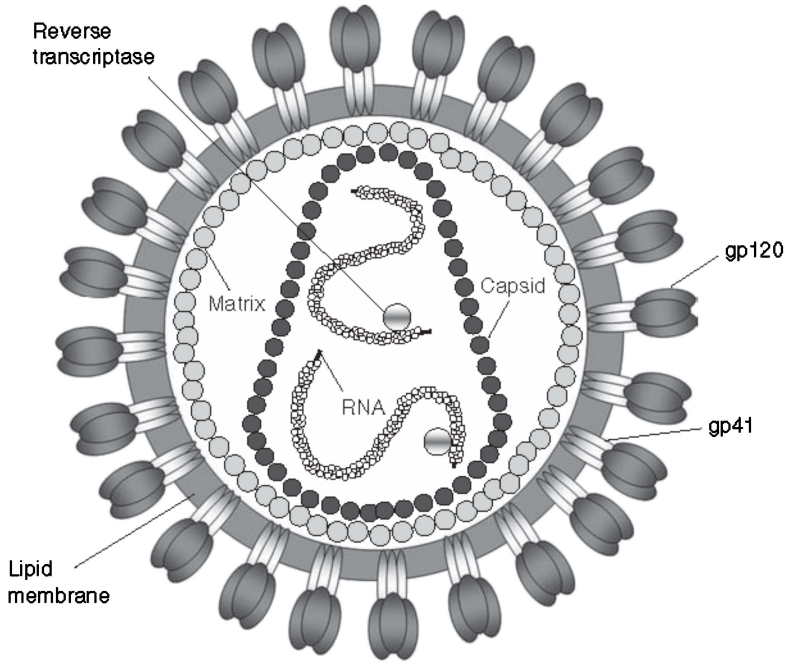


Figure 6.1 A model of the HIV virion structure. Major proteins.

virtually all HIV-1 infections, HIV-2 infection takes a much longer time before the AIDS symptoms appear, if at all (Jurriaans et al., 1994; Goudsmit, 1997; Morgan et al., 2002).

SIV is able to infect at least 33 species of African monkeys. A few SIV strains such as SIV_{cpz} (infecting chimpanzees) are more closely related to HIV-1, whereas most the others, especially SIV_{sm} and SIV_{mac} (infecting sooty mangabeys and rhesus macaque, respectively), are closer to HIV-2 (Marx et al., 1991). It should be noted that SIV does not usually cause AIDS among African nonhuman primates (Chen et al., 1995). It does, however, cause AIDS among Asian monkeys (Apetrei et al., 2006). It is believed that SIV_{sm} and SIV_{cpz} have crossed the species barrier into humans, resulting in HIV-2 and HIV-1, respectively.

FIV infects from 2.5% to 4.4% of cats worldwide (Lara et al., 2008). FIV isolates have been classified into five subtypes (A–E) based on the sequence diversity in variable regions (V3–V5) of the envelope protein (Sodora et al., 1994; Kakinuma et al., 1995; Pecoraro et al., 1996). In domestic cats, FIV infection results in an AIDS-like pathology characterized by a period of latency followed by a gradual depletion of $CD4^+$ lymphocytes, which ultimately results in immunosuppression (Willett et al., 1997). Following a prolonged asymptomatic period, some infected cats suffer from secondary or opportunistic infections and severe weight loss and sometimes show neurological signs and neoplasia (Bendinelli et al., 1995).

For the first time, a BIV was isolated in 1969 from a cow, R-29, with progressive deterioration in the physical condition reflected in clinical signs such as elevated white blood cell counts, lymphoadenopathy, evidence of central nervous system lesions, progressive weakness, and emaciation suggesting bovine leukosis (Malmquist et al., 1969). The virus was originally designated as "bovine visna-like virus" and remained unstudied until HIV was discovered. Later, it was demonstrated that the bovine R-29 isolate was a lentivirus with striking similarity to HIV (Gonda et al., 1987). Consequently, the designations "bovine immunodeficiency-like virus" and, thereafter, "bovine immunodeficiency virus" were used (St-Louis et al., 2004).

Similar to HIV, SIV, FIV, and BIV, EIAV is another retrovirus (Goudsmit, 1997; Leroux et al., 2004), which, however, spreads by insects, and the host targets are non-CD4 white blood cells such as macrophages and monocytes (Goudsmit, 1997; Beyrer, 2002). The disease caused by EIAV is not usually as fatal to its host as that of HIV, and ~90% of infected equine recover from an initial onset of symptoms (Goudsmit, 1997).

MVV may invade the lungs or the brain of sheep, causing chronic pneumonia (maedi) or encephalitis (visna), respectively. MVV infects its hosts for life. Most MVV infections are subclinical, but a minority of infected animals develop progressive, untreatable disease syndromes including dyspnea (maedi) or neurologic signs (visna). Maedi is characterized by a prolonged incubation period of more than two years, a progressive pneumonia lasting for about six months, and the inevitable death. The incubation period for visna is somewhat shorter, and symptoms can appear in sheep as young as two years (http://www.cfsph.iastate.edu/Factsheets/pdfs/maedi_visna.pdf).

CAEV is one of the only two lentiviruses that currently are known to infect sheep and goats. CAEV mostly affects goats and is closely related to MVV, which is found most often in sheep. Although documented cases of natural cross-species transmission are currently rare, CAEV can infect sheep and MVV can infect goats. Therefore, the eradication programs for either maedi-visna (ovine progressive pneumonia) or caprine arthritis and encephalitis should address both infections simultaneously. As with all other members of the Lentivirinae subfamily, infection with CAEV results in a persistent, life-long infection. Although most CAEV infections are subclinical, a minority of animals develop progressive, untreatable disease syndromes including arthritis in adult animals and encephalitis in young ones between two and six months of age. Other clinical presentations can include a hard udder or mastitis, hypogalactia, chronic interstitial pneumonia, and progressive weight loss (http://www.cfsph.iastate.edu/Factsheets/pdfs/caprine_arthritis_encephalitis.pdf).

While the search for vaccines for HIV continues to be difficult and elusive, effective vaccine for EIAV had been found 20 years ago in China (Marx et al., 1991; Beyrer, 2002; Burton et al., 2005). A major difficulty facing the search for HIV vaccines is a puzzling problem of the inability of HIV-protein-binding antibodies in eliciting an effective broad immune response (McMichael et al., 2002). While the reason for this remains largely unknown (Burton, 2002), a finding of high levels of IDPs at the surface, envelope, or perhaps, matrix could provide a mechanism by which the HIV virus evades the immune response.

6.3 MATRIX PROTEINS IN HIV-RELATED VIRUSES

6.3.1 HIV-1 Matrix Protein p17

The major structural component of all retroviruses is a Gag polyprotein, from which all the structural proteins are derived. Gag is a multidomain polypeptide that is able to assemble into viruslike particles when expressed in various cell types in the absence of other viral constituents (Gheysen et al., 1989; Wills and Craven, 1991), and Gag molecules can spontaneously assemble into spherical, immature viruslike particles *in vitro* (Campbell and Rein, 1999; Gross et al., 2000; Campbell et al., 2001). Although HIV-1 Gag contains the information necessary for tertiary and quaternary interactions, the viral particle assembly requires nonspecific RNA interactions both *in vivo* and *in vitro* and is assisted by host factors *in vivo*, including trafficking factors, assembly chaperones, and the ESCRT (endosomal sorting complex required for transport) budding pathway (Resh, 2005; Bieniasz, 2006; Klein et al., 2007; Williams and Urbe, 2007; Ganser-Pornillos et al., 2008).

Concomitant with or soon after the virion budding, HIV-1 Gag, which is synthesized as a precursor polyprotein (Pr55Gag) consisting of four major domains, is cleft by the virally encoded protease into the mature products: p17 matrix (MA), p24 capsid (CA), p7 nucleocapsid (NC), the C-terminal p6, and several small polypeptides including p1 and p2 (Gelderblom, 1991). These newly processed proteins then reassemble to form the distinct layers of the mature virion: MA remains associated with the inner viral membrane (the "matrix" layer), NC coats the viral RNA genome

(the "nucleocapsid" layer), and CA assembles into the conical capsid that surrounds the nucleocapsid and its associated enzymes, RT and IN (Ganser-Pornillos et al., 2008).

The HIV-1 matrix protein p17, constituting the N-terminal domain of the *Gag* gene product (Freed, 1998), is a 132-aa polypeptide myristoylated at N-terminus (Gottlinger et al., 1989; Bryant and Ratner, 1990). This cotranslational myristoylation of the N-terminus of the MA domain provides a targeting signal for Gag polyprotein transport to the PM (Gottlinger et al., 1989; Bryant and Ratner, 1990). p17 is directly associated with the inner leaflet of the viral membrane and forms a protective shell (Gelderblom, 1991). p17 is known to form trimers in solution (Morikawa et al., 1998; Tang et al., 2004) and in the crystals (Hill et al., 1996). It is suggested that HIV-1 p17 assembles into hexamers of trimers on membranes (Alfadhli et al., 2009). Structurally, individual p17 molecules are composed of five α-helices, a short helical stretch that forms a globular core, and a highly basic platform consisting of three β-strands (Massiah et al., 1994; Hill et al., 1996).

The matrix protein p17 possesses several important functions in the viral replication cycle, being potentially involved in nuclear import, likely via specific nuclear localization sequences, NLSs (Dingwall and Laskey, 1991; Riviere et al., 2010), and in targeting Gag polyproteins to the plasma membrane (PM) via its multipartite membrane-binding signal. In the late stage of infection, a key function of p17 is the recruitment of the viral surface/transmembrane gp120/gp41 envelope protein complex into virions. A second crucial function of p17 is to target Pr55Gag proteins to their assembly sites at the PM of infected cells (Freed, 1998). The interaction of p17 with the PM is mediated by the myristoylic moiety and by a cluster of positively charged residues located in the N-terminal region of the protein (Spearman et al., 1997). Pr55Gag direction to the PM is mediated by the p17 interaction with phosphatidylinositol-(4,5)-bisphosphate ($PI(4,5)P_2$) (Saad et al., 2006) that promotes exposure of the p17 myristate group, protein oligomerization, and virus assembly (Spearman et al., 1997).

The myristoylation signal and the NLS exert conflicting influences on the subcellular localization of the p17 matrix protein. The key regulation of these motifs might be phosphorylation of a portion of MA molecules on the C-terminal tyrosine at the time of virus maturation by virion-associated cellular tyrosine kinase.

Figure 6.2 represents a structure of the HIV-1 matrix protein p17 and indicates major domains involved in virus assembly, membrane binding, Gag targeting, postentry steps, and Env incorporation (Fiorentini et al., 2010).

6.3.2 HIV-2 Matrix Protein p17

Although HIV-1 was able to produce viruslike particles and bud from transfected yeast spheroplasts, HIV-2 was unable to assemble and bud in this system (Morikawa et al., 2007). This defect was attributed to the reduced ability of HIV-2 Gag proteins to remain stably associated with the PM at virus assembly sites (Morikawa et al., 2007). Recent NMR analysis revealed that this weaker association of the HIV-2 Gag proteins with the PM might be attributed to peculiar structural features of the

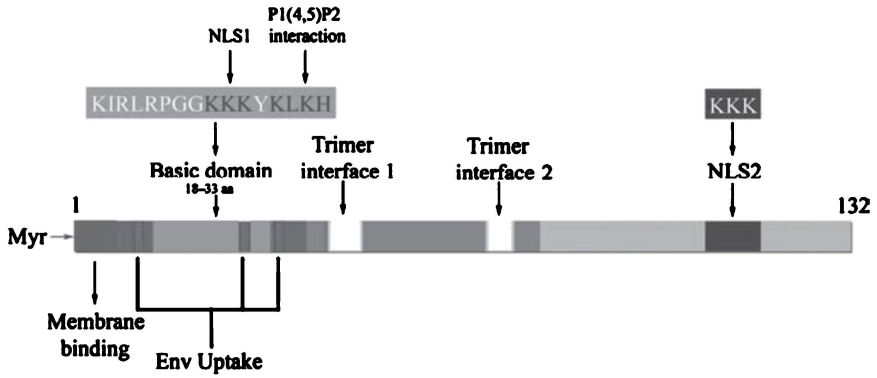


Figure 6.2 Structure of the p17, indicating the major domains involved in virus assembly, membrane binding, Gag targeting, postentry steps, and Env incorporation. *Source*: Adapted with the permission from Fiorentini et al. (2010).

HIV-2 p17, where the myristate of HIV-2 MA is more tightly sequestered within a narrow hydrophobic tunnel formed by side chains of helices 1, 2, 3, and 5 than that of the HIV-1 protein and did not exhibit concentration-dependent exposure (Saad et al., 2008).

6.3.3 SIV Matrix Protein p17

SIV matrix protein p17 is a myristoylated polypeptide of 120 residues. Similar to HIV, SIV MA proteins form the outer shell of the core of the virion, lining the inner surface of the viral membrane. In addition to myristoylation, a polybasic region within the N-terminus of the SIV p17 may also be necessary for proper membrane targeting of the Gag precursor (Gonzalez et al., 1996). The SIV MA has intrinsic information for both self-assembly and particle release, a conclusion based on the fact that when the MA protein is expressed in the absence of all other viral components by means of a vaccinia virus vector system, it assembles into lentivirus-like particles that are released into the medium of the infected cells (Gonzalez et al., 1993). Since gp120-gp41 Env complexes were found to be incorporated into the SIV MA-made particles, it has been concluded that SIV p17 contains domains that are involved in interactions with Env glycoproteins (Gonzalez et al., 1993).

6.3.4 FIV Matrix Protein p17

FIV matrix protein p17 consists of 131 residues. This *N*-myristoylated polypeptide has two main functions at different stages of the viral life cycle. At the very start of cell infection, FIV p17 plays a role in nuclear localization of the viral genome. In infected cells, this MA targets Gag and Gag-pol polyproteins to the PM via a multipartite membrane-binding signal. This matrix protein is the part of

the preintegration complex. FIV p17 utilizes a PM assembly mechanism similar to that of HIV and SIV p17, since perturbation of $PI(4,5)P_2$ levels was shown to inhibit the FIV release and the FIV MA basic patches were shown to be required for particle production (Manrique et al., 2004). FIV p17, however, is generally less basic than SIV MA or HIV-1 MA. Overall, a high degree of structural and functional conservation of FIV MA in comparison with SIV p17 or HIV p17 was emphasized (Burkala and Poss, 2007).

6.3.5 BIV Matrix Protein p16

Almost nothing is known about the BIV matrix protein p16, except for the notice that it has 126 amino acids.

6.3.6 Matrix Protein of Equine Infectious Anemia Virus

EIAV is a nonprimate enveloped virus from the *Lentivirus* genus of the Retroviridae family that causes the most important infectious disease of horses, equine infectious anemia. The wild-type EIAV MA consists of 124 residues. Contrary to HIV, SIV, and FIV MAs, EIAV MA lacks myristoylation sites. Neither the mechanism of recruitment to the membrane nor the actual packing of MA within the EIAV particle is fully understood.

Analysis of the crystal structure of EIAV MA at 2.8 Å resolution revealed that despite the lack of any apparent sequence similarity with MAs of primate lentiviruses, EIAV MA showed remarkable overall structural similarity to MAs of HIV-1 and SIV (Hatanaka et al., 2002). In fact, being aligned with the crystal structures of the primate lentivirus MAs, ~70% of the EIAV MA gave RMSD's of 1.9 and 1.8 Å for SIV and HIV-1 MAs, respectively (Hatanaka et al., 2002). C-terminal residues 110–124 were not seen in the electron density map, presumably because of their flexibility (Hatanaka et al., 2002).

It was also pointed out that although HIV-1 MA crystallized as a trimer, the EIAV MA was crystallized in a nonsymmetric dimeric unit, suggesting that this MA did not form trimers (Hatanaka et al., 2002). On the other hand, fluorescence studies revealed that EIAV MA existed as a multimer of two to three units and was able to efficiently interact with both neutral and negatively charged lipid bilayers (Provitera et al., 2000). The presence of the monomer-trimer equilibrium at micromolar concentrations was recently confirmed for EIAV MA by NMR (Chen et al., 2008).

6.3.7 Matrix Proteins of MVV and CAEV

Currently, much less is known about matrix proteins of small ruminant lentiviruses (SRLVs), the MVV, and the CAEV. MVV matrix protein p16 is a polypeptide consisting of 143 amino acids, whereas there are 146 residues in the sequence of the CAEV matrix protein p16.

6.4 INTRINSIC DISORDER IN LENTIVIRINAE MATRIX PROTEINS

6.4.1 PONDR® VLXT Analysis

Figure 6.3 compares the disorder profiles evaluated for different members of the Lentivirinae subfamily by the PONDR VLXT disorder predictor. Since UniProt contains multiple entries for the Gag and Gag/Pol polyproteins of these viruses (because of the annotation of proteins from different isolates and different strains, there are 77, 24, 27, 4, 1, 4, 4, and 1 entries for HIV-1, HIV-2, SIV, FIV, BIV, MVV, EIAV, and CAEV, respectively), we also analyzed the disorder variability within the matrix proteins of each virus. Figure 6.3 clearly shows that matrix proteins from different lentiviruses possess noticeable difference in their disorderedness. The common features for these disorder profiles are the high abundance of disorder in the N- and C-terminal regions of all the proteins and the presence of at least one middle region with relatively high disorder score. As a rule, immunodeficiency-related viruses are a bit more disordered than other lentiviruses. Another important observation is that for viruses with a large number of different matrix protein sequences (HIV, SIV, and FIV), the sequence variability was generally associated with the disorder variability. This suggests that the disordered regions of these lentiviruses are the evolutionary hot spots.

6.4.2 Percentage of Predicted Disordered Residues

Table 6.1 represents the estimations of the percentage of predicted disordered residues in the analyzed matrix and capsid proteins. Even though influenza virus is quite unrelated to lentiviruses, its M1 matrix protein is used here for comparison. It is also important to remember that the M1 protein is believed to be evolutionarily and structurally related to the p17 matrix protein of HIV. Table 6.1 shows that the amount of intrinsic disorder varies from 20% to 61% and from 0% to 40%, as evaluated by PONDR VLXT and VL3, respectively. Data for the matrix proteins with known 3D structures are further illustrated in Fig. 6.4 showing the results of the PONDR VLXT analysis as a bar chart. High level of PID in SIV and HIV-1 matrix proteins is clearly seen.

In our earlier paper (Goh et al., 2008a), the following classification of proteins characterized by X-ray crystallography but possessing various levels of predicted disorder was introduced: proteins with percentage of residues predicted to be disordered by PONDR VLXT between 20% and 29% were considered moderately disordered; those in the range of 30–39% were considered as quite disordered by prediction; and those that were disordered 40% and above were considered as very disordered by prediction. Therefore, the influenza M1 protein and EIAV p15 should be considered as moderately disordered by prediction. By the same token, the SIV_{mac} and HIV-1 p17 matrix proteins have to be considered as highly disordered by prediction.

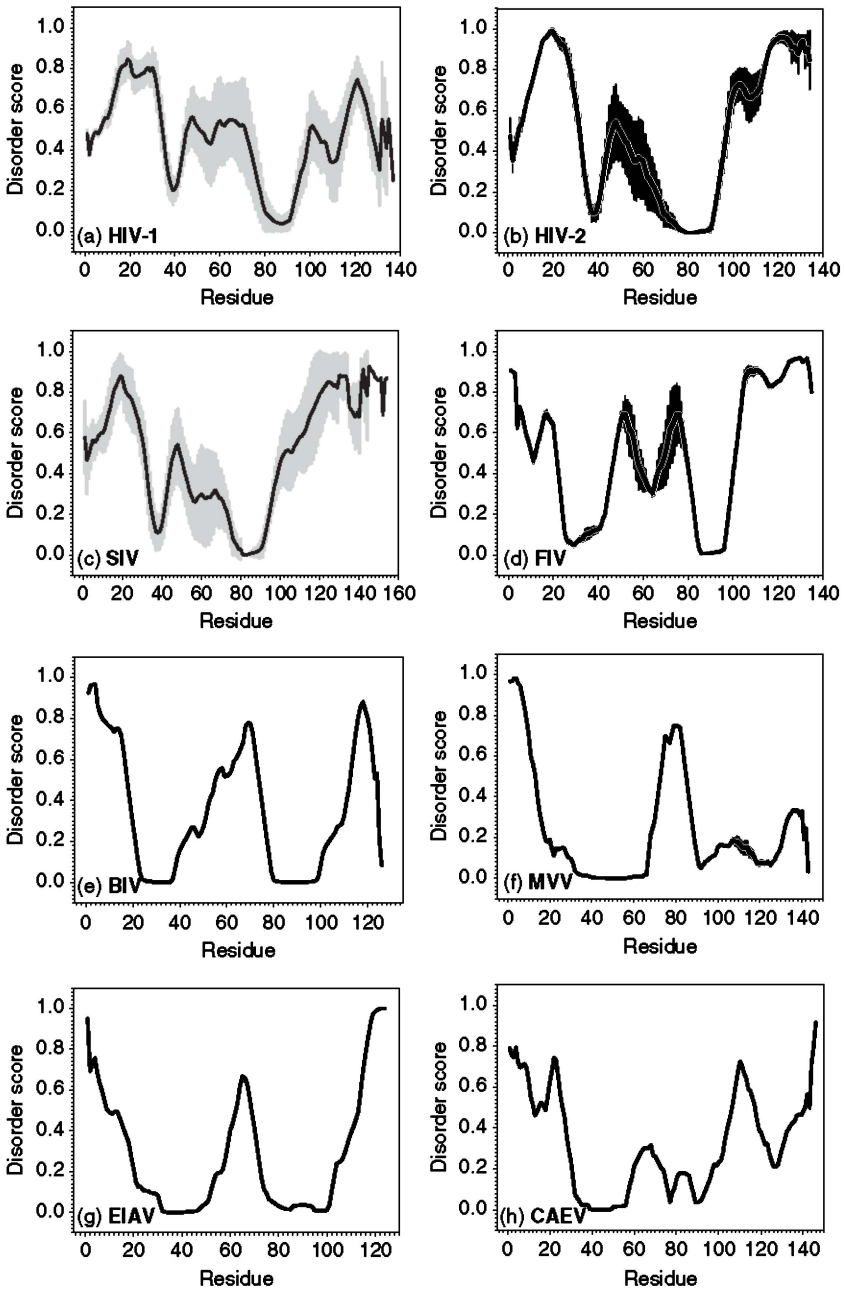


Figure 6.3 PONDR® VLXT analysis of matrix proteins from the members of Lentivirinae subfamily (a) HIV-1 (77), (b) HIV-2 (24), (c) SIV (27), (d) FIV (4), (e) BIV (1), (f) MVV (4), (g) EIAV (3), and (h) CAEI (1). Numbers in brackets represent the number of different sequences analyzed for each virus. Solid lines show averaged values, whereas average errors are shown by the gray shades.

TABLE 6.1 A Summary of Matrix Proteins and their Percentages of Disordered Residues in a Chain

Protein	Virus	Accession Number	Percentage of Predicted Disorder
M1	Influenza A	1ea3	25 (0)
Matrix (p17)	SIV_{mac}	1ed1	52 (40)
Matrix (p17)	HIV-1	1hiw	61 (39)
Matrix (p15)	EIAV	1hek	21 (12)
Capsid	HIV-1	1afv	48 (0)
Capsid	EIAV	1eia	30 (12)

All the samples analyzed were structurally characterized using X-ray crystallography. Percentage of predicted disorder corresponds to values produced by the PONDR® VLXT (VL3) analyses.

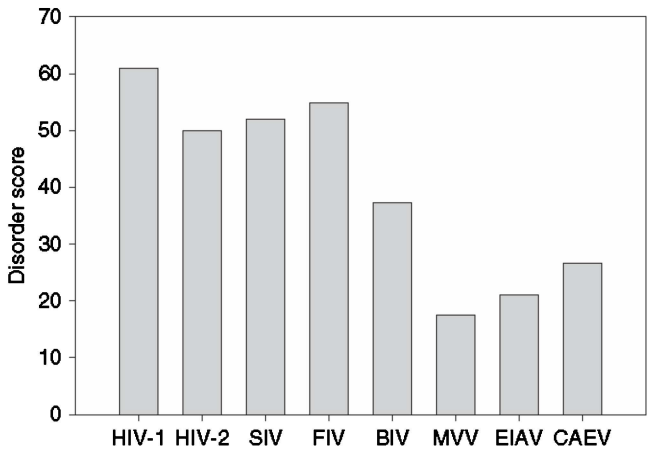


Figure 6.4 Bar chart comparing the abundance of disordered residues in matrix proteins of the Lentivirinae subfamily members. Plot represents the percentages of residues predicted to be disordered by PONDR® VLXT (i.e., the number of residues with the PONDR VLXT scores higher than 0.5).

6.4.3 PONDR Score versus B-Factor Plots and Contact Points

While Fig. 6.4 and Table 6.1 represent the predicted disorder of whole polypeptide chains, the PONDR VLXT plots in Fig. 6.5 represent per-residue distributions of disorder scores. They can be used to measure and compare factors that are not easily quantifiable. For example, Fig. 6.5 allows us to correlate the protein–protein contact sites (when such data are available) with the disorder score profiles. It also compares the normalized B-factor values (Radivojac et al., 2004) with the PONDR VLXT plots.

Analysis of Fig. 6.5 shows that contact sites (shown by thick horizontal gray lines) always correlate either with high B-factors or with high PONDR VLXT

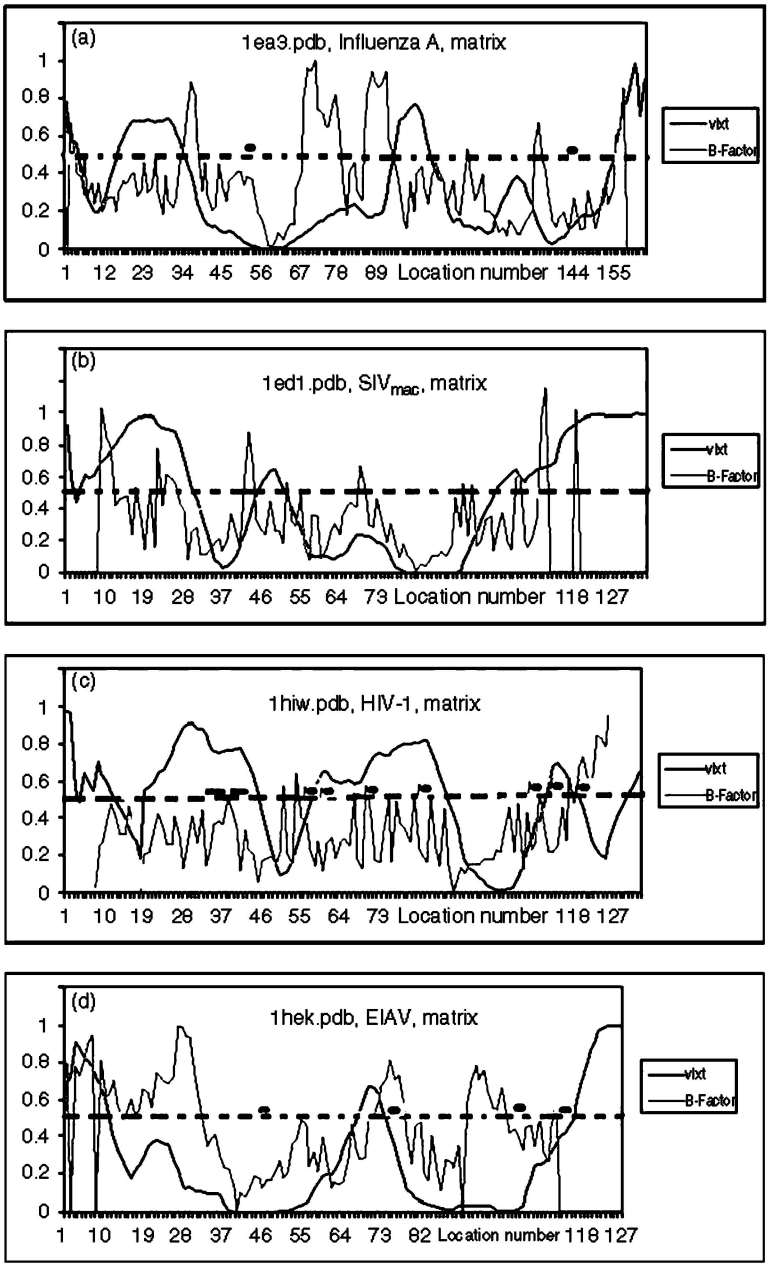


Figure 6.5 PONDR/B-factor plots of matrix proteins of various viruses. (a) The influenza A M1 protein. (b) The SIV_{mac} p17 protein. (c) HIV-1 p17 matrix protein. (d) The EIAV p15 matrix protein. Protein–protein contacts between chains are annotated by thick gray horizontal lines. The normalized B-factor values are shown by the light gray curves. PONDR® VLXT scores are shown by the blue curve in each plot. *Source*: Modified from Goh et al. (2008b). (*See insert for color representation of the figure.*)

scores, suggesting that highly flexible regions of matrix proteins are responsible for protein–protein interactions. For example, contacts between the subunits of HIV-1 p17 are located near or within regions predicted to be disordered, whereas contact sites of the EIAV p15 are mostly located in regions with high B-factor. These observations are in good agreement with earlier studies that established the usefulness of intrinsic disorder for protein–protein interaction (Garner et al., 1999; Wright and Dyson, 1999; Uversky et al., 2000, 2005; Dunker et al., 2001, 2002, 2005; Tompa, 2002; Uversky, 2002a,b, 2003; Daughdrill et al., 2005; Dyson and Wright, 2005; Oldfield et al., 2005; Cheng et al., 2007).

6.4.4 Enhancing 3D Structures with Predicted Disorder

Figure 6.6 provides 3D representations of the matrix proteins from various viruses. The areas in magenta are the protein regions predicted to be disordered by PONDR VL3 (and probably PONDR VLXT also), whereas the regions marked in red are those predicted to be disordered by PONDR VLXT. Different colors, such as blue and green in plot a, are used to denote different subunits. This presentation of structured proteins allows visualization of regions with the intrinsic propensity for being highly flexible.

6.4.5 HIV-1 versus HIV-2 and SIV_{mac}: Missing IDRs

6.4.5.1 SIV_{mac} is Very Similar to HIV-2 The HIV-2 and HIV-1 viruses, while related, differ in substantial ways in term of immune response, infection, and the onset of AIDS (Jurriaans et al., 1994; Goudsmit, 1997; Morgan et al., 2002). SIV_{mac} is a subtype of SIV, which was first found in macaques, and is known to be very closely related to HIV-2 (Marx et al., 1991). While development of AIDS symptoms are seen in virtually all HIV-1-infected patients, AIDS symptoms of HIV-2 infection appear only in a small percentage of patients. We believe that a comparative analysis of PID in related viral proteins could shed some light on the reasons behind these behaviors.

6.4.5.2 PID Rates of Matrix Proteins Correlate with the Difficulties in Finding Vaccines A brief glance at Table 6.1 and Fig. 6.4 shows that the PID rates of SIV_{mac} and HIV-1 are quite similar, even though the percentage of PID in SIV_{mac} p17 (50% by PONDR VLXT) is smaller than that in HIV-1 p17 (61%). The similarity in the level of PID is likely indicative of the ability of both viruses to evade the immune system. Further support for this hypothesis can be retrieved by analyzing the level of predicted disorder in the influenza M1 protein and in the EIAV p15 protein. Matrix proteins of both the viruses have low percentage disorder rates, 25% in M1 and 21% in p15. Interestingly, effective vaccines were developed for both these viruses, even though the mutation rates of the influenza virus is extremely high, causing well-known difficulties in the development of new vaccines.

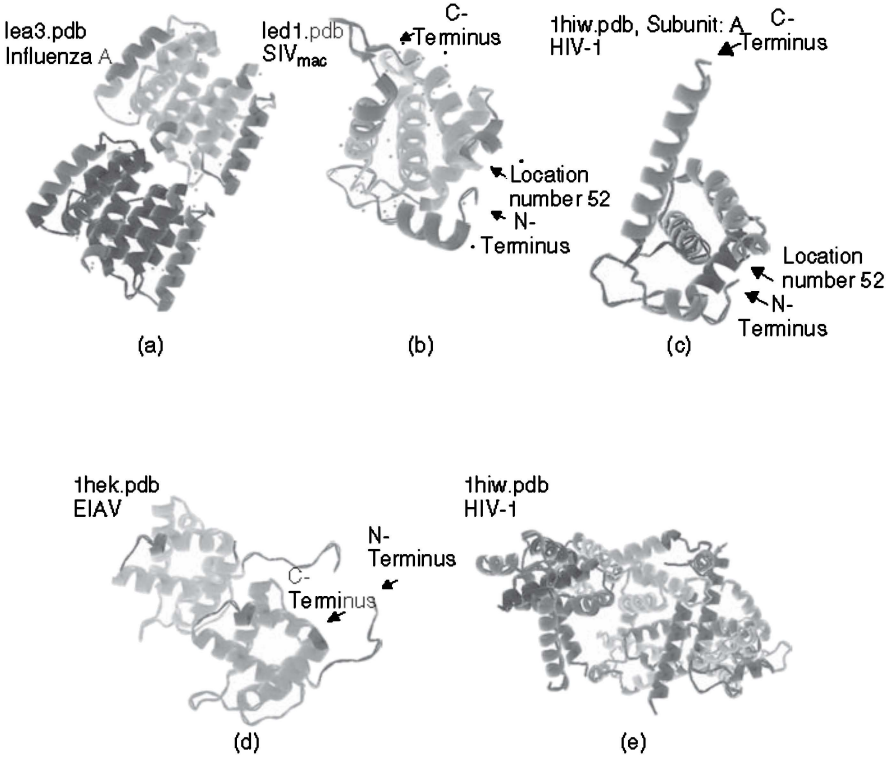


Figure 6.6 The 3D structures of the matrix proteins of the various viruses with predicted disorder annotation (predicted disorder is shown by red and magenta colors). (a) The influenza M1 protein. (b) SIV_{mac} p17 matrix protein. (c) HIV-1 p17 matrix protein shown as a monomer. (d) The EIAV matrix protein p15. (e) The HIV-1 matrix protein p17 shown as a multimer. The regions in magenta are regions predicted to be disordered by PONDR® VL3 (and probably also by PONDR VLXT). By contrast, the regions in red are areas predicted to be disordered by PONDR VLXT. *Source*: Modified from Goh et al. (2008b). (*See insert for color representation of the figure.*)

Apparently, the PID rate is a good predictor of the ease of vaccine development of a virus. This should not be surprising as our earlier study (Goh et al., 2008a) suggested that the viral matrix likely helps viruses to evade detection by the immune system because of its highly dynamic nature and constant motions. This dynamic behavior is correlated with the high propensity of matrix proteins for intrinsic disorder. Furthermore, it has been hypothesized that the role may be intertwined with the glycoprotein on the surface acting as a broom in the sweeping motion provided by the matrix (Goh et al., 2008a). This highly dynamic nature of the viral surface may explain the difficulties in the development of a vaccine for HIV.

6.4.5.3 Qualitative Differences in Predicted Disorder and Protein–Protein Interactions Even though the rates of predicted disorder in the SIV_{mac} and

HIV-1 p17 proteins seem to be similarly high, the PONDR VLXT plots revealed subtle differences in the disorder distribution within the protein sequences. Figure 5b,c shows that a long region predicted to be disordered by HIV-1 p17 (53–76 fragment) is missing in SIV_{mac} p17. Figure 6b,c, and e illustrates that in HIV-1 p17 this fragment forms an α-helix and is involved in protein–protein interactions between the subunits. In fact, residues 70–73 from one subunit contact residues 71, 60, 40, and 46 from another subunit. Analysis of Fig. 6.5c revealed that all these intersubunit contact sites are located within the PID regions. Therefore, intrinsic disorder plays a crucial role in the intersubunit interactions, which can be classified as disorder–disorder type of contact. The lack of a predicted-to-be-disordered segment in HIV-2 and SIV_{mac}, which seems to be crucial for intersubunit contacts, suggests that disorder–disorder protein–protein interactions are replaced by the order–disorder or order–order interactions.

6.4.5.4 Predicted Disorder Patterns Correlate with High B-Factors Figure 6.5 shows that, in general, there is a rather good correlation between the predicted disorder patterns and the normalized B-factor curves. For example, the 79–95 fragment of the HIV-1 matrix protein is both predicted to be disordered and is characterized by the high normalized B-factor values (Fig. 6.5c, 1hiw.pdb). In several occasions, there are noticeable lags between the PONDR VLXT and B-factor curves, as it is seen, for example, in Fig. 6.5a (M1 matrix proteins of the influenza A virus, 1ea3.pdb), where large B-factor peaks are seen in the 70–90 region, whereas the corresponding PID fragment is located in the 90–105 region.

6.4.6 HIV versus EIAV: Higher Predicted Disorder in HIV

6.4.6.1 Matrix of EIAV is Relatively Ordered Matrix protein of EIAV was predicted to be less disordered than that of HIV (see Table 6.1 and Fig. 6.4). However, even in this case, less abundant PID regions could be crucial for the intersubunit interactions. In fact, analysis of the crystal structure of the p15 protein revealed that residues 46 and 78 of the chain A are involved in interaction with the residues 114 and 105 of the chain B. All these interaction sites are shown as thick gray lines in Fig. 6.5d, which clearly indicate that the interactions between the 15 subunits are less rigorous than those of HIV-1 p17 subunits and can be ascribed to the order–disorder contact type. Since EIAV is from the same genus as HIV, that is, Lentivirinae (Beyrer, 2002; Leroux et al., 2004), these data suggest that the high PID levels are not a common characteristic of the Retroviridae family, or even the Lentivirinae genus.

Apparently, the high level of intrinsic disorder in the matrix proteins is a characteristic feature of HIV-1 and its closest relatives, SIV and HIV-2. These differences in the abundance of disorder seem to be largely constrained to the matrix proteins, as the capsids of both HIV-1 and EIAV viruses are quite disordered by prediction (48% and 30% by PONDR VLXT, see Table 6.1).

6.4.6.2 Predicted Disorder Patterns of EIAV are Closer to those of Influenza than to the Disorder Patterns of HIV/SIV Analysis of Fig. 6.5 reveals that the pattern of the predicted disorder in EIAV matrix protein is closer to that of the influenza virus than to the disorder profiles of EIAV's cousins HIV and SIV. Furthermore, matrix proteins of EIAV and influenza A are similar in their relatively low percentages of the predicted disorder (21% in EIAV and 25% in influenza A). The other similarity has to do with the interaction mode between the matrix protein subunits. In fact, contact sites of both influenza A and EIAV matrix proteins can be classified as disorder–order contacts. In the case of HIV-1, most of the contact sites between the subunits are predicted disorder–disorder interactions.

Comparison of the disorder and B-factor profiles of the HIV-1 and SIV_{mac} p17 proteins allows extrapolation of the potential modes of intersubunit interactions in SIV_{mac} p17. In fact, if potential interaction sites are distributed similarly within the amino acid sequences of HIV-1 and SIV_{mac} p17 proteins, then at least some of the SIV_{mac} p17 intersubunit interaction sites can be assigned as disorder–disorder interactions (e.g., if the residue 111 of one SIV_{mac} p17 subunit is in contact with the residue 97 from another subunit, then disorder–disorder contact takes place, as both these residues are predicted to be disordered, as seen in Fig. 6.5b).

6.5 HIGH INTRINSIC DISORDER AND IMMUNE RESPONSE

6.5.1 Potential Implications of More Disordered Matrix Proteins: Immune Evasion

The question then arose: what are the potential implications of more rigid or more disordered matrix proteins? It is likely that more rigid p17 proteins may be less effective in evading immune response. This may be a reason why HIV-2 and SIV_{mac} are less pathogenic than HIV-1. It is generally assumed that HIV-2 is less pathogenic than HIV-1 because of the fact that HIV-2 has less affinity for CD4 than HIV-1. On the other hand, our data show that there are subtle but important differences between HIV-1 and SIV_{mac} (HIV-2) in their patterns of predicted disorder distribution, which also might contribute to the virus's ability to evade the host immune system.

6.5.2 Implications to the Search for HIV Vaccines

Our findings might also have some implications to the search for HIV vaccines. One possibility is related to the use of animal models and SIV_{mac}, as in the search for HIV vaccination and drugs. SIV_{mac} and SIV_{stm} were the first subtypes found in laboratory macaques (Chen et al., 1995). Asian primates such as macaques, unlike their African cousins, developed AIDS on the average of 10 years after infection (Goudsmit, 1997). For this reason, the Asian monkeys have become the standard animal model for SIV (Hulskotte et al., 1998). However, the extrapolation of data from animal models to HIV in human remains a challenge. Our results suggest

that some of these challenges could be explained by the differences in disorder prediction between HIV-1 and SIV (or HIV-2). It is also important to remember that although the high levels of mutation caused difficulties in the development of vaccines against new strains of the influenza, there are effective vaccines against specific strains of the virus. Similarly, there are also effective vaccines available for EIAV. Note that the matrix proteins of both influenza virus and EIAV are shown in our study to contain less amount of intrinsic disorder.

6.5.3 Joint Role of Glycoproteins and Matrix Disorder

It is established that the HIV envelope glycoprotein gp120 is one of the most glycosylated proteins in Nature (Vigerust and Shepherd, 2007). Oligosaccharide moieties of viral glycoproteins often hide them from recognition by immune agents such as antibodies (Burton et al., 2005). We proposed that the abnormally disordered matrix proteins might help the surface glycoprotein in eluding immune responses. In other words, intrinsic disorder (read high dynamics) underneath the envelope would work in a tandem with envelope glycoproteins to help viruses in avoiding the induction of immune response. The question then arose: how and why would surface glycoprotein and matrix disorder work in cooperation? In a likely scenario, the oligosaccharide moieties of the glycoproteins act as an entropic brush that protects viral surface proteins such as gp120 and gp41 from contact with immune agents such as antibodies. The matrix protein could then provide the additional motion to the sweep. An advantage of motions that resemble a broom in a sweep is that it enables some regulatory roles via the matrix protein. Earlier it has been already observed that the envelope proteins are very sensitive to the behavior of the matrix proteins (Dorfman et al., 1994).

6.6 CONCLUDING REMARKS

6.6.1 Matrix Disorder of Retroviruses Vary with the Nature of the Virus

A peculiar finding in this chapter is the pattern of predicted disorder of EIAV p15 more closely matching the disorder profile of the influenza M1 protein than that of the matrix proteins of its closer relatives, namely, the HIV-1 and SIV_{mac} p17 proteins. This feature may be attributed to the ways the viruses are evolved and are transmitted to their hosts. It should be reminded that EIAV is transmitted between horses via insect vectors. In other words, the virus experiences a dramatic change in the environment during the transmission. It is likely that this mode of transmission has evolutionary requirements similar to those of the influenza virus, which is transmitted via the respiratory tract and mucus. HIV and SIV, on the other hand, spread by blood contact or sexual activities. Since there is less chance for exposure to the outside environment in the transmission mode, there is less evolutionary pressure for the matrix proteins to be ordered. This highlights a role for the matrix protein in many viruses. In many instances, the matrix acts as an

encasement for the virion, thereby protecting the virion from damage especially in adverse environments. We have also seen that disorder at the matrix is not an absolute characteristic of retroviruses.

6.6.2 Implication for the Immune System Invisibility Puzzle of HIV

A single nagging puzzle in the search for vaccines against HIV is the unknown mechanism helping the virus to evade immune response. Our study suggests that this ability might arise from the abnormal levels of intrinsic disorder at the viral matrix. This hypothesis is supported by the fact that the matrix proteins of other viruses, where vaccines have been more easily found, were predicted to be more ordered. Therefore, there are several ways how disorder predictions can be utilized in the future strategies of the vaccine development. Particularly, one of the new directions in the anti-HIV drug development could be a search for the therapeutic agents able to stabilize the HIV matrix protein.

Another puzzle of HIV viruses is the inability of virologists to account for the waves of the HIV strains seen, even after taking into account the fact that the mutation rate of HIV-1 is 25 times that of influenza. Yet another HIV puzzle is the greater pathogenicity of HIV-1 as compared to HIV-2. It has been generally understood that this is due to the fact that the HIV-1 affinity for CD4 is 28 times greater than that for HIV-2 (Chakravarty et al., 2006). Our data suggest that it is not just the affinity for CD4 that gives rise to a greater pathogenesis or viral load in HIV-1. Perhaps, it is also the differences in the abilities of the viruses in evading the immune system via disorder at the matrix. This also explains a related observation among epidemiologists that the more easily HIV-1 spreads sexually, the more virulent it becomes (Goudsmit, 1997), since the ease of transmission via blood or sexual intercourse lessens the requirements for a rigid encasement of the virion, which is used in other viruses to prevent virion damage due to harsh environmental factors.

6.6.3 Potential Implications for the Immune Evasion of Cancer Cells and Oncolysis

While this chapter has been largely focused on the study of immune evasion as applied to HIV and HIV-related viruses, it may provide a model for immune evasion by other entities, such as cancer cells. There are either very few or no studies conducted in this area. Perhaps, our results could invigorate interest in this area, given the models and approach used. Furthermore, the results of this chapter likely have novel strategic implications for experimental studies on the use of viruses as oncolytic agents, which have often been observed to be rendered ineffective by the immune system. In fact, one of the greatest problems in using the oncolytic viruses is that they are detected by the immune system very quickly so they are only useful for localized treatment of tumors (Chernajovsky et al., 2006). Our data suggest that this does not have to be always the case, and new oncolytic viruses with disordered matrix should be considered.

ACKNOWLEDGMENTS

This work was supported in part by the grant EF 0849803 (to A. K. D and V. N. U.) from the National Science Foundation, and the Program of the Russian Academy of Sciences for the "Molecular and Cellular Biology" (to V. N. U.). We gratefully acknowledge the support of the IUPUI Signature Centers Initiative. This chapter is partially based on a previous publication (Goh et al., 2008b).

ABBREVIATIONS

BIV	bovine immunodeficiency virus
CA	capsid
CAEV	caprine arthritis-encephalitis virus
EIAV	equine infectious anemia virus
IN	integrase
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
FIV	feline immunodeficiency virus
HIV	human immunodeficiency virus
MA	matrix
MVV	maedi-visna virus
NC	nucleocapsid
$PI(4,5)P_2$	phosphatidylinositol-(4,5)-bisphosphate
PID	predicted intrinsic disorder
PM	plasma membrane
PONDR	predictor of naturally disordered regions
PR	protease
RT	reverse transcriptase
SIV	simian immunodeficiency virus
SIV_{cpz}	SIV infecting chimpanzees
SIV_{mac}	SIV infecting rhesus macaque
SIV_{sm}	SIV infecting sooty mangabeys
SRLV	small ruminant lentivirus
SU	structural unit
TM	transmembrane

REFERENCES

Alfadhli A, Barklis RL, Barklis E. HIV-1 matrix organizes as a hexamer of trimers on membranes containing phosphatidylinositol-(4,5)-bisphosphate. Virology 2009;387(2):466–472.

Apetrei C, Lerche NW, Pandrea I, Gormus B, Silvestri G, Kaur A, Robertson DL, Hardcastle J, Lackner AA, Marx PA. Kuru experiments triggered the emergence of pathogenic SIVmac. AIDS 2006;20(3):317–321.

Bendinelli M, Pistello M, Lombardi S, Poli A, Garzelli C, Matteucci D, Ceccherini-Nelli L, Malvaldi G, Tozzini F. Feline immunodeficiency virus: an interesting model for AIDS studies and an important cat pathogen. Clin Microbiol Rev 1995;8(1):87–112.

Beyrer C. Injecting drug users and HIV vaccine trials: What does the science say? AIDScience 2002;2:1–6.

Bieniasz PD. Late budding domains and host proteins in enveloped virus release. Virology 2006;344(1):55–63.

Bryant M, Ratner L. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc Natl Acad Sci USA 1990;87(2):523–527.

Burkala E, Poss M. Evolution of feline immunodeficiency virus Gag proteins. Virus Genes 2007;35(2):251–264.

Burton DR. Antibodies, viruses and vaccines. Nat Rev Immunol 2002;2(9):706–713.

Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans. Proc Natl Acad Sci USA 2005;102(42):14943–14948.

Campbell S, Fisher RJ, Towler EM, Fox S, Issaq HJ, Wolfe T, Phillips LR, Rein A. Modulation of HIV-like particle assembly *in vitro* by inositol phosphates. Proc Natl Acad Sci USA 2001;98(19):10875–10879.

Campbell, S, Rein, A. *In vitro* assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. J Virol 1999;73(3):2270–2279.

Cannon PM, Matthews S, Clark N, Byles ED, Iourin O, Hockley DJ, Kingsman SM, Kingsman AJ. Structure-function studies of the human immunodeficiency virus type 1 matrix protein, p17. J Virol 1997;71(5):3474–3483.

Chakravarty J, Mehta H, Parekh A, Attili SV, Agrawal NR, Singh SP, Sundar S. Study on clinico-epidemiological profile of HIV patients in eastern India. J Assoc Phys India 2006;54:854–857.

Chen K, Bachtiar I, Piszczek G, Bouamr F, Carter C, Tjandra N. Solution NMR characterizations of oligomerization and dynamics of equine infectious anemia virus matrix protein and its interaction with PIP2. Biochemistry 2008;47(7):1928–1937.

Chen Z, Telfer P, Reed P, Zhang L, Getti A, Ho DD, Marx PA. Isolation and characterization of the first simian immunodeficiency virus from a feral sooty mangabey (Cercocebus atys) in West Africa. J Med Primatol 1995;24(3):108–125.

Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 2007;46(47):13468–13477.

Chernajovsky Y, Layward L, Lemoine N. Fighting cancer with oncolytic viruses. BMJ 2006;332(7534):170–172.

Clements JE, Zink MC. Molecular biology and pathogenesis of animal lentivirus infections. Clin Microbiol Rev 1996;9(1):100–117.

Daughdrill, GW, Pielak, GJ, Uversky, VN, Cortese, MS, Dunker, AK. Natively disordered proteins. In: Buchner J, Kiefhaber T, editors. Protein folding handbook. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA; 2005. pp.271–353.

Dingwall C, Laskey RA. Nuclear targeting sequences—a consensus? Trends Biochem Sci 1991;16(12):478–481.

Dorfman T, Mammano F, Haseltine WA, Gottlinger HG. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. J Virol 1994;68(3):1689–1696.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41(21):6573–6582.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 2005;272(20):5129–5148.

Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998;473–484.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19(1):26–59.

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Cheng JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: An update on intrinsically disordered proteins. BMC Genomics 2008;9(Suppl 2): S1.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6(3):197–208.

Fiorentini S, Giagulli C, Caccuri F, Magiera AK, Caruso A. HIV-1 matrix protein p17: a candidate antigen for therapeutic vaccines against AIDS. Pharmacol Ther 2010;128(3):433–444.

Freed EO. HIV-1 gag proteins: diverse functions in the virus life cycle. Virology 1998;251(1):1–15.

Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. Curr Opin Struct Biol 2008;18(2):203–217.

Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting binding regions within disordered proteins. Genome Inform Ser Workshop Genome Inform 1999;10:41–50.

Gelderblom HR. Assembly and morphology of HIV: potential effect of structure on viral function. AIDS 1991;5(6):617–637.

Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D, De Wilde M. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell 1989;59(1):103–112.

Goh GK-M, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genomics 2008a;9(Suppl 2): S4.

Goh GK, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. Virol J 2008b;5: 126.

Gonda MA, Braun MJ, Carter SG, Kost TA, Bess JW, Arthur LO, Van der Maaten MJ Jr. Characterization and molecular cloning of a bovine lentivirus related to human immunodeficiency virus. Nature 1987;330(6146):388–391.

Gonzalez SA, Affranchino JL, Gelderblom HR, Burny A. Assembly of the matrix protein of simian immunodeficiency virus into virus-like particles. Virology 1993;194(2):548–556.

Gonzalez SA, Burny A, Affranchino JL. Identification of domains in the simian immunodeficiency virus matrix protein essential for assembly and envelope glycoprotein incorporation. J Virol 1996;70(9):6384–6389.

Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 1989;86(15):5781–5785.

Goudsmit J. Viral sex: the nature of AIDS. New York: Oxford University Press; 1997.

Gross I, Hohenberg H, Wilk T, Wiegers K, Grattinger M, Muller B, Fuller S, Krausslich HG. A conformational switch controlling HIV-1 morphogenesis. EMBO J 2000;19(1):103–113.

Harris A, Sha B, Luo M. Structural similarities between influenza virus matrix protein M1 and human immunodeficiency virus matrix and capsid proteins: an evolutionary link between negative-stranded RNA viruses and retroviruses. J Gen Virol 1999;80(Pt 4):863–869.

Hatanaka H, Iourin O, Rao Z, Fry E, Kingsman A, Stuart DI. Structure of equine infectious anemia virus matrix protein. J Virol 2002;76(4):1876–1883.

Hearps AC, Jans DA. Regulating the functions of the HIV-1 matrix protein. AIDS Res Hum Retroviruses 2007;23(3):341–346.

Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci USA 1996;93(7):3099–3104.

Hulskotte EG, Geretti AM, Osterhaus AD. Towards an HIV-1 vaccine: lessons from studies in macaque models. Vaccine 1998;16(9–10):904–915.

Jurriaans S, Van Gemen B, Weverling GJ, Van Strijp D, Nara P, Coutinho R, Koot M, Schuitemaker H, Goudsmit J. The natural history of HIV-1 infection: virus load and virus phenotype independent determinants of clinical course? Virology 1994;204(1):223–233.

Kakinuma S, Motokawa K, Hohdatsu T, Yamamoto JK, Koyama H, Hashimoto H. Nucleotide sequence of feline immunodeficiency virus: classification of Japanese isolates into two subtypes which are distinct from non-Japanese subtypes. J Virol 1995;69(6):3639–3646.

Klein KC, Reed JC, Lingappa JR. Intracellular destinies: degradation, targeting, assembly, and endocytosis of HIV Gag. AIDS Rev 2007;9(3):150–161.

Lara VM, Taniwaki SA, Araújo Júnior JP. Occurrence of feline immunodeficiency virus infection in cats. Ciência Rural 2008;38:2245–2249.

Leroux C, Cadore JL, Montelaro RC. Equine infectious anemia virus (EIAV): what has HIV's country cousin got to tell us? Vet Res 2004;35(4):485–512.

Lyles DS, McKenzie M, Parce JW. Subunit interactions of vesicular stomatitis virus envelope glycoprotein stabilized by binding to viral matrix protein. J Virol 1992;66(1):349–358.

Malmquist WA, Van der Maaten MJ, Boothe AD. Isolation, immunodiffusion, immunofluorescence, and electron microscopy of a syncytial virus of lymphosarcomatous and apparently normal cattle. Cancer Res 1969;29(1):188–200.

Manrique ML, Gonzalez SA, Affranchino JL. Functional relationship between the matrix proteins of feline and simian immunodeficiency viruses. Virology 2004;329(1):157–167.

Marx PA, Li Y, Lerche NW, Sutjipto S, Gettie A, Yee JA, Brotman BH, Prince AM, Hanson A, Webster RG, et al. Isolation of a simian immunodeficiency virus related to human immunodeficiency virus type 2 from a west African pet sooty mangabey. J Virol 1991;65(8):4480–4485.

Massiah MA, Starich MR, Paschall C, Summers MF, Christensen AM, Sundquist WI. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. J Mol Biol 1994;244(2):198–223.

McMichael A, Mwau M, Hanke T. Design and tests of an HIV vaccine. Br Med Bull 2002;62:87–98.

Morgan D, Mahe C, Mayanja B, Whitworth JA. Progression to symptomatic disease in people infected with HIV-1 in rural Uganda: prospective cohort study. BMJ 2002;324(7331):193–196.

Morikawa Y, Goto T, Yasuoka D, Momose F, Matano T. Defect of human immunodeficiency virus type 2 Gag assembly in Saccharomyces cerevisiae. J Virol 2007;81(18):9911–9921.

Morikawa Y, Zhang WH, Hockley DJ, Nermut MV, Jones IM. Detection of a trimeric human immunodeficiency virus type 1 Gag intermediate is dependent on sequences in the matrix protein, p17. J Virol 1998;72(9):7659–7663.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(Suppl 6):566–572.

Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry 2005;44(6):1989–2000.

Pecoraro MR, Tomonaga K, Miyazawa T, Kawaguchi Y, Sugita S, Tohya Y, Kai C, Etcheverrigaray ME, Mikami T. Genetic diversity of Argentine isolates of feline immunodeficiency virus. J Gen Virol 1996;Pt 77(9):2031–2035.

Provitera P, Bouamr F, Murray D, Carter C, Scarlata S. Binding of equine infectious anemia virus matrix protein to membrane bilayers involves multiple interactions. J Mol Biol 2000;296(3):887–898.

Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, Dunker AK. Protein flexibility and intrinsic disorder. Protein Sci 2004;13(1):71–80.

Resh MD. Intracellular trafficking of HIV-1 Gag: how Gag interacts with cell membranes and makes viral particles. AIDS Rev 2005;7(2):84–91.

Riviere L, Darlix JL, Cimarelli A. Analysis of the viral elements required in the nuclear import of HIV-1 DNA. J Virol 2010;84(2):729–739.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK. Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 1998;437–448.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38–48.

Saad JS, Ablan SD, Ghanam RH, Kim A, Andrews K, Nagashima K, Soheilian F, Freed EO, Summers MF. Structure of the myristylated human immunodeficiency virus type 2 matrix protein and the role of phosphatidylinositol-(4,5)-bisphosphate in membrane targeting. J Mol Biol 2008;382(2):434–447.

Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc Natl Acad Sci USA 2006;103(30):11364–11369.

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. DisProt: the database of disordered proteins. Nucleic Acids Res 2007;35(Database issue): D786–D793.

Sodora DL, Shpaer EG, Kitchell BE, Dow SW, Hoover EA, Mullins JI. Identification of three feline immunodeficiency virus (FIV) env gene subtypes and comparison of the FIV and human immunodeficiency virus type 1 evolutionary patterns. J Virol 1994;68(4):2230–2238.

Spearman P, Horton R, Ratner L, Kuli-Zade I. Membrane binding of human immunodeficiency virus type 1 matrix protein *in vivo* supports a conformational myristyl switch mechanism. J Virol 1997;71(9):6582–6592.

St-Louis MC, Cojocariu M, Archambault D. The molecular biology of bovine immunodeficiency virus: a comparison with other lentiviruses. Anim Health Res Rev 2004;5(2):125–143.

Tang C, Loeliger E, Luncsford P, Kinde I, Beckett D, Summers MF. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. Proc Natl Acad Sci USA 2004;101(2):517–522.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27(10):527–533.

Turner BG, Summers MF. Structural biology of HIV. J Mol Biol 1999;285(1):1–32.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002a;11(4):739–756.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002b; 269(1):2–12.

Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell Mol Life Sci 2003;60(9):1852–1871.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–427.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18(5):343–384.

Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol 2007;15(5):211–218.

Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. Proteins 2003;52(4):573–584.

Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. J Proteome Res 2007;6(5):1899–1916.

Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 1996;35(43):13709–13715.

Willett BJ, Flynn JN, Hosie MJ. FIV infection of the domestic cat: an animal model for AIDS. Immunol Today 1997;18(4):182–189.

Williams RL, Urbe S. The emerging shape of the ESCRT machinery. Nat Rev Mol Cell Biol 2007;8(5):355–368.

Wills JW, Craven RC. Form, function, and use of retroviral gag proteins. AIDS 1991;5(6): 639–654.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293(2):321–331.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J Proteome Res 2007a;6(5):1917–1932.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 2007b;6(5):1882–1898.

7

STRUCTURAL DISORDER IN PROTEINS FROM INFLUENZA VIRUS

GERARD K.-M. GOH, BIN XUE, A. K. DUNKER, AND VLADIMIR N. UVERSKY

7.1 INTRODUCTION

Many proteins are intrinsically disordered; that is, they lack a rigid three-dimensional (3D) structure under physiological conditions *in vitro*, existing instead as dynamic ensembles of interconverting structures. Intrinsically disordered proteins (IDPs) (Romero et al., 2001) are also known by several other names including *intrinsically unstructured* (Wright and Dyson, 1999) and *natively unfolded* (Weinreb et al., 1996; Uversky et al., 1999, 2000). While the function of a given protein is often determined by its unique structure, comparative studies on several exceptions to the structure-to-function mechanism led to the realizations that IDPs share many sequence characteristics and so comprise a distinct cohort. These IDPs and intrinsically disordered regions (IDRs) differ from structured globular proteins and domains with regard to many attributes, including the amino acid composition, sequence complexity, hydrophobicity, charge, flexibility (Uversky et al., 2000; Romero et al., 2001), and type and rate of amino acid substitutions over evolutionary time (Brown et al., 2002). Many of these differences between ordered proteins and IDPs were utilized to develop numerous disorder predictors.

Structures and functions of a large number of viral proteins are not yet totally understood (Burton, 2002; Hilleman, 2002; McMichael et al., 2002; Reid et al., 2004a; Garcia-Sastre and Whitley, 2006). This may account for the continuous need for the development of novel computational and experimental tools suitable for

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

viral protein analysis. Although experimental techniques remain the major providers of structural and functional knowledge, often, the experiments are expensive or difficult to the point of infeasibility. The use of various bioinformatics tools to predict structure and function represents an alternative approach that is gaining significant attention. Comparative computational studies have opened a new way for easier benchmarking and functional analysis of proteins. In this chapter, we examine the usefulness of intrinsic disorder predictions for studying the viral proteins. To this end, we utilized the PONDR® VLXT and VL3 disorder predictors (Li et al., 1999; Romero et al., 2001; Obradovic et al., 2003; Vucetic et al., 2003) to address the following question: can disorder prediction be used to determine or map at least some of the functions for viral proteins?

7.2 INTRODUCING INFLUENZA VIRUS

7.2.1 Structure of the Influenza Virion

Influenza viruses belong to the Orthomyxoviridae family of negative-sense, single-stranded, segmented RNA viruses. This family contains five genera, classified by variations in nucleoprotein (NP and M) antigens into influenza A, influenza B, influenza C, thogotovirus, and isavirus (Brooks et al., 2004). Influenza A viruses, which are a major cause of influenza in humans, have multiple subtypes that are labeled according to an H number (for hemagglutinin (HA)) and an N number (for neuraminidase (NA)). Sixteen HA subtypes (or serotypes, HA antigens H1 to H16) and nine NA subtypes of influenza A virus (NA antigens N1 to N9) have been identified in different virus isolations so far for influenza A (Fouchier et al., 2005). The influenza A virus genome is contained on eight single (nonpaired) RNA strands that code for 11 proteins (HA; NA; NP; matrix proteins (M1 and M2); nonstructural proteins (NS1 and nuclear export protein (NEP)); RNA polymerase (acidic polymerase protein (PA), basic polymerase protein (PB1), and PB2); and PB1-F2 protein). The total genome size is 13,588 bases (http://www.microbiologybytes.com/virology/Orthomyxoviruses.html).

The influenza A virion (which is a complete virus particle with its RNA core and protein coat) is a globular particle sheathed in a lipid bilayer derived from the plasma membrane of its host (Fig. 7.1). Two integral membrane proteins, HA and NA, are studded in the lipid bilayer. They are distributed evenly over the virion surface, forming characteristic spike-shaped structures. Beneath the envelope, the matrix formed by matrix proteins M1 and M2 is located. This matrix encompasses eight pieces of the genomic RNA, each in association with many copies of an NP, some "nonstructural" proteins with various functions (e.g., NS1 and NEP), and several molecules of the three subunits of its RNA polymerase.

Table 7.1 lists the major functions for several viral proteins. It is important to remember that some of the functions are not totally understood or are not known at all (Burton, 2002). Multifunctionality (or moonlighting) of a protein is, of course, also possible.

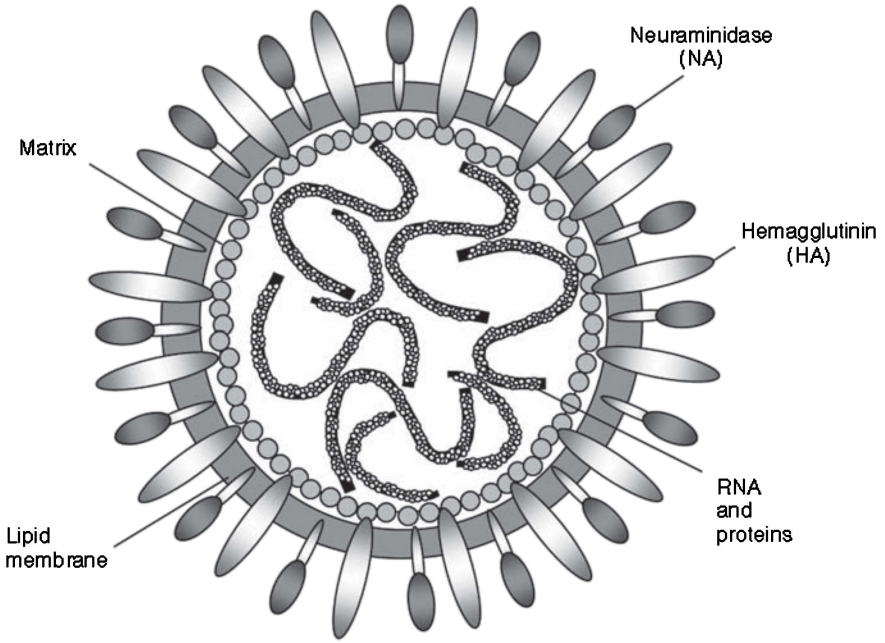


Figure 7.1 Model structure of the influenza A virion.

TABLE 7.1 Influenza Viral Proteins and Their Functions. Some of the Most Important Influenza Viral Proteins are Listed with a Brief Description of Their Functions. These Proteins are Listed by the Order of Their Location in the Virion (Clements and Zink, 1996; Turner and Summers, 1999; Lederberg, 2001; Brooks et al., 2004)

Protein	Function/Description	Location
HA	Hemagglutinin allows the attachment to host and to itself	Envelope
NA	Neuraminidase cleaves sialic acid group to allow virion release into the extracellular region	Transmembrane (TM) at the virion envelope
MA (M1, M2)	Protein assembly with membrane binding and disassociation	Matrix: beneath the envelope
NS1	Nonstructural protein involved in RNA binding. Inhibits RNA splicing.	Within capsid and beneath the matrix
NS2	Transportation of RNP to cytoplasm	—
NP	Nucleoprotein	Nucleocapsid
PB1	Main core protein	Binds to nucleocapsid

7.2.2 Hemagglutinin

Specific surface glycoproteins are used by the enveloped viruses, such as influenza, human immunodeficiency virus-1 (HIV-1), and Ebola, to enter target cells via fusion of the viral membrane with the target cellular membrane (Skehel and Wiley, 1998, 2000b; Eckert and Kim, 2001). One of the most well-studied membrane fusion proteins is the influenza virus HA, which is a homotrimeric type I transmembrane (TM) surface glycoprotein responsible for virus binding to the host receptor, internalization of the virus, and subsequent membrane-fusion events within the endosomal pathway in the infected cell. HA is also the most abundant antigen on the viral surface and harbors the primary neutralizing epitopes for antibodies. Each 70-kDa HA subunit contains two disulfide-linked polypeptide chains, HA_1 and HA_2, created by proteolytic cleavage of the precursor protein HA_0 (Wiley and Skehel, 1987). Such a cleavage is very crucial for membrane fusion (Wiley and Skehel, 1987). During membrane fusion, HA binds the virus to sialic acid receptors on the host cell surface and, following endocytosis, the acidic pH (pH 5–6) of endosomal compartments induces dramatic and irreversible reorganization of the HA structure (Skehel et al., 1982).

The HA trimer has a tightly intertwined "stem" domain at its membrane-proximal base, which is composed of HA_1 residues 11–51 and 276–329 and HA_2 residues 1–176. The dominant feature of this stalk region in the HA trimer is the three long, parallel α-helices (~50 amino acids in length each), one from each monomer, that associate to form a triple-stranded coiled coil. The membrane-distal domain consists of a globular "head" which is formed by HA_1 and which can be further subdivided into the R region (residues 108–261), containing the receptor-binding site (RBS) and major epitopes for neutralizing antibodies, and the E region (residues 56–108 and 262–274), with close structural homology to the esterase domain of influenza C HA esterase fusion (HEF) protein (Stevens et al., 2004). The HA_2 chain contains two membrane-interacting hydrophobic peptide sequences: an N-terminal "fusion peptide" (residues 1–23) that interacts with the target membrane bilayer (Durrer et al., 1996) and a C-terminal TM segment that passes through the viral membrane.

Crystallographic studies suggested that the interaction with the host cell involves a dramatic structural reorganization of HA_2, which moves the fusion peptide from the interior ~100 Å toward the target membrane (Wilson et al., 1981; Bullough et al., 1994). In this process, the middle of the original long α-helix unfolds to form a reverse turn, jackknifing the C-terminal half of the long α-helix backward toward the N terminus. These molecular rearrangements place the N-terminal fusion peptide and the C-terminal TM anchor at the same end of the rod-shaped HA_2 molecule (Weber et al., 1994; Wharton et al., 1995), facilitating membrane fusion by bringing the viral and cellular membranes together.

7.2.3 Neuraminidase

Viral NA is found on the surface of influenza viruses. NA enables the virus to be released from the host cell. NAs are the enzymes that cleave sialic acid groups

from glycoproteins and glycolipids and are required for influenza virus replication. Sialic acid groups are found on various glycoproteins at the host cell surface, and the virus exploits these groups to bind to the host cell utilizing another virus surface protein, HA. The sialic acid-bound virion is endocytosed into an acidified endosomal compartment. There, the M2 ion channel facilitates acidification of the virion and dissociation of the ribonucleoprotein (RNP) complexes from M1 (Pinto et al., 1992; Pinto and Lamb, 2006). The low pH of the endosome induces conformational changes in HA that promote mixing of viral and endosomal lipids, membrane fusion, and entry of the RNP complexes (Wiley and Skehel, 1987; Bullough et al., 1994; Skehel and Wiley, 2000a).

Similar to HA and M2, NA is incorporated into the envelope of the budding virion and plays a crucial role in virion release (Kilbourne et al., 1968; Wagner et al., 2002). In fact, in order for the virus to be released from the cell, NA must enzymatically cleave the sialic acid groups from host glycoproteins and glycolipids, since host glycoproteins and glycolipids with intact sialic acids would bind HA and prevent the virion from escaping the virus-producing cell (Seto and Rott, 1966; Moscona, 2005). NA also cleaves sialic acid residues from viral proteins, preventing newly assembled viruses from aggregating with each other through HA–sialic acid interactions. The action of NA promotes the release of progeny viruses and the spread of the virus from the host cell to uninfected surrounding cells. Therefore, NA is the crucial enzyme that helps viruses to be released from a host cell. Not surprisingly, this viral proteins is one of the primary targets for the anti-influenza drugs (Mitrasinovic, 2010).

Finally, by removing the sialic acid receptors from the cell surface, NA is involved in the suppression of the superinfection; that is, in the infection of a target cell by more than one virion (Huang et al., 2008). Such inhibition of superinfection is important for the evolution of segmented viruses because it limits the frequency of reassortment of viral genes (Webster et al., 1982).

NA is a mushroom-shaped tetramer of identical subunits, with the head of the mushroom suspended from the virus membrane on a thin ~60-Å-long stalk, a length that is variable among virus strains. Each of the subunits that form the head of the mushroom is made up of a six-bladed propellerlike structure, the blades of which are formed by four antiparallel strands of β-structure (Colman et al., 1983; Varghese et al., 1983; Burmeister et al., 1992). The enzyme active site, containing a number of conserved charged amino acid residues, is located at roughly the center of each subunit (Colman et al., 1983; Varghese et al., 1983; Burmeister et al., 1992).

7.2.4 Matrix Proteins M1 and M2

Two matrix proteins (M1 and M2) are encoded in one of the eight RNA segments. Both proteins are synthesized from the same RNA segment by using different reading frames. About 3000 matrix protein molecules are needed to make one virion.

7.2.4.1 M1 Protein The M1 protein forms a layer under the patches of host cell membrane that are rich in viral HA, NA, and M2 TM proteins and facilitates

budding of the mature viruses. This protein binds to the viral RNA nonspecifically via a peptide sequence rich in basic amino acids. The M1 protein also has multiple regulatory functions, performed by interaction with the components of the host cell. The mechanisms regulated include a role in the export of the viral RNP (vRNP) from the host cell nucleus, inhibition of viral transcription, and a role in virus assembly and budding. The incoming RNP at the beginning of infection is transported into the nucleus only after detachment from M1, where RNP is involved in transcription/replication of the viral genome. In the late stage of infection, M1 inhibits viral RNA (vRNA) polymerase activity by binding to RNP, which may be a signal for RNP transport from the nucleus to the cell surface. Finally, M1 mediates the association of RNP with viral envelope glycoproteins on the inner surface of the cytoplasmic membrane, which then promotes virion formation and budding. Thus, M1 is a multifunctional protein that plays important roles in various steps of virus replication (Itoh and Hotta, 1997).

7.2.4.2 M2 Protein The M2 protein has an important role in the life cycle of the influenza A virus. It is located in the viral envelope as a homotetrameric proton-selective ion channel. This channel is activated by low pH. It enables hydrogen ions to enter the viral particle (virion) from the endosome, thus lowering the pH inside the virus. This causes dissociation of the viral matrix protein M1 from the RNP and represents a crucial step in uncoating of the virus and exposing its content to the cytoplasm of the host cell. The M2 protein unit consists of the following three protein domains: (i) the 24 amino acids in the N-terminal end, exposed to the outside environment; (ii) the 22 mostly hydrophobic amino acids in the TM region; and (iii) the 52 amino acids in the C-terminal domain, exposed to the inside of the viral particle (Schnell and Chou, 2008; Stouffer et al., 2008).

7.2.5 Nonstructural Proteins NS1 and NEP

7.2.5.1 NS1 Protein NS1 is one of the nonstructural proteins of influenza A viruses. NS1 is not a structural component of the virion but is expressed at very high levels in infected cells (Krug and Etkind, 1973). It is a nonessential virulence factor that has multiple accessory functions during viral infection. In general, the viral NS1 protein is considered as an important factor by which all influenza A viruses antagonize host interferon (IFN)-α/β-mediated antiviral responses (Egorov et al., 1998; Garcia-Sastre et al., 1998; Kochs et al., 2007). However, NS1 is a multifunctional protein that performs a plethora of activities, which may additionally contribute toward efficient virus replication and virulence during infection. These include (i) temporal regulation of vRNA synthesis; (ii) control of viral mRNA splicing; (iii) enhancement of viral mRNA translation; (iv) regulation of viral particle morphogenesis; (v) suppression of host immune/apoptotic responses; (vi) activation of phosphoinositide 3-kinase (PI3K); and (vii) involvement in strain-dependent pathogenesis. All these functions of NS1 rely on its ability to participate in a multitude of protein–protein and protein–RNA interactions (Hale et al., 2008).

Based on the phylogenetic analysis of the NS1 amino acid sequences, it has been concluded that there are two major groups of NS1 proteins, alleles A and B (Treanor et al., 1989; Ludwig et al., 1991). Allele A includes a number of NS1 proteins from avian influenza viruses together with those of all human, swine, and equine influenza viruses, whereas allele B exclusively includes proteins from avian viruses. The level of homology within each allele is 93–100%; however, between alleles it can be as little as 62% (Hale et al., 2008).

Structurally and functionally, NS1 is divided into two domains: an N-terminal RNA-binding domain (residues 1–73), which *in vitro* binds with low affinity to several RNA species in a sequence-independent manner, and a C-terminal "effector" domain (residues 74–230), which predominantly mediates interactions with host cell proteins and also functionally stabilizes the RNA-binding domain (Hale et al., 2008). The RNA-binding domain is a symmetrical homodimer in which each monomer consists of three α-helices (Chien et al., 1997). The C-terminal effector domain of NS1 protein (residues 74–230) is also a homodimer, in which each monomer consists of three α-helices and seven β-strands which form a twisted, crescentlike, antiparallel β-sheet around a long, central α-helix (Bornholdt and Prasad, 2006).

*7.2.5.2 **Nuclear Export Protein*** NEP is another nonstructural protein of influenza virus. Similar to NS1, this protein is encoded in the eight RNA segments of the influenza A virus. In fact, two mRNAs are produced from this vRNA segment. The first of these encodes the NS1 protein, whereas the splicing of this NS1 mRNA produces another mRNA encoding a 121 amino acid-long protein NEP (Lamb and Lai, 1980). NEP interacts with the virus protein M1 and the nuclear export machinery during viral replication and is directly involved in the nuclear export of vRNPs during the viral life cycle, acting as an adapter molecule between the nuclear export machinery and the vRNP complex (O'Neill et al., 1998). At the N terminus of NEP, there is a Rev-like nuclear export signal (NES) that is able to interact with nucleoporins of the host cell (Paragas et al., 2001).

Structural analysis of the purified NEP revealed that this protein exists as a monomer in solution and adopts a compact, but very flexible, conformation with the characteristics of the molten globule state under near physiological conditions (Lommer and Luo, 2002). Analysis of the proteolytic sensitivity suggested that NEP possesses a rather heterogeneous structure, where the C terminus adopted a relatively rigid conformation, but the N terminus, being recognized by the nuclear export machinery, was highly mobile and exposed (Lommer and Luo, 2002).

7.2.6 Nucleoprotein

Nucleoprotein gene is located at segment 5 of the segmented negative-strand RNA genome. It codes for a single protein of 498 amino acid residues. NP is a multifunctional protein: it encapsidates vRNA (Kobayashi et al., 1994), forms homo-oligomers to maintain vRNP structure (Prokudina-Kantorovich and Semenova, 1996), and serves as the key adaptor for virus and host cell interactions (Portela

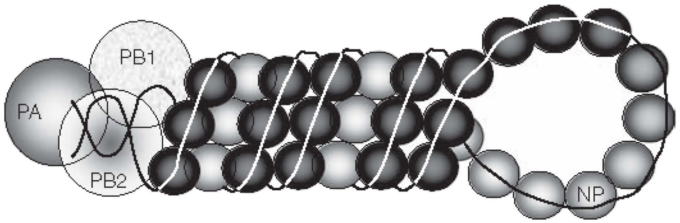


Figure 7.2 Model structure of nucleoprotein particle.

and Digard, 2002; Ng et al., 2009). NP interacts with PB1 and PB2 subunits in the viral RNA polymerase in forming the RNP (Biswas et al., 1998). In vRNP, NP forms a core around which the vRNA is helically wrapped (Baudin et al., 1994) (Fig. 7.2). Approximately 24 nucleotides associate with each NP (Compans et al., 1972). Because each vRNA is about 890–2341 nucleotides long (Lamb and Krug, 2001), each influenza vRNP has 37–97 copies of NP (Wu et al., 2007).

Sequence comparison of multiple influenza A NPs revealed that there are 30.1% of polymorphic residues in this protein. Structurally, NP can be divided into three regions, the head domain, the body domain, and the tail loop/linker region. The head domain is more conserved than the body domain, as revealed from the structure-based sequence alignment. NP oligomerization is mediated by the insertion of the nonpolymorphic and structurally conserved tail loop of one NP molecule into a groove of another NP. The different form of NP oligomers is because of the flexibility of the polymorphic linkers that join the tail loop to the rest of the protein. The RNA-binding property of NP is known to involve the protruding element and the flexible basic loop between the head and body domains, both having high degrees of primary sequence conservation. To bind RNA, NP may first capture the RNA by the flexible basic loop and then the RNA is clamped by the protruding element (Ng et al., 2009).

7.2.7 RNA Polymerase Complex

Viral RNA polymerase is a heterotrimeric RNA-dependent RNA polymerase complex with multiple enzymatic activities for catalyzing vRNA transcription and replication, which is composed of PB1, PB2, and PA. PB1 forms the core structure of the heterotrimeric RNA polymerase complex. The N-terminal region of PB1 binds to the C-terminal region of PA, and C-terminal region of PB1 binds to the N-terminal region of PB2 (He et al., 2008; Obayashi et al., 2008). The RNA polymerase plays a central role in the viral life cycle and is directly responsible for RNA synthesis for both viral replication and transcription, being able to replicate its viral genome RNA via two steps. The replication step (vRNA → cRNA → vRNA) involves synthesis of complementary RNA (cRNA) from a vRNA template, followed by the synthesis of vRNA from cRNA to complete the genomic RNA replication process. The transcription step (vRNA → mRNA) involves viral mRNA transcription from vRNA by snatching capped primers from nascent host cell mRNA (Liu et al., 2009).

In RNA polymerase complex, PB1 is the central component that interacts directly with both PA and PB2. Functionally, PB1 is the central subunit for the RNA polymerase activity because it binds to the viral promoter and is responsible for the vRNA elongation and cap RNA cleavage activities (Biswas and Nayak, 1994). PB2 is required for transcription of vRNA (Li et al., 2001) and can bind to the methylated cap-1 structure of host RNAs for cleavage by the PB1 subunit (Fechter et al., 2003). PA is required for replication and transcription of vRNA and for the endonuclease cleavage of the cap RNA primer, induces proteolysis of viral and host proteins, and is involved in virus assembly (Liu et al., 2009).

7.2.8 PB-F2 Protein

PB1-F2 is the most recently discovered influenza A virus protein, which contributes to viral pathogenesis by inducing apoptosis, being localized to the inner and outer mitochondrial membranes through a nontraditional mitochondrial targeting sequence at the C-terminal region of the protein (Gibbs et al., 2003). In the mitochondrial membrane, PB1-F2 interacts with ANT3 and VDAC1, punches holes in the mitochondrial membrane, and causes cytochrome c release (Zamarin et al., 2005).

7.2.9 Influenza Pandemics

Influenza pandemics (i.e., epidemics of the influenza virus) of the recent era include the Spanish influenza (H1N1, 1918), the Asian influenza (H2N2, 1958), and the Hong Kong influenza (H3N2, 1968) (Reid et al., 2001b, 2004a; Garcia-Sastre and Whitley, 2006; Kilbourne, 2006a). The death of humans resulting from the pandemics has been very large. The Spanish influenza pandemic of 1918–1919 alone caused acute illness in 25–30% of the world's population and resulted in the death of 40 million people worldwide. The numbers of deaths in the United States have been estimated at 500,000, 70,000 and 34,000, respectively, for the Spanish influenza, the Asian influenza, and the Hong Kong influenza pandemics (Reid et al., 2004a; Garcia-Sastre and Whitley, 2006; Kilbourne, 2006a) (for more information see also http://www.kdheks.gov/flu/download/avian_flu_facts.pdf).

More recently, various subtypes have afflicted smaller numbers of humans mainly via avian species, especially poultry. The subtypes include H5N1, H7N3, H7N7, and H9N2. For instance, in 2003, an outbreak of H7N7 occurred in the Netherlands, where 89 people were confirmed to have H7N7 influenza virus infection following an outbreak in poultry on several farms. One death was recorded, whereas most of the 89 patients only had conjunctivitis and mild influenza symptoms. As for the related H7N3, which resulted in quarantine of 18 North American farms in 2004 to halt the spread of the virus, only two human cases have been known, but none resulted in fatalities (Tweed et al., 2004). H5N1 subtypes, on the other hand, generally have higher fatality rates of above 50% (Garcia-Sastre and Whitley, 2006). For example, in 1997, the avian H5N1 virus was transmitted from poultry to humans in Hong Kong and caused 18 human cases of influenza with a

high mortality rate (Centers for Disease Control and Prevention, 1998). In February 2003, two cases of H5N1 infection in humans occurred in Hong Kong, one of them fatal (Wuethrich, 2003). In 2004–2005, during an extensive outbreak of H5N1 infection in poultry in eight countries of Southeast Asia, over 100 human cases occurred in Vietnam, Thailand, and Cambodia, 50% of them fatal (World Health Organization, 2005). A few human cases of H9N2 have been reported (Guan et al., 1999). It has been reported that H9N2 bears striking genetic resemblance to H5N1 (Lin et al., 2000a).

7.2.10 Causes of Influenza Virulence Remain Largely Elusive

The understanding of the causes of virulence for the respective subtypes of influenza viruses remains largely incomplete. This is clearly illustrated by the example of the H1N1 virus, for which several hypotheses have been put forth to explain its virulence. One theory proposes that the H1N1 virulence could be attributed to the effective inhibition of type I interferon by the NS1 protein from the 1918 H1N1 virus. However, this model is unable to account for the virulence of H5N1 subtypes that have dissimilar NS1 sequences (Taubenberger et al., 2001). Multiple gene mutations in various viral proteins (HA, NA, and polymerases) have also been suggested as a requirement for high virulence, especially in mice (Garcia-Sastre and Whitley, 2006). Adding to the confusion is the finding that viruses with 1918 H1N1 HA are highly pathogenic to mice (Kobasa et al., 2004; Garcia-Sastre and Whitley, 2006). The reason for this remains elusive. We hope that by using a relatively new set of computational tools (e.g., protein disorder prediction) we can provide answers to some of these questions.

7.2.11 Other Puzzles of Influenza

There is yet another mystery related to influenza, and it is connected to the mysterious way in which the 1918 H1N1 virus suddenly disappeared by 1920, after raging relentlessly in 1918 and 1919 (Reid et al., 2001b, 2005a; Kilbourne, 2006a). Several strains of H1N1 appeared but none of them was as virulent as the 1918 virus (Taubenberger et al., 2001; Reid et al., 2001a; Kilbourne, 2006b). How did these reappearing strains differ from the 1918 H1N1 virus? How did they evolve? Are the H1N1 and the H5N1 virulences coming from the same protein, e.g., HA? We believe that disorder prediction can shed at least some light on these mysteries too.

7.3 PREDICTED INTRINSIC DISORDER IN INFLUENZA VIRUS

7.3.1 Average Percentage of Predicted Disordered Residues Enable Analysis of the Viral Protein Disorder

Table 7.2 lists the average percentages of predicted disordered residues (the percentage disorder rate) that have been found in influenza viral proteins with known

TABLE 7.2 Abundance of Intrinsic Disorder in Various Datasets Table Presents the Means of Predicted Disorder by PONDR® VLXT. The Corresponding PONDR VL3 Are Shown in Brackets. The Total Number of Influenza a Proteins Used in this Analysis was 27

Protein	% Predicted Disordered X-Ray[a]	% Predicted Disordered NMR
PDBS90	24 ± 2 (14 ± 2)	34 ± 2 (32 ± 1)
Influenza	21 ± 2 (10 ± 2)	34 ± 3 (40 ± 3)

[a]Percentage of residues that are predicted to be disordered that have been characterized by X-ray crystallography.

3D structure studied by nuclear magnetic resonance (NMR) or X-ray crystallography. These viral proteins are compared with the PDB Select 90 proteins (Hobohm et al., 1992; Berman et al., 2000). Table 7.2 shows that the percentage of residues predicted to be disordered by PONDR VLXT in both protein datasets is relatively similar. The values given in this table are the average percentages of disordered residues in a given dataset, not the average percentages of disordered residues in each chain. The former provides a better gauge of the mean because the number of influenza proteins available in PDB (Berman et al., 2000) is relatively small.

Table 7.2 provides a simple measure to classify a given protein as ordered or disordered by prediction. For example, the averaged predicted disorder rate for proteins from PDB Select 90 is $24 \pm 2(15 \pm 2)$. If this value is used as a benchmark for labeling a given protein as moderately disordered or mostly structured, any protein that falls close to this number with respect to percentage of predicted disordered residues can safely be classified as "moderately disordered" by prediction. The information in Table 7.2 provides the benchmarks for the analysis of the results shown in Table 7.3, which categorizes proteins found in the influenza virus by their function and the abundance of disorder and arranges data by protein location in the virion. The envelope of the orthomyxovirus contains HA and NA proteins, levels of the predicted intrinsic disorder in which vary significantly by the subtype (Table 7.4). M1 is a matrix protein that provides a link between the surface protein and the capsid. M1 is predicted to be moderately disordered. Both nonstructural proteins of the influenza A viruses are predicted to be rather disordered. Similarly, both nucleoprotein and main core protein are predicted to contain significant percentage of disordered residues.

7.3.2 Predicted Disorder Varies with Location of Protein in the Virion

There is an interesting correlation between the percentage disorder rate and the protein localization within the virion. This phenomenon is especially clear from Fig. 7.3, which clearly shows that the closer the protein gets to the core the higher is the level of predicted disorder it possesses. This can easily be explained by the increased likelihood of colocalization of the RNA-binding proteins and the genomic RNA in the viral core. The exceptions to this general trend are several enzymes,

TABLE 7.3 Summary of the Predicted Disorder Rates in Influenza Proteins

Protein	Accession Number	% VLXT(VL3)[a]	Method
HA1 (H1)	1ruz.pdb, Subunit: H	12 (0)	X-ray
HA2 (H1)	1ruz.pdb, Subunit: I	12 (0)	—
HA1 (H3)	1mqn.pdb, Subunit: A	25 (0)	—
HA2 (H3)	1mqn.pdb, Subunit: B	19 (2)	X-ray
NA (N1)	2hu0.pdb, Subunit: S	8 (0)	X-ray
M1	1ea3.pdb, Subunit: A	25 (0)	X-ray
NS1	1xeq.pdb, Subunit: A	69 (63)	X-ray
NS2	1pd3.pdb, Subunit: A	60 (23)	X-ray
PB1	2hn8.pdb, Subunit: A	47 (NA)	NMR
NP	2iqh.pdb, Subunit: A	44 (0)	X-ray

[a]The percentage of residues that are predicted ordered by VLXT. The parenthesized values are data obtained by VL3.

TABLE 7.4 Predicted Intrinsic Disorder in Surface Proteins of Influenza Virus

	Disorder prediction in various NA subtypes			
Protein		Accessions	%VLXT(VL3)	Method
N1		2hu0.pdb, Subunit: A	8 (0)	X-ray
N2		2f10.pdb, Subunit: A	25 (13)	X-ray
N4		2htw.pdb, Subunit: A	4 (0)	X-ray
N6		2w20.pdb, Subunit: A	15 (27)	X-ray
N8		2htr.pdb, Subunit: A	4 (48)	X-ray
N9		1jsn.pdb, Subunit: S	15	X-ray
	Disorder predictions in various HA subtypes			
Protein	Subunit	Accessions	%VLXT(VL3)	Method
H1	HA_1	1ruz.pdb, Subunit: H	12 (0)	X-ray
	HA_2	1ruz.pdb, Subunit: I	12 (0)	
H3	HA_1	1mqn.pdb, Subunit: A	25 (0)	X-ray
	HA_2	1mqn.pdb, Subunit: B	19 (1)	
H5	HA_1	2ibx.pdb, Subunit: A	12 (2)	X-ray
	HA_2	2ibx.pdb Subunit: B	13 (0)	—
H7	HA_1	1ti8.pdb, Subunit: A	22 (11)	X-ray
	HA_2	1ti8.Pdb, Subunit: B	30 (0)	

which are located in close proximity to the core. However, enzymes need to be structured so that the active site can provide a catalytic surface, so this result is entirely consistent with previous work (Vucetic et al., 2007; Xie et al., 2007a,b).

Correlation between the predicted disorder and functions of viral proteins is of special interest since it correlates nicely with the disorder/function trends found in proteins in general. In fact, RNA-binding proteins are known to show a strong

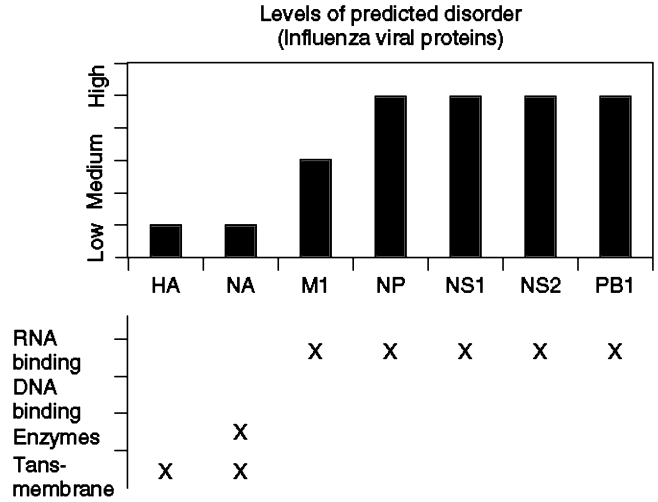


Figure 7.3 Summary of the percentages of residues predicted to be disordered in the influenza A virus proteins. Top of the panel is a chart of proteins with percentages of residues that are predicted to be disordered by PONDR® VLXT. The types of proteins are also summarized. It should be noted that the percentages of predicted to be disordered residues of HA vary with subtypes, but HA is essentially predicted to be quite ordered. More information is available in Table 7.4.

tendency to be highly disordered, both experimentally and computationally. Single-span membrane proteins also contain significant amount of intrinsic disorder, except for the segment that crosses the membrane, which is typically predicted to be ordered. Finally, various enzymes as well as TM proteins (e.g., pores) are among the polypeptides with the least amount of intrinsic disorder (Vucetic et al., 2007; Xie et al., 2007a,b).

Figure 7.3 presents data for disorder distribution within the major influenza proteins. The proteins are labeled as "predicted to be ordered," "moderately disordered," or "very disordered." The qualitative values are assigned according to the PONDR VLXT values by comparing with the mean values of Table 7.2. Here, the proteins whose percentages of predicted disordered residues are between 30 and 40% were considered as moderately disordered. The proteins with the disorder scores above 40% and below 20% were considered as very disordered or ordered, respectively. Figure 7.3 shows that the amount of predicted intrinsic disorder increases as the protein becomes located closer to the viral core, which is the site of the genomic RNA. In general, Fig. 7.3 supports the notion that viral proteins in general follow the trend in the distribution of intrinsic disorder among functional classes previously described for other proteins.

*7.3.2.1 **Transmembrane Proteins are More Ordered*** Figure 7.3 shows that the surface proteins HA and NA are generally predicted to be ordered. HA is crucial to the entry of the virus into the host. A cleavage at the disulfide bond

between subunits, HA_1 and HA_2, has to occur before the viral entry can take place. NA, on the other hand, plays an enzymatic role at the other end of the viral process. This protein cleaves the sialic acid group from the oligosaccharide portion. This cleavage step is needed for the virion to be released into the extracellular region (Colman, 1994). It should also be noted that both HA (HA_2) and NA are TM proteins (Colman, 1994; Barman et al., 2004). Previously, TM fragments of channels and pores were predicted to be highly ordered (Vucetic et al., 2007; Xie et al., 2007a,b). However, the situation might be quite different for membrane proteins with relatively large extra- and intracellular domains, for example, for fusion proteins. We now have an opportunity to analyze the predicted disorder rate of TM proteins that are involved in the membrane fusion.

HA is a TM glycoprotein, a member of the class I viral fusion proteins that mediate viral entry into cells. Class I viral fusion proteins are thought to fold into a prefusion, metastable conformation, which is then activated to undergo a large conformational rearrangement to a lower energy state, thereby providing the energy needed to accomplish membrane fusion (Earp et al., 2005; Kielian, 2006; Kielian and Rey, 2006). The role of HA as a fusion protein and the associated large-scale conformational changes may help to account for the slightly higher predicted disorder rates observed for HA in many subtypes (Table 7.4).

7.3.2.2 More Disorder at the Core The matrix proteins, which form a layer below the lipid envelope, produced interesting data. The influenza A virus matrix is relatively disordered (Table 7.3). The matrix protein M1 is predicted to be moderately disordered (or somewhat ordered) by PONDR VLXT. In the influenza virion, proteins that are even closer to the core include NS1, NP, and PB1. All these proteins are predicted to be highly disordered. While M1 is known to bind RNA, proteins that are located closer to the core are even more likely to interact with the vRNA. This may account for the trend that, for proteins that are closer to the core, the amount of predicted disorder increases.

7.4 INTRINSIC DISORDER OF VIRAL PROTEINS AND VIRAL INFECTIVITY

7.4.1 HA Predicted Disorder Correlates with the Viral Infectivity

7.4.1.1 Loss of Infectivity for Specific Virus Subtypes via Fatty Acid Deprivation Attachment and membrane fusion of the influenza virus and the host cell are mediated by the viral HA. HA is a homotrimer, and each monomer comprises an ectodomain with about 510 amino acid residues, a TM domain with 27 residues, and a cytoplasmic domain with 10–11 residues. The HA monomer is synthesized as a single polypeptide chain and cleaved into two subunits, HA_1 and HA_2, by proteolytic enzymes after virus budding or during intracellular transport. The HA_1 and HA_2 subunits are functionally specialized. HA_1 carries receptor-binding activity, and HA_2 mediates membrane fusion (Skehel and Wiley, 2000a). As discussed

briefly above, the amount of intrinsic disorder in HA_1 and HA_2 varies with the viral subtype (Table 7.4).

Analysis of previous experimental data in comparison with the disorder predictions in Table 7.4 suggests that variations in the infectivity of the virus (Oldfield et al., 2005), assembly of the HA proteins, and correlation between protein–membrane interaction and the lipid raft motion may all be related to the amount of predicted disorder. For example, one of the HA functions is to assemble proteins, including those involved into the formation of pores. Acetylation of the HA molecules often affects this function, which is crucial for the infectivity of the virus. However, it has been shown that H1, H3, and H7 behave differently when the sites that are normally palmitylated are mutated. Viral subtypes with H1 proteins were most affected by the mutations, whereas the virions did not lose much of their infectivity in the case of H3 and H7 (Jin et al., 1996; Sakai et al., 2002). Table 7.4 shows that, among the viral subtypes analyzed, H1 proteins possess the least amount of predicted disorder, whereas H3 and H7 proteins were predicted to be essentially more disordered. We showed elsewhere that enzyme-mediated posttranslational modifications usually occur with disordered regions (Iakoucheva et al., 2004; Vucetic et al., 2007; Xie et al., 2007a,b). The increased predictions of intrinsic disorder in HA are associated with the increased infectivity of influenza virus, perhaps via changes in posttranslational modification, the ease of which may depend on the tendency to be disordered.

7.4.1.2 Intrinsic Disorder May Provide a Bypass to the Lipid Raft Requirement

For all enveloped viruses, the envelope is derived from the host cell during the process of virus budding. In the case of influenza virus, budding takes place at the apical plasma membrane and is heavily dependent on the presence of lipid microdomains, or "rafts" (Brown and London, 1998; Scheiffele et al., 1999; Zhang et al., 2000). Lipid rafts, also known as *detergent-insoluble glycosphingolipid-enriched domains*, are specific domains on plasma membranes that are enriched in detergent-insoluble glycolipids (DIGs), cholesterol, and sphingolipids (Mukherjee and Maxfield, 2000; Simons and Ikonen, 2000; Simons and Toomre, 2000). Levels of cholesterol and sphingolipids can vary amongst individuals, which alters the extent of the raft formation. Lipid rafts play an important role in several biological processes, including signal transduction, T-cell activation, protein sorting, and virus assembly and budding (Simons and Toomre, 2000).

Such enveloped viruses incorporate some integral membrane proteins; among the best studied are the influenza virus HA and NA (Simons and Ikonen, 1997). Acetylation of the envelope proteins and also palmitoylation are important for these viral proteins to be targeted to the lipid raft microdomains on the cell surface (Melkonian et al., 1999). C-terminal domains of both HA and NA of influenza virus are crucial for association with rafts, and this interaction constitutes a part of the signaling machinery necessary for apical targeting in polarized cells. In fact, the cytoplasmic tails of HA and NA are so important for assembly that the information contained in these tails is partially redundant (Jin et al., 1997). For example, the removal of the cytoplasmic tail or mutation of the three palmitoylated cysteine residues in the TM

domain and the cytoplasmic tail of influenza virus HA was shown to decrease the association of HA with lipid rafts, decrease the incorporation of HA into virions (Zhang et al., 2000), and modulate the incorporation of cholesterol into the viral envelope.

The level of the envelope cholesterol has been shown to play a crucial role in the HA-mediated fusion of the influenza virus with the host cell (Sun and Whittaker, 2003). These data were obtained for the WSN (H1N1) strain of influenza virus, and the authors proposed that differences may exist with other virus strains. Perhaps the virion cholesterol is important for the organization of influenza virus HA trimers into fusion-competent domains, and perhaps the depletion of cholesterol also inhibits virus infectivity because of inefficient fusion (Sun and Whittaker, 2003). Here we suggest that variations in intrinsic disorder in the surface proteins may play a similar role. In fact, Table 7.5 shows that H1 is predicted to be ordered, whereas H3 and H7 are predicted to be more disordered. This increased level of disorder might offer a mechanism for proteins to bypass the lipid raft requirement. Studies on chimera proteins with specific swapping of regions predicted to be ordered or disordered could be used to test this proposed mechanism.

7.4.2 Disorder/Disorder or Order/Order Pairing of HA and NA may be Intertwined with the Evolution of the Influenza Viruses

As has already been mentioned, 16 HA serotypes and 9 NA subtypes of influenza A virus are known. Among the three influenza types, type A viruses are the most virulent human pathogens that cause the most severe disease. The list of some influenza A virus serotypes with the largest known human pandemic deaths includes H1N1 ("Spanish flu"), H2N2 ("Asian flu"), H3N2 ("Hong Kong flu"), and H5N1 ("Avian flu"). Table 7.5 illustrates an interesting correlation between the amounts of predicted intrinsic disorder in HA and NA proteins from the different influenza A virus serotypes: in H1N1 and H5N1 subtypes, both HA and NA are predicted to be ordered, whereas H3N2 serotype is characterized by more disordered HA and NA. Perhaps such a combination is not coincidental but is instead evolutionarily preferred.

TABLE 7.5 Observed Pairing of Predicted Disorder of HA–NA in Subtypes Involved in Major Epidemics. Quantitative Details of the Respective Predicted Disorder Values can be Found in Table 7.4

Subtype	Description	HA (Predicted Disorder/Order)	NA (Predicted Disorder/Order)
H1N1	"Spanish Flu" (1918)	Ordered	Ordered
H2N2	"Asian Flu" (1957)	NA[a]	More disordered
H3N2	"Hong Kong Flu" (1968)	More Disordered	More disordered
H5N1	"Avian Flu" (1997)	Ordered	Ordered

[a]PDB structure of H2 has not been found, presumably because the H2N2 subtype evolved quickly into H3N2.

7.5 DISORDER IN THE 1918 H1N1 AND H5N1 VIRUSES

7.5.1 H1N1 1918 Viral Strain Has Predicted Disorder at a Crucial Region of HA

The 1918 pandemic caused global devastation in 1918–1919, but it became nonexistent by 1920 (Taubenberger et al., 2001; Reid et al., 2001a; Reid, 2005b; Kilbourne, 2006b). However, the H1N1 subtype of influenza A was still present in 1934. To answer the question why the 1918 strain was highly virulent, we compared the intrinsic disorder distributions in amino acid sequences of HA proteins from the 1918 strain and later strains. Table 7.6 provides an overview of the percentage of predicted disorder found in HA_1 and HA_2 chains of various influenza A subtypes whose crystal structures were resolved. Table 7.6 shows that H1N1 and H5N1 subtypes have several variants, such as 1918 H1N1 and 1930 H1N1. Analysis of the data in Table 7.6 shows that the overall percentage of predicted disorder does not vary much for the different strains of H1N1 and that all the HA proteins could be considered as ordered by prediction.

However, a comparison of 3D models enhanced by the PONDR VLXT annotations revealed marked differences between various HA subtypes. This is illustrated by Fig. 7.4, which represents a set of models generated by the combination of X-ray data and the results of the PONDR VLXT analysis. The HA_1 and HA_2 subunits are shown in light turquoise and dark blue colors. The areas predicted

TABLE 7.6 Summary of the Percentage of Predicted Disorder in Each Chain of the Influenza Hemagglutinins

Viral Subtype	Accessions	Strain	Description	%VLXT HA_1	%VLXT HA_2[a]	Predicted Disorder (HA_1)[b]
H1N1	1ruz	A/South Carolina/1/18	Spanish Flu 1918	12	12	Ordered
	1ruy	A/Swine// Iowa/15/30	Swine 1930	15	5	
	1rvx	A/Puerto Rico/3/34	1934	11	6	
H3N2	1mqn	A/Duck/ Ukraine/63	Hong Kong Flu 1968	25	19	More Disordered
H5N1	2fk0	A/Vietnam/ 1203/2004	Avian Flu	15	18	Ordered
	2ibx	A/Vietnam/ 1119/2004		12	12	Ordered
	1jsn	A/Duck/ Singapore/3/97		15	19	Ordered
H9N2	1jsd	A/Swine/ HongKong/9/98	Swine 1998	15	20	Ordered

[a]Percentage of residues that are predicted to be disordered by PONDR VLXT in a given chain.
[b]Qualitative description based on predicted disorder for HA_1 chain.

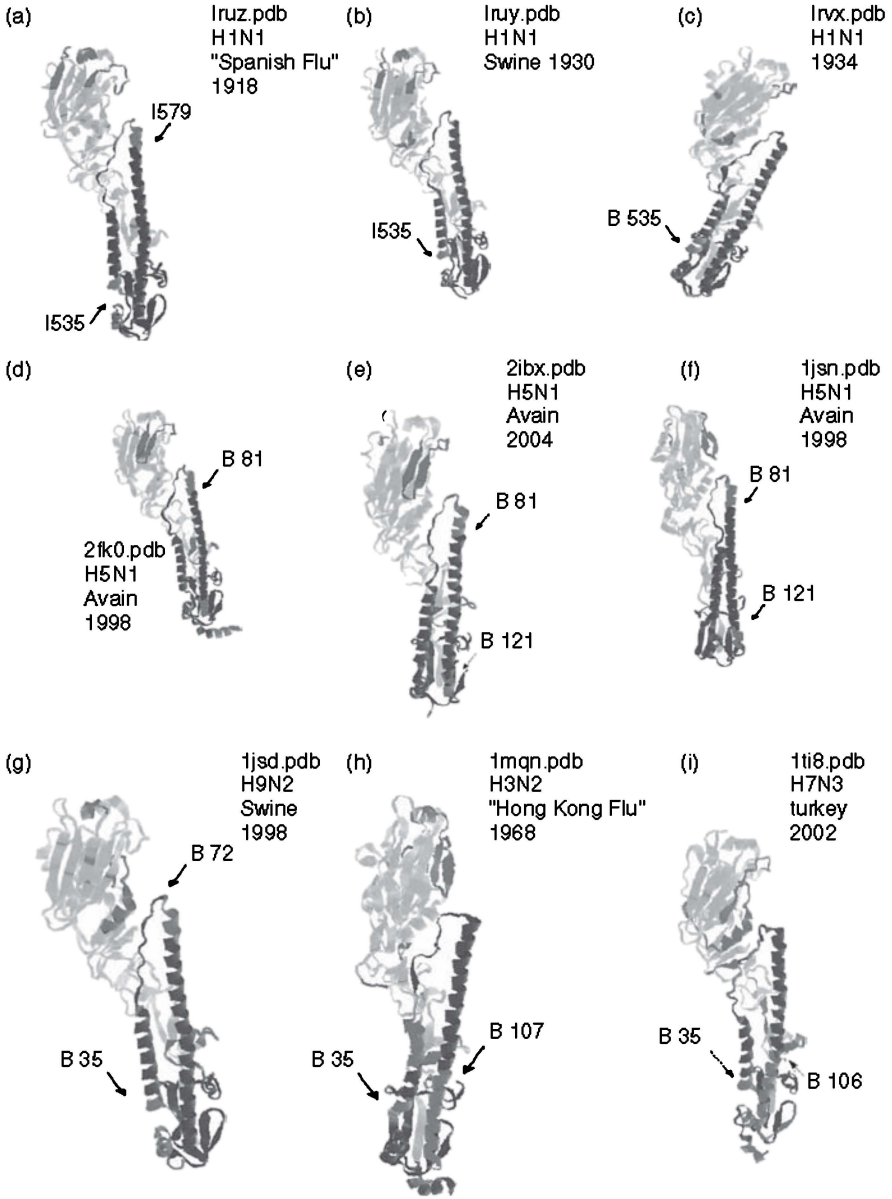


Figure 7.4 The 3D representation of the HA with predicted disorder annotation. The regions in red represent areas predicted to be disordered by VLXT. Conversely, the blue areas represent regions predicted to be ordered. (a) 1918 H1N1 HA_2, 1ruz.pdb. (b) 1930 H1N1 1ruy.pdb. (c) 1934 H1N1, 1rvx.pdb. (d) H5N1, 2fk0.pdb. (e) H5N1, 2ibx.pdb. (f) H5N1, 1jsn.pdb. (g) H9N2.pdb, Swine. (h) H3N2, "Hong Kong Flu-1968" Progenitor, sample from 1963. (i) H7N3, avian. (*See insert for color representation of the figure.*)

to be disordered by PONDR VLXT are marked by red color. Figure 7.4 clearly shows that the various HA subtypes differ dramatically in the amount and localization of the predicted intrinsic disorder. As another way to represent this data, Fig. 7.5 compares the distributions of predicted disorder within the sequences of HA_2 subunits from several influenza A subtypes and strains with the distributions of normalized B-factor scores evaluated from X-ray structures of the corresponding subunits. A residue with a PONDR VLXT score of 0.5 or more is considered to be disordered. For convenience, some areas predicted to be disordered are marked by arrows and the corresponding labels point to the residues. Figure 7.5 shows that there is a general correlation between the high level of predicted disorder in a given HA region and its high B-factor scores. This indicates that HA regions of predicted disorder can gain some fixed structure during crystallization, although these regions still possess high mobility even in the crystal structures.

Figures 7.4 and 7.5 clearly show that the HA_2 protein of the H1N1 1918 strain has two distinct regions of predicted disorder (starting at positions 35 and 68, respectively, see Figs. 7.4a and 7.5a), whereas two other strains from the 1930s reveal only one disordered region (starting at position 35, see Figs. 7.4b–c and 7.5b–c). The region that was predicted to be disordered in the 1918 strain but not in the H1N1 strains after 1919 is a region at or around position number 68 of the HA_2 protein. The high level of predicted disorder in the region around position 35 seems to be a common feature among strains of the subtype H1N1. This region lies at the base of the HA stalk. The missing region of predicted disorder among the H1N1 strains in the 1930s, around position 68, is located at the higher end (the "tip") of the stalk. A superficial analysis of the HA structure (Fig. 7.4) suggests that this region is likely to be located in an area close to the HA center of gravity, suggesting that motions of HA_2 at this region are likely to have greatest impact on the HA_1 subunits, where the exposed area of the protein lies.

7.5.2 The Crucial Region of Predicted Disorder is Associated with Virulent Strains

7.5.2.1 A Specific Region of Predicted Disorder Observed in 1918 H1N1 is seen in both H5N1 and H9N2 A noticeable feature of the HA_2 subunit in the H5N1 subtype is the presence of two specific regions of predicted disorder. Again, the first region lies at the base of the stalk, even though on a different α-helix (Fig. 7.4). The other region of predicted disorder coincides with the disordered fragment seen in the 1918 virus; that is, this region is positioned at the tip of the HA_2 stalk (around residue 68) and is quite similar in size to that in the 1918 strain. This region is also seen in the H9N2 subtype (Figs. 7.4e and 7.5e), but is noticeably shorter (residues 72–79).

7.5.2.2 The Crucial Region of Predicted Disorder in 1918 H1N1 is Replaced by More Ordered Regions in Many Less-Virulent Strains Given the data discussed thus far, one might suggest that the region around position 68 could be commonly predicted to be disordered among all influenza subtypes. However, the

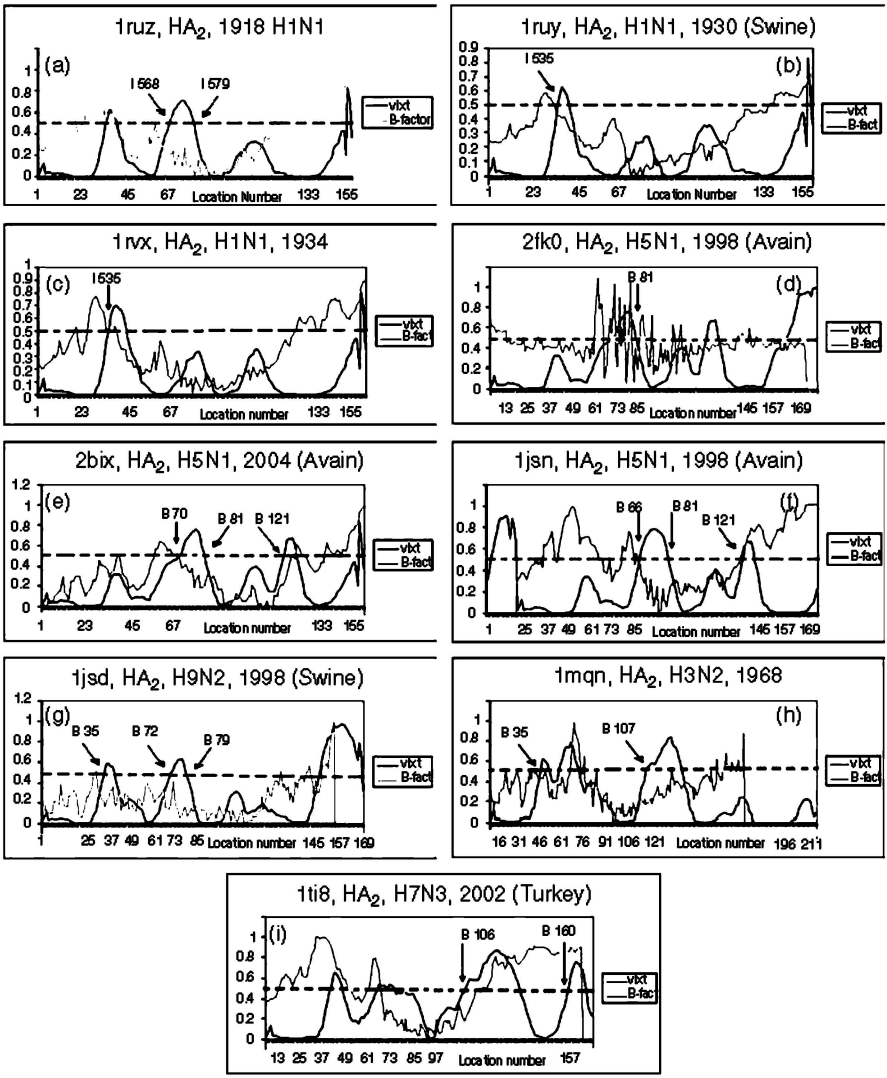


Figure 7.5 PONDR® VLXT and B-factor plots for the various HA_2. The gray curves represent the VLXT score, while the black ones represent the normalized B-factor. (a) 1918 H1N1 HA2, 1ruz.pdb. (b) 1930 H1N1 1ruy.pdb. (c) 1934 H1N1, 1rvx.pdb. (d) H5N1, 2fk0.pdb. (e) H5N1, 2ibx.pdb. (f) H5N1, 1jsn.pdb. (g) H9N2, 1jsd.pdb, swine. (h) H3N2, "Hong Kong Flu-1968" Progenitor, sample from 1963, 1mqn.pdb. (i) H7N3, Avian, H7N3, Italy, Avian, 1it8.

H7N3 subtype shown in Figs. 7.4g and 7.5g does not have any predicted disorder near position 68. This is also the case for the H3N2 subtype shown in Figs. 7.4f and 7.5f. Interestingly, the Hong Kong H3N2 virus and the avian H7N3 virus, both of which have ordered helical tips of HA_2, were less virulent than the Spanish influenza strain, in which this tip is predicted to be disordered.

7.5.2.3 The Low-Pathogenic Avian Influenza H7N3 Strain Does Not Have Predicted Disorder at the Helical Tip Infection of humans by avian influenza H7 subtypes has been described in the Section 7.1. The human symptoms of the H7N3 have been largely mild, suggesting that in the majority of cases we are dealing with low-pathogenic avian influenza (LPAI) strains. Although some highly pathogenic avian influenza (HPAI) strains do exist, the particular sample available via PDB (1ti8.pdb, see Figs. 7.4i and 7.5i), that is, A/Turkey/Italy/02/, is a representative of the LPAI strains (Di Trani L et al., 2004; Russell et al., 2004) (Table 7.6). A visual inspection of the 3D structure (Fig. 7.4i) reveals some interesting characteristics. The crucial region of predicted disorder observed in the highly virulent 1918 H1N1 strain (residues 68–79) is absent in this H7 subtype. An inspection of the regions in close proximity to the helical tip (HA_2 residues 68–79) reveals that the adjacent loop (residues 61–68, see Fig. 7.5i) is predicted to be disordered. This suggests that H7-related viruses could become more pathogenic with mutations that extend the disordered region to include the helical tip.

7.5.3 The H9N2 Puzzle and Modulation of Virulence

7.5.3.1 A Modulating "Switch" in H9N2: Infectivity versus Immune Evasion The H9N2 virus is generally nonvirulent for mammals (Wan et al., 2008). However, our analysis revealed that it has a region of predicted disorder at the tip of its helical stalk. This region of predicted disorder in the H9N2 HA_2 is noticeably shorter than that in the 1918 H1N1 protein. On the other hand, studies on H9N2 strains with low virulence revealed that a single point mutation can make these viruses more virulent (Wan et al., 2008). In fact, Leu226 in the HA RBS, responsible for human viruslike receptor specificity, was found to be important for transmission of the H9N2 viruses in ferrets (Wan et al., 2008). Interestingly, earlier it has been shown that changes in the HA RBS of H9N2 isolates allow adaptation to mammalian-type sialic receptors (Matrosovich et al., 2001) and that a change at position 227 of the HA_1 subunit of H5N1 had a noticeable effect on its virulence in mice (Hatta et al., 2001). This suggests a mechanism by which the virus can potentially modulate its virulence. The virulence arising from mutations near the receptor-binding site may work in tandem with the immune evasion arising from tip of the HA_2 stalk. One way that the former can modulate the effects of the latter is by controlling the infectivity of the virus. A virus that binds to its host with great efficiency and is also effective in evading the immune systems creates the "perfect storm" for virulence.

7.5.3.2 A Secondary Switch for Virulence of H9N2 and H5N1 via Oligosaccharides It has been suggested that immune pressure caused by the use of a

vaccine could create a survival advantage for those influenza viruses that undergo antigenic variation (Guan et al., 2004). This evolution produces low-virulence escape mutants, many of which were associated with the acquisition of new glycosylation sites in the HA_1 subunit, at positions 198 and 131 in H9N2 and H5N1, respectively (Kaverin et al., 2002, 2004). It has been reported that the restored pathogenicity of low-virulence H5 and H9 escape mutants by lung-to-lung passage in mice could be attributed to reacquisition of the wild-type HA gene sequence and was also associated either with the removal of a glycosylation site (the one acquired previously by the escape mutant) without the exact restoration of the initial wild-type amino acid sequence, or, for an H5 escape mutant that had no newly acquired glycosylation sites, with an additional amino acid change in a remote part of the HA molecule (at position 156 of HA_1) (Rudneva et al., 2005). This clearly shows the crucial role of the loss of glycosylation sites in restoration of the virulence of H9 and H5 readaptants. It should be noted that mutations affecting the glycosylation of HA are likely to affect virulence because it has been proposed that it is common for the glyco-conjugate to act in tandem with intrinsic disorder of viral proteins in immune evasion (Goh et al., 2002, 2008c). Therefore, it should not be surprising that modulation of the virulence will often involve changes at the glyco-conjugate.

7.5.3.3 Possible Mechanism for Initial Wave of Nonvirulent Strain in 1918

By analogy, this may also help explain why and how the first wave of the 1918 H1N1 virus was not as virulent as the second (Taubenberger et al., 2001; Reid et al., 2001a). As little as a single mutation could have set off the virulent potential of the second wave of the 1918 H1N1 virus.

7.5.4 Shuffling and Grouping by Residues of the Same Polarity around Position 68 of HA_2

An analysis of the HA_2 sequence of the 1918 H1N1 strain shows that a set of polar residues located in the vicinity of residue 68 make this segment more polar and, thus, potentially more disordered. Figure 7.5 shows that the normalized B-factor values vary with the PONDR VLXT values. However, the effects seem to be lagged in many cases. For example, it can be seen that a peak in the normalized B-factor curve near position 60 in the 1918 H1N1 HA_2 plot (Fig. 7.5a, 1ruz.pdb), corresponds to the PONDR VLXT maximum in the vicinity of residues 68–79. This crucial PONDR VLXT maximum, seen in the 1918 H1N1 and H5N1 HA_2, is also seen on a smaller scale in the 1930s H1N1 HA_2. The peak of the 1918 H1N1 HA_2 is relatively higher than those from the 1930s, suggesting that this region in the 1918 H1N1 HA_2 is more flexible. This is further supported by the corresponding B-factor curves (Fig. 7.5).

7.5.5 Swine versus Avian Influenza Viruses: Patterns of Predicted Disorder

Another puzzle is the presence of predicted disorder at the bottom of the stalk. Often the predicted disorder appears at the bottom of the longer α-helix, but sometimes

it appears at the base of the smaller helix. An example of the former is the H1N1 subtype, whereas an example of the latter is H5N1. In yet other cases, such as H7N3 and H3N2, predicted disorder appears at the bases of both helices. A hint for a possible explanation of this behavior is provided by the analysis of the swine variant of the H9N2 subtype (Figs. 7.5g) (Ha et al., 2001). In this instance, a region of predicted disorder appears at the bottom of the longer α-helix, just as in H1N1. This is of particular interest because H1N1 is believed to be of swine-host origin (Ha et al., 2001; Reid et al., 2001a). Therefore, we suggest that the position of this predicted disorder region at the stalk bottom is dependent on the host type. Interestingly, the 1918 H1N1 is suspected to be of both avian and swine origins (Webster et al., 1995; Taubenberger et al., 2001; Gamblin et al., 2004; Reid et al., 2004b) just as the swine strain of the H9N2 subtype is of avian origin (Lin et al., 2000b; Ha et al., 2001). The pattern of predicted disorder of the 1918 H1N1 HA, which resembles that of H9N2, may therefore cast more light on the origin of the 1918 H1N1 virus. The theory that the 1918 H1N1 virus is of avian origin but evolved in swine (Taubenberger et al., 2001; Reid et al., 2004b) seems to be supported here.

7.6 CONCLUDING REMARKS

This chapter illustrates the usefulness of the intrinsic disorder prediction for the comparative analysis of viral proteins. This approach offers several advantages, including the opportunity to map proteins by functionality, predicted disorder, and locality across viral species, strains, and subtypes. Furthermore, it provides useful benchmarks for the evaluation of the intrinsic disorder concept and for the analysis of various disorder predictors. Using this comparative study of predicted disorder, several interesting patterns in the behavior of viral proteins from influenza A virus were uncovered. We have shown that the patterns of predicted disorder can be mapped and related to the functions of the various viral proteins. There is evidence that the functions and the amount of disorder of the proteins are related to their physical location in the virion. Some of the key findings of this paper are further outlined below.

Intrinsic disorder is unevenly distributed within the virions, with the least predicted disorder being observed for the surface proteins and the most predicted disorder being characteristic for the proteins at the virion core.

Significant variations in the amount of predicted disorder by HA subtypes in influenza A virus were observed. This might provide an explanation for the variations in the functionality and infectivity of specific viral subtypes. Furthermore, NA and HA of major influenza A pandemics, tend to pair in such a way that both tend to be predicted either ordered–ordered or disordered–disordered. Such behavior might be linked to the evolutionary advantages of being ordered or disordered, but more experiments are needed to test this conjecture.

All the virulent influenza A strains analyzed in this study have a region of predicted disorder at the top of the largest α-helix of the HA stalk (near the residues

68–79). This is the case for the 1918 H1N1 strain and for all the virulent samples of the H5N1 and H9N2 subtypes. Although the similarity between H9N2 and H5N1 is striking, the pathogenesis of H9N2 is still relatively unknown because few human cases have been reported so far. Our data suggest that the H9 subtypes are likely to be highly virulent as their HA_2 contains the predicted disorder region similar to the segment of predicted disorder seen in all the virulent strains studied. Conversely, the subtypes and strains that are known to be less virulent than the 1918 H1N1 and H5N1 have no predicted disorder at the tip of the stalk. Examples of this behavior are H3N2 and the particular strain of H7N3.

The H9N2 subtype is potentially able to modulate its virulent nature by affecting the efficiency of binding to the host cells, thus controlling its ability to evade the immune system. This is supported by the observation that a single mutation near the receptor-binding site is generally sufficient to convert H9N2 and some H5N1 strains from nonvirulent to highly pathogenic. This also represents a potential mechanism by which the second wave of the 1918 H1N1 virus became more virulent than the first wave.

We propose that the high mobility of the exposed region of the HA trimer accounts for the evasion of the initial immune response. This highly dynamic nature of the potentially immunogenic region weakens the binding of antibodies and other immune response-related molecules to the HA molecule. In this way, the virus buys some time to invade the host. It is also possible that the immune system does not elicit the adequate immune response throughout all stages of the disease as tight binding to the highly dynamic HA is difficult. Evidence supporting both hypotheses can be found in the observation that when the 1918 H1N1 virus was introduced to mice, unusually high viral loads were seen within a short period (Kobasa et al., 2004). The crucial region of predicted disorder at the helical tip is seen in all HAs of influenza viruses of avian origin analyzed thus far. It is highly plausible that this disordered region is needed for the virus to move between various species of birds by allowing sufficient copies of the virion to avoid immune detection in the host's body in order to increase the odds of infections.

Development of vaccines for highly virulent viruses such as 1918 H1N1 or H5N1 may be fraught with dangers as a result of the ability of their proteins to trigger fatal immune responses (Loo and Gale, 2007). The results here provide new potential strategies to develop vaccines for virulent influenza strains. Relatively nonvirulent immunogenic proteins could be developed in either of the following two ways. One way is to use disorder predictors such as PONDR VLXT to quickly identify variants of the virulent viruses that are predicted to be nonvirulent by lacking the region of predicted disorder at the tip of the stalk. Another strategy would be to mutate the crucial region in the viruses themselves at the crucial sequence in the HA_2 protein.

ACKNOWLEDGEMENTS

This work was supported in part by the grant EF 0849803 (to A.K.D and V.N.U.) from the National Science Foundation and the Program of the Russian Academy

of Sciences for the "Molecular and Cellular Biology" (to V.N.U.). We gratefully acknowledge the support of the IUPUI Signature Centers Initiative. This chapter is partially based on previous publications (Goh et al., 2008ab, 2009).

ABBREVIATIONS

BIV	bovine immunodeficiency virus
CA	capsid
CAEV	caprine arthritis-encephalitis virus
EIAV	equine infectious anemia virus
IN	integrase
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
FIV	feline immunodeficiency virus
HIV	human immunodeficiency virus
MA	matrix
MVV	maedi-visna virus
NC	nucleocapsid
$PI(4,5)P_2$	phosphatidylinositol-(4,5)-bisphosphate
PID	predicted intrinsic disorder
PM	plasma membrane
PONDR	predictor of natural disordered regions
PR	protease
RT	reverse transcriptase
SIV	simian immunodeficiency virus
SIV_{cpz}	SIV infecting chimpanzees
SIV_{mac}	SIV infecting rhesus macaque
SIV_{sm}	SIV infecting sooty mangabeys
SRLV	small ruminant lentivirus
SU	structural unit
TM	transmembrane

REFERENCES

Centers for Disease Control and Prevention. Update: isolation of avian influenza A(H5N1) viruses from humans–Hong Kong, 1997–1998. MMWR 1998;46(52–53):1245–1247.

Barman S, Adhikary L, Chakrabarti AK, Bernas C, Kawaoka Y, Nayak DP. Role of transmembrane domain and cytoplasmic tail amino acid sequences of influenza a virus neuraminidase in raft association and virus budding. J Virol 2004;78(10):5258–5269.

Baudin F, Bach C, Cusack S, Ruigrok RW. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. EMBO J 1994;13(13):3158–3165.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28(1):235–242.

Biswas SK, Boutz PL, Nayak DP. Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. J Virol 1998;72(7):5493–5501.

Biswas SK, Nayak DP. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J Virol 1994;68(3):1819–1826.

Bornholdt ZA, Prasad BV. X-ray structure of influenza virus NS1 effector domain. Nat Struct Mol Biol 2006;13(6):559–560.

Brooks GF, Butel JS, Morse SA. Jawetz, Melnick and Adelberg's medical microbiology. 23rd ed. New York: Lange/McGraw-Hill; 2004.

Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 2002;55(1):104–110.

Brown DA, London E. Functions of lipid rafts in biological membranes. Annu Rev Cell Dev Biol 1998;14:111–136.

Bullough PA, Hughson FM, Skehel JJ, Wiley DC. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature 1994;371(6492):37–43.

Burmeister WP, Ruigrok RW, Cusack S. The 2.2 A resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. EMBO J 1992;11(1):49–56.

Burton DR. Antibodies, viruses and vaccines. Nat Rev Immunol 2002;2(9):706–713.

Chien CY, Tejero R, Huang Y, Zimmerman DE, Rios CB, Krug RM, Montelione GT. A novel RNA-binding motif in influenza A virus non-structural protein 1. Nat Struct Biol 1997;4(11):891–895.

Clements JE, Zink MC. Molecular biology and pathogenesis of animal lentivirus infections. Clin Microbiol Rev 1996;9(1):100–117.

Colman PM. Influenza virus neuraminidase: structure, antibodies, and inhibitors. Protein Sci 1994;3(10):1687–1696.

Colman PM, Varghese JN, Laver WG. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature 1983;303(5912):41–44.

Compans RW, Content J, Duesberg PH. Structure of the ribonucleoprotein of influenza virus. J Virol 1972;10(4):795–800.

Di Trani L BB, Cordioli P, Muscillo M, Vignolo E, Moreno A, Tollis M. Molecular characterization of low pathogenicity H7N3 avian influenza viruses isolated in Italy. Avian Dis 2004;482(2):376–383.

Durrer P, Galli C, Hoenke S, Corti C, Gluck R, Vorherr T, Brunner J. H^+-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. J Biol Chem 1996;271(23):13417–13421.

Earp LJ, Delos SE, Park HE, White JM. The many mechanisms of viral membrane fusion proteins. Curr Top Microbiol Immunol 2005;285:25–66.

Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem 2001;70:777–810.

Egorov A, Brandt S, Sereinig S, Romanova J, Ferko B, Katinger D, Grassauer A, Alexandrova G, Katinger H, Muster T. Transfectant influenza A viruses with long deletions in the NS1 protein grow efficiently in Vero cells. J Virol 1998;72(8):6437–6441.

Fechter P, Mingay L, Sharps J, Chambers A, Fodor E, Brownlee GG. Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. J Biol Chem 2003;278(22):20381–20388.

Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, Smith D, Rimmelzwaan GF, Olsen B, Osterhaus AD. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. J Virol 2005;79(5):2814–2822.

Gamblin, SJ, Haire, LF, Russell, RJ, Stevens, DJ, Xiao, B, Ha, Y, Vasisht, N, Steinhauer, DA, Daniels, RS, Elliot, A, Wiley, DC, Skehel JJ. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science 2004;303:1838–1842.

Garcia-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. Virology 1998;252(2):324–330.

Garcia-Sastre A, Whitley RJ. Lessons learned from reconstructing the 1918 influenza pandemic. J Infect Dis 2006;194(Suppl 2):S127–S132.

Gibbs JS, Malide D, Hornung F, Bennink JR, Yewdell JW. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. J Virol 2003;77(13):7214–7224.

Goh GK, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. Virol J 2008a;5:126.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genom 2008b;9(Suppl 2):S4.

Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. Virol J 2009;6:69.

Goh GKM, Dunker AK, Uversky V. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genom 2008c;9:S4.

Goh GKM, Uversky V, Dunker AK. A comparative analysis of viral matrix proteins using disorder predictor Virol J 2008;5:126.

Guan Y, Poon LL, Cheung CY, Ellis TM, Lim W, Lipatov AS, Chan KH, Sturm-Ramirez KM, Cheung CL, Leung YH, Yuen KY, Webster RG, Peiris JS. H5N1 influenza: a protean pandemic threat. Proc Natl Acad Sci USA 2004;101(21):8156–8161.

Guan Y, Shortridge KF, Krauss S, Webster RG. Molecular characterization of H9N2 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? Proc Natl Acad Sci USA 1999;96(16):9363–9367.

Ha Y, Stevens DJ, Skehel JJ, Wiley DC. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. Proc Natl Acad Sci USA 2001;98:11181–11186.

Hale BG, Randall RE, Ortin J, Jackson D. The multifunctional NS1 protein of influenza A viruses. J Gen Virol 2008;89(Pt 10):2359–2376.

Hatta M, Gao P, Halfmann P, Kawaoka Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science 2001;293(5536):1840–1842.

He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z, Liu Y. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. Nature 2008;454(7208):1123–1126.

Hilleman MR. Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. Vaccine 2002;20(25–26):3068–3087.

Hobohm U, Scharf M, Schneider R, Sander C. Selection of representative protein data sets. Protein Sci 1992;1(3):409–417.

Huang IC, Li W, Sui J, Marasco W, Choe H, Farzan M. Influenza A virus neuraminidase limits viral superinfection. J Virol 2008;82(10):4834–4843.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004;32(3):1037–1049.

Itoh M, Hotta H. Structure, function and regulation of expression of influenza virus matrix M1 protein. Nippon Rinsho 1997;55(10):2581–2586.

Jin H, Leser GP, Zhang J, Lamb RA. Influenza virus hemagglutinin and neuraminidase cytoplasmic tails control particle shape. EMBO J 1997;16(6):1236–1247.

Jin H, Subbarao K, Bagai S, Leser GP, Murphy BR, Lamb RA. Palmitylation of the influenza virus hemagglutinin (H3) is not essential for virus assembly or infectivity. J Virol 1996;70(3):1406–1414.

Kaverin NV, Rudneva IA, Ilyushina NA, Lipatov AS, Krauss S, Webster RG. Structural differences among hemagglutinins of influenza A virus subtypes are reflected in their antigenic architecture: analysis of H9 escape mutants. J Virol 2004;78(1):240–249.

Kaverin NV, Rudneva IA, Ilyushina NA, Varich NL, Lipatov AS, Smirnov YA, Govorkova EA, Gitelman AK, Lvov DK, Webster RG. Structure of antigenic sites on the haemagglutinin molecule of H5 avian influenza virus and phenotypic variation of escape mutants. J Gen Virol 2002;83(Pt 10):2497–2505.

Kielian M. Class II virus membrane fusion proteins. Virology 2006;344(1):38–47.

Kielian M, Rey FA. Virus membrane-fusion proteins: more than one way to make a hairpin. Nat Rev Microbiol 2006;4(1):67–76.

Kilbourne ED. Influenza pandemics of the 20th century. Emerg Infect Dis 2006a;12(1): 9–14.

Kilbourne ED. Influenza pandemics of the 20th century. Emerg Infect Dis 2006b;12(1): 9–14.

Kilbourne ED, Laver WG, Schulman JL, Webster RG. Antiviral activity of antiserum specific for an influenza virus neuraminidase. J Virol 1968;2(4):281–288.

Kobasa D, Takada A, Shinya K, Hatta M, Halfmann P, Theriault S, Suzuki H, Nishimura H, Mitamura K, Sugaya N, Usui T, Murata T, Maeda Y, Watanabe S, Suresh M, Suzuki T, Suzuki Y, Feldmann H, Kawaoka Y. Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. Nature 2004;431(7009):703–707.

Kobayashi M, Toyoda T, Adyshev DM, Azuma Y, Ishihama A. Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. J Virol 1994;68(12):8433–8436.

Kochs G, Koerner I, Thiel L, Kothlow S, Kaspers B, Ruggli N, Summerfield A, Pavlovic J, Stech J, Staeheli P. Properties of H7N7 influenza A virus strain SC35M lacking interferon antagonist NS1 in mice and chickens. J Gen Virol 2007;88(Pt 5):1403–1409.

Krug RM, Etkind PR. Cytoplasmic and nuclear virus-specific proteins in influenza virus-infected MDCK cells. Virology 1973;56(1):334–348.

Lamb RA, Krug RM. Orthomyxoviridae: The virus and their replication. In: Knipe DM, Howley PM, editors. Fields virology. Philadelphia, PA: Lippincott Williams & Wilkins; 2001, pp. 1487–1532.

Lamb RA, Lai CJ. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell 1980;21(2):475–485.

Lederberg J. H1N1-influenza as Lazarus: genomic resurrection from the tomb of an unknown. Proc Natl Acad Sci USA 2001;98(5):2115–2116.

Li ML, Rao P, Krug RM. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. EMBO J 2001;20(8):2078–2086.

Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. Genome Inform Ser Workshop Genome Inform 1999;10:30–40.

Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, Subbarao K, Guan Y, Krauss S, Shortridge K, Webster R, Cox N, Hay A. Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci USA 2000a;97(17):9654–9658.

Lin YP, Shaw M, Gregory V, Cameron K, Lim W, Klimov A, Subbarao K, Guan Y, Krauss S, Shortridge K, Webster R, Cox N, Hay A. Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci USA 2000b;97(17):9654–9658.

Liu Y, Lou Z, Bartlam M, Rao Z. Structure-function studies of the influenza virus RNA polymerase PA subunit. Sci China C Life Sci 2009;52(5):450–458.

Lommer BS, Luo M. Structural plasticity in influenza virus protein NS2 (NEP). J Biol Chem 2002;277(9):7108–7117.

Loo YM, Gale M. Influenza: Fatal immunity and the 1918 virus. Nature 2007;445(7125): 267–268.

Ludwig S, Schultz U, Mandler J, Fitch WM, Scholtissek C. Phylogenetic relationship of the nonstructural (NS) genes of influenza A viruses. Virology 1991;183(2):566–577.

Matrosovich MN, Krauss S, Webster RG. H9N2 influenza A viruses from poultry in Asia have human virus-like receptor specificity. Virology 2001;281(2):156–162.

McMichael A, Mwau M, Hanke T. Design and tests of an HIV vaccine. Br Med Bull 2002;62:87–98.

Melkonian KA, Ostermeyer AG, Chen JZ, Roth MG, Brown DA. Role of lipid modifications in targeting proteins to detergent-resistant membrane rafts. Many raft proteins are acylated, while few are prenylated. J Biol Chem 1999;274(6):3910–3917.

Mitrasinovic PM. Advances in the structure-based design of the influenza A neuraminidase inhibitors. Curr Drug Targets 2010;11(3):315–326.

Moscona A. Neuraminidase inhibitors for influenza. N Engl J Med 2005;353(13): 1363–1373.

Mukherjee, S, Maxfield, FR. Role of membrane organization and membrane domains in endocytic lipid trafficking. Traffic 2000;1(3):203–211.

Ng AK, Wang JH, Shaw PC. Structure and sequence analysis of influenza A virus nucleoprotein. Sci China C Life Sci 2009;52(5):439–449.

O'Neill RE, Talon J, Palese P. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J 1998;17(1):288–296.

Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, Nagata K, Tame JR, Park SY. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. Nature 2008;454(7208):1127–1131.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(Suppl 6):566–572.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44(37):12454–12470.

Paragas J, Talon J, O'Neill RE, Anderson DK, Garcia-Sastre A, Palese P. Influenza B and C virus NEP (NS2) proteins possess nuclear export activities. J Virol 2001;75(16):7375–7383.

Pinto LH, Holsinger LJ, Lamb RA. Influenza virus M2 protein has ion channel activity. Cell 1992;69(3):517–528.

Pinto LH, Lamb RA. The M2 proton channels of influenza A and B viruses. J Biol Chem 2006;281(14):8997–9000.

Portela A, Digard P. The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. J Gen Virol 2002;83(Pt 4):723–734.

Prokudina-Kantorovich EN, Semenova NP. Intracellular oligomerization of influenza virus nucleoprotein. Virology 1996;223(1):51—385–6.

Reid A. The effects of the 1918–1919 influenza pandemic on infant and child health in Derbyshire. Med Hist 2005a;49(1):29–54.

Reid A. The effects of the 1918–1919 influenza pandemic on infant and child health in Derbyshire. Med Hist 2005b;49(1):29–54.

Reid AH, Taubenberger JK, Fanning TG. The 1918 Spanish influenza: Integrating history and biology. Microb Infect 2001a;3:81–87.

Reid AH, Taubenberger JK, Fanning TG. The 1918 Spanish influenza: integrating history and biology. Microb Infect 2001b;3(1):81–87.

Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. Nat Rev Microbiol 2004a;2(11):909–914.

Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. Nat Rev Microbiol 2004b;2(11):909–914.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38–48.

Rudneva I, Ilyushina N, Timofeeva T, Webster R, Kaverin N. Restoration of virulence of escape mutants of H5 and H9 influenza viruses by their readaptation to mice. J Gen Virol 2005;86:2831–2838.

Russell RJ, Gamblin SJ, Haire LF, Stevens DJ, Xiao B, Ha Y, Skehel JJ. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. Virology 2004;325(2):287–296.

Sakai T, Ohuchi R, Ohuchi M. Fatty acids on the A/USSR/77 influenza virus hemagglutinin facilitate the transition from hemifusion to fusion pore formation. J Virol 2002;76(9):4603–4611.

Scheiffele P, Rietveld A, Wilk T, Simons K. Influenza viruses select ordered lipid domains during budding from the plasma membrane. J Biol Chem 1999;274(4):2038–2044.

Schnell JR, Chou JJ. Structure and mechanism of the M2 proton channel of influenza A virus. Nature 2008;451(7178):591–595.

Seto JT, Rott R. Functional significance of sialidose during influenza virus multiplication. Virology 1966;30(4):731–737.

Simons K, Ikonen E. Functional rafts in cell membranes. Nature 1997;387(6633): 569–572.

Simons K, Ikonen E. How cells handle cholesterol. Science 2000;290(5497):1721–1726.

Simons K, Toomre D. Lipid rafts and signal transduction. Nat Rev Mol Cell Biol 2000; 1(1):31–39.

Skehel JJ, Bayley PM, Brown EB, Martin SR, Waterfield MD, White JM, Wilson IA, Wiley DC. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc Natl Acad Sci USA 1982;79(4):968–972.

Skehel JJ, Wiley DC. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell 1998;95(7):871–874.

Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 2000a;69:531–5–69.

Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 2000b;69:531–569.

Stevens J, Corper AL, Basler CF, Taubenberger JK, Palese P, Wilson IA. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science 2004;303(5665):1866–1870.

Stouffer AL, Acharya R, Salom D, Levine AS, Di Costanzo L, Soto CS, Tereshko V, Nanda V, Stayrook S, DeGrado WF. Structural basis for the function and inhibition of an influenza virus proton channel. Nature 2008;451(7178):596–599.

Sun X, Whittaker GR. Role for influenza virus envelope cholesterol in virus entry and infection. J Virol 2003;77(23):12543–12551.

Taubenberger JK, Reid AH, Janczewsk TA, Fanning TG. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. Philos Trans R Soc Lond B Biol Sci 2001;356(1416):1829.

Treanor JJ, Snyder MH, London WT, Murphy BR. The B allele of the NS gene of avian influenza viruses, but not the A allele, attenuates a human influenza A virus for squirrel monkeys. Virology 1989;171(1):1–9.

Turner BG, Summers MF. Structural biology of HIV. J Mol Biol 1999;285(1):1–32.

Tweed SA, Skowronski DM, David ST, Larder A, Petric M, Lees W, Li Y, Katz J, Krajden M, Tellier R, Halpert C, Hirst M, Astell C, Lawrence D, Mak A. Human illness from avian influenza H7N3, British Columbia. Emerg Infect Dis 2004;10(12):2196–2199.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–427.

Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. Biochemistry 1999;38(45):15009–15016.

Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. Nature 1983;303(5912):35–40.

Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. Proteins 2003;52(4):573–584.

Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. J Proteome Res 2007;6(5):1899–1916.

Wagner R, Matrosovich M, Klenk HD. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol 2002;12(3):159–166.

Wan H, Sorrell E, Song H, Hossain M, Ramirez-Nieto G, Monne I, Stevens J, Cattoli G, Capua I, Chen L, Donis R, Busch J, Paulson J, Brockwell C, Webby R, Blanco J, Al-Natour M, Perez DR. Replication and transmission of H9N2 influenza viruses in ferrets: evaluation of pandemic potential. PLoS ONE 2008;3:e2923.

Weber T, Paesold G, Galli C, Mischler R, Semenza G, Brunner J. Evidence for H(+)-induced insertion of influenza hemagglutinin HA2N-terminal segment into viral membrane. J Biol Chem 1994;269(28):18353–18358.

Webster RG, Laver WG, Air GM, Schild GC. Molecular mechanisms of variation in influenza viruses. Nature 1982;296(5853):115–121.

Webster RG, Sharp GB, Claas EC. Interspecies transmission of influenza viruses. Am J Respir Crit Care Med 1995;152:S25–S23.

Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 1996;35(43):13709–13715.

Wharton SA, Calder LJ, Ruigrok RW, Skehel JJ, Steinhauer DA, Wiley DC. Electron microscopy of antibody complexes of influenza virus haemagglutinin in the fusion pH conformation. EMBO J 1995;14(2):240–246.

Wiley DC, Skehel JJ. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem 1987;56:365–394.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 1981;289(5796):366–373.

World Health Organization. Avian influenza, Viet Nam–update. Wkly Epidemiol Rec 2005;80(27):233–234.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293(2):321–331.

Wu WW, Sun YH, Pante N. Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. Virol J 2007;4:49.

Wuethrich B. Infectious disease. An avian flu jumps to people. Science 2003;299(5612): 1504.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J Proteome Res 2007a;6(5):1917–1932.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 2007b;6(5):1882–1898.

Zamarin D, Garcia-Sastre A, Xiao X, Wang R, Palese P. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. PLoS Pathog 2005;1(1):e4.

Zhang J, Pekosz A, Lamb RA. Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins. J Virol 2000;74(10):4634–4644.

8

MAKING ORDER IN THE INTRINSICALLY DISORDERED REGIONS OF HIV-1 Vif PROTEIN

TALI H. REINGEWERTZ, DEBORAH E. SHALEV, AND ASSAF FRIEDLER

8.1 INTRODUCTION

8.1.1 HIV-1: General Background

Human immunodeficiency virus-1 (HIV-1) causes the acquired immunodeficiency syndrome (AIDS) [1,2], which is widely known as an epidemic and is recently becoming a manageable chronic, though a still incurable, disease [3,4]. HIV-1 is a retrovirus from the primate *Lentivirus* genus. Being a retrovirus, its genome has two single-stranded positive-sense RNA molecules and expresses the typical genes *Gag*, yielding the structural proteins; *Pol*, yielding the HIV-1 enzymes; and *Env*, encoding the envelope glycoproteins [5,6]. As a lentivirus, the HIV-1 encodes six additional accessory proteins: Tat and Rev, which are major players in the regulation of HIV-1 gene expression [7], and Vif (viral infectivity factor), Vpu, Vpr, and Nef, which play a role in changing the cell environment to enable virus infectivity [8]. HIV-1 mature virion, presented in Fig. 8.1 [5,6,9,10], has different layers surrounded by an envelope derived from the lipid bilayer of the producer cell, from which the *Env* glycoproteins (SU-gp120), connected to the transmembrane glycoproteins (TM-gp41), extend on the surface. The inner layers contain proteins that are products of *Gag* proteolysis. These include the matrix (MA-p17) protein, which forms the matrix shell at the inner side of the membrane and the capsid (CA-p27) protein that forms a cone-shaped capsid in the center, which encloses the nucleocapsid (NC-p7) proteins that coat the viral genomic RNA. The capsid

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

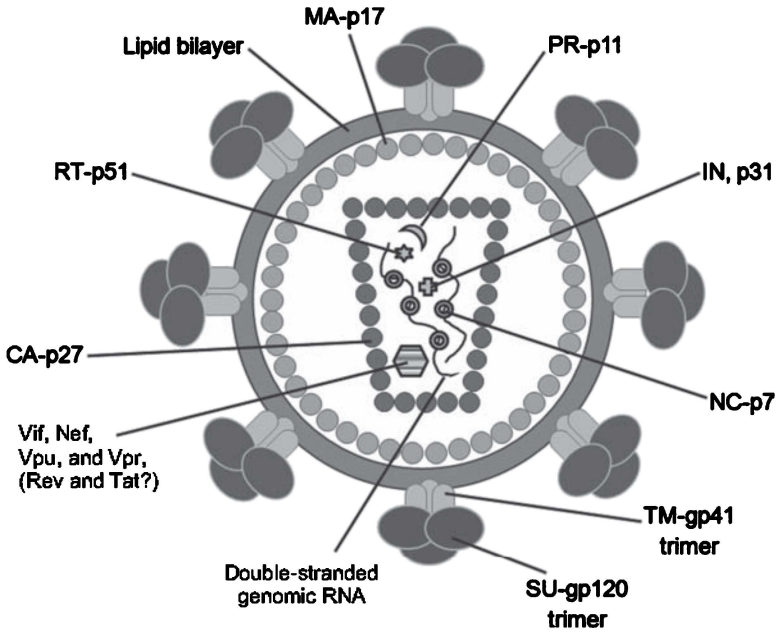


Figure 8.1 Schematic representation of the HIV-1 structure, indicating the arrangement of viral components in the different layers (see Section 8.1.1 in the text for details).

also contains the *Pol*-derived enzymes, reverse transcriptase (RT-p51), integrase (IN, p31), and protease (PR-p11), and some of the accessory proteins including Vif, Nef, Vpu, and Vpr.

The HIV-1 replication cycle begins by the virus fusing with the cell membrane when its envelope proteins bind the cellular receptor CD4, followed by interactions with the α and β chemokines, CCR5 and CXCR4 [11,12]. Binding the coreceptor induces a conformational change in the structure of gp120 and gp41, triggering membrane fusion and subsequent release of the viral core into the cytoplasm of the infected cell [13]. Reverse transcription of viral RNA into a double-stranded DNA is catalyzed by reverse transcriptase (RT). Then, the viral cDNA is transported into the nucleus in the form of a preintegration complex together with IN, MA, RT, Vpr, and other proteins [14]. In the host cell nucleus, IN catalyzes the integration of the viral DNA into the host chromosome, where it can reside for years in a latent form. The viral genes are expressed using the host cell protein production apparatus. Transcription of the viral DNA results in spliced and unspliced mRNA, which is translated into the different viral polyprotein precursors, Pr55Gag and Pr160GagPol, and the Env precursor, gp160. The RNA also serves as the genomic RNA for the newly formed virions. Virus assembly into preliminary noninfectious virions is initiated by targeting Pr55Gag to the cell membrane and acquiring a lipid envelope while budding out of the host cell. At the same time, the viral PR is activated, cleaving the precursor polyproteins into the viral components that rearrange to form the mature infectious particles [10].

8.1.2 The HIV-1 Vif Protein

HIV-1 infection of many target cells requires Vif, a 23-kDa protein comprising 192 residues. Vif overcomes the innate antiviral cellular defense and is conserved in almost all lentiviruses [15,16]. Vif is synthesized during the late phase of infection and is both virion- and cell-associated [17,18]. During productive infection, newly synthesized virions contain considerable amounts of Vif (60–100 copies per virion), while chronically infected cells are Vif depleted (4–6 copies per virion) [18]. Vif is incorporated into the virions by an interaction with the viral genomic RNA [18,19]. Although Vif is a cytosolic protein, it also exists as a membrane-associated protein on the cytoplasmic side of the membrane [20–22]. Vif is in dynamic equilibrium among various oligomeric forms, which plays a part in its antiviral activity [23–26]. Other possible regulation modes of Vif include multiple phosphorylations on serine and threonine residues [27–29], as well as processing by the viral protease that releases a Vif fragment including residues 151–192 from the Vif CTD [30].

Vif is only required for HIV-1 growth in certain cells termed *restrictive* or *non-permissive*. In cells such as $CD4^+$ T-cell lines CEM and H9 and the peripheral blood T lymphocytes and macrophages, the virus cannot produce infectious particles in the absence of Vif. In permissive cells, such as SupT1, C8166, and Jurkat T-cell lines, HIV-1 infection can proceed without the presence of Vif [16,31–33]. Restrictive cells express the cellular cytosine deaminase APOBEC-3G (A3G) [34], a core protein in the antiviral mechanism. A3G inserts dC → dU mutations in the viral DNA during its synthesis from the viral RNA template [35–37]. This leads, in many cases, to subsequent degradation of the mutated DNA [38] and therefore the production of noninfectious particles. While A3G inhibition is the central role of Vif [39–43], the exact mechanism has not been resolved. Several modes of action have been postulated for the inhibitory effect of Vif, as illustrated in Fig. 8.2: (i) Vif induces A3G degradation by the proteasome–ubiquitin pathway [41,44–46]; (ii) Vif prevents the inclusion of A3G into newly synthesized virions [24,47]; (iii) Vif directly inhibits A3G enzymatic activity [48,49]; and (iv) Vif impairs the translation of A3G mRNA [50,51]. A3G degradation is mediated by an E3 cullin-RING ubiquitin ligase complex in which Vif serves as a recognition substrate of A3G. This results in polyubiquitination of A3G and its subsequent degradation in the proteasome [41,45,46]. The E3 complex includes Vif, Cul5, Rbx2, ElonginB, and ElonginC. The latter four proteins are capable of binding many other specificity factors that have a suppressor of cytokine signaling box (SOCS-box) domain [52].

Vif interacts with several viral and host factors to exert its antiviral activity. Vif is divided into two regions and further subdomains. The N-terminal region, comprising residues 1–100, mainly interacts with A3G through several binding motifs [53–57]. It also interacts with the viral RNA [58,59] and the viral protease [60–62]. The N-terminal region of Vif also interacts with the MDM2 E3 ligase, which targets Vif for degradation by the ubiquitin–proteasome pathway [63].

The focus of this chapter is on the C-terminal region of Vif, containing residues 101–192. This region is composed of two intrinsically disordered regions (IDRs): the HCCH Zn^{2+}-binding domain and the C-terminal domain (CTD). Figure 8.3

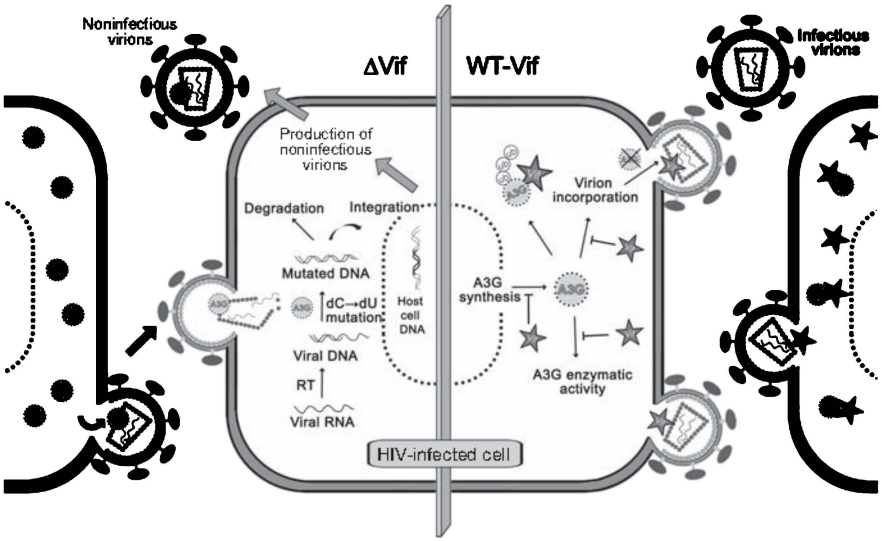


Figure 8.2 Proposed mechanism of Vif action representing several optional mechanisms attributed to Vif inhibitory effect on A3G antiviral activity. A3G activity is illustrated in the absence of Vif (see Section 8.1.2 in the text for details).

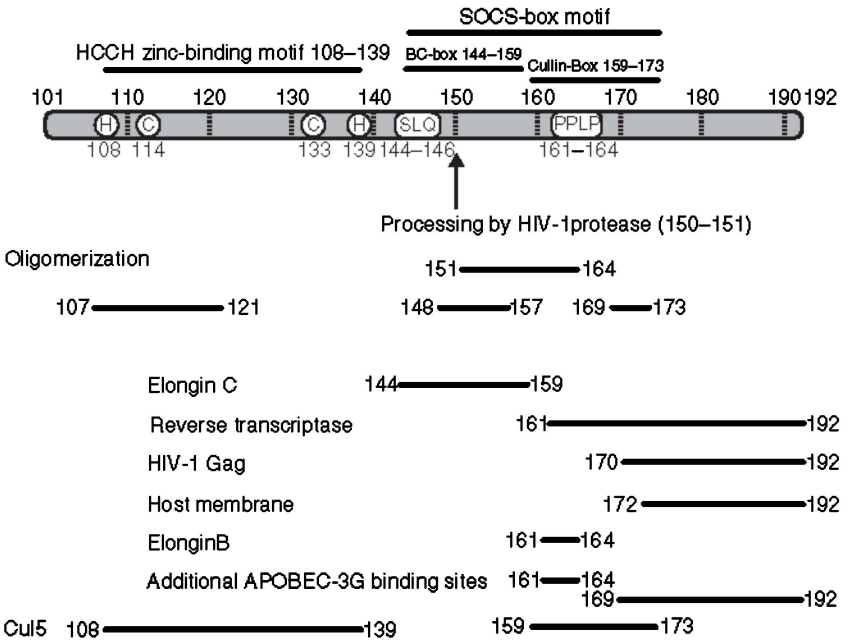


Figure 8.3 Schematic representation of Vif sequence and domain mapping, based on Vif interactions with its partner ligands.

maps the functional domains of the C-terminal region that contains multiple protein-binding domains. These domains mediate the formation of the E3 ligase complex and many other important interactions of Vif. The HCCH Zn^{2+}-binding domain (residues 108–139) directly interacts with Cul5 in a Zn^{2+}-enhanced mode [64–66]. It is followed by a SOCS-box domain (residues 144–173), which contains a BC-box region (residues 144–159) that binds EloC. It also contains a putative cullin box (residues 159–173), which contains a conserved ^{161}PPLP^{164} motif that is crucial for its function [67]. This domain is specifically involved in different interactions of Vif, including an interaction with the Cul5 [68] and EloB [69] components of the E3 complex, as well as interactions with A3G [70,71], the viral RT [72], and Vif oligomerization [25,26]. The terminal region of the CTD participates in the interaction with the cytoplasmic side of the host cell membrane (residues 172–192) [20,21], binds A3G (residues 169–192) [48], and is important for RT enzymatic activity (residues 169–192) [72]. The vast biochemical and biophysical data that describe Vif function and interactions, together with several suggested inhibitors of its function, emphasize both the importance and complexity of this protein. Extensive structural analysis of Vif is necessary to completely understand its mechanism of action and will probably become the focus of research on Vif.

8.1.3 HIV-1 and Intrinsic Disorder

The term *intrinsic disorder* (ID) was introduced by Uversky et al. [73] to describe intrinsically disordered proteins (IDPs) or IDRs in proteins that lack stable tertiary structures under physiological conditions *in vitro*. The highly flexible, extended conformation of IDPs is composed of an ensemble of conformations interchanging dynamically and distinguishable from loops and other regions in folded proteins that lack secondary structures [73,74]. Various levels of disorder can appear in IDPs, including or depleted of secondary structure elements implied by the appearance of residual structure [75,76]. Dynamic transitions among several structures are highly dependent on factors in the surrounding local environment, such as different solution composition, solutes, pH, ionic strength, and the proximity of the binding partners [77,78]. ID and dynamic transitions can still exist even in a crowded environment, such as inside cells, as was demonstrated using cell NMR [79–81]. Such experiments imply the existence and importance of IDPs and the role of ID in living cells and not just the result of incompatible experimental folding conditions.

The unfolded character of IDPs, as well as the structure of globular proteins, is encoded in the primary amino acid sequence of the protein [82–85]. ID characteristics include low complexity of defined segments in the sequence, indicating low variation and frequent repetition of residues [82]. These sequences have a high content of polar and charged residues, referred to as *disorder-promoting residues*, as opposed to bulky hydrophobic residues and aromatic residues that usually promote the formation of the core of ordered proteins [82,85,86]. These fundamental differences were used to develop various computer algorithms for predicting ID in proteins, as previously summarized [86,87].

The functionality of IDPs has long been acknowledged [88–90]. The functions of IDPs encompass four major functional classes: molecular recognition, molecular assembly, protein modification, and entropic chains [91]. Proteins from the first two categories are mostly involved in signaling, control, or regulation functions that typically involve multiple partners and can benefit from the dynamic interplay among the different structures adopted by IDPs [91–93]. The ability to bind multiple partners is one of the important characteristics of IDPs [94]. The interplay between specificity and affinity of interaction can be addressed differently by IDPs. Disordered regions can bind partners with both high specificity and low affinity, enabling the interactions to be specific but easily reversed [73,94]. On the other hand, disorder-to-order transitions on binding may tighten up the interactions or the assembly of an overall complex [92], resulting in enhanced affinity coupled with broad specificity [74,95], as demonstrated by the self-assembly of macromolecules [96] and in the role of induced fit in the process of RNA–protein recognition [97]. Upon binding, the flexible disordered region may or may not gain structure, and different ligands may induce different structures. Synergistic folding on recognition and interaction of the different binding partners has also been ascribed to IDPs [89].

Several proteins in the HIV-1 are IDPs or contain IDRs. The Tat protein is a well-characterized IDP that interacts with a wide range of intra- and extracellular molecules [98–100]. Tat has been shown to be disordered both in its unbound state [99] and when complexed with specific fragments of antibodies, exhibiting only a limited gain of structure [98]. Evidence for ID in the HIV-1 Rev protein comes from the crystal structure of Rev with a specifically engineered monoclonal Fab fragment [101]. While the NTD of the Rev protein is folded into a helix-loop-helix motif, the CTD of Rev remains disordered. The third variable loop (V3) of the HIV-1 envelope glycoprotein gp120 was shown to adopt different conformations when bound to different anti-V3 antibodies [102–104]. This loop includes the principle neutralizing determinant of HIV-1 that serves as a key target for an anti-HIV-1 neutralizing antibody and was suggested to be conformationally heterogeneous in the unbound, full-length gp120 [104]. Comparative computational analysis of the HIV-1 and HIV-related proteins [105] predicted a high content of IDRs in the MA and CA proteins (61% and 48%, respectively) using PONDR (predictor of naturally disordered regions, summarized in [82]). The accessory proteins Tat (100% calculated disordered regions) and Vpr (39%) were considered to be disordered, while Nef and Vpu (both 26%) were assigned as moderately disordered together with IN (26%). PR and RT were predicted to be the least disordered. Vif and Rev were excluded from this analysis. A subsequent analysis compared matrix proteins related and unrelated to HIV-1 and found an important role for ID in the interactions of different MA proteins forming the matrix shells since the intersubunit contact sites were located within the predicted IDR [106]. The intrinsically disordered nature of the HIV-1 proteins may enhance virus infectivity by evading antibodies of the immune system and by being able to adapt the structure and function of the IDPs to different constraints imposed on the virus.

Structure determination and biophysical research of Vif is mainly conducted by using segments or domains of the molecule. This is mainly due to difficulties

in expressing and purifying full-length Vif and obtaining the high concentrations required for structural and biophysical studies. Therefore, there is no complete experimental structure of the full-length Vif. However, even if the problems of yield could be overcome, disordered conformations are invisible to X-ray and NMR by their nature. The HCCH Zn^{2+}-binding domain and the CTD of Vif have been shown to be intrinsically disordered in their unbound states. The ID contributes to the modularity of the function and interactions of Vif, as discussed in this chapter [107].

8.2 COMPUTATIONAL BASED STRUCTURAL ANALYSIS OF Vif

8.2.1 Order and Disorder Predictions

High resolution experimental 3D structures of Vif have not yet been solved and represent a great challenge. Structural data of the full-length Vif can be derived from computational predictions and modeling based on the biochemical nature of the protein. Secondary structure predictions of the Vif structure roughly divide the protein into two halves, the mainly ordered N-terminal region (residues 1–100) that is mostly a β-sheet and has one α-helix and a more disordered C-terminal region (residues 101–192) with helical elements. In total, about half (51.5%) of the secondary structure of Vif was predicted to be a random coil [108,109].

The Vif sequence was submitted to an array of 19 ID prediction algorithms on publicly available servers. These methods were based on different criteria for IDP predictions [108] and gave results consistent with previously reported predictions [23,110] and with the secondary structure predictions [108,109]. Most of the algorithms predicted the N-terminal region of Vif to be mainly ordered in residues 1–100, while in the second region of Vif, residues 100–150 were predicted to be mostly ordered and the CTD, residues 150–192, was predicted to be disordered.

8.2.2 Homology Modeling

Two homology models of full-length Vif have been calculated based on the published structural folds in the PDB [111,112]. The two models utilize secondary structure prediction and molecular dynamics based on template proteins that were chosen according to the similarities in their secondary elements. Thus, while the secondary elements in the two models are common, the overall tertiary arrangements of the structures are distinct. The first model [112] presents a detailed structure prediction available in the PDB (code 1VZF) based on the structure of IPNS (PDB code 1BK0). Among the five α-helices presented in the model, two are in the N-terminal region, comprising residues 59–86 and 93–97. These interact with another helix from the CTD, residues 176–183. This interaction contributes to the specific tertiary arrangement of the protein according to the model. The CTD of Vif is presented in the model as a mainly disordered, exposed region with one helix spanning from residues 176 to 183 and an additional short β-sheet element.

Disorder in the CTD was further confirmed by the low score of the Kyte–Doolittle hydropathy plot, suggesting that the CTD is a surface-exposed region.

The second model [111] presented a structure of Vif that also represented a specific conformation in complex with EloBC. The authors used VHL (PDB code 1LM8) in complex with EloBC [113] as a basis for the Vif SOCS-box structure prediction (residues 142–177). They further used Narl (PDB code 1A04) for structure prediction of the rest of the N-terminal region (residues 1–143). The last 15 residues of Vif CTD were not included in the prediction because of the lack of structural similarity to any template. According to the calculated model, Vif CTD was mostly ordered when bound to EloBC, including the two major α-helices at the interface with EloC. The BC box and cullin box formed two hydrophobic clusters, among which, the first directly interacted with EloC and the second stabilized the interaction. Several mutations were shown to disrupt the suggested interface between Vif and EloBC, including SLQ to AAA, A149L, L163S, and L169S, which may indicate their importance for Vif action and virus infectivity [67,114].

The crystal structure of VHL in the E3 complex was the basis for other previously reported homology models of the local structure of Vif SOCS box in complex with EloBC [28,114]. Another SOCS-box template for Vif used the crystal structure of SOCS2 with EloBC [115] to model the interface of the Vif-Cul5 interaction [68]. This model suggested two potential interfaces, the first involving the PPLP motif in the cullin box and the second was found in the HCCH Zn^{2+} domain.

8.3 THE Vif C-TERMINAL DOMAIN AND THE SOCS-box DOMAIN

8.3.1 The Vif BC box Undergoes Induced Folding on Binding

The CTD of Vif (residues 141–192) is a major interaction interface with several of the Vif-binding partners. It undergoes different interactions mediated by overlapping regions that are crucial for Vif action, as shown in Fig. 8.3. Several biophysical and structural studies have focused on the structure and interactions of the SOCS-box domain with the EloBC complex [68,69,110,116,117], indicating that Vif CTD is intrinsically disordered in its unbound state and undergoes induced folding on interacting with the EloBC complex. The only currently available crystal structure of any of the Vif domains showed a helix-loop conformation at residues 140–155 from the Vif BC box, as part of a Vif 139–176 fragment in interaction with a hydrophobic face of the EloBC complex [68]. The remaining residues in this fragment were presumably unstructured as no electron density was observed for them. Two different studies showed the CTD of Vif to be disordered in the unbound state [69,110]. In the first study [69], a Vif construct containing the SOCS-box region (residues 139–176) fused to a 56-residue solubility enhancement tag [118] showed a structural change from a disordered to an ordered conformation by NMR spectroscopy. The structural change in the Vif SOCS box was seen by the disappearance of the peaks corresponding to the BC box, indicating which residues participated in the interaction with the EloBC complex. In the second study [110], an unfolded conformation with a residual α-helical structure was seen by biophysical analysis

of a synthetic peptide spanning residues 141–192 using analytical size exclusion chromatography, CD, and NMR spectroscopy. Incubating the CTD with a peptide derived from the binding interface of EloC (residues 76–111) did not show a structural change in the CTD, probably because of the use of peptides and not the full proteins [110,117].

In vitro studies have quantified the interaction between segments from the Vif CTD and E3 ligase components, demonstrating high affinity interactions that form a stable complex with a 1:1:1 stoichiometry [69,116,117]. The Vif-mediated E3 complexes were formed *in vitro* without the need for cofactors, as opposed to other SOCS-box motifs containing specificity factors that form complexes with EloBC [119,120]. The dissociation constants (K_d) reported for this interaction ranged from several micromolars to fractions of nanomolars, which may be due to using different segments of Vif and different research methods, as well as different experimental conditions such as ionic strength.

Analysis of the interaction with components of the E3 complex [116] implies very strong binding, $K_d = 0.4$ nM, between Vif CTD 139–192 and EloBC using isothermal titration calorimetry (ITC) and fluorescence polarization (FP). Mutating L145 to alanine caused a nearly 10^4-fold decrease in the binding affinity to 4.6 μM. Another study that used ITC [69] reported a K_d of 2.2 μM for the EloBC complex binding to Vif 139–176 fused to a 56-residue solubility enhancement tag [118]. Differences in the reported dissociation constants (K_d) may stem from either the different experimental conditions or the difference in the Vif fragments used. The unfolded nature of this domain enables a wide range of weak interactions mediated by numerous residues in the Vif CTD [73,94]. Shortening the Vif fragment used may result in loss of important interactions or residues that stabilize the binding conformation and hence result in loss of affinity. Using small segments of Vif, although advantageous, has a significant drawback at the structural level, particularly when it comes to IDPs since structural elements that are important for intermolecular interactions may be truncated. Structural transitions from disorder to α-helical conformation on binding may increase the extent of the interaction between Vif CTD and EloC and/or strengthen the assembly of the overall E3 complex, as one of the multiple binding abilities of IDPs that have been proposed [92].

Different K_d values were observed from the same assay in the case of deuterium exchange monitored by mass spectroscopy [117]. A segment of Vif (135–158) showed an increase of almost 10-fold in the binding affinity to EloBC ($K_d = 0.23$ μM) as compared to that of full-length Vif ($K_d = 1.9$ μM). Moreover, while full-length Vif did not exhibit significant global conformational changes on binding to the EloBC complex, Vif 135–158 dramatically changed its conformation on binding to this complex. The decrease in deuterium exchange of the folded region may have been compensated for by an increase resulting from unfolding of other regions in the full Vif protein. The intrinsically disordered nature of the full CTD of Vif may have weakened the interaction either by steric hindrance or by entropic barriers caused by disordered-to-ordered conformational changes of the IDP. Additional binding partners may be required for binding full-length Vif, such as Cul5. The K_d of 11 μM for Vif CTD 141–192 binding to a peptide derived

from EloC 76–111, without an evident structural change on binding [110], may result from steric hindrance, the absence of other components of the E3 complex, or other binding determinants in EloC. Induced folding of EloC was shown to be required for synergistic folding of both EloC and Vif CTD [69] as was proposed for IDPs [89]. In this case, the CTD would not be able to undergo a structural change in the absence of EloC folding.

8.3.2 Multiple Binding Partners of the PPLP Motif in the Cullin Box

In the cellular SOCS-box proteins, the SOCS motif contains a cullin box downstream to the BC box, which determines the recruitment of a specific cullin in the E3 complex [121] and apparently forms two α-helices as was shown in the crystal structures of several SOCS-box domains [113,115,122]. The putative Vif cullin box slightly resembled the cellular cullin box in the crystal structure with the EloBC complex [68,121], although it did not adopt the same α-helical conformation. Different studies reported contradicting results regarding the involvement of the Vif cullin box in the interaction with Cul5. The cullin box–Cul5 interaction has been shown, by coimmunoprecipitation, to be impaired in Vif mutants L163S and L169S [68], while alanine mutations of ^{161}PPL^{163} to AAA did not result in impaired Cul5 binding [70]. Moreover, a peptide derived from the HCCH Zn^{2+}-binding domain in the absence of the cullin box interacted with Cul5 [65,123]. Cul5 bound to Vif 139–192 in complex with EloBC *in vitro*, as shown by ITC experiments resulting in a K_d of ~670 nM, although no interaction occurred between Cul5 and a complex of Vif 139–154 and EloBC [116]. The same region was implicated in an interaction with EloB by NMR analysis and ITC binding experiments, demonstrating additional binding of the EloBC complex to the cullin box but only when the EloBC complex had already bound the BC box [69]. The exact role of the cullin box in the assembly of the E3 complex is yet unclear, and its elucidation would greatly improve understanding of the mechanism of action of Vif.

The cullin box of Vif has also been shown to be important for numerous interactions with other proteins, particularly through the conserved PPLP motif. Originally this domain was related to the ability of Vif to form oligomers, since deleting residues 151–164, and specifically the PPLP motif, significantly impaired Vif oligomerization [25,26]. Peptides derived from the Vif cullin box were able to inhibit the Vif–Vif interaction, as did other PXP-motif-containing peptides [25]. An opposing study found that alanine mutations of ^{161}PPLP^{164} to AALA did not impair Vif oligomerization [68]. Additional oligomerization regions were assigned using mass spectrometry analysis of cross-linked peptides, indicating regions of dimerization and/or trimerization of Vif [23]. These include residues 169–173, the remaining part of the cullin box following the PPLP motif, which participates in both dimerization and trimerization, while residues 148–157 from the BC box and residues 107–121 from the HCCH Zn^{2+}-binding motif are important for trimerization only. The PPLP (161–164) motif was not indicated as important for oligomerization and is located amid regions determined to participate in this process: 148–157 and 169–173. This fact suggests a possible role for PPLP as a

structural element that mediates the tertiary arrangement of flanking regions that are important for the interactions of Vif such as oligomerization. Vif oligomerization may regulate the interactions it forms since adding Cul5 to Vif complexed with EloBC shifted Vif from its dimeric to monomeric state [116].

The importance of the PPLP motif for interacting with A3G was demonstrated using the alanine mutations of ^{161}PPL^{163} to AAA, which resulted in decreased binding to A3G and A3G degradation [70]. Recently, dominant negative variants of Vif have been described, including mutations in the PPLP motif to AALA or deletion of the cullin-box Δ155–164 [71]. These nonfunctional mutants were unable to bind A3G and interfered with the activity of WT-Vif when expressed in trans, resulting in higher cellular and viral levels of A3G. The viral RT binds Vif at the same PPLP motif [72]. The PPLP motif was reported to be an additional interaction point with the tyrosine kinase, Hck, that inhibited HIV-1 infectivity in a Vif-dependent manner [124]. Peptides bearing the PPLP sequence were able to disrupt this interaction [25].

The Vif PPLP motif, situated in the middle of an IDR, clearly is important for its activity, as evident in its numerous interactions and from the effect conferred on virus infectivity by deleting or mutating these residues. Binding to multiple partners in the same region is one of the basic characteristics of IDPs and is possible due to the dynamic interplay among the different conformations [91–94]. The structural role of the rigid PPLP motif may be based on the secondary-structure-breaking properties of the proline residues. Increased conformational freedom resulting from P to A mutations may affect the flanking regions in Vif.

8.4 THE HCCH Zn^{2+}-BINDING DOMAIN IS IN EQUILIBRIUM WITH DIFFERENT CONFORMATIONS

The HCCH Zn^{2+}-binding motif is central in Vif function and A3G degradation and mediates the interaction with the Cul5 component of the E3 ubiquitin ligase complex [64,65,114]. Each of the H^{108}–C^{114}–C^{133}–H^{139} residues has been shown to be important for interaction with Cul5 and for binding Zn^{2+} [65,125]. The specific arrangement of the HCCH motif and the exact spacing among these residues are also important for Vif function, together with additional conserved residues [66]. These conserved hydrophobic residues, including residues F115, I120, A123, and L124, create a stable hydrophobic interface between Vif and Cul5. This interface is important for A3G degradation, but with no direct binding of the HCCH motif to A3G [65,125].

The Vif HCCH Zn^{2+}-binding motif is conserved in all primate lentiviral Vif proteins but distinct from other known classes of Zn^{2+} finger motifs [65,125]. Zn^{2+} binding to the HCCH motif is important for Vif function as was demonstrated by Zn^{2+} chelation coupled to the loss of Vif activity [126]. Zn^{2+} binding to a peptide derived from Vif residues 101–142 (HCCHp) was shown to be specific and reversible and to induce a conformational change from α-helix to β-sheet with subsequent aggregation [127]. Zn^{2+} binding also increased the affinity between the

Vif HCCHp and Cul5 [123]. Recently, direct binding between Cul5 and a complex of EloBC with Vif 100–154, Vif 100–176, or Vif 100–192 gave K_d of ~90 nM [116]. In contrast to the α-helical conformation previously reported [126] for the unbound HCCHp, two recent studies demonstrated using several biophysical methods that in the absence of Zn^{2+}, the HCCHp mainly adopted a random conformation that fits the pattern of IDPs [123,128]. The possible interplay among different conformations correlates with the ability of HCCHp to form a stable α-helical conformation in the presence of 2,2,2-trifluoroethanol (TFE). The increase in the α-helical content correlated with a decrease in the random coil and β-sheet content [128]. Analysis of the impact of His/Cys mutations on the Zn^{2+}-binding affinities and the conformation of the HCCHp showed that while a C133S mutation and C114/C133 double mutation to serine reduced the Zn^{2+}-binding affinity and increased the aggregation propensity of the peptides, mutants C114S, H139A, and H108A behaved more like the WT-HCCHp with the same degree of Zn^{2+} binding and only a small effect on aggregation as compared to WT-HCCHp [123]. The H127A mutation, which does not participate in Zn^{2+} coordination in the proposed HCCH motif, caused a reduction in the Zn^{2+}-binding affinity and reduced aggregation of Vif compared to the wild type. The mutant peptide adopted an α-helical conformation. These results suggest that Zn^{2+} coordination can occur via more than one set of ligand residues in the Vif HCCH motif. Zn^{2+} binding at the HCCH domain is postulated to play an important structural role in Vif–Cul5 binding. Zn^{2+} binding may stabilize HCCH in the correct conformation for Cul5 binding by positioning the hydrophobic residues at the binding interface [65,125]. In the absence of Cul5, the hydrophobic surface induced by Zn^{2+} can interact with a parallel hydrophobic Vif HCCH domain, which may cause the observed aggregation of the HCCHp.

The unfolded nature of the HCCH domain makes it possible to adopt several conformations, as indicated by several biophysical studies [123,127,128]. A model describing the structural changes of the HCCH domain is presented in Fig. 8.4. The suggested model addresses the contradictions in the reported conformations of the HCCH domain: (i) the unfolded conformation of the HCCH is in equilibrium with a partially folded conformation that may include structural elements from any of the HCCH folds, presumably a partially folded α-helix; (ii) adding TFE stabilized the α-helix conformation, which may be induced by membrane association in a similar manner; (iii) when Zn^{2+} binds the unfolded or partially folded HCCH domain, hydrophobic residues are exposed. These participate in the interaction with Cul5, which may result either in (iv) a stable conformation induced by binding Cul5 or (v) aggregation that includes elements of β-sheet. The disordered conformation represents the basis of this model, enabling the dynamic interchange between the suggested conformations.

8.5 CONCLUDING REMARKS

Vif-isolated domains are useful for biophysical and structural studies of Vif *in vitro*, as the full-length Vif is unavailable at the high concentrations required for such

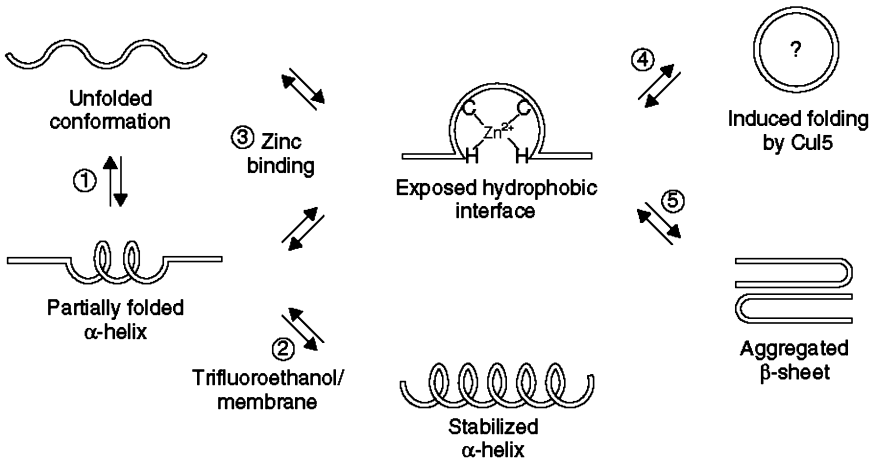


Figure 8.4 A model describing structural changes of the HCCH domain available by the dynamic conformational changes of the intrinsically disordered domain (see Section 8.4 for details).

experiments. At the same time, the effect of other domains on the function and structural organization of the studied region is lost. The involvement of IDRs in the self-assembly of macromolecules demonstrates their importance for the correct interactions among different domains in the protein [96]. Several such examples are present throughout the studies of Vif. Deleting Vif CTD (residues 141–192) impaired the ability of the resultant HCCH-containing Vif construct to bind a Zn^{2+} affinity column [72], although direct binding of Zn^{2+} to an intact HCCH peptide including residues 101–142 has been demonstrated. This suggests that the CTD, which does not bind Zn^{2+} directly, still affects Zn^{2+} binding of the full protein.

Vif is an example of an IDP that has a variety of active sites that can inhibit or enhance each other, providing a pathway for self-regulation, a versatility of actions, and a diversity of interactions that give this protein the potential to be particularly efficient. For technical reasons, this family of proteins is only recently being investigated as appropriate tools are developed for gleaning an insight into their intricate mechanisms of action.

ACKNOWLEDGMENTS

T. H. R. is supported by a fellowship from the ISEF foundation. A. F. is supported by a grant from the US—Israel Binational Science Foundation (BSF).

ABBREVIATIONS

ID	intrinsic disorder
IDPs	intrinsically disordered proteins

IDRs	intrinsically disordered regions
HIV-1	human immunodeficiency virus type 1
Vif	viral infectivity factor
CTD	C-terminal domain
NTD	N-terminal domain
A3G	APOBEC-3G, apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G
CD	circular dichroism
NMR	nuclear magnetic resonance
ITC	isothermal titration calorimetry
FP	fluorescence polarization
PDB	protein data bank
TFE	2,2,2-trifluoroethanol

REFERENCES

1. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 1983;220:868–871.
2. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 1984;224:497–500.
3. Kallings LO. The first postmodern pandemic: 25 years of HIV/AIDS. J Intern Med 2008;263:218–243.
4. Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. Nat Med 2003;9:867–873.
5. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 1998;67:1–25.
6. Freed EO. HIV-1 replication. Somat Cell Mol Genet 2001;26:13–33.
7. Nekhai S, Jeang KT. Transcriptional and post-transcriptional regulation of HIV-1 gene expression: role of cellular factors for Tat and Re. Future Microbiol 2006;1:417–426.
8. Malim MH, Emerman M. HIV-1 accessory proteins—ensuring viral survival in a hostile environment. Cell Host Microbe 2008;3:388–398.
9. Turner BG, Summers MF. Structural biology of HIV. J Mol Biol 1999;285:1–32.
10. Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. Curr Opin Struct Biol 2008;18:203–217.
11. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 1999;17:657–700.
12. Doms RW, Peiper SC. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. Virology 1997;235:179–190.
13. Doms RW. Beyond receptor expression: the influence of receptor conformation, density, and affinity in HIV-1 infection. Virology 2000;276:229–237.

14. Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol 1997;71: 5382–5390.
15. Sodroski J, Goh WC, Rosen C, Tartar A, Portetelle D, Burny A, Haseltine W. Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions. Science 1986;231:1549–1553.
16. Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA. The HIV 'A' (sor) gene product is essential for virus infectivity. Nature 1987;328:728–730.
17. Simon JH, Miller DL, Fouchier RA, Malim MH. Virion incorporation of human immunodeficiency virus type-1 Vif is determined by intracellular expression level and may not be necessary for function. Virology 1998;248:182–187.
18. Kao S, Akari H, Khan MA, Dettenhofer M, Yu XF, Strebel K. Human immunodeficiency virus type 1 Vif is efficiently packaged into virions during productive but not chronic infection. J Virol 2003;77:1131–1140.
19. Khan MA, Aberham C, Kao S, Akari H, Gorelick R, Bour S, Strebel K. Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. J Virol 2001;75:7252–7265.
20. Goncalves J, Jallepalli P, Gabuzda DH. Subcellular localization of the Vif protein of human immunodeficiency virus type 1. J Virol 1994;68:704–712.
21. Goncalves J, Shi B, Yang X, Gabuzda D. Biological activity of human immunodeficiency virus type 1 Vif requires membrane targeting by C-terminal basic domains. J Virol 1995;69:7196–7204.
22. Simon JH, Fouchier RA, Southerling TE, Guerra CB, Grant CK, Malim MH. The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. J Virol 1997;71:5259–5267.
23. Auclair JR, Green KM, Shandilya S, Evans JE, Somasundaran M, Schiffer CA. Mass spectrometry analysis of HIV-1 Vif reveals an increase in ordered structure upon oligomerization in regions necessary for viral infectivity. Proteins 2007;69:270–284.
24. Miller JH, Presnyak V, Smith HC. The dimerization domain of HIV-1 viral infectivity factor Vif is required to block virion incorporation of APOBEC3G. Retrovirology 2007;4:81.
25. Yang B, Gao L, Li L, Lu Z, Fan X, Patel CA, Pomerantz RJ, DuBois GC, Zhang H. Potent suppression of viral infectivity by the peptides that inhibit multimerization of human immunodeficiency virus type 1 (HIV-1) Vif proteins. J Biol Chem 2003;278:6596–6602.
26. Yang S, Sun Y, Zhang H. The multimerization of human immunodeficiency virus type I Vif protein: a requirement for Vif function in the viral life cycle. J Biol Chem 2001;276:4889–4893.
27. Yang X, Goncalves J, Gabuzda D. Phosphorylation of Vif and its role in HIV-1 replication. J Biol Chem 1996;271:10121–10129.
28. Mehle A, Goncalves J, Santa-Marta M, McPike M, Gabuzda D. Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. Genes Dev 2004;18:2861–2866.
29. Yang X, Gabuzda D. Mitogen-activated protein kinase phosphorylates and regulates the HIV-1 Vif protein. J Biol Chem 1998;273:29879–29887.

30. Khan MA, Akari H, Kao S, Aberham C, Davis D, Buckler-White A, Strebel K. Intravirion processing of the human immunodeficiency virus type 1 Vif protein by the viral protease may be correlated with Vif function. J Virol 2002;76:9112–9123.
31. Borman AM, Quillent C, Charneau P, Dauguet C, Clavel F. Human immunodeficiency virus type 1 Vif- mutant particles from restrictive cells: role of Vif in correct particle assembly and infectivity. J Virol 1995;69:2058–2067.
32. Strebel K. HIV accessory genes Vif and Vpu. Adv Pharmacol 2007;55:199–232.
33. Gabuzda DH, Lawrence K, Langhoff E, Terwilliger E, Dorfman T, Haseltine WA, Sodroski J. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J Virol 1992;66:6489–6495.
34. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002;418:646–650.
35. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. DNA deamination mediates innate immunity to retroviral infection. Cell 2003;113:803–809.
36. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 2003;424:99–103.
37. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 2003;424:94–98.
38. Yang B, Chen K, Zhang C, Huang S, Zhang H. Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. J Biol Chem 2007;282: 11667–11675.
39. Kao S, Khan MA, Miyagi E, Plishka R, Buckler-White A, Strebel K. The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. J Virol 2003;77:11398–11407.
40. Kremer M, Schnierle BS. HIV-1 Vif: HIV's weapon against the cellular defense factor APOBEC3G. Curr HIV Res 2005;3:339–344.
41. Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem 2004;279:7792–7798.
42. Wedekind JE, Dance GS, Sowden MP, Smith HC. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet 2003;19:207–216.
43. Wissing S, Galloway NL, Greene WC: HIV-1 Vif versus the APOBEC3 cytidine deaminases: An intracellular duel between pathogen and host restriction factors. Mol Aspects Med 2010;31(5):383–397.
44. Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 2003;9:1398–1403.
45. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science 2003;302: 1056–1060.

46. Shirakawa K, Takaori-Kondo A, Kobayashi M, Tomonaga M, Izumi T, Fukunaga K, Sasada A, Abudu A, Miyauchi Y, Akari H, et al. Ubiquitination of APOBEC3 proteins by the Vif-Cullin5-ElonginB-ElonginC complex. Virology 2006;344(2): 263–266.

47. Yamashita T, Nomaguchi M, Miyake A, Uchiyama T, Adachi A. Status of APOBEC3G/F in cells and progeny virions modulated by Vif determines HIV-1 infectivity. Microbes Infect 2010;12:166–171.

48. Santa-Marta M, da Silva FA, Fonseca AM, Goncalves J. HIV-1 Vif can directly inhibit apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G-mediated cytidine deamination by using a single amino acid interaction and without protein degradation. J Biol Chem 2005;280:8765–8775.

49. Santa-Marta M, Aires da Silva F, Fonseca AM, Rato S, Goncalves J. HIV-1 Vif protein blocks the cytidine deaminase activity of B-cell specific AID in E. coli by a similar mechanism of action. Mol Immunol 2007;44:583–590.

50. Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell 2003;12:591–601.

51. Mercenne G, Bernacchi S, Richer D, Bec G, Henriet S, Paillart JC, Marquet R: HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. Nucleic Acids Res 2010;38:633–646.

52. Mahrour N, Redwine WB, Florens L, Swanson SK, Martin-Brown S, Bradford WD, Staehling-Hampton K, Washburn MP, Conaway RC, Conaway JW. Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases. J Biol Chem 2008;283:8005–8013.

53. Schrofelbauer B, Senger T, Manning G, Landau NR. Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with nonhuman primate APOBEC3G. J Virol 2006;80:5984–5991.

54. Russell RA, Pathak VK. Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. J Virol 2007;81:8201–8210.

55. Pery E, Rajendran KS, Brazier AJ, Gabuzda D. Regulation of APOBEC3 proteins by a novel YXXL motif in human immunodeficiency virus type 1 Vif and simian immunodeficiency virus SIVagm Vif. J Virol 2009;83:2374–2381.

56. Mehle A, Wilson H, Zhang C, Brazier AJ, McPike M, Pery E, Gabuzda D. Identification of an APOBEC3G Binding Site in HIV-1 Vif and Inhibitors of Vif-APOBEC3G Binding. J Virol 2007;81:13235–13241.

57. He Z, Zhang W, Chen G, Xu R, Yu XF. Characterization of conserved motifs in HIV-1 Vif required for APOBEC3G and APOBEC3F interaction. J Mol Biol 2008;381:1000–1011.

58. Dettenhofer M, Cen S, Carlson BA, Kleiman L, Yu XF. Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. J Virol 2000;74:8938–8945.

59. Zhang H, Pomerantz RJ, Dornadula G, Sun Y. Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. J Virol 2000;74:8252–8261.

60. Baraz L, Friedler A, Blumenzweig I, Nussinuv O, Chen N, Steinitz M, Gilon C, Kotler M. Human immunodeficiency virus type 1 Vif-derived peptides inhibit the viral protease and arrest virus production. FEBS Lett 1998;441:419–426.
61. Friedler A, Blumenzweig I, Baraz L, Steinitz M, Kotler M, Gilon C. Peptides derived from HIV-1 Vif: a non-substrate based novel type of HIV-1 protease inhibitors. J Mol Biol 1999;287:93–101.
62. Baraz L, Hutoran M, Blumenzweig I, Katzenellenbogen M, Friedler A, Gilon C, Steinitz M, Kotler M. Human immunodeficiency virus type 1 Vif binds the viral protease by interaction with its N-terminal region. J Gen Virol 2002;83:2225–2230.
63. Izumi T, Takaori-Kondo A, Shirakawa K, Higashitsuji H, Itoh K, Io K, Matsui M, Iwai K, Kondoh H, Sato T, et al. MDM2 is a novel E3 ligase for HIV-1 Vif. Retrovirology 2009;6:1.
64. Luo K, Xiao Z, Ehrlich E, Yu Y, Liu B, Zheng S, Yu XF. Primate lentiviral virion infectivity factors are substrate receptors that assemble with cullin 5-E3 ligase through a HCCH motif to suppress APOBEC3G. Proc Natl Acad Sci USA 2005;102:11444–11449.
65. Mehle A, Thomas ER, Rajendran KS, Gabuzda D. A zinc-binding region in Vif binds Cul5 and determines cullin selection. J Biol Chem 2006;281:17259–17265.
66. Xiao Z, Xiong Y, Zhang W, Tan L, Ehrlich E, Guo D, Yu XF. Characterization of a novel cullin5 binding domain in HIV-1 Vif. J Mol Biol 2007;373:541–550.
67. Simon JH, Sheehy AM, Carpenter EA, Fouchier RA, Malim MH. Mutational analysis of the human immunodeficiency virus type 1 Vif protein. J Virol 1999;73:2675–2681.
68. Stanley BJ, Ehrlich ES, Short L, Yu Y, Xiao Z, Yu XF, Xiong Y. Structural Insight into the HIV Vif SOCS Box and its role in human E3 ubiquitin ligase assembly. J Virol 2008;82:8656–8663.
69. Bergeron JR, Huthoff H, Veselkov DA, Beavil RL, Simpson PJ, Matthews SJ, Malim MH, Sanderson MR: The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS Pathog6:e1000925.
70. Donahue JP, Vetter ML, Mukhtar NA, D'Aquila RT. The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation. Virology 2008;377:49–53.
71. Walker RC Jr, Khan MA, Kao S, Goila-Gaur R, Miyagi E, Strebel K. Identification of dominant negative human immunodeficiency virus type 1 Vif mutants that interfere with the functional inactivation of APOBEC3G by virus-encoded Vif. J Virol 2010;84:5201–5211.
72. Kataropoulou A, Bovolenta C, Belfiore A, Trabatti S, Garbelli A, Porcellini S, Lupo R, Maga G. Mutational analysis of the HIV-1 auxiliary protein Vif identifies independent domains important for the physical and functional interaction with HIV-1 reverse transcriptase. Nucleic Acids Res 2009;37:3660–3669.
73. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18:343–384.
74. Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15:35–41.
75. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 2004;338:1015–1026.

76. Bussell R, Eliezer D Jr. Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein. J Biol Chem 2001;276:45996–46003.
77. Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.
78. Johansson J, Gudmundsson GH, Rottenberg ME, Berndt KD, Agerberth B. Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. J Biol Chem 1998;273:3718–3724.
79. Bryant JE, Lecomte JT, Lee AL, Young GB, Pielak GJ. Protein dynamics in living cells. Biochemistry 2005;44:9275–9279.
80. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, Young GB, Macdonald JM, Pielak GJ. Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. J Am Chem Soc 2008;130:6310–6311.
81. Pielak GJ, Li C, Miklos AC, Schlesinger AP, Slade KM, Wang GF, Zigoneanu IG. Protein nuclear magnetic resonance under physiological conditions. Biochemistry 2009;48:226–234.
82. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.
83. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
84. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.
85. Lise S, Jones DT. Sequence patterns associated with disordered regions in proteins. Proteins 2005;58:144–150.
86. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007;92:1439–1456.
87. Ferron F, Longhi S, Canard B, Karlin D. A practical overview of protein disorder prediction methods. Proteins 2006;65:1–14.
88. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.
89. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.
90. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579:3346–3354.
91. Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem 2002;62:25–49.
92. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41:6573–6582.
93. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 2008;9(Suppl 2):S1.
94. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 2005;272:5129–5148.
95. Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc Natl Acad Sci USA 2007;104:8311–8315.

96. Namba K. Roles of partly unfolded conformations in macromolecular self-assembly. Genes Cells 2001;6:1–12.
97. Leulliot N, Varani G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. Biochemistry 2001;40:7947–7956.
98. Foucault M, Mayol K, Receveur-Brechot V, Bussat MC, Klinguer-Hamour C, Verrier B, Beck A, Haser R, Gouet P, Guillon C. UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. Proteins 2010;78:1441–1456.
99. Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 2006;281:8347–8356.
100. Shojania S, O'Neil JD. Intrinsic disorder and function of the HIV-1 Tat protein. Protein Pept Lett 2010;17:999–1011.
101. DiMattia MA, Watts NR, Stahl SJ, Rader C, Wingfield PT, Stuart DI, Steven AC, Grimes JM. Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element. Proc Natl Acad Sci USA 2010;107:5810–5814.
102. Balbach JJ, Yang J, Weliky DP, Steinbach PJ, Tugarinov V, Anglister J, Tycko R. Probing hydrogen bonds in the antibody-bound HIV-1 gp120V3 loop by solid state NMR REDOR measurements. J Biomol NMR 2000;16:313–327.
103. Stanfield R, Cabezas E, Satterthwait A, Stura E, Profy A, Wilson I. Dual conformations for the HIV-1 gp120V3 loop in complexes with different neutralizing fabs. Structure 1999;7:131–142.
104. Weliky DP, Bennett AE, Zvi A, Anglister J, Steinbach PJ, Tycko R. Solid-state NMR evidence for an antibody-dependent conformation of the V3 loop of HIV-1 gp120. Nat Struct Biol 1999;6:141–145.
105. Goh GK, Dunker AK, Uversky VN. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genomics 2008;9(Suppl 2):S4.
106. Goh GK, Dunker AK, Uversky VN. A comparative analysis of viral matrix proteins using disorder predictors. Virol J 2008;5:126.
107. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008;37:215–246.
108. Reingewertz TH, Shalev DE, Friedler A. Structural disorder in the HIV-1 Vif protein and interaction-dependent gain of structure. Protein Pept Lett 2010;17:988–998.
109. Barraud P, Paillart JC, Marquet R, Tisne C. Advances in the structural understanding of Vif proteins. Curr HIV Res 2008;6:91–99.
110. Reingewertz TH, Benyamini H, Lebendiker M, Shalev DE, Friedler A. The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. Protein Eng Des Sel 2009;22:281–287.
111. Lv W, Liu Z, Jin H, Yu X, Zhang L, Zhang L. Three-dimensional structure of HIV-1 VIF constructed by comparative modeling and the function characterization analyzed by molecular dynamics simulation. Org Biomol Chem 2007;5:617–626.
112. Balaji S, Kalpana R, Shapshak P. Paradigm development: comparative and predictive 3D modeling of HIV-1 Virion Infectivity Factor (Vif). Bioinformation 2006;1:290–309.

113. Stebbins CE, Kaelin WG, Pavletich NP Jr. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science 1999;284:455–461.

114. Yu Y, Xiao Z, Ehrlich ES, Yu X, Yu XF. Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. Genes Dev 2004;18:2867–2872.

115. Bullock AN, Debreczeni JE, Edwards AM, Sundstrom M, Knapp S. Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. Proc Natl Acad Sci USA 2006;103:7637–7642.

116. Wolfe LS, Stanley BJ, Liu C, Eliason WK, Xiong Y. Dissection of the HIV Vif interaction with human E3 ubiquitin ligase. J Virol 2010;84:7135–7139.

117. Marcsisin SR, Engen JR. Molecular Insight into the conformational dynamics of the elongin BC complex and its interaction with HIV-1 Vif. J Mol Biol 2010;402(5):892–904.

118. Zhou P, Lugovskoy AA, Wagner G. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR 2001;20:11–14.

119. Babon JJ, Sabo JK, Soetopo A, Yao S, Bailey MF, Zhang JG, Nicola NA, Norton RS. The SOCS box domain of SOCS3: structure and interaction with the elonginBC-cullin5 ubiquitin ligase. J Mol Biol 2008;381:928–940.

120. Feldman DE, Thulasiraman V, Ferreyra RG, Frydman J. Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. Mol Cell 1999;4:1051–1061.

121. Kamura T, Maenaka K, Kotoshiba S, Matsumoto M, Kohda D, Conaway RC, Conaway JW, Nakayama KI. VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. Genes Dev 2004;18:3055–3065.

122. Bullock AN, Rodriguez MC, Debreczeni JE, Songyang Z, Knapp S. Structure of the SOCS4-ElonginB/C complex reveals a distinct SOCS box interface and the molecular basis for SOCS-dependent EGFR degradation. Structure 2007;15:1493–1504.

123. Giri K, Scott RA, Maynard EL. Molecular structure and biochemical properties of the HCCH-Zn^{2+} site in HIV-1 Vif. Biochemistry 2009;48:7969–7978.

124. Hassaine G, Courcoul M, Bessou G, Barthalay Y, Picard C, Olive D, Collette Y, Vigne R, Decroly E. The tyrosine kinase Hck is an inhibitor of HIV-1 replication counteracted by the viral vif protein. J Biol Chem 2001;276:16885–16893.

125. Xiao Z, Ehrlich E, Yu Y, Luo K, Wang T, Tian C, Yu XF. Assembly of HIV-1 Vif-Cul5 E3 ubiquitin ligase through a novel zinc-binding domain-stabilized hydrophobic interface in Vif. Virology 2006;349:290–299.

126. Xiao Z, Ehrlich E, Luo K, Xiong Y, Yu XF. Zinc chelation inhibits HIV Vif activity and liberates antiviral function of the cytidine deaminase APOBEC3G. Faseb J 2007;21:217–222.

127. Paul I, Cui J, Maynard EL. Zinc binding to the HCCH motif of HIV-1 virion infectivity factor induces a conformational change that mediates protein-protein interactions. Proc Natl Acad Sci USA 2006;103:18475–18480.

128. Giri K, Maynard EL. Conformational analysis of a peptide approximating the HCCH motif in HIV-1 Vif. Biopolymers 2009;92:417–425.

9

ORDER FROM DISORDER: STRUCTURE, FUNCTION, AND DYNAMICS OF THE HIV-1 TRANSACTIVATOR OF TRANSCRIPTION

SHAHEEN SHOJANIA AND JOE D. O'NEIL

9.1 INTRODUCTION

According to the World Health Organization, human immunodeficiency virus type 1 (HIV-1) is the leading infectious killer of humans: 25 million people have lost their lives to HIV-1 since 1981, about 2 million people die annually, and 33.4 million people are currently infected (UNAIDS and WHO, 2009). Highly active antiretroviral therapy (HAART) can suppress but not eradicate the virus because of the existence of a reservoir of quiescent, latently infected cells, the most abundant being resting $CD4^+$ memory T lymphocytes (Chun et al., 2010; Trono et al., 2010). The HIV-1 transactivator of transcription (Tat) plays a central role in provirus transcription activation, and in its absence, the virus cannot replicate. Recently, there has been significant progress in elucidating the molecular details by which this intrinsically disordered protein (IDP) interacts with host proteins to activate transcription, and with this understanding, a possible new route to suppressing the virus is emerging. In this chapter, the role of Tat in the HIV-1 life cycle is discussed, with emphasis on the dynamic behavior of this IDP and its implications for therapeutics.

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

9.2 THE HUMAN IMMUNODEFICIENCY VIRUS

Despite some disagreements in naming and credit for discovery, HIV has been understood for more than 25 years to be the causative agent of acquired immunodeficiency syndrome (AIDS) (Barre-Sinoussi et al., 1983; Popovic et al., 1984; Coffin et al., 1986). HIV is a *Lentivirus*, a genus of retroviruses, displaying a characteristically long latency period (Turner and Summers, 1999; Emini, 2002). The retroviral genome consists of two copies of positive-sense single-stranded RNA (Fig. 9.1), which are reverse transcribed and then integrated into the DNA of an infected cell (Cullen, 1991). The major cells infected and depleted by HIV are the $CD4^+$ T lymphocytes that play a critical role in immune response (Kingsman and Kingsman, 1996). Macrophages, hematopoietic stem cells, and microglial cells can also become infected (Kingsman and Kingsman, 1996; Emini, 2002) and may serve as a source of latent infection (Carter et al., 2010).

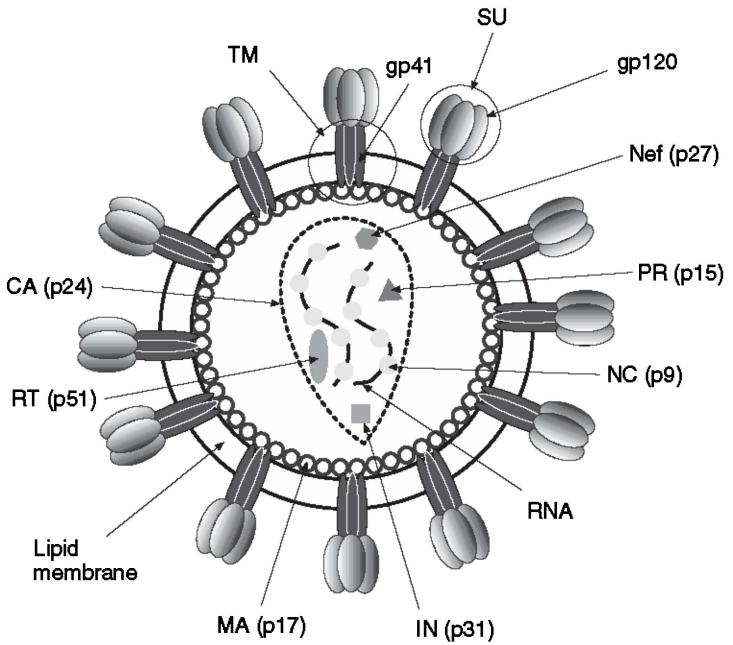


Figure 9.1 The characteristic features of the HIV-1 virion showing the conical capsid composed of ~1500 copies of the capsid protein (CA or p24). The core contains the viral diploid single-stranded RNA (Haseltine, 1991), nucleocapsid protein (NC or p9), protease (PR), integrase (IN), negative factor (Nef), integrase (IN or p31), and reverse transcriptase (RT or p51). The capsid core is enclosed in a protein matrix (MA or p17). The matrix is enveloped by a lipid bilayer derived from the host cell along with some host cell proteins. The surface unit (SU) is composed of trimers of glycoprotein 120 (gp120), which are anchored to the envelope via the transmembrane (TM) complex consisting of a trimer of glycoproteins 41 (gp41).

9.2.1 HIV-1 Virus Structure

The mature HIV-1 virion, depicted in Fig. 9.1, is distinguished by its cone-shaped capsid core (Turner and Summers, 1999) that is surrounded by a lipid bilayer derived from its host cell. The surface of the virion is covered with glycoproteins (gps) involved in binding to host cell surface receptors, the most common of which is the immunoglobulin-like protein CD4 (Frankel and Young, 1998). The surface unit (SU) is made of a trimeric complex of gp120 and is attached to the virion through a transmembrane (TM) trimeric complex of gp41 (Kwong et al., 1998; Zwick et al., 2004). The TM complex may also aid in the fusion of the virion and host cell initiated through an N-terminal glycine-rich peptide (Frankel and Young, 1998). Beneath the lipid bilayer is the matrix (MA) complex of proteins (p17) (Chapter 6). The conical capsid is a complex of approximately 1500 capsid (CA) proteins (p24) (Garzon et al., 2004). The capsid contains the viral genome (two single strands of RNA) along with several viral enzymes necessary for early replication steps (Haseltine, 1991). The inner core contains the viral enzyme reverse transcriptase (RT or p51) that processes the viral RNA into viral DNA inside the host cell, integrase (IN or p31) that inserts the viral DNA into the host DNA in the nucleus, the nucleocapsid (NC or p9) that functions to deliver unspliced RNA for assembly of new virions, protease (PR) that cleaves viral polyproteins into their functional units, and the negative factor (Nef) that is primarily involved in the downregulation of CD4 surface expression (Aiken et al., 1994) but it may also serve to enhance the envelope (Env) incorporation into new virions and facilitate viral budding and release (Frankel and Young, 1998).

There are nine open reading frames in the HIV-1 genome (Frankel and Young, 1998) as depicted in Fig. 9.2. The group-specific antigen (*gag*) gene encodes a polyprotein (Gag) that contains the major structural components of the virus core (matrix, capsid, and NC). The *pol* gene encodes a polyprotein (Pol) containing RT, IN and PR. PR (also contained in the mature virion) cleaves the Gag and Gag-Pol polyproteins into the individual protein units (Turner and Summers, 1999). The envelope (*env*) gene encodes the Env proteins gp120 and gp41, which make up

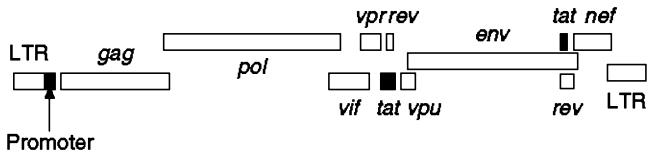


Figure 9.2 Open reading frames of the HIV-1 genome. The HIV-1 long terminal repeat (LTR) has an inducible promoter (Karn, 1999) followed by the genes for group-specific antigen (gag) encoding a polyprotein containing the major structural components of the matrix, capsid, and nucleocapsid complexes; polyprotein (pol) encoding another polyprotein containing the viral enzymes protease, reverse transcriptase, and integrase; accessory proteins, namely, viral infectivity factor (vif), viral protein R (vpr), viral protein U (vpu), and negative factor (nef); envelope (env) encoding the surface and transmembrane glycoproteins; and the regulatory proteins, namely, transactivator of transcription (tat) and regular expression of virus (rev). Both the regulatory proteins are encoded by two exons.

the SU and TM complexes (Frankel and Young, 1998). The six additional reading frames are the genes for the regulatory proteins transactivator of transcription (Tat) and regular expression of virus (Rev), as well as the accessory proteins: viral infectivity factor (Vif) (Chapter 8), viral protein U (Vpu), viral protein R (Vpr), and negative factor (Nef) (Turner and Summers, 1999).

9.2.2 HIV-1 Life Cycle

An outline of the HIV-1 life cycle is depicted in Fig. 9.3 for infection of a $CD4^+$ T lymphocyte. Upon gp120 recognition and binding to the cell surface receptor (in this case CD4), the virion attaches itself to the cell. Additional interactions between

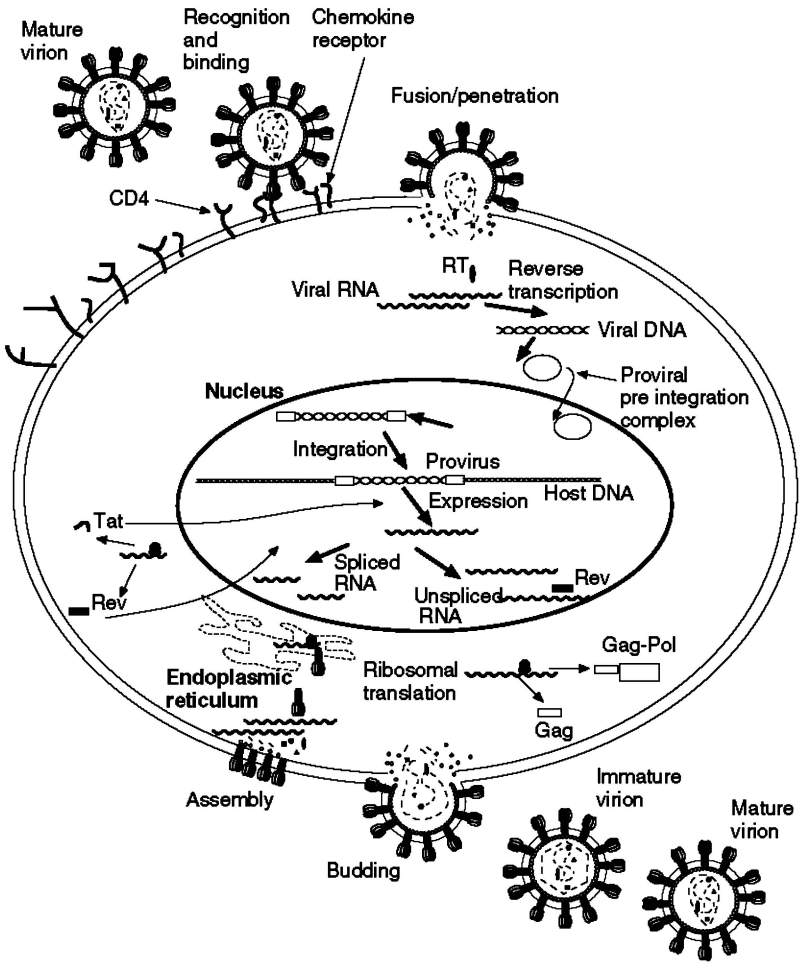


Figure 9.3 The general features of the HIV-1 life cycle on infection of a $CD4^+$ T lymphocyte. See text for details.

host cell surface chemokine receptors (CXCR4 or CCR5) induce a conformational change in the CD4 receptor that allows for fusion of the viral envelope with the cell plasma membrane (Turner and Summers, 1999; Freed, 2001). Once the diploid viral RNA is released into the cell, it is replicated by the viral enzyme RT yielding double-stranded DNA. A preintegration complex results from the association of the viral DNA with a complex that contains at least IN, matrix protein, and RT (Turner and Summers, 1999). The preintegration complex crosses the nuclear membrane and is integrated into the host DNA by IN and becomes the provirus.

Expression of the virus results in a number of mRNAs of varying length. These viral RNAs fall into three categories: unspliced, partially spliced, and multiply spliced (Freed, 2001). The full-length unspliced RNAs can exit the nucleus for translation or assemble at the cell membrane for packaging into a new virion. Early in infection, multiply spliced mRNAs are transported to the cytoplasm and translated into the viral regulatory proteins Tat and Rev along with the accessory protein Nef. The Rev protein ensures that the full-length viral transcripts leave the nucleus and enter the cytoplasm for Gag and Gag-Pol synthesis and assembly of new virions (Turner and Summers, 1999). The viral regulatory protein Tat is essential for virus expression, and its function is described in Section 9.4.

9.3 INTRINSICALLY DISORDERED PROTEINS

A current view of proteins suggests that functional proteins inhabit a wide spectrum from tightly folded to completely disordered including such diverse states as molten globules (Ohgushi and Wada, 1983) and amyloid fibrils (Dobson, 2004; Uversky et al., 2005; Maji et al., 2009). Biological activity nearly always involves changes in protein conformation that can range from subpicosecond atomic vibrations to multisecond, 100-Å refolding events (Henzler-Wildman and Kern, 2007; Baldwin and Kay, 2009). Recent bioinformatics analyses suggest that up to 30% of the eukaryotic proteome is intrinsically disordered under physiological conditions (Chen et al., 2006); that is, in the absence of binding partners, these proteins, or protein segments, do not fold into a stable conformation but rather exist in a more or less restricted ensemble of conformations determined by the amino acid sequence. Although tremendous resources have been expended to understand the folded end of the protein structure spectrum, we know much less about IDPs and the key role that disorder plays in effecting biological function. This has significant implications including the wide assumption that *rational* drug design is possible only for proteins with known high-resolution structures and the lack of tools and concepts available to quantify dynamic proteins on which IDP-directed drug design can be effected.

9.3.1 The Role of Disorder in Function

Early insight into IDP function came from its classification into six broad categories (Tompa, 2002, 2005 Tompa and Csermely, 2004): *effectors*, proteins that activate,

or more commonly inhibit, activities such as transcription initiation; *assemblers* of macromolecular complexes such as ribosomes; *chaperones* that provide a protective function to proteins and RNA; *display sites* for posttranslational modification; *scavengers* of small ligands, such as calcium phosphate by casein; and *entropic chains* that can be dynamic linkers or regulators of access to binding sites, active sites, and pores (Brown and Hoh, 1997; Hoh, 1998; Dunker et al., 2001; Tompa, 2002; Uversky, 2002; Denning et al., 2003; Rout et al., 2003; Mukhopadhyay et al., 2004; Tompa and Csermely, 2004; Tompa, 2005). All these categories of protein, with the exception of the entropic chains, involve some form of molecular recognition, a key element in drug design. Further insight came with the realization that the five categories of IDP can be classified based on whether their interactions with a binding partner are transient (display sites and chaperones) or permanent (effectors, scavengers, and assemblers) (Tompa, 2005).

Several advantages have been proposed for disordered proteins over folded proteins (Schrader et al., 2009; Tsvetkov et al., 2009; Radivojac et al., 2010). IDPs have significantly greater surface areas than folded proteins of the same length, endowing them with the ability to bind multiple partners and to have multiple functions (Bussell and Eliezer, 2001; Uversky et al., 2005; Hegde and Rao, 2007). This binding promiscuity, which can modulate the activity of different targets, has been observed for several IDPs (Dunker et al., 2001; Tompa, 2002; Radivojac et al., 2006) and can have opposing effects on the same target (Tompa, 2002). However, in all the categories, except for the entropic chains, interactions between a disordered protein or segment and its target typically result in some degree of disorder-to-order transition (Tompa, 2002). The large surface areas of disordered proteins can contribute to high specificity of binding, but the large entropy loss on folding will, in principle, reduce binding affinity (Demchenko, 2001). Surprisingly, a survey of the binding affinities of folded and IDPs shows that the range of binding affinities is remarkably similar between the two classes (Huang and Liu, 2009). High affinity binding in IDPs may be achieved in part by the counteracting solvent entropy gain that occurs on burial of a large hydrophobic surface. It has also been suggested that decoupling folding and binding permits faster dissociation of high affinity IDPs than is possible for folded proteins (Tompa, 2002; Huang and Liu, 2009). In this regard, recent dynamics analyses of several proteins (Henzler-Wildman and Kern, 2007; Baldwin and Kay, 2009) and RNA (Zhang et al., 2007) suggest that many biomolecular interactions entail conformational selection of a target that samples the bound state in its unbound ensemble. However, more work is needed to dissect the multiple enthalpic and entropic contributions that explain IDP folding transitions and expression of biological activity (Demchenko, 2001).

Recent analysis supports earlier observations that IDPs are particularly abundant in processes involving the binding of DNA and RNA in the regulation of cell cycle, transcription, differentiation, and growth (Tompa, 2002; Dunker et al., 2008). Genome analysis suggests that disordered segments of 30 or more consecutive residues occur in 2.0% of archaean, 4.2% of eubacterial, and 33.0% of eukaryotic proteins (Ward et al., 2004). These analyses also indicate that viruses have the highest fraction of proteins containing a disordered segment (37%)

compared to bacteria (30%), eukaryotes (21%), and archaea (19%) (Chen et al., 2006). Furthermore, it appears that protein disorder is important in cell signaling or regulation only in eukaryotes and viruses (Chen et al., 2006).

Most viruses have very simple genomes that encode only a small number of proteins and therefore have limited biological machinery. Viruses must therefore "hijack" the infected host cell machinery in order to replicate. One solution to this problem is multifunctional viral proteins capable of interacting with multiple host proteins. A recent survey of interactions between human proteins and the proteins of human pathogens, including HIV-1, indicates that viruses and bacteria mainly manipulate host cell cycle, apoptosis, immune response, and nuclear transport (Dyer et al., 2008). Currently, the HIV-1 Human Protein Interaction Database (Ptak et al., 2008; Fu et al., 2009; Pinney et al., 2009) at the National Center for Biological Interaction (NCBI) lists 1443 interactions between host proteins and the 18 proteins of HIV-1. This type of binding promiscuity has sometimes been referred to as *moonlighting* (Jeffery, 2009), and IDPs appear well designed to effectively interact with multiple host proteins to control host cell physiology and replicate new virions.

9.3.2 Disorder Prediction

A number of algorithms have been developed in recent years to predict the disordered segments of proteins based on amino acid properties such as net charge, hydropathy, secondary structure propensity, and their frequency of occurrence throughout a protein (Romero et al., 1997, 2001; Li et al., 1999; Obradovic et al., 2003; Peng et al., 2005; Yang et al., 2005; Dosztányi et al., 2005a,b). Low sequence complexity is often an indicator of protein disorder (i.e., low variability of the 20 amino acids within a segment of the protein and repetition of amino acids) (Uversky et al., 2005). Disordered proteins often exhibit a compositional bias against bulky or nonpolar amino acids (Dyson and Wright, 2005) and display a high content of polar or charged residues (Uversky et al., 2005). Gly and Ala are often found to be present in higher proportions because their small side chains favor a flexible backbone (Dyson and Wright, 2005). Recently, progress has been made in predicting those regions of IDPs that undergo a disorder-to-order transition on binding to partners. Such regions are sometimes observed in disorder predictions as regions of intermediate order flanked by disordered regions (Cheng et al., 2006, 2007). Regions that undergo a disorder-to-order transition are sometimes referred to as *molecular recognition features* (MoRFs). Detection of MoRFs in IDPs might become an important step in the design of small molecule inhibitors of protein and nucleic acid binding (Metallo, 2010) because IDPs are highly represented in the human disease network (Cheng et al., 2006; Midic et al., 2009).

Disorder predictions using VLXT and VL3 algorithms of the web server program PONDR® (Romero et al., 1997, 2001; Dunker et al., 1998; Li et al., 1999; Dunker and Obradovic, 2001) of the proteins of the HIV-1 genome (Goh et al., 2008) identified several HIV-1 proteins with varying degrees of disorder. Goh et al. suggest that the predicted disorder in matrix (MA) and capsid (CA) proteins is high, as well

as for the accessory protein Vpr. Nef and IN are suggested to be moderately disordered (26%), and Tat is found to be entirely (100%) disordered by their analyses. Rev is omitted from the table of results in (Goh et al., 2008), but it too is predicted to be disordered (98% by VLXT and 100% by VL3 algorithms). The PONDR VLXT predictions for the HIV-1 proteins are shown in Fig. 9.4 for the three polyproteins (Gag, Pol, and Env) and Fig. 9.5 for the 15 individual protein units.

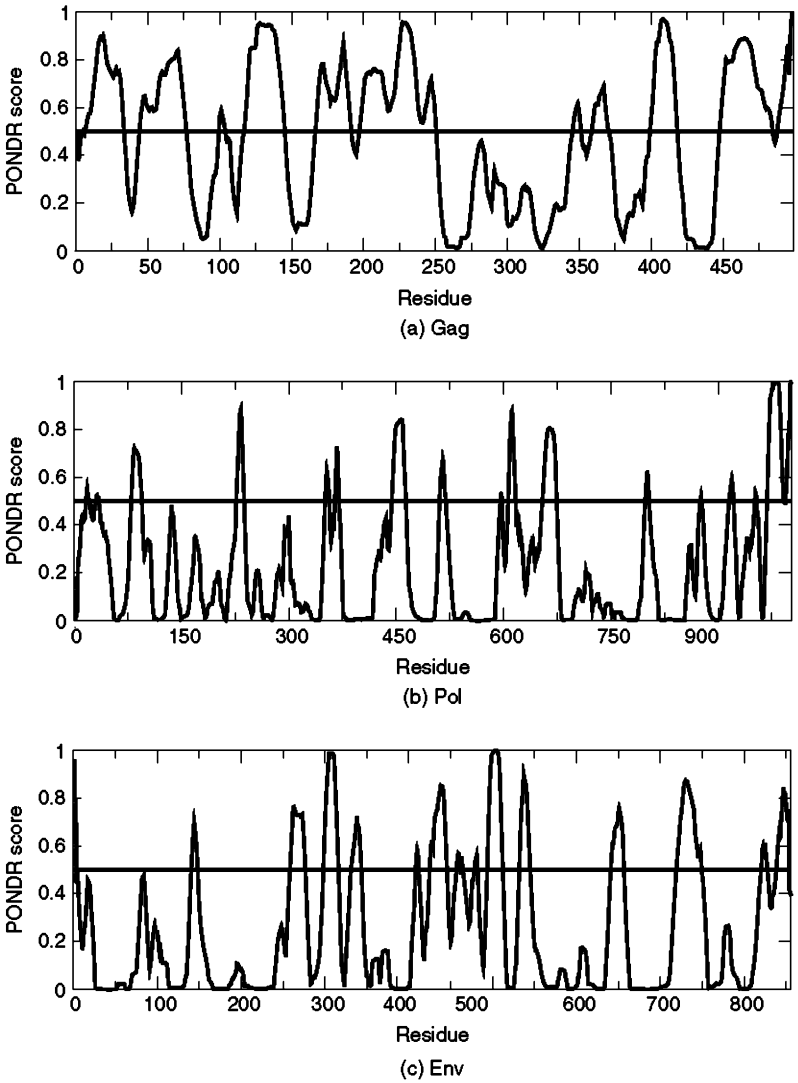


Figure 9.4 Disorder predictions using the PONDR® VLXT algorithm for the HIV-1 polyproteins. Scores >0.5 over a region of several residues suggest disorder in the segment.

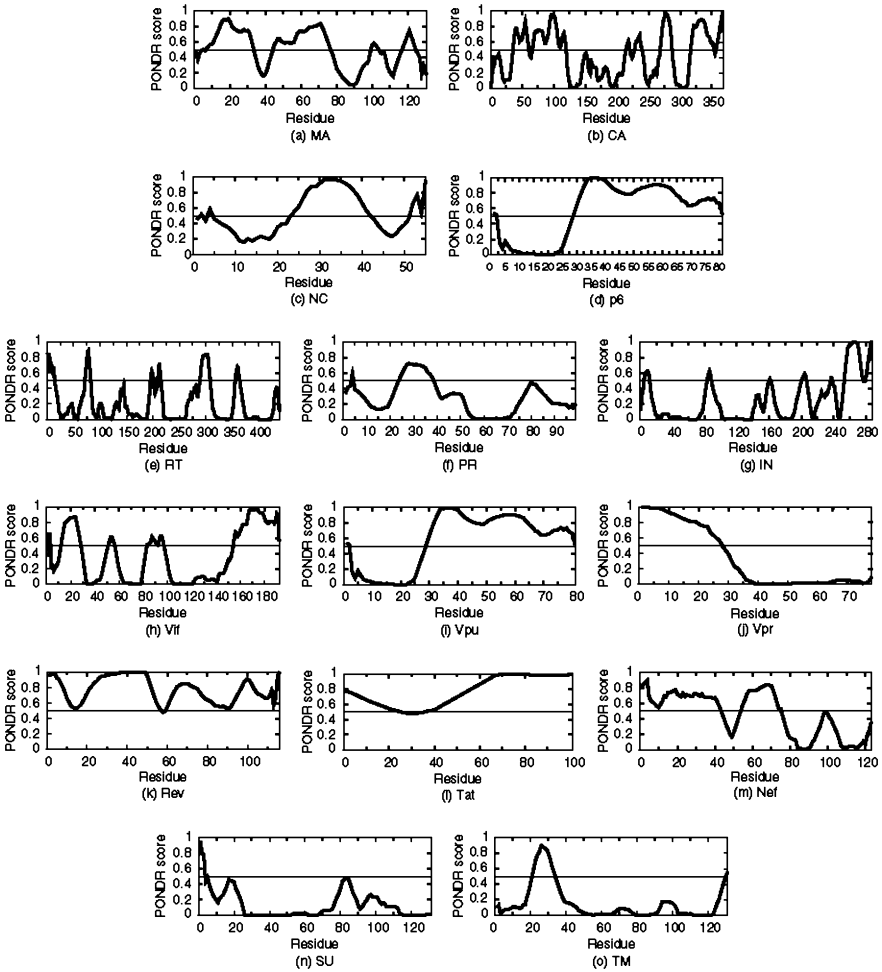


Figure 9.5 Disorder predictions using the PONDR® VLXT algorithm for the HIV-1 proteins. Scores >0.5 over a region of several residues suggest disorder in the segment.

9.4 THE HIV-1 TRANSACTIVATOR OF TRANSCRIPTION

9.4.1 Role of Tat in Transcription Activation

The transactivator of transcription (Tat) is a regulatory protein found in all lentiviruses including those of primates, sheep, cattle, horses, goats, and cats (Clements and Zink, 1996; Coffin et al., 1997). Although sequence homology among the Tat proteins from different lentiviruses is low, the function is highly conserved (Clements and Zink, 1996).

Transcription of the HIV-1 genome is regulated by promoter-proximal pausing of RNA polymerase II (RNAP II), a feature it shares with approximately 30% of

all human genes (Guenther et al., 2007; Core and Lis, 2008). It has long been recognized that transcription activation is accomplished by phosphorylation at Ser-2 of the carboxy terminal domain of RNAP II by the positive transcription elongation factor b (P-TEFb) (Wei et al., 1998). Specificity for the HIV-1 provirus is determined by the transactivator of transcsitpion (Tat) protein that binds to both P-TEFb and a stem loop structure in the nascent RNA referred to as the *transactivation response* (TAR) element. The TAR element is located downstream of the HIV-1 long terminal repeat (LTR) and spans nucleotides +1 to +59 of the nascent RNA. In addition, P-TEFb, a heterodimeric complex of a regulatory Cyclin T and cyclin-dependent kinase 9 (CDK9), also phosphorylates the components of negative transcription elongation factor (N-TEF) and the transcription elongation factor Spt5 (Yamaguchi et al., 1999; Bourgeois et al., 2002; Kim et al., 2002), relieving their inhibition.

Over the past year, significant new details have been discovered that explain why Tat is such a powerful activator of HIV-1 transcription. Because P-TEFb can activate transcription of a vast array of genes, it is highly regulated by both positive and negative regulators, and it has now been revealed that Tat can recruit both inactive and active P-TEFb to the LTR (Barboric and Lenasi, 2010; D'Orso and Frankel, 2010; Sobhian et al., 2010). Two groups have shown that Tat and transcription cofactor AF4 assemble a large, multifunctional transcription elongation complex composed of P-TEFb, PAF1, CDC73, EAF1, ELL2, AF1, AF9, AF4, and ENL (He et al., 2010; Sobhian et al., 2010) (Fig. 9.6) that permits multiple levels of elongation regulation. For example, ELL2 promotes elongation by preventing RNAP II from sliding backward a few nucleotides (Sims et al., 2004). In addition, Tat can also bind to 7SK snRNP-bound and inactive P-TEFb, forming a stress-resistant particle also containing Hexim1, LARP7, SART3, and the 7SK-capping enzyme MePCE (Sobhian et al., 2010). This complex appears to be able to bind to the HIV-1 LTR *before* TAR is expressed, suggesting that P-TEFb is activated by the displacement of the 7SK snRNA by TAR (D'Orso and Frankel, 2010). Recent results suggest that Tat activates P-TEFb by displacing Hexim1 (hexamethylene bisacetamide-inducible protein 1) from its Cyclin T1 binding site (Schulte et al., 2005; Tahirov et al., 2010), but it may also displace Hexim1 from the 7SK snRNP (Barboric and Lenasi, 2010). Tat itself is also regulated as its affinity for TAR is modulated through Tat acetylation by histone acetyl transferase (HAT) (Mujtaba et al., 2002; Bannwarth and Gatignol, 2005).

9.4.2 Other Biological Functions of Tat

In addition to its role as a transcriptional regulator of HIV-1 gene expression, Tat has been implicated in a number of intracellular and extracellular activities unrelated to transcription activation, including supporting endothelial cell proliferation and contributing to the development of Kaposi's sarcoma (Ensoli et al., 1990; Albini et al., 1996a,b), inducing apoptosis of T cells (Goldstein, 1996), inducing cell death of neurons (Nath et al., 1996; Pocernich et al., 2005), decreasing expression of tight junction proteins (András et al., 2003), disrupting

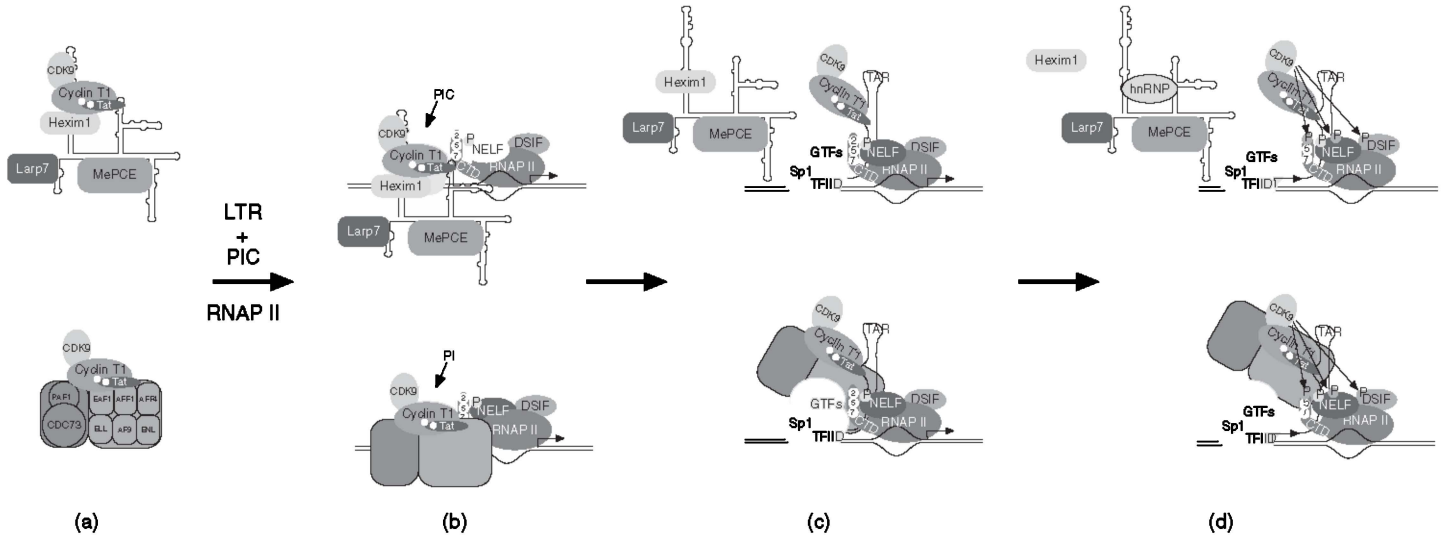


Figure 9.6 Models of transcription activation (Barboric and Lenasi, 2010; D'Orso and Frankel, 2010; Sobhian et al., 2010). P-TEFb can be recruited in multiple active and inactive forms. (a) Top: inactive Tat-P-TEFb (Cyclin T1-CDK9) is shown in complex with 7SK snRNP (Hexim1, Larp7, 7SK snRNA) bound to it through Cyclin T1 and Hexim1 interactions; bottom: active Tat-P-TEFb and AF4 form a large transactivation complex including PAF1, CDC73, and ELL. (b) The Tat-P-TEFb complexes recruited to the preinitiation complex (PIC) of RNAP II and the core promoter complex (Sp1, TFIID, and GTFs). The Tat-P-TEFb 7SK snRNP (top) and AF4 (bottom) complexed with the PIC assembly remain in a paused state until TAR formation. (c) Top: Tat recognizes and binds to a UCU stem loop bulge on the TAR element, thus completely disrupting the Hexim1–Cyclin T1 interaction, dissociating the 7SK snRNP, and activating the P-TEFb complex; bottom: the P-TEFb-AF4 complex is more active and efficient than P-TEFb alone, but specificity for the HIV-1 provirus results from the Tat–TAR interaction. (d) CDK9 of P-TEFb phosphorylates Ser2 in the CTD of RNAPII, DSIF (Spt4-Spt5), and NELF. The hyperphosphorylated RNAP II is able to continue synthesizing the full HIV-1 transcript. Hexim1 may completely dissociate from the rest of the 7SK RNP complex because it is not stably bound. (*See insert for color representation of the figure.*)

the blood-brain barrier (BBB) (Banks et al., 2005), and inducing oxidative stress (Westendorp et al., 1995; Pocernich et al., 2005). Tat may also be involved in derepression of heterochromatin, in transcription initiation (Pumfery et al., 2003), and in reverse transcription (Guo et al., 2003). Absence of Tat and low levels of Tat-associated host factors (CDK9 and Cyclin T1) in resting $CD4^+$ T cells are all implicated in the latency of the virus (Lassen et al., 2004). In general, the pathological activities of Tat contribute to both immune and nonimmune dysfunction, resulting in an overall increase in the burden of the viral infection. More recently, Tat has been shown to regulate capping of HIV-1 mRNA (Chiu et al., 2002), interact with Dicer and suppress the production of small interfering RNA (Bennasser and Jeang, 2006), and act as a nucleic acid chaperone (Kuciak et al., 2008).

A recent proteomics study to determine the range of cellular targets of Tat within the nucleus of Jurkat T cells revealed 183 interacting proteins, 90% of which had not been previously reported (Gautier et al., 2009). The identified nuclear targets covered a range of biological processes, namely, transcription, RNA processing, translation, nuclear organization, cell cycle, DNA replication, and signaling, with transcription being the most highly represented process (39%). The range of processes covered by the nuclear interaction targets emphasizes the multiplicity of binding partners of this small IDP.

9.4.3 Tat Amino Acid Sequence and Properties

The HIV-1 Tat protein is a 101-residue RNA-binding protein encoded by two exons and expressed during the early stages of viral infection (Liang and Wainberg, 2002).The first *tat* exon defines amino acids 1–72 (shown in Fig. 9.7). Several early reports suggested that the 72-residue segment of Tat can activate transcription with the same proficiency as the full-length protein (Garcia et al., 1988; Kuppuswamy et al., 1989; Jeang et al., 1999; Smith et al., 2003). Recently however, HIV-1 expressing the full-length 101-residue protein was shown to have enhanced NFκB

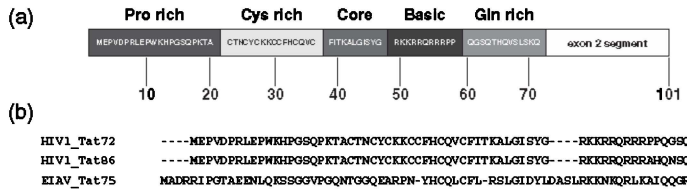


Figure 9.7 (a) The HIV-1 Tat sequence (BH10 isolate) encoded by exon 1. The 72-residue segment encompasses an N-terminal proline-rich region (1–21) containing the only three acidic residues, a cysteine-rich region (22–37), a core (38–47), a basic region (48–57), and a Gln-rich segment (58–72). Residues 73–101 are encoded by exon 2. (b) Sequence alignment of the first exon-encoded region of HIV-1 Tat_{1-72} of the Tat BH10 isolate, the 86-residue Tat (HXB2 isolate; Swiss-Prot P04608) (Tahirov et al., 2010), and the 75-residue Tat protein from EIAV (Swiss-Prot accession number 20920) (Anand et al., 2008).

activation, transactivation, and replication in T cells compared to the exon 1 product (Mahlknecht et al., 2008), strongly suggesting that future work on the protein includes the segment from exon 2. There are several laboratory strains of the Tat protein with 86 residues that may originate from the HXB2 strain (subtype B) commonly found in Europe and North America (Opi et al., 2004). These 86-residue variants are not found in natural viral isolates (Jeang, 1996). It has been suggested that the 86-residue form of Tat was a consequence of tissue culture passaging, and a single nucleotide correction in the laboratory genomes yielded the expected 101-residue protein from the Tat coding frame (Neuveut and Jeang, 1996).

The product of the first Tat exon can be divided into several regions (Fig. 9.7): an acidic and proline-rich N-terminus (1–21), a cysteine-rich region (22–37), a core (38–47), a basic region (48–57), and a Gln-rich segment (58–72) (Derse et al., 1991). Residues 1–24 form the coactivator and acetyltransferase CBP (CREB binding protein) KIX domain binding site (Vendel and Lumb, 2003). Cyclin T1 is now known to interact with the Cys-rich region of Tat (Bieniasz et al., 1998; Tahirov et al., 2010). Mutation of any one of the cysteine residues in the protein, except for C31, results in loss of transactivation (Jeang et al., 1999). C31 is common in most major subtypes of HIV-1, but it is not present in the HIV-1 subtype C strain (which has serine in its place), which accounts for more than 50% of the infections worldwide (Campbell et al., 2007). The end of the Cys-rich region and the core are involved in mitochondrial apoptosis of bystander noninfected cells through their ability to bind tubulin and prevent its depolymerization (Chen et al., 2002). The basic region is critical for TAR RNA binding (Weeks et al., 1990; Anand et al., 2008) and nuclear localization; peptide fragments corresponding to the region of Tat spanning Tyr-47 and Arg-57 have been used to transport a large variety of materials including proteins, DNA, drugs, imaging agents, liposomes, and nanoparticles across cell and nuclear membranes (Gupta et al., 2005). The Gln-rich region has been implicated in mitochondrial apoptosis of T cells (Campbell et al., 2004).

The second tat exon defines residues 73–101 and encodes an RGD motif that may mediate Tat binding to cell surface integrins (Avraham et al., 2004). The function of the second exon-encoded polypeptide has, thus far, been difficult to determine (Guo et al., 2003; Smith et al., 2003). Studies have shown that the peptide encoded by the second exon is involved in repressing expression of the major histocompatibility complex (MHC) class I molecules whose presence at the cell surface serves as targets for cytotoxic T lymphocytes (Howcroft et al., 1993; Carroll et al., 1998; Weissman et al., 1998). This repressive function from the exon-2-encoded polypeptide may contribute to HIV-infected cells escaping an immune response (Carroll et al., 1998; Weissman et al., 1998). Recently, the second exon product has been implicated in major changes to the T-cell cytoskeleton, chemotaxis, and migration and the downregulation of several cell surface receptors (Lopez-Huertas et al., 2010). The identification of different biological functions with short amino acid sequences in the protein highlights the "moonlighting" advantage of IDPs over folded proteins.

The Tat amino acid sequence has a low overall hydrophobicity and a high net positive charge, and analyses by several disorder prediction algorithms suggest that it is intrinsically disordered (Goh et al., 2008) with a potential for order in the cysteine region. Early CD spectropolarimetry experiments suggested a lack of secondary structural elements but showed minor conformational changes in the presence of zinc (Vendel and Lumb, 2003).

9.5 STRUCTURAL BIOLOGY OF Tat

9.5.1 Preparation of Reduced Tat

There have been several attempts to determine the solution conformations of Tat and its segments, both alone and in complexes, and these were recently reviewed (Shojania and O'Neil, 2010). All previous homonuclear NMR studies of Tat showed that amide proton chemical shifts of the protein are within the range characteristic of disordered proteins (Gregoire et al., 2001; Peloponese et al., 2000; Freund et al., 1995; Shojania and O'Neil, 2006). However, it is challenging to obtain a solution of monomeric reduced protein at concentrations amenable to NMR spectroscopy because the readily oxidized Cys-rich region leads to mixtures of soluble and insoluble aggregates. Tat regulates transcription in the interior of a cell that is a reduced environment where most proteins do not contain intramolecular disulfide bonds. In the rare cases where intracellular disulfides have been observed, they are usually involved in redox cycling, are transiently formed, and are not required to maintain the structure of the protein (Mallick et al., 2002). In agreement with this, evidence suggests that the reduced Tat monomer is the form of the protein that is active in transactivation (Siddappa et al., 2006). Although a number of condensed globular protein structures of Tat have been published, it remains possible that some of these contain disulfide-cross-linked monomers or multimers and that the intrinsic folding propensity of the polypeptide can only be revealed in strictly anaerobic reducing conditions.

Preparation of fully reduced HIV-1 Tat_{1-72} uniformly labeled with ^{13}C and ^{15}N isotopes for multidimensional heteronuclear NMR studies involved bacterial expression in *Escherichia coli* and extraction under denaturing conditions with guanidine hydrochloride in the presence of the strong reducing agent tris(carboxyethyl)phosphine (TCEP). Tat_{1-72} was purified by immobilized metal affinity chromatography followed by the removal of the denaturants and reducing agent by dialysis (Shojania and O'Neil, 2006; Shojania et al., 2010). All NMR experiments were done at pH 4.1 in order to maintain the protein in a reduced monomeric state, and the implications of this are discussed below.

9.5.2 Multinuclear NMR Studies of Tat Dynamics

Multinuclear labeling of Tat was essential to obtaining correct chemical shift assignments of backbone resonances and for high resolution dynamics experiments

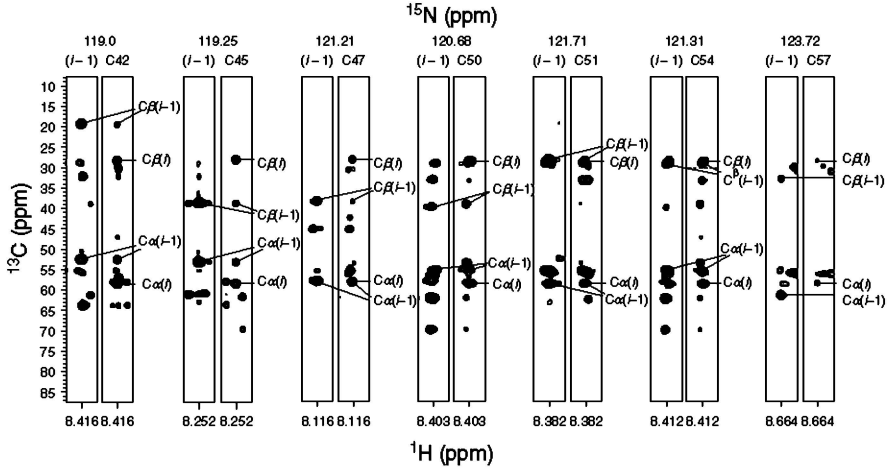


Figure 9.8 Strip plots extracted from three-dimensional, amide-detected, heteronuclear NMR experiments for the Cys-rich region of $^{13}C/^{15}N$-labeled His-tagged Tat_{1-72}. Inter- and intraresidual correlations are obtained from an HNCACB (Wittekind and Mueller, 1993) spectrum correlating $H^N(i)$ and $N(i)$ with $C\alpha(i)$, $C\alpha(i-1)$, $C\beta(i)$, and $C\beta(i-1)$ resonances and a CBCA(CO)NH (Grzesiek and Bax, 1992) spectrum correlating $H^N(i)$ and $N(i)$ with $C\alpha(i-1)$ and $C\beta(i-1)$ resonances. Indicated are the chemical shifts of the $C\alpha$ and $C\beta$ resonances for the seven Cys residues, C42, C45, C47, C50, C51, C54, and C57 from left to right, and the residues preceding the cysteines.

(Shojania and O'Neil, 2006). The assignments of the Cys residues are particularly informative, as they confirmed that all of the Cys residues in the NMR sample were reduced (Fig. 9.8); all the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts observed in a three-dimensional HNCACB (Wittekind and Mueller, 1993) spectrum are in the range of the random coil chemical shifts of reduced cysteine (58.6 and 28.3 ppm) (Schwarzinger et al., 2000), differing significantly from those of oxidized cysteine (55.6 and 41.2 ppm) involved in disulfide bonds. The dispersion of the resonances in ^{15}N-^{1}H HSQC spectra of Tat_{1-72} (Fig. 9.9) is typical for a disordered protein lacking regular secondary structure (Freund et al., 1995; Dyson and Wright, 1998; Peloponese et al., 2000; Gregoire et al., 2001; Peti et al., 2001) and agrees with previous homonuclear NMR studies. The resonance linewidths of 15 ± 5 Hz for ^{1}H and 6 ± 1 Hz for ^{15}N are slightly broader than those measured for folded proteins of a comparable size (6–9 and 3 Hz, respectively) (Cavanagh et al., 1996). The broad linewidths suggest backbone conformational exchange on the micro- to millisecond timescale. Furthermore, multiple minor resonances are detected for many residues, suggesting the presence of multiple, likely disordered, conformers in equilibrium (Shojania and O'Neil, 2006).

The use of the chemical shift as an indicator of secondary structure (Wishart and Sykes, 1994) has been successfully applied to the analysis of proteins under denaturing conditions to identify regions of residual structure (Garcia et al., 2001; Yao et al., 2001; Teilum et al., 2002) but required adaptation to account for local

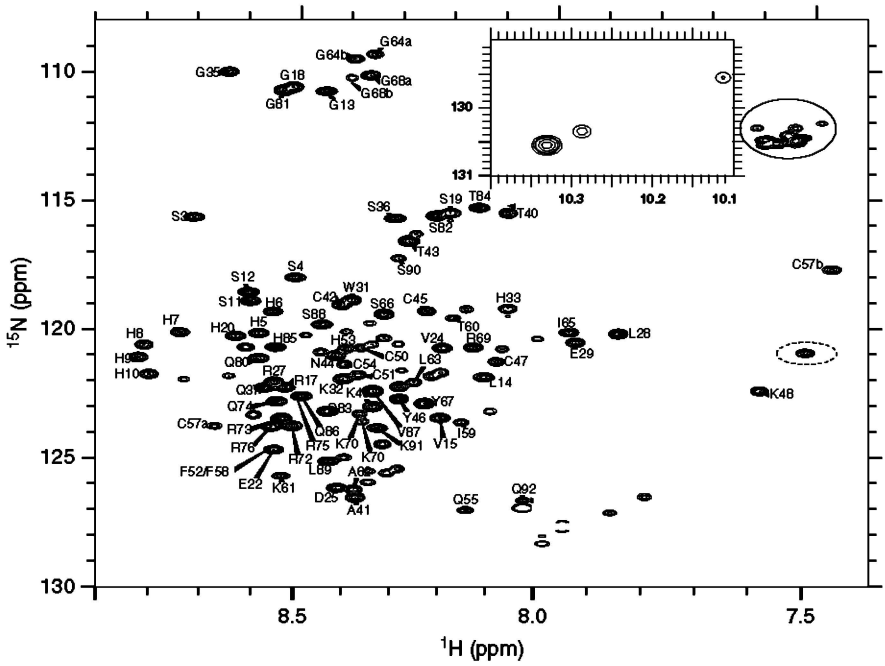


Figure 9.9 Two-dimensional $^{1}H/^{15}N$-HSQC spectrum (Kay et al., 1992) of Tat at pH 4.1 and 293K. Backbone amide proton resonances are narrowly dispersed over the range of 7.4–8.9 ppm. Backbone resonance assignments of 80 of the 84 nonproline resonances are shown. The inset spectrum shows the three peaks associated with the side chain of the single Trp residue indicating one major and two minor conformers. Also labeled are several glycine conformers. Side-chain Asn and Gln NH_2 resonances are outlined with a solid ellipse, and side-chain Arg resonances are outlined in a dashed ellipse.

sequence effects on the reference shifts for the random coil (Schwarzinger et al., 2000, 2001, 2002). In some cases, a nonnative residual structure has been identified, which may indicate transient intermediates along a protein-folding pathway (Yao et al., 2001; Teilum et al., 2002). Marsh et al. recently described a method for combining the chemical shift differences from many nuclei into a single probe score to obtain a secondary structure propensity (SSP) especially suited to detecting transient structure and structural propensity in IDPs (Marsh et al., 2006). In this procedure, consecutive residues with SSP scores of $+1$ or -1 indicate a fully formed α- or β-structural element, respectively. Lower scores correspond to the fraction of the ensemble that is in a particular structural state. The SSP scores for Tat are shown in Fig. 9.10 (Shojania and O'Neil, 2006). For the entire protein, the expected fractions of α- and β-structure are below 20%. Particularly noteworthy in the context of the two X-ray diffraction structures discussed below (Anand et al., 2008; Tahirov et al., 2010) are the SSP scores for the region centered on residue 52, which indicate a propensity toward β structure followed by a propensity to α-helix for a region centered on residue 66. The two regions correspond to the

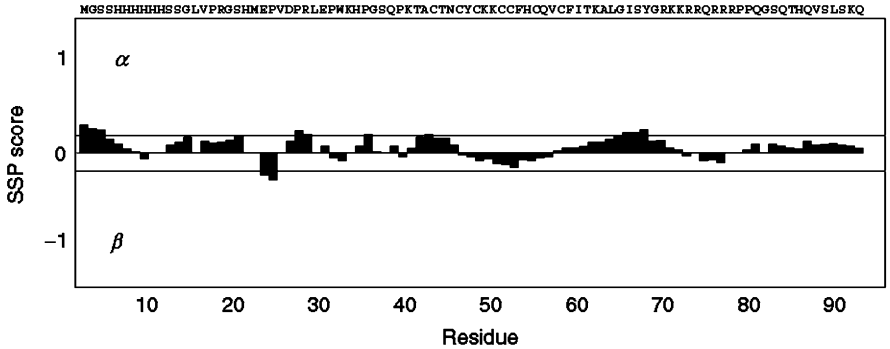


Figure 9.10 Secondary structure propensity (SSP) scores (Marsh et al., 2006) based on the H^N, N, Cα, Cβ, C′, and Hα chemical shifts. Upper and lower lines correspond to 20% of the conformers in the ensemble being in either the α or β states. Note that the 20-residue affinity tag has been added to the sequence in the figure, whereas it is absent in Fig. 9.7.

Cys-rich region and a region that encompasses the core and basic regions shown in Fig. 9.7.

Relaxation data were obtained for 77% and 72% of the observable resonances at 600 and 800 MHz spectrometer frequencies, respectively. The steady-state heteronuclear ^{15}N-^{1}H nuclear Overhauser effects (NOEs) exhibited a flattened bell-shaped variation with amino acid sequence (Shojania and O'Neil, 2006), which is diagnostic of a completely disordered polypeptide (Hu et al., 1990). The lack of any significant variation in the ^{15}N-^{1}H NOEs, coupled with their negative values, are good indicators of the degree of uniform disorder throughout the protein backbone on the nano- to picosecond timescale. Longitudinal relaxation rates (R1) observed for Tat (Shojania and O'Neil, 2006) were consistent with slower relaxation and faster dynamics at the nano- to picosecond timescale than for a folded protein of similar size, and both the ^{15}N-^{1}H NOEs and R1 values were very similar to those measured for a guanidine-denatured protein (Farrow et al., 1997). The rotating frame longitudinal relaxation rates (R1ρ) were slightly higher than average at the end of the Cys-rich region (Cys-57 to Ile-59), possibly indicating contributions from slow conformational exchange.

Analysis of the relaxation rates by reduced spectral density mapping (Peng and Wagner, 1992; Lefevre et al., 1996; Farrow et al., 1997) showed little variation over the length of the sequence and showed a range of values similar to those observed for the guanidine-denatured state of drkN SH3 (Farrow et al., 1997), acid-denatured apomyoglobin (Yao et al., 2001), apomyoglobin denatured by low pH/urea (Peng and Wagner, 1992), and the intrinsically disordered propeptide of subtilisin (Buevich et al., 2001). Interestingly, the acid-urea-unfolded state of apomyoglobin shows minima in the high frequency plots that appear to be diagnostic of residual structure because they correspond to maxima in buried surface area in the folded protein (Yao et al., 2001). The lack of definition in the plots of high frequency motion for reduced Tat_{1-72} implies a lack of formation of any transient structure at pH 4.1.

However, residues Asn-44, Lys-61, and Ala-62 were found to have faster than average motions in the mid-frequency plots. Schwarzinger et al. (2002) made a similar observation in their studies of acid-urea-denatured apomyoglobin and noted that high proportions of glycine and alanine are present in the most flexible regions of the protein. The region of Tat following the Cys-rich region contains one alanine and two glycines (Ala-62, Gly-64, and Gly-68), which may explain the higher dynamics. Calculation of Lipari–Szabo order parameters from the relaxation data gave an average value of 0.5, where a value of 1 represents order and 0 represents disorder and where most unfolded proteins show values in the range of 0.4–0.6 (Farrow et al., 1997; Buevich et al., 2001).

9.5.3 Implications of the NMR Data

Two weaknesses in the NMR experiments are that they were measured at pH 4 and that the exon 2 product is absent from the protein and it could be critical for folding. The cross-peaks in the HSQC spectra of Tat_{1-72} (Fig. 9.11) diminish in intensity as the pH is raised from 3.3 to 6.7, whereas the chemical shifts of the resonances move very little. For the most part, the loss of resonance intensity can be explained by the pH-dependent increase in hydrogen exchange rates and suggests that the protein does not fold at the elevated pH because buried resonances would be expected to exchange more slowly and remain visible in the spectrum. However, it cannot be ruled out that the protein forms a molten globule or other condensed conformation at elevated pH, in which conformational exchange causes broadening of the resonances. And although the dynamics experiments and chemical shift analysis suggest that the protein is in a random coil conformation at pH 4, it is also conceivable that measurement of ^{1}H-NOEs might detect minor condensed conformations in the ensemble.

The observation of intrinsic disorder in HIV-1 Tat_{1-72} at pH 4 was recently corroborated by small angle X-ray scattering (SAXS) and dynamic light scattering (DLS) measurements of the hydrodynamic radius of Tat_{1-101} in solution (Foucault et al., 2010). SAXS measurements showed a hydrodynamic radius of 33 ± 1.5 Å, and DLS yielded a radius of 30 ± 3 Å, whereas a folded globular protein of 101 residues would be expected to have a Stokes radius of about 18 Å (Foucault et al., 2010). The data were interpreted to indicate that the 101-residue protein is mostly in an extended conformation at neutral pH. These observations confirm the predictions of disorder in the C-terminal product of exon 2 (Fig. 9.5l) and prove that the C-terminus does not form a folding nucleus that catalyzes folding of the exon 1 protein. However, none of the solution experiments address the likely influence of zinc ions on the folding of the Cys-rich region of the protein.

9.5.4 X-Ray Diffraction Studies of EIAV Tat

X-ray diffraction studies have been reported for EIAV Tat in complex with Cyclin T1 and TAR. EIAV Tat contains only two Cys residues in contrast to the six

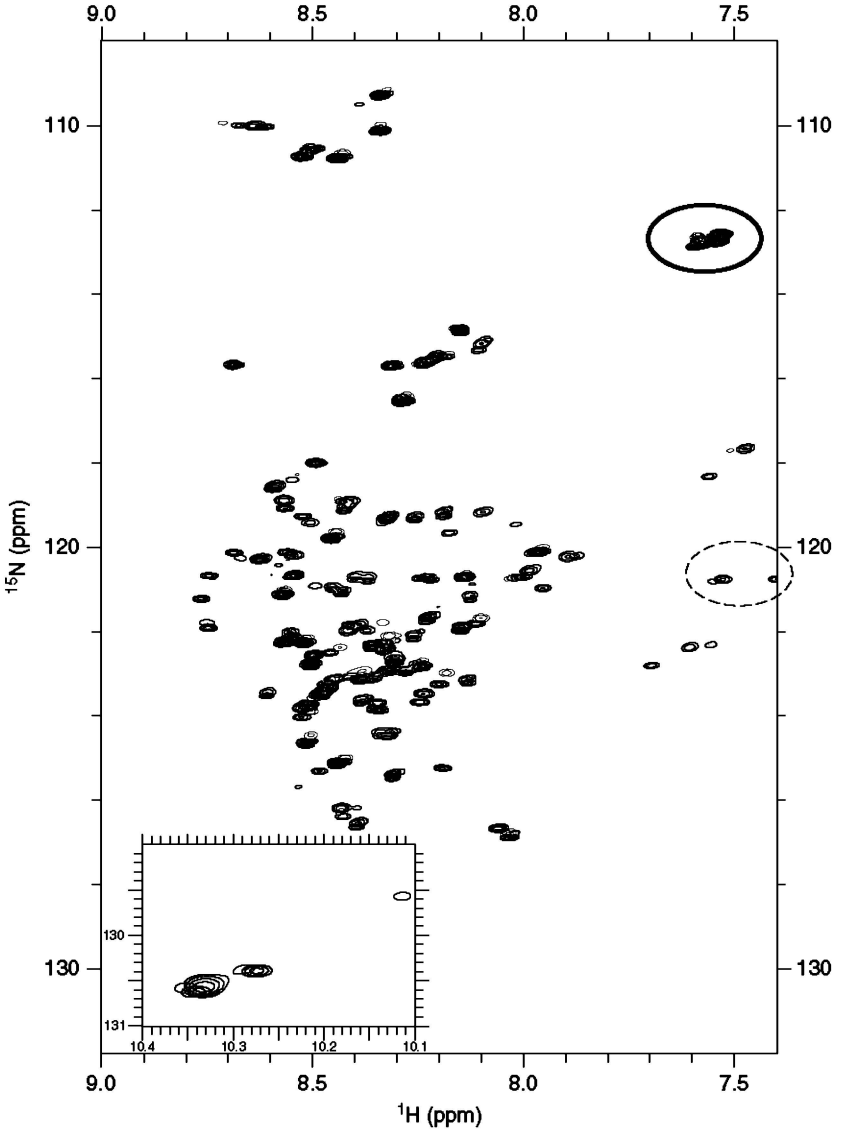


Figure 9.11 Two-dimensional $^{1}H/^{15}N$-HSQC spectra of Tat_{1-72} at 293 K observed at pH 3.3 (red), pH 4.1 (orange), pH 5.3 (purple), pH 5.8 (cyan), and pH 6.7 (blue). The solution is approximately 1 mM, and all spectra were obtained from a single expression/purification. Each spectrum was collected with 32 transients, 2048 × 256 complex points, and sweep widths of 10 ppm in F2(^{1}H) and 24 ppm in F1(^{15}N). Side chain Asn and Gln NH2 resonances are outlined with a solid ellipse. (*See insert for color representation of the figure.*)

(type C) or seven (type B) present in the HIV-1 proteins, considerably simplifying expression and purification of the protein. A 57-residue segment of EIAV Tat (residues 22–78) fused with a flexible linker to the cyclin box domain of human Cyclin T1 formed crystals with electron density observed for residues 8–263 of the Cyclin T1 (Anand et al., 2007). No electron density was observed for EIAV Tat, indicating that Cyclin T1 does not induce a stable conformation in Tat (Anand et al., 2007). Crystallization of an EIAV Cyclin-T1 Tat fusion protein in the presence of TAR RNA resulted in the detection of electron density for residues 41–69 of Tat (Anand et al., 2008) (Fig. 9.12a). Residues 41–47 in the core region (Fig. 9.7) exist in an extended conformation interacting mostly with Cyclin T1. Residues 48–59 in the basic segment (Fig. 9.7) form a helix that binds to the major groove of the TAR stem loop structure (Fig. 9.12a). This region is followed by a tight turn and an extended segment that inserts the C-terminal Leu-68 and Leu-69 into a hydrophobic groove on Cyclin T1. It is interesting to note that the strongest indication of a propensity to fold based on the NMR chemical shift analysis shown in Fig. 9.10 is found in the basic region of the Tat protein (Shojania and O'Neil, 2006). The structure of the tripartite complex explains the specificity of Tat proteins for T-type cyclins, the binding specificities of Cyclin T1 for Tat–TAR complexes from different species, the structural implication of the acetylation of Lys-50 and likely reveals the binding site for Hexim1 on Cyclin T1. The Cyclin-T1 Tat interface encompasses 806 $Å^2$, the Tat–TAR interface covers 582 $Å^2$, and the Cyclin-T1 TAR complex buries only 193 $Å^2$. Clearly the role of Tat is to bring the two molecules together. Although this structure does not directly reveal the dynamics of Tat and TAR, it clearly illustrates the large surface area available to a disordered protein, permitting high affinity and high specificity of binding. The structure also shows the power of molecular structures to illuminate biological activity even in the realm of IDPs and suggests that the crystallization of fusion proteins will be a rich source of information on IDPs.

9.5.5 X-Ray Diffraction Studies of HIV-1 Tat

A more recent X-ray diffraction study of HIV-1 Tat (86 residues) in complex with P-TEFb (Tahirov et al., 2010) is highly complementary to the EIAV Tat structure. It revealed residues 1–49 of Tat in an extended conformation along the interface of the CDK9-Cyclin-T1 complex (Fig. 9.12b). The region observed in the complex is thought to encompass the minimal surface for activation of Cyclin T1 and includes the acidic, cysteine-rich, and core regions of Tat. The basic region is involved in RNA binding, and in the absence of bound TAR, residues 50–86 of Tat do not show any defined electron density and are therefore disordered. The interaction between Tat and Cyclin T1 also revealed the expected Zn bridge between Tat and Cys-261 of Cyclin T1 (Garber et al., 1998). Although residues 1–49 are defined in the electron density maps of the complex, only two small segments adopt any regular secondary structure along with a small segment coordinating two zinc ions; the remainder is essentially disordered. Residues 29–33 form a small 3_{10}-helix, followed by an α-helix spanning residues 35–43. One zinc ion coordinates with Cys-22, His-33,

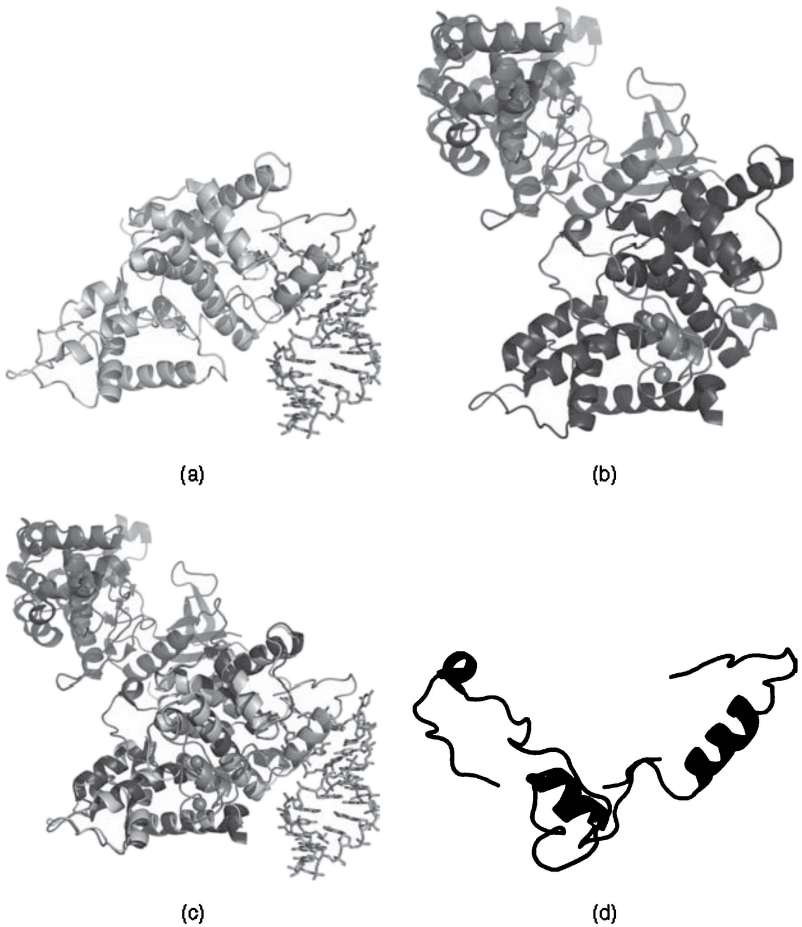


Figure 9.12 X-ray diffraction structures of (a) the Cyclin-T1-Tat-TAR complex from EIAV (Anand et al., 2008) showing the C-terminal region of EIAV Tat residues 41–69 (orange) bound to the cyclin box repeat of Cyclin T1 (cyan) and the major groove and loop region of TAR generated from PDB id 2W2H; (b) the human P-TEFb complex of Cyclin T1 (blue) and CDK9 (red) showing HIV-1 Tat residues 1–49 (violet) and the zinc ions in gray (Tahirov et al., 2010) generated from PDB id: 3MI9. (c) Alignment of the two structures in (a) and (b) with the Cyclin T1 proteins superimposed. (d) The structured portions of HIV-1 Tat (1–49) and EIAV Tat (41–69) extracted from the alignment in (c). The remainder of the HIV-1 and EIAV Tat proteins is missing from the electron density maps. (*See insert for color representation of the figure.*)

Cys-34, and Cys-37, and the other coordinates with Cys-25, Cys-27, and Cys-30 of Tat and Cys-261 of Cyclin T1. Thus, only the Cys-rich and core regions of Tat gain in structural order on formation of the complex, a result that complements the NMR chemical shift and relaxation analyses described earlier (Shojania and O'Neil, 2006). Tahirov et al. also suggest a possible mechanism for the displacement of

Hexim1 by Tat by revealing a small α-helix (HC) on Cyclin T1, which becomes disordered on binding with Tat. Helix HC is likely near the binding interface of the Hexim1 7SK snRNA dimer, and its order-to-disorder transition on Tat binding to Cyclin T1 may disrupt the binding of the Hexim1-7SK dimer (Tahirov et al., 2010). The structure may also be showing how some of the entropy loss on Tat binding can be paid for by disordering its binding partner.

The two X-ray diffraction structures each provide only fragmentary information on the structure of Tat; however, together the two structures yield information on nearly the entire exon 1 protein (HIV-1 Tat residues 1–49 and EIAV Tat residues 41–69). By aligning the two complexes by superimposing the Cyclin T1 structures (Fig. 9.12c), a picture of Tat emerges (Fig. 9.12d), which shows a predominantly extended conformation with two small helical segments and a zinc finger conformation that forms on complex formation. The folded segments may be targets of small molecule antagonists of P-TEFb or TAR binding as discussed below.

9.6 THERAPEUTIC IMPLICATIONS OF DISORDER

In the United States, more than 25 drugs have been approved by the FDA to combat HIV-1 infection, including PR, RT, IN, and fusion inhibitors, as well as CCR5 antagonists (Hughes, 2010). Because of its essential role in the life cycle of the virus, proviral transcription is a highly attractive target for inhibitor development, yet no drugs are currently on the market. Tat has been the target of several attempts to develop a vaccine for the prevention and treatment of HIV-1 infection. A recent phase I trial showed significant promise; the vaccine is safe and highly immunogenic and is reported to be proceeding to a phase II trial (Bellino et al., 2009). An important implication of these studies is that the absence of a single dominant protein conformation does not preclude the development of high affinity antigen–antibody interactions. This runs counter to the current pursuits in vaccination research, which suggest that conformational restriction of peptides can lead to higher immunogenicity (Kao and Hodges, 2009; Moseri et al., 2010), and indeed, in the vast majority of cases, peptides have so far proven to be less immunogenic than the proteins from which they were derived (Azizi and Diaz-Mitoma, 2007). In this regard, it is very likely that Tat immunogens are disulfide cross-linked, and it may be these stabilized segments that form the primary epitopes.

The Tat–TAR interaction has also been the target of multiple attempts to develop inhibitors of transcription activation, including large-scale small molecule screens (Bannwarth and Gatignol, 2005). Many of the compounds under development target the bulge in the HIV-1 TAR RNA (Baba, 2006), with fewer studies targeting the Tat protein (Montembault et al., 2004; Chen et al., 2009). Conceptually, structure-based drug design aimed at IDPs may at first glance appear impossible because the proteins do not adopt a single conformation in the absence of a binding partner. Adding to the challenge, high affinity high specificity small molecule inhibitors of protein–protein interactions have been notoriously difficult to design (Wells and McClendon, 2007). However, recently, a number of small molecule inhibitors

targeting IDPs have been developed, suggesting that IDPs may be the source of an abundance of new drugs for a wide variety of disorders.

The transcription activator cMyc contains a basic helix-loop-helix leucine zipper domain that is a disordered monomer but undergoes a coupled folding and binding transition in the presence of its partner Max. Recently, Metallo and colleagues (Hammoudeh et al., 2009) identified several small-molecule-binding sites in the intrinsically disordered regions of cMyc. Small molecules bind simultaneously and independently to three different sites inducing very little order in the IDP yet inhibiting cMyc–Max interactions potently. This study also suggests that MoRFs are likely sites of drug interaction in IDPs. Although the small molecule affinity is on the micromolar scale, it was suggested that higher affinity could be obtained by introducing multilvalency into the drug candidate. Optimization allowed specificity in disruption of the cMyc–Max interaction (Metallo, 2010).

The previously discussed X-ray diffraction structures (Anand et al., 2008; Tahirov et al., 2010) show that the small disordered Tat protein can provide a large binding surface bringing together two proteins and RNA in a high affinity high specificity complex for efficient transcription activation. The large binding interface should provide ample opportunity for the development of highly specific inhibitors of proviral transcription activation, a process that is essential to viral replication. Specifically, based on the new X-ray diffraction structures of the HIV-1 and EIAV Tat complexes mentioned earlier, residues 29–33, 35–43, and 48–59 adopt some local secondary structure (MoRFs) that may be targets for drug development.

ACKNOWLEDGMENTS

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Manitoba; the Canadian Foundation for Innovation provided funding for the NMR facility at the University of Manitoba.

ABBREVIATIONS

AIDS	acquired immunodeficiency syndrome
BBB	blood–brain barrier
CA	capsid
CBP	cAMP response element binding protein
CCR5	CC chemokine receptor 5
CD	circular dichroism
CD4	cluster of differentiation 4
CDK9	cyclin-dependent kinase 9
CREB	cAMP response element binding
CTD	carboxy terminal domain
CXCR4	CXC chemokine receptor 4
DSIF	DRB sensitivity-inducing factor

EIAV	equine infectious anemia virus
ELL	eleven-nineteen lysine-rich leukemia gene
Env	envelope protein
Gag	group-specific antigen
gp	glycoprotein
GTFs	general transcription factors
HAART	highly active antiretroviral therapy
HAD	HIV-associated dementia
HAT	histone acetyl transferase
Hexim1	hexamethylene bisacetamide-inducible protein 1
HIV-1	human immunodeficiency virus type 1
HIVE	HIV-associated encephalitis
hnRNP	heterogeneous nuclear ribonucleoprotein
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IN	integrase
LTR	long terminal repeat
MA	matrix
MePCE	methylphosphate capping enzyme
MHC	major histocompatibility complex
MLL-fps	MLL fussion proteins
MoRF	molecular recognition fragment
N-TEF	negative transcription elongation factor
NC	nucleocapsid
NCBI	National Center for Biological Interaction
Nef	negative factor
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
P-TEFb	positive transcription elongation factor b
PCAF	p300/CBP-associated factor
PIC	preinitiation complex
Pol	polyprotein
PR	protease
R1	longitudinal relaxation rate
$R1\rho$	rotating frame longitudinal relaxation rate
R2	transverse relaxation rate
Rev	regular expression of virus
RGD	arginine-glycine-aspartic acid
RNAP II	RNA polymerase II
RNP	ribonucleoprotein
RT	reverse transcriptase
snRNA	small nuclear RNA
Spt5	suppressor of tyrosine 5
SSP	secondary structure propensity
SU	surface unit

TAR	transactivation response
Tat	transactivator of transcription
TCEP	tris(carboxyethyl)phosphine
TFIID	transcription factor IID
TM	transmembrane
Vif	viral infectivity factor
Vpr	viral protein r
Vpu	viral protein u

REFERENCES

Aiken C, Konner J, Landau NR, Lenburg ME, Trono D. Nef induces CD4 endocytosis: Requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. Cell 1994;76(5):853–864.

Albini A, Benelli R, Presta M, Rusnati M, Ziche M, Rubartelli A, Paglialunga G, Bussolino F, Noonan D. HIV-tat protein is a heparin-binding angiogenic growth factor. Oncogene 1996a;12(2):289–297.

Albini A, Soldi R, Giunciuglio D, Giraudo E, Benelli R, Primo L, Noonan D, Salio M, Camussi G, Rockl W, Bussolino F. The angiogenesis induced by HIV-1 tat protein is mediated by the Flk-1/KDR receptor on vascular endothelial cells. Nat Med 1996b;2(12):1371–1375.

Anand K, Schulte A, Fujinaga K, Scheffzek K, Geyer M. Cyclin box structure of the P-TEFb subunit Cyclin T1 derived from a fusion complex with EIAV Tat. J Mol Biol 2007;370(5):826–836.

Anand K, Schulte A, Vogel-Bachmayr K, Scheffzek K, Geyer M. Structural insights into the Cyclin T1-Tat-TAR RNA transcription activation complex from EIAV. Nat Struct Mol Biol 2008;15(12):1287–1292.

András IE, Pu H, Deli MA, Nath A, Hennig B, Toborek M. HIV-1 Tat protein alters tight junction protein expression and distribution in cultured brain endothelial cells. J Neurosci Res 2003;74(2):255–265.

Avraham HK, Jiang S, Lee T-H, Prakash O, Avraham S. HIV-1 Tat-mediated effects on focal adhesion assembly and permeability in brain microvascular endothelial cells. J Immunol 2004;173(10):6228–6233.

Azizi A, Diaz-Mitoma F. Viral peptide immunogens: current challenges and opportunities. J Pept Sci 2007;12(12):776–786.

Baba M. Recent status of HIV-1 gene expression inhibitors. Antiviral Res 2006;71(2–3): 301–306.

Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol 2009;5(11):808–814.

Banks WA, Robinson SM, Nath A. Permeability of the blood-brain barrier to HIV-1 Tat. Exp Neurol 2005;193(1):218–227.

Bannwarth S, Gatignol A. HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. Curr HIV Res 2005;3(1):61–71.

Barboric M, Lenasi T. Kick-sTARting HIV-1 transcription elongation by 7SK snRNP deporTATion. Nat Struct Mol Biol 2010;17(8):928–930.

Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 1983;220:868–871.

Bellino S, Francavilla V, Longo O, Tripiciano A, Paniccia G, Arancio A, Fiorelli V, Scoglio A, Collacchi B, Campagna M, Lazzarin A, Tambussi G, Tassan Din C, Visintini R, Narciso P, Antinori A, D'Offizi G, Giulianelli M, Carta M, Di Carlo A, Palamara G, Giuliani M, Laguardia ME, Monini P, Magnani M, Ensoli F, Ensoli B. Parallel conduction of the phase I preventive and therapeutic trials based on the Tat vaccine candidate. Rev Recent Clin Trials 2009;4(3):195–204.

Bennasser Y, Jeang K. HIV-1 Tat interaction with Dicer: requirement for RNA. Retrovirology 2006;3:95–101.

Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR. Recruitment of a protein complex containing Tat and Cyclin T1 to TAR governs the species specificity of HIV-1 Tat. EMBO J 1998;17(23):7056–7065.

Bourgeois CF, Kim YK, Churcher MJ, West MJ, Karn J. Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences. Mol Cell Biol 2002;22(4):1079–1093.

Brown HG, Hoh JH. Entropic exclusion by neurofilament sidearms: A mechanism for maintaining interfilament spacing. Biochemistry 1997;36(49):15035–15040.

Buevich AV, Shinde UP, Inouye M, Baum J. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. J Biomol NMR 2001;20(3):233–249.

Bussell R Jr, Eliezer D Residual structure and dynamics in parkinson's disease-associated mutants of alpha-synuclein. J Biol Chem 2001;276(49):45996–46003.

Campbell GR, Pasquier E, Watkins J, Bourgarel-Rey V, Peyrot V, Esquieu D, Barbier P, de Mareuil J, Braguer D, Kaleebu P, Yirrell DL, Loret EP. The glutamine-rich region of the HIV-1 Tat protein is involved in T-cell apoptosis. J Biol Chem 2004;279(46):48197–48204.

Campbell GR, Watkins JD, Singh KK, Loret EP, Spector SA. Human immunodeficiency virus type 1 subtype C Tat fails to induce intracellular calcium flux and induces reduced tumor necrosis factor production from monocytes. J Virol 2007;81(11):5919–5928.

Carroll IR, Wang J, Howcroft TK, Singer DS. HIV Tat represses transcription of the beta2-microglobulin promoter. Mol Immunol 1998;35(18):1171–1178.

Carter CC, Onafuwa-Nuga A, McNamara LA, Riddell J, Bixby D, Savona MR, Collins KL. HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs. Nat Med 2010;16(4):446–451.

Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ. Protein NMR spectroscopy: principles and practice. San Diego: Academic Press; 1996.

Chen D, Wang M, Zhou S, Zhou Q. HIV-1 Tat targets microtubules to induce apoptosis, a process promoted by the pro-apoptotic Bcl-2 relative Bim. EMBO J 2002;21(24): 6801–6810.

Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. J Proteome Res 2006;5(4):879–887.

Chen S, Chen R, He M, Pang R, Tan Z, Yang M. Design, synthesis, and biological evaluation of novel quinoline derivatives as HIV-1 Tat-TAR interaction inhibitors. Bioorg Med Chem 2009;17(5):1948–1956.

Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van Y-YJ, Romero P, Cortese MS, Uversky VN, Dunker AK. Rational drug design via intrinsically disordered protein. Trends Biotechnol 2006;24(10):435–442.

Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 2007;46(47):13468–13477.

Chiu Y-L, Ho CK, Saha N, Schwer B, Shuman S, Rana TM. Tat stimulates cotranscriptional capping of HIV mRNA. Mol Cell 2002;10(3):585–597.

Chun T-W, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, Herma M. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature 2010;387(6629):183–188.

Clements JE, Zink MC. Molecular biology and pathogenesis of animal lentivirus infections. Clin Microbiol Rev 1996;9(1):100–117.

Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P, Weiss RA. What to call the AIDS virus? Nature 1986;321(6065):10.

Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Woodbury, NY: Cold Spring Harbor Laboratory Press: 1997.

Core LJ, Lis JT. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 2008;319(5871):1791–1792.

Cullen B. Regulation of HIV-1 gene expression. Faseb J 1991;5(10):2361–2368.

D'Orso I, Frankel AD. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol 2010;17(7):815–821.

Demchenko AP. Recognition between flexible protein molecules: induced and assisted folding. J Mol Recogn 2001;14(1):42–61.

Denning DP, Patel SS, Uversky V, Fink AL, Rexach M. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. Proc Natl Acad Sci USA 2003;100(5):2450–2455.

Derse D, Carvalho M, Carroll R, Peterlin BM. A minimal lentivirus Tat. J Virol 1991;65(12):7012–7015.

Dobson CM. Principles of protein folding, misfolding and aggregation; protein misfolding and human disease and developmental biology of the retina. Semin Cell Dev Biol 2004;15(1):3–16.

Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005a;21(16):3433–3434.

Dosztányi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 2005b;347(4):827–839.

Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998;3:471–482.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW. Intrinsically disordered protein. J Mol Graphics Model 2001;19(1):26–59.

Dunker AK, Obradovic Z. The protein trinity—linking function and disorder. Nat Biotechnol 2001;19(9):805–806.

Dunker AK, Oldfield C, Meng J, Romero P, Yang J, Chen J, Vacic V, Obradovic Z, Uversky V. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genom 2008;9(Suppl 2):S1.

Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathogens 2008;4(2):e32.

Dyson HJ, Wright PE. Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol 1998;5:499–503.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6(3):197–208.

Emini EA, editor. The human immunodeficiency virus: biology, immunology, and therapy. Princeton, N.J., Chichester: Princeton University Press; 2002.

Ensoli B, Barillari G, Salahuddin SZ, Gallo RC, Wong-Staal F. Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. Nature 1990;345(6270):84–86.

Farrow NA, Zhang O, Forman-Kay JD, Kay LE. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. Biochemistry 1997;36(9):2390–2402.

Foucault M, Mayol K, Receveur-Bréchot V, Bussat MC, Klinguer-Hamour C, Verrier B, Beck A, Haser R, Gouet P, Guillon C. UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. Proteins 2010;78(6):1441–1456.

Frankel AD, Young JAT. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 1998;67(1):1–25.

Freed EO. HIV-1 Replication. Somat Cell Mol Genet 2001;26(1):13–33.

Freund J, Vertesy L, Koller KP, Wolber V, Heintz D, Kalbitzer HR. Complete 1H nuclear magnetic resonance assignments and structural characterization of a fusion protein of the alpha-amylase inhibitor tendamistat with the activation domain of the human immunodeficiency virus type 1 Tat protein. J Mol Biol 1995;250(5):672–688.

Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG. Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic Acids Res 2009;37(Database issue):D417–D422.

Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, Rice AP, Littman DR, Jones KA. The interaction between HIV-1 Tat and human Cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. Genes Dev 1998;12(22):3512–3527.

Garcia JA, Harrich D, Pearson L, Mitsuyasu R, Gaynor RB. Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. EMBO J 1988;7(10):3143–3147.

Garcia P, Serrano L, Durand D, Rico M, Bruix M. NMR and SAXS characterization of the denatured state of the chemotactic protein Che Y: Implications for protein folding initiation. Prot Sci 2001;10(6):1100–1112.

Garzon MT, Lidon-Moya MC, Barrera FN, Prieto A, Gomez J, Mateu MG, Neira JL. The dimerization domain of the HIV-1 capsid protein binds a capsid protein-derived peptide: A biophysical characterization. Prot Sci 2004;13(6):1512–1523.

Gautier V, Gu L, O'Donoghue N, Pennington S, Sheehy N, Hall W. In vitro nuclear interactome of the HIV-1 Tat protein. Retrovirology 2009;6(1):47.

Goh G, Dunker AK, Uversky V. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genom 2008;9(Suppl 2):S4.

Goldstein G. HIV-1 Tat protein as a potential AIDS vaccine. Nat Med 1996;2(9):960–964.

Gregoire C, Peloponese J, Esquieu D, Opi S, Campbell G, Solomiac M, Lebrun E, Lebreton J, Loret E. Homonuclear 1H-NMR assignment and structural characterization of human immunodeficiency virus type 1 Tat Mal protein. Biopolymers 2001;62:324–335.

Grzesiek S, Bax A. Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc 1992;114(16): 6291–6293.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007;130(1):77–88.

Guo X, Kameoka M, Wei X, Roques B, Gotte M, Liang C, Wainberg MA. Suppression of an intrinsic strand transfer activity of HIV-1 Tat protein by its second-exon sequences. Virology 2003;307(1):154–163.

Gupta B, Levchenko TS, Torchilin VP. Intracellular delivery of large molecules and small particles by cell-penetrating proteins and peptides. Adv Drug Deliv Rev 2005;57(4):637–651.

Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J Am Chem Soc 2009;131(21):7390–7401.

Haseltine W. Molecular biology of the human immunodeficiency virus type 1. Faseb J 1991;5(10):2349–2360.

He N, Liu M, Hsu J, Xue Y, Chou S, Burlingame A, Krogan NJ, Alber T, Zhou Q. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol Cell 2010;38(3):428–438.

Hegde ML, Rao KSJ. DNA induces folding in alpha-synuclein: Understanding the mechanism using chaperone property of osmolytes. Arch Biochem Biophys 2007;464(1): 57–69.

Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007;450(7172): 964–972.

Hoh JH. Functional protein domains from the thermally driven motion of polypeptide chains: A proposal. Proteins 1998;32(2):223–228.

Howcroft T, Strebel K, Martin M, Singer D. Repression of MHC class I gene promoter activity by two-exon Tat of HIV. Science 1993;260(5112):1320–1322.

Hu Y, Macinnis JM, Cherayil BJ, Fleming GR, Freed KF, Perico A. Polypeptide dynamics—experimental tests of an optimized rouse-zimm type model. J Chem Phys 1990;93(1):822–836.

Huang Y, Liu Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "Fly-Casting" mechanism. J Mol Biol 2009;393(5):1143–1159.

Hughes V. The outlook for a cure. Nature 2010;466(7304):S11–S13.

Jeang K, Xiao H, Rich E. Multifaceted activities of the HIV-1 transactivator of transcription, Tat. J Biol Chem 1999;274:28837–28840.

Jeang K-T. HIV-1 Tat: structure and function. In: Myers G, Korber BT, Foley BT, Jeang K-T, Mellors JW, Wain-Hobson S, editors. Human retroviruses and AIDS 1996: A compilation and analysis of nucleic acid and amino acid sequences. Los Alamos: Los Alamos National Laboratories; 1996. pp.11–26.

Jeffery CJ. Moonlighting proteins—an update. Mol BioSyst 2009;5(4):345–350.

Kao DJ, Hodges RS. Advantages of a synthetic peptide immunogen over a protein immunogen in the development of an anti-pilus vaccine for *Pseudomonas aeruginosa*. Chem Biol Drug Des 2009;74(1):33–42.

Karn J. Tackling tat. J Mol Biol 1999;293(2):235–254.

Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 1992;114(26):10663–10665.

Kim YK, Bourgeois CF, Isel C, Churcher MJ, Karn J. Phosphorylation of the RNA Polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. Mol Cell Biol 2002;22(13):4622–4637.

Kingsman SM, Kingsman AJ. The regulation of human immunodeficiency virus type-1 gene expression. Eur J Biochem 1996;240(3):491–507.

Kuciak M, Gabus C, Ivanyi-Nagy R, Semrad K, Storchak R, Chaloin O, Muller S, Mely Y, Darlix J-L. The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro. Nucleic Acids Res 2008;36(10):3389–3400.

Kuppuswamy M, Subramania T, Srinivasan A, Chinnadurai G. Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. Nucleic Acids Res 1989;17:3551–3561.

Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 1998;393(6686):648–659.

Lassen K, Han Y, Zhou Y, Siliciano J, Siliciano RF. The multifactorial nature of HIV-1 latency. Trends Mol Med 2004;10(11):525–531.

Lefevre JF, Dayie KT, Peng JW, Wagner G. Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. Biochemistry 1996;35(8):2674–2686.

Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. Genome Inform 1999;10:30–40.

Liang C, Wainberg MA. The role of Tat in HIV-1 replication: an activator and/or a suppressor? AIDS Rev 2002;4(1):41–49.

Lopez-Huertas MR, Callejas S, Abia D, Mateos E, Dopazo A, Alcami J, Coiras M. Modifications in host cell cytoskeleton structure and function mediated by intracellular HIV-1 Tat protein are greatly dependent on the second coding exon. Nucleic Acids Res 2010;38(10):3287–3307.

Mahlknecht U, Dichamp I, Varin A, Van Lint C, Herbein G. NF-kappaB-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells. J Leuk Biol 2008;83(3):718–727.

Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, Singru PS, Nilsson KPR, Simon R, Schubert D, Eisenberg D, Rivier J, Sawchenko P, Vale W, Riek R. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science 2009;325(5938):328–332.

Mallick P, Boutz DR, Eisenberg D, Yeates TO. Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. Proc Natl Acad Sci USA 2002;99(15):9679–9684.

Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structural propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Prot Sci 2006;15:2795–2804.

Metallo SJ. Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol 2010;14(4):481–488.

Midic U, Oldfield C, Dunker AK, Obradovic Z, Uversky V. Protein disorder in the human diseasome: unfoldomics of human genetic diseases. BMC Genom 2009;10(Suppl 1):S12.

Montembault M, Vo-Thanh G, Deyine A, Fargeas V, VilliÈras M, Adjou A, Dubreuil D, Esquieu D, GrÈgoire C, Opi S, PÈloponËse J-M, Campbell G, Watkins J, de Mareuil J, Aubertin A-M, Bailly C, Loret E, Lebreton J. A possible improvement for structure-based drug design illustrated by the discovery of a Tat HIV-1 inhibitor. Bioorg Med Chem Lett 2004;14(6):1543–1546.

Moseri A, Tantry S, Sagi Y, Arshava B, Naider F, Anglister J. An optimally constrained V3 peptide is a better immunogen than its linear homolog or HIV-1 gp120. Virology 2010;401(2):293–304.

Mujtaba S, He Y, Zeng L, Farooq A, Carlson JE, Ott M, Verdin E, Zhou MM. Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. Mol Cell 2002;9(3):575–586.

Mukhopadhyay R, Kumar S, Hoh JH. Molecular mechanisms for organizing the neuronal cytoskeleton. BioEssays 2004;26(9):1017–1025.

Nath A, Psooy K, Martin C, Knudsen B, Magnuson DS, Haughey N, Geiger JD. Identification of a human immunodeficiency virus type 1 Tat epitope that is neuroexcitatory and neurotoxic. J Virol 1996;70(3):1475–1480.

Neuveut C, Jeang KT. Recombinant human immunodeficiency virus type 1 genomes with tat unconstrained by overlapping reading frames reveal residues in Tat important for replication in tissue culture. J Virol 1996;70(8):5572–5581.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(S6):566–572.

Ohgushi M, Wada A. Molten-globule state: a compact form of globular proteins with mobile side-chains. FEBS Lett 1983;164(1):21–24.

Opi S, Péloponèse J-M, Esquieu D, Watkins J, Campbell G, De Mareuil J, Jeang KT, Yirrell DL, Kaleebu P, Loret EP. Full-length HIV-1 Tat protein necessary for a vaccine. Vaccine 2004;22(23–24):3105–3111.

Peloponese J-M, Gregoire C, Opi S, Esquieu D, Sturgis J, Lebrun E, Meurs E, Collette Y, Olive D, Aubertin A-M, Witvrow M, Pannecouque C, De Clercq E, Bailly C, Lebreton J, Loret EP. 1H-13C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. Compt Rendus Accad Sci III 2000;323(10):883–894.

Peng JW, Wagner G. Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. Biochemistry 1992;31(36):8571–8586.

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 2005;3(1):35–60.

Peti W, Smith LJ, Redfield C, Schwalbe H. Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. J Biomol NMR 2001;19(2):153–165.

Pinney JW, Dickerson JE, Fu W, Sanders-Beer BE, Ptak RG, Robertson DL. HIV-host interactions: a map of viral perturbation of the host system. AIDS 2009;23(5): 549–554.

Pocernich CB, Sultana R, Mohmmad-Abdul H, Nath A, Butterfield DA. HIV-dementia, Tat-induced oxidative stress, and antioxidant therapeutic considerations. Brain Res Rev 2005;50(1):14–26.

Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 1984;224(4648):497–500.

Ptak RG, Fu W, Sanders-Beer BE, Dickerson JE, Pinney JW, Robertson DL, Rozanov MN, Katz KS, Maglott DR, Pruitt KD, Dieffenbach CW. Cataloguing the HIV-1 human protein interaction network. AIDS Res Hum Retroviruses 2008;24(12):1497–1502.

Pumfery A, Deng L, Maddukuri A, de la Fuente C, Li H, Wade JD, Lambert P, Kumar A, Kashanchi F. Chromatin remodeling and modification during HIV-1 Tat-activated transcription. Curr HIV Res 2003;1(3):343–362.

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. Proteins 2010;78(2):365–380.

Radivojac P, Vucetic S, O'Connor TR, Uversky VN, Obradovic Z, Dunker AK. Calmodulin signaling: Analysis and prediction of a disorder-dependent molecular recognition. Proteins 2006;63(2):398–410.

Romero P, Obradovic Z, Dunker AK. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform 1997;8:110–124.

Romero P, Obradovic Z, Li X, Garner E, Brown C, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.

Rout MP, Aitchison JD, Magnasco MO, Chait BT. Virtual gating and nuclear transport: the hole picture. Trends Cell Biol 2003;13(12):622–628.

Schrader EK, Harstad KG, Matouschek A. Targeting proteins for degradation. Nat Chem Biol 2009;5(11):815–822.

Schulte A, Czudnochowski N, Barboric M, Schonichen A, Blazek D, Peterlin BM, Geyer M. Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. J Biol Chem 2005;280(26):24968–24977.

Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 2001; 123(13):2970–2978.

Schwarzinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ. Random coil chemical shifts in acidic 8M urea: implementation of random coil shift data in NMRView. J Biomol NMR 2000;18(1):43–48.

Schwarzinger S, Wright PE, Dyson HJ. Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. Biochemistry 2002;41(42):12681–12686.

Shojania S, Henry GD, Chen VC, Vo TN, Perreault H, O'Neil JD. High yield expression and purification of HIV-1 Tat1-72 for structural studies. J Virol Meth 2010;164(1–2):35–42.

Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 2006;281(13):8347–8356.

Shojania S, O'Neil JD. Intrinsic disorder and function of the HIV-1 Tat protein. Protein Pept Lett 2010;17(8):999–1011.

Siddappa N, Venkatramanan M, Venkatesh P, Janki M, Jayasuryan N, Desai A, Ravi V, Ranga U. Transactivation and signaling functions of Tat are not correlated: biological and immunological characterization of HIV-1 subtype-C Tat protein. Retrovirology 2006;3(1):53.

Sims RJ, Belotserkovskaya R, Reinberg D. Elongation by RNA polymerase II: the short and long of it. Genes Dev 2004;18(20):2437–2468.

Smith SM, Pentlicky S, Klase Z, Singh M, Neuveut C, Lu CY, Reitz MS, Yarchoan R, Marx PA, Jeang KT. An in vivo replication-important function in the second coding exon of Tat is constrained against mutation despite cytotoxic T lymphocyte selection. J Biol Chem 2003;278(45):44816–44825.

Sobhian B, Laguette N, Yatim A, Nakamura M, Levy Y, Kiernan R, Benkirane M. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. Mol Cell 2010;38(3):439–451.

Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature 2010;465(7299):747–751.

Teilum K, Kragelund BB, Poulsen FM. Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. J Mol Biol 2002;324(2):349–357.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27(10):527–533.

Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579(15):3346–3354.

Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. Faseb J 2004;18(11):1169–1175.

Trono D, Van Lint C, Rouzioux C, Verdin E, Barre-Sinoussi F, Chun T-W, Chomont N. HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals. Science 2010;329(5988):174–180.

Tsvetkov P, Reuven N, Shaul Y. The nanny model for IDPs. Nat Chem Biol 2009; 5(11):778–781.

Turner BG, Summers MF. Structural Biology of HIV. J Mol Biol 1999;285(1):1–32.

UNAIDS and WHO. AIDS Epidemic Update 09. Joint United Nations Program on HIV/AIDS and World Health Organization, Geneva, 2009.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Prot Sci 2002;11(4):739–756.

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 2005;18(5):343–384.

Vendel AC, Lumb KJ. Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. Biochemistry 2003;42(4):910–916.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337(3):635–645.

Weeks KM, Ampe C, Schultz SC, Steitz TA, Crothers DM. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. Science 1990;249(4974):1281–1285.

Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell 1998;92(4):451–462.

Weissman JD, Brown JA, Howcroft TK, Hwang J, Chawla A, Roche PA, Schiltz L, Nakatani Y, Singer DS. HIV-1 Tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes. Proc Natl Acad Sci USA 1998;95(20):11601–11606.

Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 2007;450(7172):1001–1009.

Westendorp MO, Shatrov VA, Schulze-Osthoff K, Frank R, Kraft M, Los M, Krammer PH, Droge W, Lehmann V. HIV-1 Tat potentiates TNF-induced NF-kappaB activation and cytotoxicity by altering the cellular redox state. EMBO J 1995;14(3):546–554.

Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994;4(2):171–180.

Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the Alpha- and beta-carbon resonances in proteins. J Magn Reson B 1993;101(2):201–205.

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 1999;97(1):41–51.

Yang ZR, Thomson R, McMeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005;21(16):3369–3376.

Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. Biochemistry 2001;40(12):3561–3571.

Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 2007;450(7173):1263–1267.

Zwick MB, Saphire EO, Burton DR. gp41: HIV's shy protein. Nat Med 2004;10(2): 133–134.

10

INTRINSICALLY DISORDERED DOMAINS OF SESBANIA MOSAIC VIRUS ENCODED PROTEINS

SMITA NAIR, M. R. N. MURTHY, AND H. S. SAVITHRI

10.1 INTRODUCTION

Sobemoviruses (Hull and Fargette, 2006) are spherical, positive-sense single-stranded RNA viruses named after their type member southern bean mosaic virus (SBMV). The first sobemovirus was isolated in 1943 (Zaumeyer and Harter, 1943). Although these viruses have a narrow host range, they can infect plant species from at least 15 different families (Hull, 1977) belonging to both monocots and dicots. A typical sobemoviral infection is marked by necrotic lesions, stunting, mild or severe chlorosis, mottling, and vein clearing. Sobemoviruses are transmitted by vectors (aphids, beetles, leaf miners, leaf hoppers) or seeds. They can also be readily propagated by mechanical inoculation using infected leaf sap or purified virus.

Sobemoviruses are icosahedral particles of about 30 nm diameter. These virus particles have a sedimentation coefficient of 110–120S (Hull, 1988). Electron microscopic data suggests that the particles are compact nucleocapsids. They are stable to treatments such as repeated freezing and thawing, although they swell upon treatment with chelating agents such as EDTA/EGTA because of the removal of capsid-associated divalent metal ions. The three-dimensional structure of five sobemoviruses, namely, southern cowpea mosaic virus (SCPMV), *Sesbania* mosaic virus (SeMV), ryegrass mottle virus (RGoMV), rice yellow mottle virus (RYMV), and cocksfoot mottle virus (CfMV) have been determined

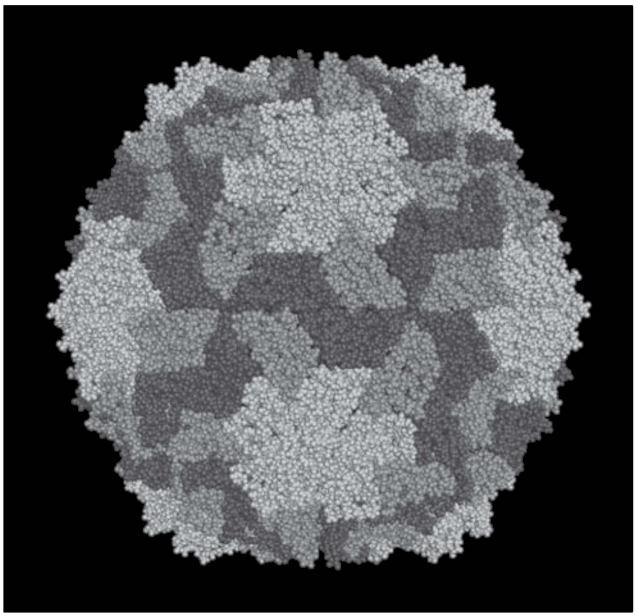
Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

(Abad-Zapatero et al., 1980; Bhuvaneshwari et al., 1995; Opalka et al., 2000; Tars et al., 2003; Plevka et al., 2007). Like other $T = 3$ viruses, SeMV capsid consists of 180 coat protein (CP, 29 kDa) subunits (Fig. 10.1a). Although the 180 subunits of the capsid are identical, they occur in three slightly different bonding environments. Therefore, these capsids contain 60 copies each of A-, B-, and C-type subunits occupying quasi-equivalent positions (Fig. 10.1b). The A subunits form pentamers at the icosahedral fivefold axes, while the B and C subunits form hexamers at the icosahedral threefold (quasi-sixfold) axes in the capsid. An icosahedral twofold relates two C-type subunits, while a quasi-twofold relates A and B subunits. Thus, the capsid consists of 30 symmetric C/C dimers and 60 quasi-symmetric A/B dimers. Protein–nucleic acid as well as metal-ion-mediated protein–protein interactions contribute to the stability of SeMV capsids. The CP subunit can be broadly divided into two domains, a random (R) domain comprising 71 N-terminal residues and a shell (S) domain. The shell domain corresponds to the core of the CP molecule, and it has the canonical eight-stranded antiparallel β-sheet/barrel topology called as the *jelly-roll motif* and five α-helices. The jelly-roll motif is commonly found in many viral CPs. The first 73 residues of the R domain are disordered in the A and B subunits, whereas in the C subunit, residues are ordered from 44 and 48–58 to form a β-annulus structure. More recently, the structure of SeMV-like particles obtained by expression of the CP in *Escherichia coli* and several of the deletion and site-specific mutants have been determined. On the basis of an analysis of these structures and biophysical studies, a model for the assembly of the virus was proposed (Savithri and Murthy, 2010).

The particles of sobemoviruses contain a single-stranded positive-sense genomic RNA (gRNA) of size 4–4.5 kb and a subgenomic RNA (sgRNA) of 1 kb. In addition to their gRNA, some sobemoviruses are known to encapsidate viroidlike satellite RNAs (sat RNAs) of size 220–390 nucleotides. sat RNAs occur both in linear and circular forms and have been reported from RYMV (Collins et al., 1998), lucerne transient streak virus (LTSV) (Forster and Symons, 1987), subterranean clover mottle virus (SCMoV) (Davies et al., 1990), and so on. These sat RNAs contain a conserved hammerhead structure, which is involved in self-cleavage (ribozyme) (Forster et al., 1987; Forster and Symons, 1987). Studies reveal the dependence of these sat RNAs on the helper virus and the host plant (Jones and Mayo, 1984; Paliwal, 1984; Sehgal et al., 1993).

10.1.1 gRNA Organization

To date, the complete genome sequences of 12 sobemoviruses and 1 unclassified sobemoviruses have been determined (Wu et al., 1987; Yassi et al., 1994; Jeffries et al., 1995; Makinen et al., 1995; Lee and Anderson, 1998; Lokesh et al., 2001; Zhang et al., 2001; Callaway and Lommel, 2002; Dwyer et al., 2003; McGavin et al., 2009; Sereme et al., 2008; Arthur et al., 2010; Chen et al., 2010; Eastwell et al., 2010). Both gRNA and sgRNA have a viral protein (viral protein genome linked, VPg) covalently attached to their 5′ end and lack a poly A tail or a t-RNA-like structure at their 3′ end. The genome codes for four overlapping open



(a)

(b)

Figure 10.1 Structure of SeMV. (a) $T = 3$ icosahedral shell of SeMV. (b) Icosahedral asymmetric unit showing three different conformations. A-type subunit is represented in green, B in red, and C in blue. Calcium ions at the interface of AB, BC, and CA subunits are depicted as yellow circles. (*See insert for color representation of the figure.*)

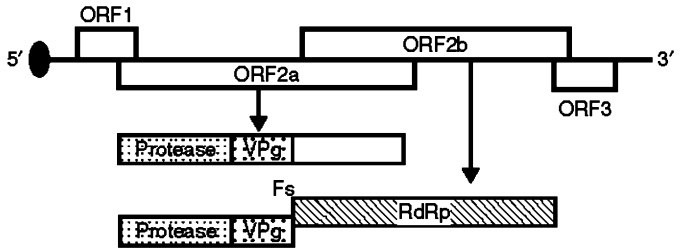


Figure 10.2 Schematic representation of the genome organization in sobemoviruses. VPg attached at the 5′ end is denoted by a black oval. The translation products of the central ORFs are also shown. Fs corresponds to the region where −1 ribosomal frame shift occurs. The RdRp domain translated via −1 ribosomal frame shifting is represented by block shifted upwards.

reading frames (ORFs), of which the ones proximal to the 5′ and 3′ end code for the movement protein (MP) and the CP, respectively. The central two ORFs, 2a and 2b, code for the polyprotein 2a and RNA-dependent RNA polymerase (RdRp), respectively (Fig. 10.2).

SeMV infects the Fabaceae plant, *Sesbania grandiflora* pers agathi, and was first identified in Tirupati, Andhra Pradesh, India (Solunke et al., 1983). SeMV can be transmitted by mechanical inoculation. Infected plants show yellowish green irregular patches on leaves 10 days post inoculation (d.p.i.). After 40–50 d.p.i., the plants appear stunted with reduction in the size of the leaflets (Subramanya et al., 1993) (Fig. 10.3). As in other sobemoviruses, SeMV ORF 1 codes for the MP and ORF 2a codes for polyprotein 2a, which was shown to have a domain arrangement of membrane-anchor-protease-VPg-P10-P8. ORF 2b is translated by −1 ribosomal frame shift into polyprotein 2ab (membrane-anchor-protease-VPg-RdRp) (Nair and Savithri, 2010a). The N-terminal serine protease domain was shown to be responsible for processing of the polyproteins into functional domains (Satheshkumar et al., 2004a; Nair and Savithri, 2010a). Among the seven protein domains encoded by SeMV genome, VPg and P8 are completely disordered, while CP, MP, and RdRp have some unstructured segments. This chapter describes the investigations that led to the identification of the intrinsically disordered protein domains encoded by SeMV genome and their functional significance.

10.2 BIOINFORMATIC ANALYSIS OF INTRINSICALLY DISORDERED DOMAINS OF SeMV-ENCODED PROTEINS

FoldIndex analysis of SeMV-encoded proteins was carried out to locate intrinsically disordered segments/domains (Fig. 10.4). As apparent from Fig. 10.4a, ORF-1-encoded MP has disordered segments in the N- and C-terminal regions. Recently, using yeast two-hybrid assays, it was demonstrated that SeMV MP interacts with other viral proteins such as CP, VPg, and P10 via the same disordered N-terminal domain (Chowdhury and Savithri, 2011a,b). FoldIndex analysis of the

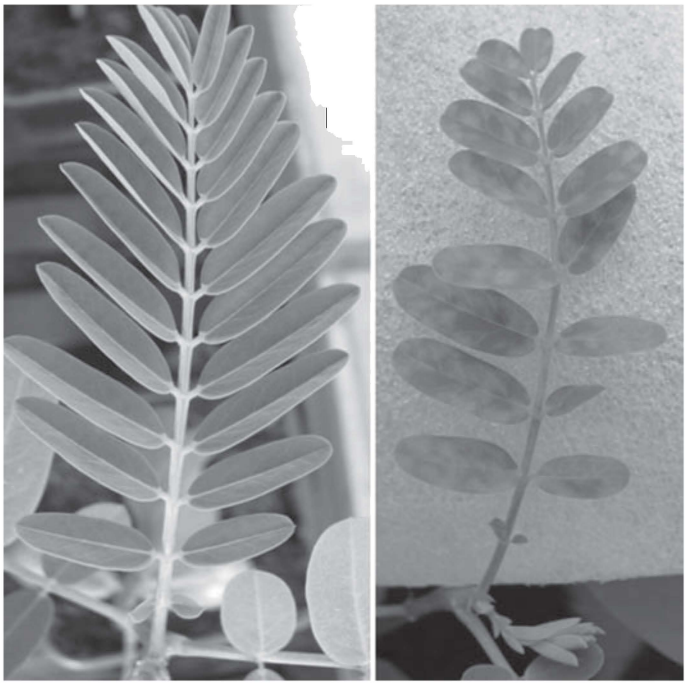


Figure 10.3 Symptoms of SeMV infection. (a) Uninfected and (b) infected *Sesbania grandiflora* leaves.

SeMV polyprotein 2a revealed that it has an alternating arrangement of ordered and disordered domains, with protease and P10 regions being ordered and VPg and P8 intrinsically disordered (Nair and Savithri, 2010b) (Fig. 10.4b). Interestingly, FoldIndex analysis of the polyprotein 2a of all known sobemoviruses and the related luteo and poleroviruses also shows a similar pattern of alternating ordered and disordered domains, indicating the plausible functional importance of such a pattern. FoldIndex analysis of polyprotein 2ab (Fig. 10.4c) showed that VPg and C-terminal 43 residues of RdRp are disordered. The results of FoldIndex analysis of SeMV CP (Fig. 10.4d) are consistent with the structures of sobemoviruses, in which N-terminal domain of CP is disordered.

The following sections discuss the biophysical and functional characterization of the intrinsically disordered proteins VPg and P8. The role of the disordered R domain in controlling the assembly of SeMV is also described.

10.3 INTRINSICALLY DISORDERED VPg MODULATES THE PROTEASE FUNCTION

SeMV VPg is a 9-kDa protein and has a theoretical pI of 4.5. This was the first protein to be identified as intrinsically disordered among all known sobemoviruses

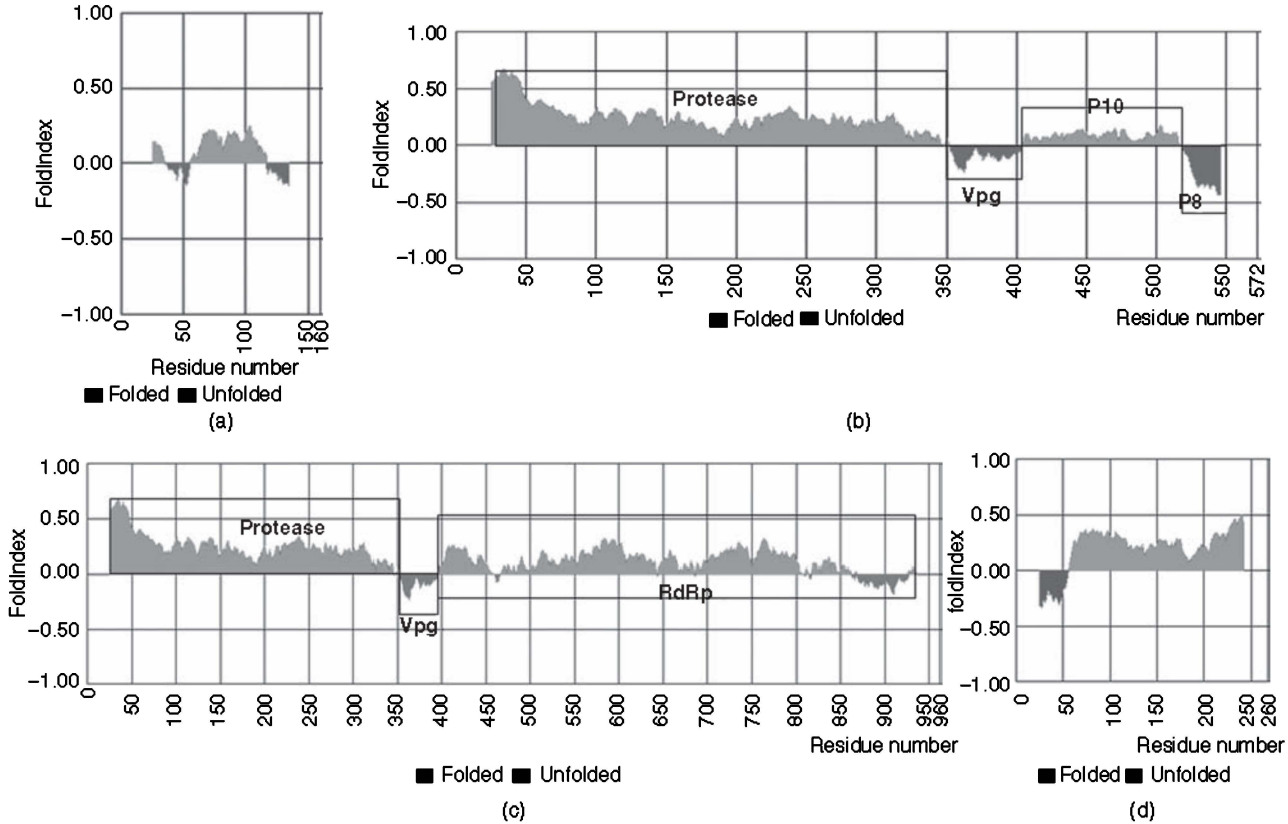


Figure 10.4 Prediction of disordered segments in SeMV-encoded proteins. The disordered regions were predicted using the FoldIndex program. Results of FoldIndex analysis for (a) MP; (b) polyprotein 2a; (c) polyprotein 2ab; and (d) CP are shown.

and related polero and luteoviruses. As shown in the FoldIndex analysis of polyprotein 2a (Fig. 10.4), unlike the protease domain, VPg was completely disordered. Predictions using PONDR also revealed the disordered nature of VPg (Fig. 10.5a). Far-UV CD spectrum provides information on the secondary structure of proteins. The VPg CD spectrum revealed maximum negative ellipticity at 200 nm, a characteristic of random coil (Fig. 10.5b). Fluorescence spectrum of VPg showed maximum emission at 357 nm, which remained unchanged upon the addition of 8*M* urea, confirming that VPg lacked tertiary structure (Fig. 10.5c). In gel filtration studies, VPg eluted at a position corresponding to a higher molecular size, suggesting its open and random conformation (Fig. 10.5d).

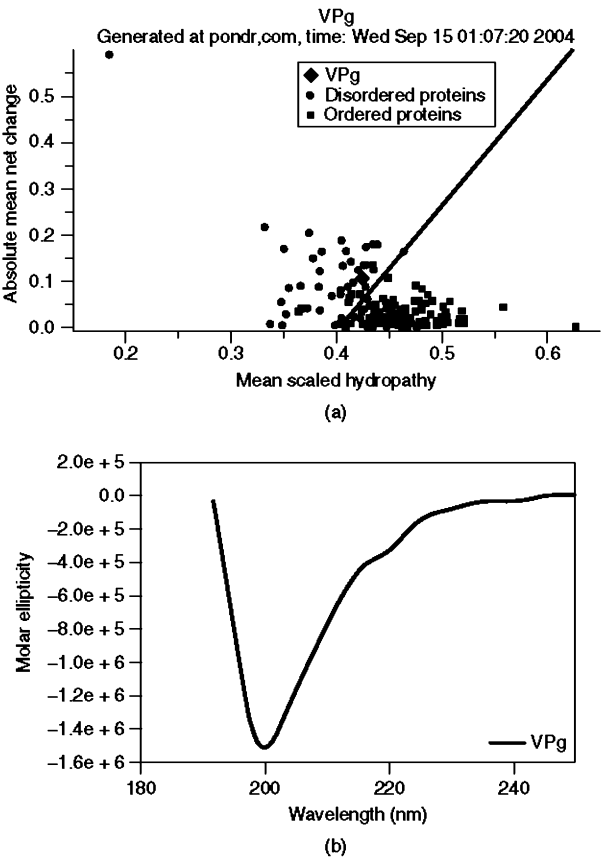
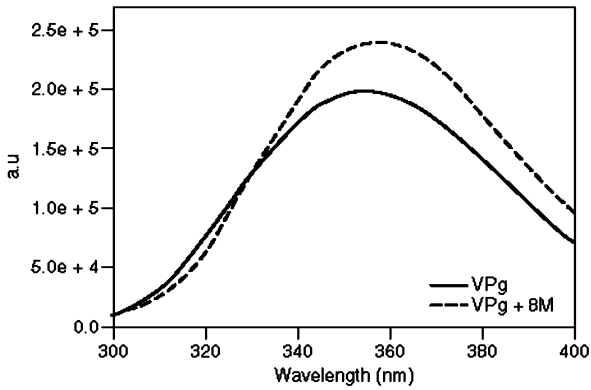
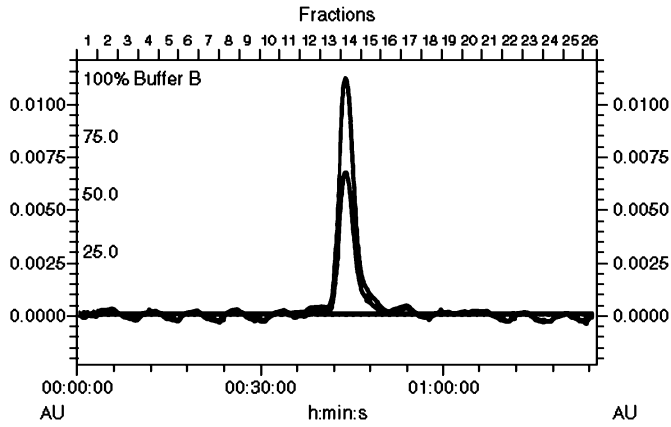


Figure 10.5 Biophysical characterization of VPg. (a) Protein fold prediction obtained using PONDR with default values. (b) Far-UV CD spectral analysis showing the unfolded nature of VPg. The CD spectrum was recorded using 0.5 mg/mL of purified protein from 190 to 250 nm. (c) Intrinsic fluorescence spectrum of VPg. The spectrum was recorded between 300 and 400 nm after excitation at 280 nm under native and denaturing conditions. (d) Gel filtration analysis. VPg was analyzed on a FPLC system (Amersham Biosciences) using a Superdex S-200 analytical gel filtration column.



(c)



(d)

Figure 10.5 (*Continued*)

10.3.1 Functional Role

Intrinsically disordered proteins are known to attain significantly ordered conformations on interaction with their specific targets or ligands. Higher structural flexibility allows them to interact with a larger milieu of targets. This may serve to regulate a number of cellular processes, and therefore, these proteins have an edge over structurally ordered proteins in terms of protein–protein interactions and associated functions.

VPgs of RNA viruses have been shown to serve as protein primers during viral replication. Interestingly, in SeMV, this natively unfolded VPg has a role apart from priming. SeMV VPg was shown to play an important role in the proteolytic activity of the protease (Satheshkumar et al., 2005a). The N-terminal serine protease domain is responsible for processing the polyproteins 2a and 2ab into functional

domains (Satheshkumar et al., 2004a; Nair and Savithri, 2010a). *In vitro* studies demonstrated that the protease domain, when expressed alone, was devoid of any proteolytic activity (Satheshkumar et al., 2005a). However, when it was expressed as a protease-VPg fusion protein, it was active both in cis and trans. Far-UV CD spectral analysis sheds light on the possible mechanism of how this natively unfolded VPg could render the otherwise inactive protease active. The cleavage site mutant of protease-VPg (E325A) exhibited a positive peak at 230 nm, which was absent in the spectrum of either protease or VPg alone (Fig. 10.6). Such a peak at 230 nm might result from the presence of polyproline helix like structures or from aromatic stacking interactions between protease and VPg domains. By systematic deletions from the C-terminus of the protease-VPg and site-specific mutagenesis, it was shown that interaction of a single tryptophan residue at position 43 in VPg domain of protease-VPg contributes significantly to the positive CD peak at 230 nm. The mutation of W43 of VPg to F in protease-VPg not only resulted in the loss of the positive peak (Fig. 10.6) but also abolished the activity of protease-VPg.

The crystal structure of SeMV protease was determined, revealing a characteristic chymotrypsin fold (Gayathri et al., 2006). Examination of the 3D structure revealed a stack of aromatic residues exposed to the solvent (F269, W271, Y315, and Y319) (Fig. 10.7). Mutational analysis of these residues and H275 (although not a part of the exposed stack, but close to W271) was carried out to examine the importance of these residues in protease–VPg interaction and activation of the protease. When the wild-type protease-VPg was expressed in *E. coli*, it was cleaved to protease and VPg domains (Fig. 10.8, lane 1). W271A and H275A mutants of protease-VPg were completely inactive (Fig. 10.8, lanes 3 and 5). These mutants also did not exhibit the positive CD peak at 230 nm (Fig. 10.9). Y315A and Y319A mutants were as active as the wild-type protease-VPg (Fig. 10.8, lanes 6 and 8), and they also retained the positive CD peak at 230 nm (Fig. 10.9). F269A mutant

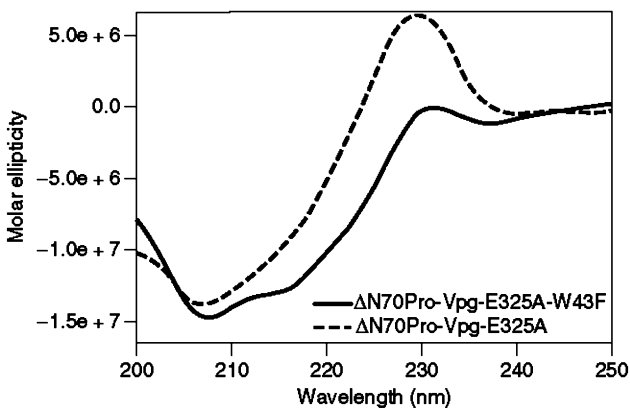


Figure 10.6 Far-UV CD spectrum of protease-VPg-E325A (PVEA) cleavage site mutant. The spectrum was recorded using 0.5 mg/mL of purified protein from 190 to 250 nm. A characteristic positive peak at 230 nm was observed. Also shown is the spectrum of PVEA-W43F mutant where the peak at 230 nm is abolished.

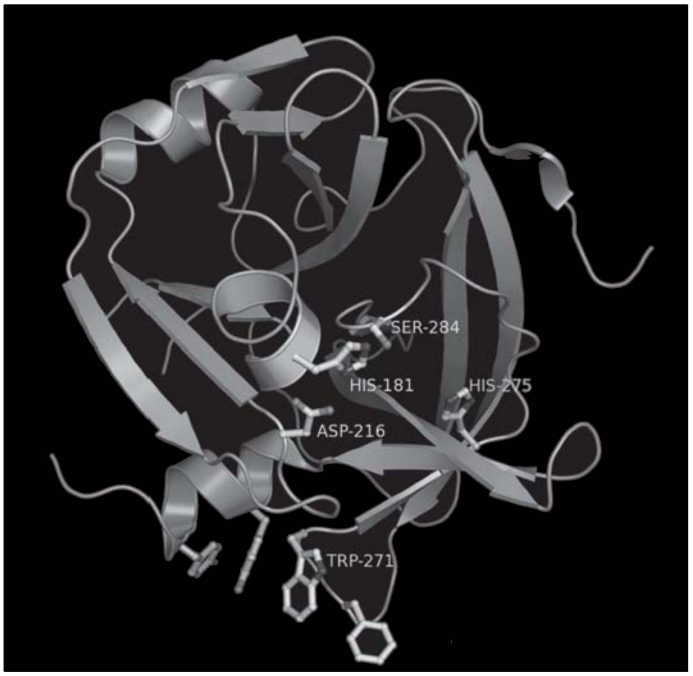


Figure 10.7 Ribbon diagram of SeMV protease domain (green) showing the exposed aromatic amino acids F269, W271, Y315 and Y319, and H275. Active site residues S284, H181, and D216 are also shown (yellow). (*See insert for color representation of the figure.*)

was partially active (Fig. 10.8, lane 4) and showed reduced CD peak at 230 nm (Fig. 10.9). These results demonstrated that W43 of VPg is likely to interact with residues W271 and H275 in the protease domain to mediate this stacking interaction responsible for the activation of the protease domain (Nair et al., 2008). Although the structure of active protease-VPg is not available, based on the biophysical and biochemical studies performed, we envisage alterations in the conformation of the substrate-binding pocket of the protease due to the aromatic stacking interaction of W43 of VPg with W271 and H275 of the protease, such that it can now cleave the substrate. This is also supported by the fact that the H275, which is otherwise stably H-bonded in the substrate-binding pocket of protease, is also the residue that interacts with the W271 of protease and W43 of VPg. The crystal structure of protease shows a well-formed active site and an oxyanion hole; yet, the protease requires VPg as its C-terminal fusion to be catalytically active. Interestingly, addition of purified VPg in trans to the protease cannot render the latter active. This suggests that only when covalently linked, the two domains can interact to bring about conformational changes in the protease so that it can cleave the substrate.

In the case of sobemoviruses, the role of VPg as a primer in RNA replication has not been established yet. But preliminary results from our laboratory suggest that presence of unfolded VPg at the N-terminus of RdRp has an inhibitory effect

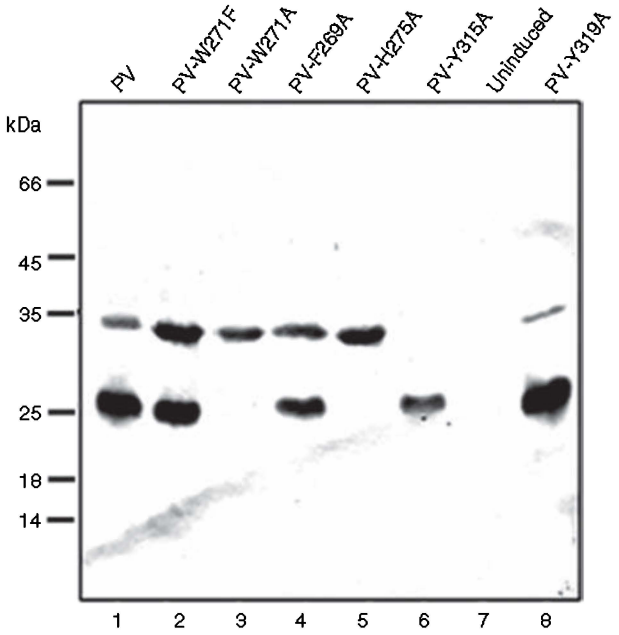


Figure 10.8 *In vivo* cleavage assay. Protease-VPg (PV) and the mutants PV-W271F, PV-W271A, PV-F269A, PV-H275A, PV-Y315A, and PV-Y319A were expressed in *Escherichia coli* BL21 (DE3) pLys S cells and analyzed on 12% SDS-PAGE followed by western analysis using polyclonal antibodies to the protease domain as the primary antibody.

on the *in vitro* polymerase activity of the latter. This is also true of the unfolded region present at the C-terminus of RdRp itself. The deletion of this domain seems to enhance the *in vitro* polymerase activity (Govind and Savithri, unpublished data).

10.4 INTRINSICALLY DISORDERED, NUCLEIC-ACID-BINDING P8 DOMAIN ACTIVATES P10 ATPASE

P8 is released as a result of the cleavage at E498-S499 at the C-terminus of P10 domain of polyprotein 2a (Nair and Savithri, 2010a). Cleavage at this site occurred only in cis, that is when polyprotein 2a was expressed in *E. coli* and not when purified E325A protease-VPg was added in trans to the polyprotein. Also, the cleavage at this site depends on the cleavage at the E132-S133 site (removes the membrane anchor from protease domain), as P8 was not released in the E132A mutant of polyprotein 2a. Thus, the release of this domain seems to be highly regulated both in terms of the order and mode of cleavage. The 8-kDa P8 has a theoretical pI of 11.5 and moves abnormally on SDS-PAGE as a 14-kDa protein owing to its highly basic nature (Fig. 10.10a). It is interesting to note that this highly basic region is present in all the known sobemoviruses, poleroviruses, and also the

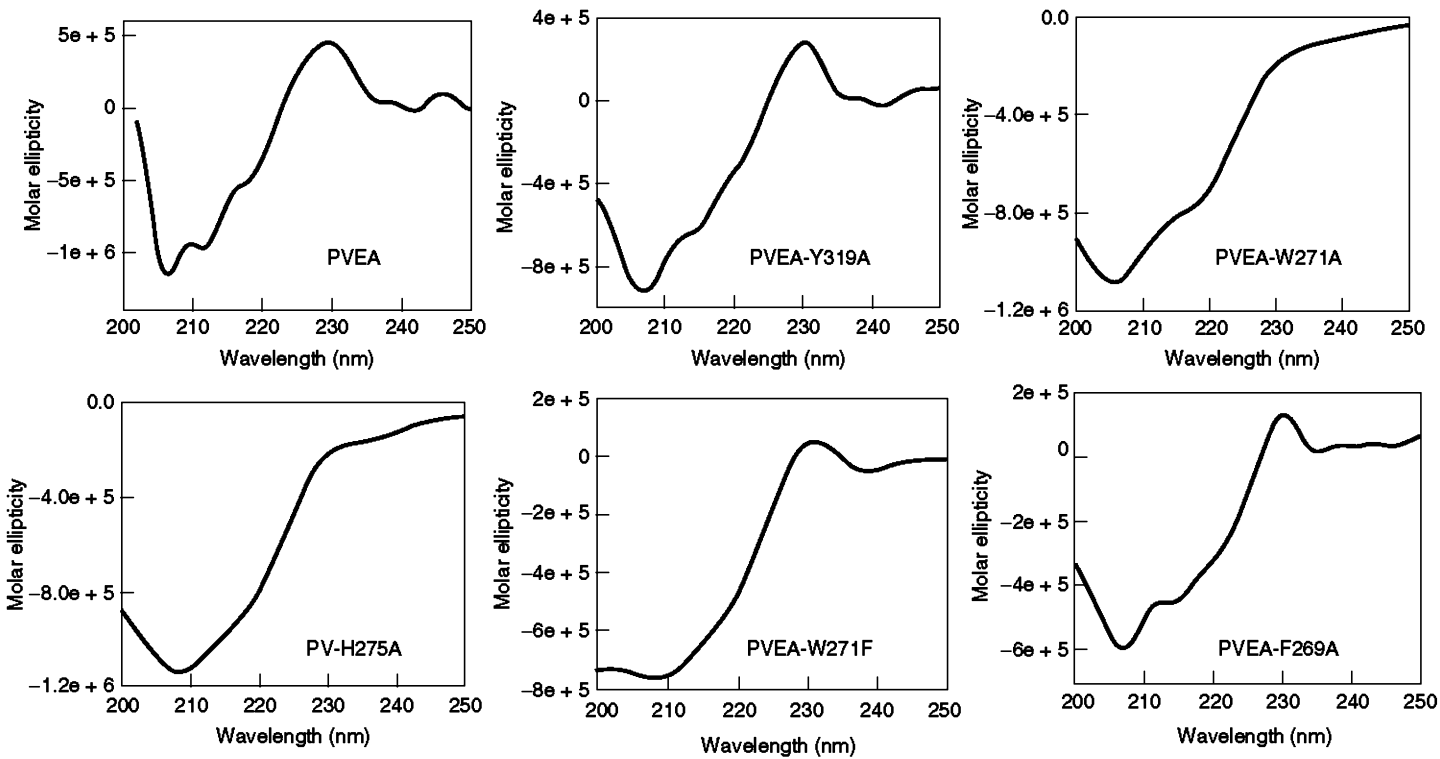


Figure 10.9 Far-UV CD spectra of protease-VPg mutants. The ellipticity was monitored from 200 to 250 nm at 25°C using 0.2 mg/mL of the purified protein in 20 mM Tris HCl pH 8.0. PVEA stands for Protease-VPg-E325A mutant.

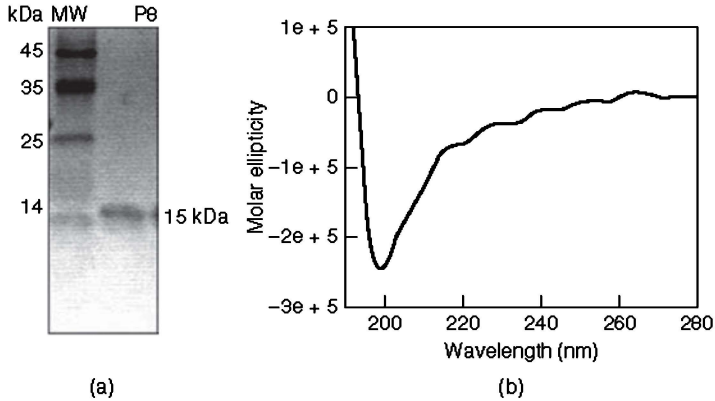


Figure 10.10 Biophysical characterization of P8. (a) Purified His-tagged P8 on a 15% SDS-PAGE moving abnormally at 15 kDa instead of the expected 9.7 kDa. (b) The far-UV CD spectrum of P8 protein. The ellipticity was monitored from 190 to 280 nm using 400 μg/mL of the protein in 20 mM Tris-HCl pH 8.0.

fungi-infecting barnaviruses although the cleavage site for P8 remains to be established in these viruses. The far-UV CD spectrum of P8 revealed a negative ellipticity at 200 nm, suggesting that P8 too was intrinsically disordered (Nair and Savithri, 2010b) (Fig. 10.10b). Like VPg, P8 also showed an abnormal elution profile on the gel filtration column, suggesting its disordered nature.

10.4.1 Functional Role

Owing to the highly basic nature of P8, its ability to bind nucleic acids was investigated. The results revealed that P8 could bind to single- or double-stranded nucleic acid in a sequence-nonspecific manner (Nair and Savithri, 2010b) (Fig. 10.11). Such a nucleic-acid-binding region has also been identified at the C-terminus of VPg in polyprotein P1 of potato leafroll virus (PLRV) (luteovirus) and the polyprotein 2a of CfMV (Dirk Prufer et al., 1999; Tamm and Truve, 2000). Together, these findings suggest that the positively charged, disordered region at the C-terminus of VPg in the polyprotein of these viruses may play an essential role in the viral life cycle. One possibility could be that polyprotein 2a binds to the gRNA via the C-terminal P8 RNA-binding domain, and this might aid in correctly positioning the natively unfolded VPg for nucleotidylylation by RdRp. Analysis of the amino acid sequence of P8 shows that it may have a bipartite nuclear localization signal. It would be interesting to examine if P8 or any of its precursors could enter the nucleus of the host cell and act as transcriptional regulators. P8 or its precursors, because of their natively unfolded structure, might interact with a large number of host factors and modulate their functions.

Recently, one of the precursors of P8, P10-P8/P18 was shown to possess inherent Mg^{2+}-dependent ATPase activity that was inhibited in the presence of poly A (Nair and Savithri, 2010b) (Fig. 10.12a). Although, P10 alone could hydrolyze ATP,

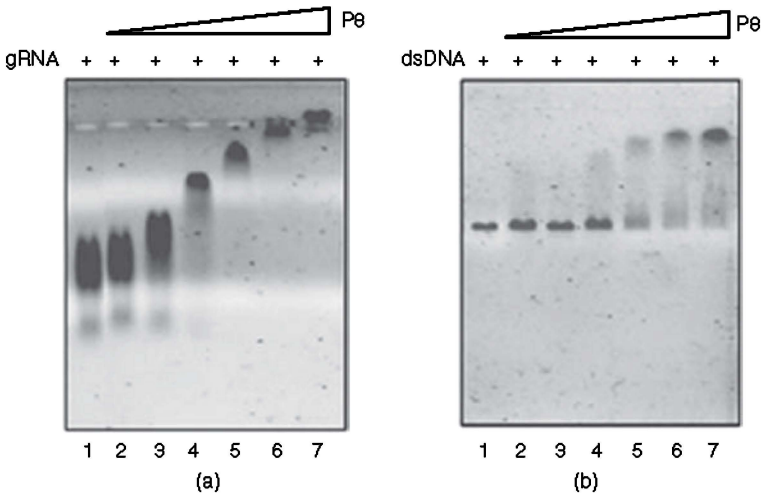


Figure 10.11 Gel retardation assay demonstrating nucleic acid binding by purified P8 protein. Increasing amounts of purified P8 were incubated with (a) genomic RNA and (b) DNA (VPg-P10-P8 PCR product), in the presence of 1X binding buffer (50 mM NaCl, 25 mM Tris-HCl pH 7.4, 25 mM KCl, 2 mM EDTA) at room temperature (RT). The samples were analyzed by agarose gel electrophoresis and stained with EtBr.

in the presence of P8 domain, there was an increase in the k_{cat} of P10 ATPase by 4.3-fold, V_{max} by 2.5-fold, and K_m by 3-fold (Fig. 10.12b). In other words, the ATPase activity of P10 domain was enhanced by the disordered P8 domain. Therefore, like VPg, P8 also seems to regulate the function of its neighboring domain. More intriguing is the fact that both VPg and P8 positively regulate the activities of the domains (protease and P10, respectively), present at their N-terminus. Whereas, VPg has an inhibitory effect on the *in vitro* polymerase activity of RdRp present at its C-terminus (Govind and Savithri, unpublished data). These *in vitro* studies suggest that the virus manages to fine-tune the activity of the various domains

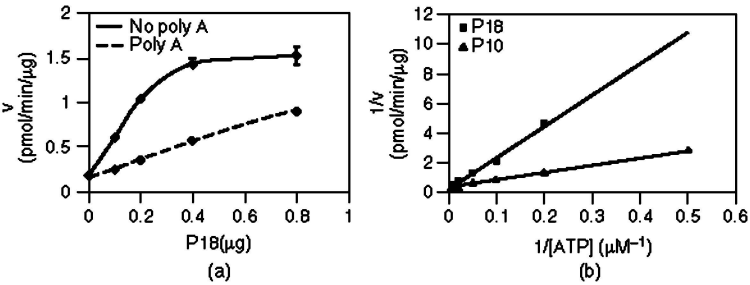


Figure 10.12 ATPase assay. The assay was performed with increasing amounts of purified His-tagged P18 protein. (a) ATPase activity of P18 in the absence and presence of 0.5 mg/mL of poly A. (b) 1/v versus 1/[ATP] Lineweaver–Burk plot for ATPase activities of purified P10 and P18 proteins.

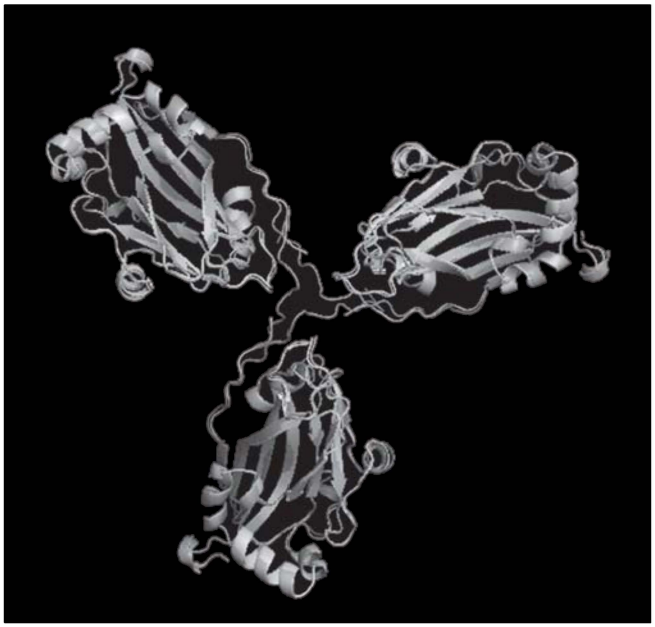


Figure 10.13 Superposition of the C subunits of the native $T = 3$ particles and rCP-Δ 48-59 recombinant capsids where the residues forming the β-annulus were deleted. The native and mutant structures are represented in red and green, respectively. (*See insert for color representation of the figure.*)

by encoding intrinsically disordered regions/domains that are capable of higher flexibility and hence a plethora of interactions.

10.5 INTRINSICALLY DISORDERED SEGMENT OF SeMV CP CONTROLS ASSEMBLY

The polypeptide fold of SeMV CP consists of two domains, the S and the R domain. The S domain mediates protein–protein interactions, and the R domain binds to the nucleic acid and controls the size of the assembled capsids (Lokesh et al., 2002). The R domain encompasses two important motifs, an arginine-rich motif (ARM) and the β-annulus. The ARM is present in several viruses, although its precise position in the polypeptide is variable. ARM is found to be disordered in all the sobemovirus structures determined thus far, suggesting that the flexible nature of this motif is essential for promoting protein–RNA interactions and probably for RNA encapsidation (Choi et al., 2000). In SCPMV and SeMV, the ARM has been shown to possess RNA-binding property (Lee and Anderson, 1998).

10.5.1 Functional Role of the Disordered ARM

Overexpression of the SeMV CP gene in *E. coli* was shown to result in the assembly of $T = 3$ viruslike particles (VLPs) resembling the native particles (Lokesh et al.,

2002). These VLPs encapsidate CP mRNA and *E. coli* 23S rRNA. In the native virus particles, the N-terminal 43 residues are disordered in all the three subunits of the icosahedral asymmetric unit. Residues 44–73 are ordered only in the C subunits (Bhuvaneshwari et al., 1995). Mutational analysis carried out in SeMV CP showed that deletion of 22 amino acids from the N-terminus (CP-ΔN22), which retains the ARM, does not affect the $T = 3$ capsid assembly. However, the CP-ΔN36 mutant, in which the ARM is absent but the sequences corresponding to the β-annulus and the βA strand (residues 67–70) are intact, assembled mostly into $T = 1$ particles. Further, the deletion of the N-terminal 65 amino acid residues (CP-NΔ65) resulted in the exclusive formation of $T = 1$ particles. The three-dimensional structures of the wild-type and the mutant VLPs have been determined, and an analysis of these structures revealed that the disordered N-terminal segment of the CP controls particle assembly (Sangita et al., 2002, 2005a,b; Satheshkumar et al., 2004b). Also, the interaction between the disordered ARM and RNA is crucial for RNA encapsidation. Substitution of all the arginine residues (CP 28-36E) with glutamate in the ARM led to the formation of less stable but empty $T = 3$ capsids (Satheshkumar et al., 2005b).

10.5.2 Disordered Region Folds to Form Structured β-Annulus

In the SeMV capsid, the N-terminal amino acid residues 48–58 of C subunits form the β-annulus structure. The β-annulus is formed by hydrogen-bonding interactions of residues 48–52 from one C subunit with residues 55–58 of the neighboring subunit. Since β-annulus confers quasi-equivalence to the CP subunits, it was considered to be the switch that controls the size and curvature of the assembled capsids (Olson et al., 1983). Surprisingly, rCP-Δ48-59 VLP, which lacked the whole segment forming the β-annulus, assembled into $T = 3$ particles. In the crystal structure of this mutant (Pappachan et al., 2008), the β-annulus was absent (Fig. 10.13). Therefore, the disordered segment corresponding to the β-annulus assumes an ordered structure upon interaction with the C-type CP subunit and hence is a consequence of assembly into $T = 3$ particles. On the basis of an analysis of these structures and biophysical studies, a model for the assembly of the virus was proposed (Savithri and Murthy, 2010).

In conclusion, disordered segments of SeMV-encoded proteins play crucial structural and functional roles and regulate the function of neighboring folded domains.

ACKNOWLEDGMENTS

We thank M. Bhuvaneswari, H. S. Subramanya, K. Gopinath, G. L. Lokesh, T. D. S. Gowri, P. S. Satheshkumar, K. Govind, V. Sangita, P. Gayathri, A. Papacchan, V. Saravanan, C. S. Vijay, S. Chinnathambi, and K. Prasad for their contributions. We acknowledge the Council of Scientific and Industrial Research (CSIR), Department of Biotechnology (DBT), Department of Science and Technology (DST),

Government of India, INDO-Finnish grant, and the Indian Institute of Science, Bangalore, for financial support.

ABBREVIATIONS

Å	Angstrom (S)
ARM	arginine-rich motif
CD	circular dichroism
CP	coat protein
3D	three dimensional
EDTA	ethylenediaminetetraaceticacid
EGTA	ethyleneglycol-bis(aminoethyl ether)-tetraacetic acid
EM	electron microscopy, electron micrographs
FPLC	fast-performance liquid chromatography
gRNA	genomic RNA
MP	movement protein
ORF	open reading frame
P8	protein of size 8 kDa
P10	protein of size 10 kDa
PAGE	polyacrylamide gel electrophoresis
poly A	polyadenylated
PTGS	posttranscriptional gene silencing
RdRp	RNA-dependent RNA polymerase
SatRNA	satellite RNA
SDS	sodium dodecyl sulfate
S	Svedberg sedimentation coefficient
sgRNA	subgenomic RNA
T	triangulation number
UV	ultraviolet
VLP	viruslike particle
VPg	viral protein genome linked

LIST OF VIRUSES

CfMV	cocksfoot mottle virus
LTSV	lucerne transient streak virus
PLRV	potato leaf roll virus
RGMoV	ryegrass mottle virus
RYMV	rice yellow mottle virus
SBMV	southern bean mosaic virus
SCMoV	subterranean clover mottle virus
SCPMV	southern cowpea mosaic virus
SeMV	*Sesbania* mosaic virus

REFERENCES

Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AGW, Rayment I, Rossmann MG, Suck D, Tsukihara T. Structure of southern bean mosaic virus at 2.8 AA resolution. Nature 1980;286:33–39.

Arthur K, Dogra S, Randles JW. Complete nucleotide sequence of Velvet tobacco mottle virus, isolate K1. Arch Virol 2010;155(11):1893–1896.

Bhuvaneshwari M, Subramanya HS, Gopinath K, Savithri HS, Nayudu MV, Murthy MRN. Structure of sesbania mosaic virus at 3 A resolution. Structure 1995;3(10):1021–1030.

Callaway AS, Lommel SA. Turnip rosette virus: a Sobemovirus infecting Arabidopsis thaliana. Direct Submission 2002. Accession No. AY177608.

Chen H, Chen Q, Dong J, Li G, Zhu S, Chen H. Direct Submission 2010. Accession No. GQ845002.

Choi YG, Grantham GL, Rao AL. Molecular studies on bromovirus capsid protein. Virology 2000;270(2):377–385.

Chowdhury SR, Savithri HS. Interaction of Sesbania mosaic virus movement protein with the coat protein: implications for viral spread. FEBS J 2011;278(2):257–272.

Chowdhury SR, Savithri HS. Interaction of Sesbania mosaic virus movement protein with VPg and P10: implication to specificity of genome recognition. PLoS One 2011;6(1):e15609.

Collins RF, Gellatly DL, Sehgal OP, Abouhaidar MG. Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. Virology 1998;241(2):269–275.

Davies C, Haseloff J, Symons RH. Structure, self-cleavage, and replication of two viroid-like satellite RNAs (virusoids) of subterranean clover mottle virus. Virology 1990;177(1):216–224.

Dirk Prufer LK, Michael Monecke SabineNowok, Fisher Rainer, Rohde Wolfgang. Immunological analysis of Potato leafroll luteovirus (PLRV), PI identifies the expression of 25kDa RNA binding domain derived via P1 processing. Nucleic acid Res 1999;27(2):421–425.

Dwyer GI, Njeru R, Williamson S, Fosu-Nyarko J, Hopkins R, Jones RA, Waterhouse PM, Jones MG. The complete nucleotide sequence of Subterranean clover mottle virus. Arch Virol 2003;148(11):2237–2247.

Eastwell KC, Villamor DV, McKinney CL, Druffel KL. Characterization of an isolate of Sowbane mosaic Virus. Arch Virol 2010;155(12):2065–2067.

Forster AC, Jeffries AC, Sheldon CC, Symons RH. Structural and ionic requirements for self-cleavage of virusoid RNAs and trans self-cleavage of viroid RNA. Cold Spring Harb Symp Quant Biol 1987;52:249–259.

Forster AC, Symons RH. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. Cell 1987;49(2):211–220.

Gayathri P, Satheshkumar PS, Prasad K, Nair S, Savithri HS, Murthy MR. Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. Virology 2006;346(2):440–451.

Hull R. The grouping of small spherical plant viruses with single RNA components. J Gen Virol 1977;36:289–295.

Hull, R. The sobemovirus group. Volume 3, In: Koenig R, editor. The plant viruses, polyhedral virions with monopartite RNA genomes. New York: Plenum Press; 1988. pp.113–146.

Hull, R, Fargette, D. Sobemoviruses. In: Osmond B, editor. ICTVdB- The universal virus database version 4. New York: Columbia University; 2006.

Jeffries AC, Rathjen JP, Symons RH. Direct Submission. 1995. Accession No. NC_001696.

Jones AT, Mayo MT. Satellite nature of the viriod-like RNA-2 of Solanumnodiflorum mottle virus and the ability of other plant viruses to support the replication of viriod-like RNA molecules. J Gen Virol 1984;65:1713–1721.

Lee L, Anderson EJ. Nucleotide sequence of a resistance breaking mutant of southern bean mosaic virus. Arch Virol 1998;143(11):2189–2201.

Lokesh GL, Gopinath K, Satheshkumar PS, Savithri HS. Complete nucleotide sequence of Sesbania mosaic virus: a new virus species of the genus Sobemovirus. Arch Virol 2001;146(2):209–223.

Lokesh GL, Gowri TD, Satheshkumar PS, Murthy MR, Savithri HS. A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. Virology 2002;292(2):211–223.

Makinen K, Tamm T, Naess V, Truve E, Puurand U, Munthe T, Saarma M. Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. J Gen Virol 1995;76(Pt 11):2817–2825.

McGavin W, Wright F, MacFarlane S. Rubus chlorotic mottle virus, a new sobemovirus infecting raspberry and bramble. Virus Res 2009;139(1):10–13.

Nair S, Gayathri P, Murthy MRN, Savithri HS. Stacking interactions of W271 and H275 of SeMV serine protease with W43 of natively unfolded VPg confer catalytic activity to protease. Virology 2008;382:83–90.

Nair S, Savithri HS. Processing of SeMV polyproteins revisited. Virology 2010a;396(1): 106–117.

Nair S, Savithri HS. Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. FEBS Lett 2010b;584(3):571–576.

Olson AJ, Bricogne G, Harrison SC. Structure of tomato busy stunt virus IV. The virus particle at 2.9 Å resolution. J Mol Biol 1983;171(1):61–93.

Opalka N, Tihova M, Brugidou C, Kumar A, Beachy RN, Fauquet CM, Yeager M. Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction. J Mol Biol 2000;303(2):197–211.

Paliwal YC. Interaction of the viriod-like RNA 2 of lucerne transient streak virus with southern bean mosaic virus. Can J Plant Pathol 1984;6:93–97.

Pappachan A, Subashchandrabose C, Satheshkumar PS, Savithri HS, Murthy MRN. Structure of recombinant capsids formed by the beta-annulus deletion mutant—rCP (Delta 48-59) of Sesbania mosaic virus. Virology 2008;375(1):190–196.

Plevka P, Tars K, Zeltins A, Balke I, Truve E, Liljas L. The three-dimensional structure of ryegrass mottle virus at 2.9 Å resolution. Virology 2007;369(2):364–374.

Sangita V, Lokesh GL, Satheshkumar PS, Saravanan V, Vijay CS, Savithri HS, Murthy MR. Structural studies on recombinant T = 3 capsids of Sesbania mosaic virus coat protein mutants. Acta Crystallogr D Biol Crystallogr 2005a;61(Pt 10):1402–1405.

Sangita V, Parthasarathy S, Toma S, Lokesh G, Gowri T, Satheshkumar P, Savithri H, Murthy M. Determination of the structure of the recombinant T = 1 capsid of Sesbania mosaic virus. Curr Sci 2002;82:1123–1131.

Sangita V, Satheshkumar PS, Savithri HS, Murthy MR. Structure of a mutant T = 1 capsid of Sesbania mosaic virus: role of water molecules in capsid architecture and integrity. Acta Crystallogr D Biol Crystallogr 2005b;61(Pt 10):1406–1412.

Satheshkumar PS, Gayathri P, Prasad K, Savithri HS. "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. J Biol Chem 2005a;280(34): 30291–30300.

Satheshkumar PS, Lokesh GL, Murthy MR, Savithri HS. The role of arginine-rich motif and beta-annulus in the assembly and stability of Sesbania mosaic virus capsids. J Mol Biol 2005b;353(2):447–458.

Satheshkumar PS, Lokesh GL, Sangita V, Saravanan V, Vijay CS, Murthy MR, Savithri HS. Role of metal ion-mediated interactions in the assembly and stability of Sesbania mosaic virus T = 3 and T = 1 capsids. J Mol Biol 2004b;342(3):1001–1014.

Satheshkumar PS, Lokesh GL, Savithri HS. Polyprotein processing: cis and trans proteolytic activities of Sesbania mosaic virus serine protease. Virology 2004a;318(1): 429–438.

Savithri HS, Murthy MRN. Structure and assembly of Sesbania mosaic virus. Curr Sci 2010;98(3):346–351.

Sehgal OP, Sinha RC, Gellatly DL, Ivanov I, AbouHaidar MG. Replication and encapsidation of the viriod like satellite RNA of lucerne streak virus are supported in divergent hosts by cockfoot mottle virus and turnip rosette virus. J Gen Virol 1993;74:785–788.

Sereme D, Lacombe S, Konate M, Pinel-Galzi A, Traore VS, Hebrard E, Traore O, Brugidou C, Fargette D, Konate G. Biological and molecular characterization of a putative new sobemovirus infecting Imperata cylindrica and maize in Africa. Arch Virol 2008;153(10):1813–1820.

Solunke BS, Eranna S, Nayudu MV. Sesbania mosaic-a new virus disease. Indian Phytopathology 1983;36:568–570.

Subramanya HS, Gopinath K, Nayudu MV, Savithri HS, Murthy MRN. Structure of Sesbania mosaic virus at 4.7 Å resolution and partial sequence of the coat protein. J Mol Biol 1993;229(1):20–25.

Tamm T, Truve E. RNA-binding activities of cocksfoot mottle sobemovirus proteins. Virus Res 2000;66(2):197–207.

Tars K, Zeltins A, Liljas L. The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution. Virology 2003;310(2):287–297.

Wu SX, Rinehart CA, Kaesberg P. Sequence and organization of southern bean mosaic virus genomic RNA. Virology 1987;161(1):73–80.

Yassi MN, Ritzenthaler C, Brugidou C, Fauquet C, Beachy RN. Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. J Gen Virol 1994;75(2):249–257. Pt

Zaumeyer WJ, Harter LL. Inheritance of symptom expression of bean mosaic virus 4. J Agric Res 1943;67:295–300.

Zhang FY, Toriyama S, Takahashi M. complete nucleotide sequence of ryegrass mottle virus: A new species of the genus sobemovirus. J Gen Plant Pathol 2001;67:63–68.

11

INTRINSIC DISORDER IN GENOME-LINKED VIRAL PROTEINS VPgs OF POTYVIRUSES

JADWIGA CHROBOCZEK, EUGÉNIE HÉBRARD, KRISTIINA MÄKINEN, THIERRY MICHON, AND KIMMO RANTALAINEN

11.1 INTRODUCTION

In a supergroup of picornaviruslike RNA viruses, the genus *Potyvirus* (family Potyviridae) with more than 110 species is the biggest genus of plant-infecting viruses. Potyviruses cause diseases in crop plants that result in significant economic losses (for review on potyviruses see Shukla et al. (1994)). Their single-stranded messenger-polarity RNA genome of about 9.7 kb contains at the 5′ end a covalently attached viral genome linked (VPg) protein and a poly(A) tail at the 3′ end (Fig. 11.1a). The viral RNA contains one large open reading frame (ORF) encoding a polyprotein that is cleaved into 10 functional proteins by three virus-encoded proteinases (Merits et al., 2002). Recently, an eleventh protein, PIPO (pretty interesting Potyviridae ORF), which is encoded by a short ORF embedded within the P3 cistron of the polyprotein, was found (Chung et al., 2008). In the polyprotein, the VPg of picornaviruslike viruses in families of Picornaviridae, Potyviridae, Comoviridae, Caliciviridae, and possibly also Sequiviridae is located upstream of the chymotrypsinlike cysteine proteinase domain. In the genera *Sobemovirus* (not assigned to any family), *Polerovirus*, and *Enamovirus* (family Luteoviridae), the VPg is located downstream of the chymotrypsinlike serine proteinase in the viral genome. In the first situation, VPg forms the N-terminal part of VPg-proteinase (VPg-Pro, also called NIa), which is one of the first products released from the

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

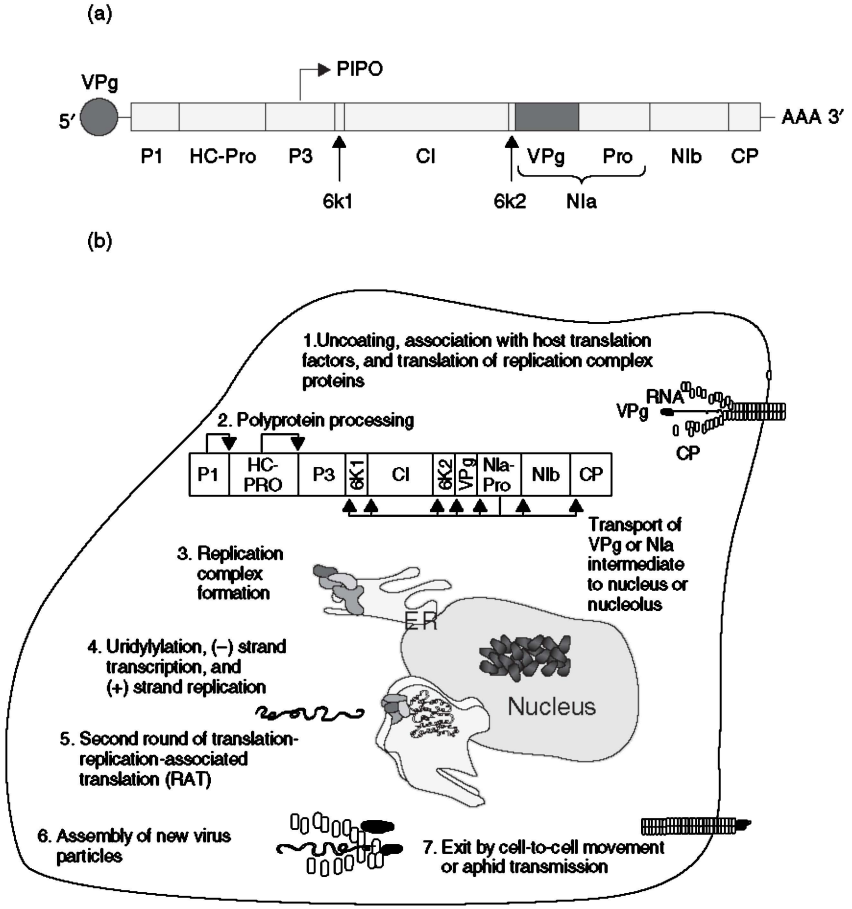


Figure 11.1 Potyviral infection cycle from the VPg perspective. (a) Potyvirus genome. (b) VPg is involved in all essential phases of the infection cycle—translation, replication and movement. *Source*: Modified from PhD thesis of K. Rantalainen (http://urn.fi/URN:ISBN:978-952-10-6014-4).

polyprotein (Fig. 11.1). Progress of the viral infection requires that the cleavage between VPg and Pro is slow, which is achieved via a suboptimal amino acid sequence at the cleavage site (Schaad et al., 1996). Maturation of VPg results in a protein of approximately 183–194 residues (22–26 kDa). Processing intermediate 6K2-VPg-Pro mediates the VPg binding to membranes at sites of RNA replication (Restrepo-Hartwig and Carrington, 1994; Léonard et al., 2004). In addition to the VPg-Pro and free VPg forms, several phosphorylated forms of VPg can be found in infected cells (Hafren and Mäkinen, 2008). The potyviral VPg contains an N-terminally located two-partite nuclear localization signal (NLS) (Carrington et al., 1991), and indeed, the VPg-Pro protein is translocated to the host cell nucleus and nucleolus (Restrepo et al., 1990; Carrington et al., 1991; Schaad et al., 1996;

Beauchemin et al., 2007; Rajamäki and Valkonen, 2009). Mutations in the NLS sequence are critical for RNA replication and virus accumulation in infected plant tissues and/or systemic movement in plants (Schaad et al., 1996, Rajamäki and Valkonen, 2009). It seems that the presence of VPg in nucleus might be related to the suppression of gene silencing (Rajamäki and Valkonen, 2009). At the end of the infection cycle, VPg or NIa fusion protein is packed into the virions alongside the viral RNA (Urcuqui-Inchima et al., 2001), where it is externally located, thus capable of taking part in interactions with the host, aphid vectors, and other viral proteins (Puustinen et al., 2002). VPg protein is required for potyvirus infectivity. The protein is covalently linked to the 5′ terminus of the viral genomic RNA via a conserved Tyr residue (Murphy et al., 1996; Oruetxebarria et al., 2001; Anindya et al., 2005). Its removal or substitution is lethal for the virus. Moreover, VPg removal by treatment of genomic RNA with proteinase K abolishes viral RNA infectivity (Herbert et al., 1997). The VPg protein interacts with potyviral RNA polymerase NIb and is involved in the genome replication (Hong et al., 1995; Schaad et al., 1996; Li et al., 1997; Fellers et al., 1998). It was shown that the NIb polymerase attaches the α-phosphate of uridylic acid to VPg in a uridylylation reaction (Puustinen and Mäkinen, 2004) and that the resulting VPg-UMP probably acts as a primer for viral RNA synthesis, similarly as described in poliovirus and human norovirus (Paul et al., 1998; Belliot et al., 2008). In potato virus A (PVA) VPg, the fragment between residues 38 and 44 binds nucleotide triphosphates (NTPs), and its deletion debilitates VPg uridylylation *in vitro* (Puustinen and Mäkinen, 2004). In addition, VPg plays a role in the potyvirus movement, and it is a determinant of potyvirus systemic infection in certain host plants (Schaad et al., 1997; Keller et al., 1998; Rajamäki and Valkonen, 1999, 2002; Borgstrøm et al., 2001; Lellis et al., 2002).

Several studies have reported the interaction between the potyvirus VPg protein and the host eukaryotic translation initiation factor eIF4E, which is the mRNA 5′ cap-binding protein. After the first demonstrations of an interaction between the eIF4E and eIF(iso)4E of *Arabidopsis thaliana* and turnip mosaic virus (TuMV) VPg (Wittmann et al., 1997; Léonard et al., 2000), this interaction has received a lot of attention since it seems to be crucial for a productive virus cycle; natural plant resistance to potyvirus infection has been shown to stem from the inability of VPg to interact with eIF4E, as a result of amino acid polymorphisms either in VPg (Borgstrøm and Johansen, 2001) or in eIF4E or eIF(iso)4E (Duprat et al., 2002; reviewed in Ruffel et al., 2002; Robaglia and Caranta, 2006; Charron et al., 2008). Indeed, disruption of genes encoding eIF4E or eIF(iso)4E leads to loss of susceptibility to several potyviruses (Duprat et al., 2002; Lellis et al., 2002; Sato et al., 2005; Decroocq et al., 2006).

The molecular mechanism of eIF4E involvement in potyvirus multiplication is actively studied. It has been shown that the presence of tobacco etch virus (TEV) VPg stimulates the *in vitro* translation of uncapped internal ribosome entry site (IRES) containing RNA (viral messengers), whereas translation of capped RNA (host messengers) is inhibited (Khan et al., 2006) most probably by the removal of eIF4E from the available pool through interaction with VPg. The inhibition of

in vitro translation was also demonstrated in potato virus Y (PVY) and TuMV VPgs (Cotton et al., 2006; Grzela et al., 2006; Miyoshi et al., 2008). Another role of VPg/eIF4E interaction in the virus cell-to-cell movement has been suggested for the pea seed-borne mosaic virus (PSbMV) VPg (Gao et al., 2004). In this hypothesis, VPg employs known interactions of the cytoskeletal structures with the translational machinery (Bokros et al., 1995). Recent findings have revealed that eIF(iso)4F and eIF4F, which are binary complexes between the coupled initiation factors eIF(iso)4E/eIF(iso)4G or eIF4E/eIF4G, respectively, interact with TuMV and lettuce mosaic virus (LMV) VPgs (Khan et al., 2006, 2008; Michon et al., 2006). In this interaction, it has been suggested that eIF4G recruitment is mediated through VPg–eIF4E association (Michon et al., 2006).

Potyvirus VPgs interact also with other host proteins. VPg of TuMV has been shown to interact with a plant protein called Potyvirus VPg-interacting protein (PVIP) through the first 16 amino acids of the VPg N-terminus (Dunoyer et al., 2004). A reduced expression of PVIP has been observed to diminish the susceptibility to TuMV infection, and knocking down the interaction with VPg resulted in reduced virus cell-to-cell and systemic movements (Dunoyer et al., 2004). In addition, it has been found that the PVA VPg can be phosphorylated by plant kinases (Ivanov et al., 2001). It was hypothesized that this could trigger the disassembly of infecting virions and the subsequent initiation of potyvirus protein synthesis in the infected cell (Puustinen et al., 2002). The plum pox virus (PPV) VPg was shown to interact with two DEAD-box helicaselike proteins from peach and *Arabidopsis*, respectively, and this interaction is necessary for infection with PPV and also TuMV (Huang et al., 2010). Finally, the PVA VPg is able to interact with fibrillarin, a nucleolar protein, and its depletion reduces PVA accumulation (Rajamäki and Valkonen, 2009).

All these data show that potyviral VPg is a virulence factor and that it is a multifunctional protein. We demonstrate below that the structure of VPgs from several potyviruses has the characteristics of intrinsically disordered proteins (IDPs), which may help to explain the multiple functions of VPg in potyvirus life cycle. A large degree of flexibility would allow these IDPs to combine high specificity with low affinity to ensure faster association and dissociation rates, enable the binding of numerous structurally distinct targets, and provide the ability to overcome steric restrictions, thus leading to larger surfaces of interaction between partners (Dunker et al., 2001; Uversky, 2002, Tompa, 2005). It is relevant in this respect that adaptation and survival of viruses that contain only restricted genetic information depend on the multifunctionality of their components. Viruses have evolved to maximize their genomic information by producing multifunctional proteins.

11.2 EXPERIMENTAL PROBING OF INTRINSIC DISORDER IN POTYVIRAL VPgs

The studies that led to the discovery of the unstructured nature of PVY VPg were carried out with the VPg sequence of 188 amino acids derived from the polyprotein of PVY strain O, British isolate (accession number Z29526), causing mild symptoms (mottle mosaic, chlorosis). In order to be able to analyze its interaction

with the plant initiation factor eIF4E, the recombinant PVY VPg was prepared. During manipulations of VPg, some interesting features were observed, suggesting the unstructured nature of the protein. For example, it could not be easily precipitated and showed an unusual thermal stability. Multiple attempts to crystallize the PVY (similarly as LMV) VPg were unsuccessful. Furthermore, the protein samples gave quite a low absorbance signal at 280 nm related to its low content of tryptophan, an order-promoting residue.

Before any experimental evidence on the unstructured nature of potyviral VPg existed, data on the structure of two other, albeit quite small, VPgs suggested that the unordered state may be a common feature of this class of proteins. NMR studies on cowpea mosaic virus (CPMV) VPg implied absence of an ordered structure (van de Ven et al., 1990). Similarly, CD and fluorescence spectroscopic analyses performed on *Sesbania* mosaic virus (SeMV, genus *Sobemovirus*) VPg indicated a lack of tertiary structure (Satheshkumar et al., 2005). However, as CPMV VPg and SeMV VPg consist of 28 and 77 amino acid residues, respectively, no general conclusions could be drawn from such a small sample of unrelated sequences and rather short proteins when compared to potyviral VPgs.

11.2.1 Effect of Temperature on VPg

Structured proteins are generally easily denatured by heat, while IDPs are rather insensitive to heat denaturation (Receveur-Brechot et al., 2006). PVY VPg remained soluble on heating at 100°C, while its partner eIF4E (which is a highly structured globular protein) was completely aggregated, implying denaturation, when maintained at 100°C for 15 min (Fig. 11.2).

Slow unfolding or denaturation due to changes in pH or temperature can be monitored by dynamic light scattering (DLS). The protein melting point temperature T_m obtained from DLS analysis is indicative of thermal stability. The change in mean particle size that accompanies protein denaturation was measured using a 12–65°C thermal gradient. The reference globular protein eIF4E produced a typical denaturation profile with a slow increase in particle size beginning at about 55°C, followed by a rapid increase above 60°C, while the PVY VPg preparation did not show any signs of denaturation (Fig. 11.2.) It is noteworthy that such a high heat resistance is not a common property of all potyviral VPg structures. For example, the T_m for disruption of heat-sensitive secondary and tertiary structures of PVA VPg is about 42°C (Rantalainen et al., 2008). These changes may be attributable for dissociation of dimers or changes in the structured environment of the monomer (Section 2.2). Although a clear increase in the proportion of unfolded structure is obvious, a portion of the secondary structures is retained even at 80°C and the protein remains soluble, similar to PVY VPg. The heat denaturation of PVA VPg is reversible, which is typical of IDPs.

11.2.2 VPg Oligomerization

Some of the IDPs show a propensity for dimerization (Alonso et al., 2002; Sigalov et al., 2004). A 49-kDa protein could be detected in disrupted, RNase-treated PVA

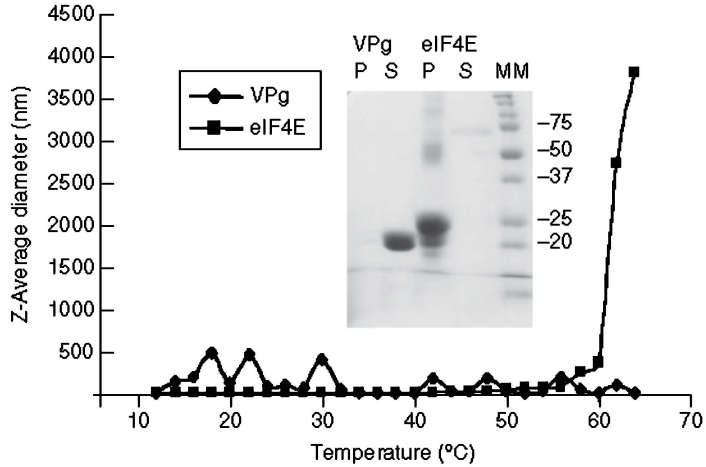


Figure 11.2 Thermal stability of PVY VPg. Dynamic light scattering analysis of VPg (1 mg/mL) and eIF4E (0.1 mg/mL) was done as described in Grzela et al. (2008) Inset: samples of VPg and eIF4E (2 μg each) were boiled at 100°C for 15 min, and the resulting pellets (P) and supernatants (S) were recovered by centrifugation and analyzed by SDS-PAGE followed by staining with CBB.

virions with anti-VPg antibodies but not with antibodies to NIa-Pro, and it was concluded that it possibly represents a dimeric form of VPg (Oruetxebarria et al., 2001). In the same study, analysis with yeast two-hybrid system showed the capacity of PVA VPg for self-interaction. PVA VPg oligomerization was also demonstrated with asymmetric flow field-flow fractionation (AF4), a method suitable for separation of biopolymers, proteins, macromolecules, and particles of different kind (viruses, liposomes, even small bacteria), that is done with the aid of different force fields applied (e.g., flow field). The separation range for particles is up to 100 μm. The advantage of AF4 method over size exclusion chromatography is that no matrix is needed for separation. When recombinant PVA VPg preparations that differ in their storage time and protein concentrations were subjected to AF4 separation, monomers, dimers, and higher oligomers were detected (Moldenhauer, Klein, Rantalainen and Mäkinen, unpublished data). On the basis of the retention times, the size of PVA VPg monomer was determined to be 23 kDa; dimer, 49 kDa; and the higher aggregate, 85 kDa. A two-month-old preparation stored at 4°C contained 85% of dimers. Surprisingly, more monomers were detected in a more concentrated preparation of PVA VPg (1.2 mg/mL) than in a low concentration preparation (0.2 mg/mL). The propensity for oligomerization has been observed also for VPg proteins of clover yellow vein virus (ClYVV) and TuMV (Yambao et al., 2003; Beauchemin et al., 2007).

When PVY VPg was prepared under nonreducing conditions, in addition to a major band (22 kDa) corresponding to the monomeric protein, a slowly migrating band of 44 kDa, consistent with the molecular weight calculated for PVY VPg

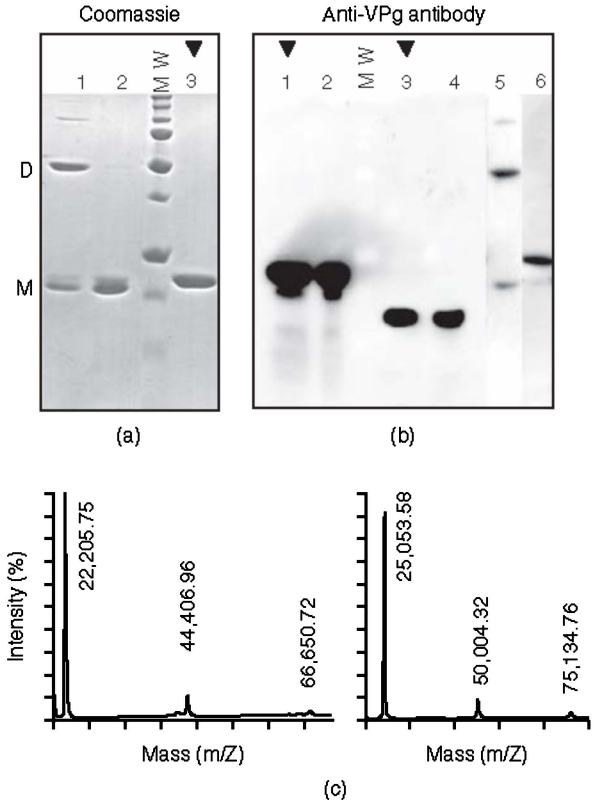


Figure 11.3 PVY VPg dimerization. VPg, C-terminal deletion mutant VPg139, and his VPg (Grzela et al., 2006) were analyzed by SDS-PAGE and stained with CBB (a) or revealed with anti-VPg antibody (b). (A) Lanes 1 and 2, nondenatured VPg (3 μg); lane 2, VPg treated with 5 mM DTT. Lane 3, denatured VPg (3 μg) treated with 50 mM NEM. (b) Lanes 1 and 2,VPg (3 μg); lane 3 and 4, VPg139 (1.5 μg); lane 5, VPg (2 μg); and lane 6, HF (High Five, *Trichoplusia ni*) cells (10^5) expressing hisVPg. Samples in lanes 1 and 3 were denatured and treated with NEM (marked by ▼), and in lanes 2, 4, 5, and 6, SDS was omitted in Laemmli solution. (c) MALDI-TOF mass spectrometry analysis performed for recombinant VPg and his-tagged VPg.

dimer, could be seen on semidenaturing SDS-PAGE (Fig. 11.3a, lane 1 and with MALDI-TOF mass spectrometry analysis (Fig. 11.3c)). The same preparation run on denaturing PAGE showed only one band migrating at 22 kDa (Fig. 11.3a, lane 3). The amount of the 44-kDa form increased with the time of storage at 4°C but disappeared when VPg was incubated with 5 mM dithiothreitol (DTT) (Fig. 11.3a, lane 2). When *N*-ethylmaleimide (NEM) was added at the beginning of VPg purification, the protein appeared in a monomeric form only (Fig. 11.3b, lanes 1 and 2). NEM forms covalent bonds with sulfhydryl groups of cysteines and prevents S–S bond formation. Similar gel analysis performed on VPg C-terminally

truncated at residue 139, and thus devoid of the unique cysteine 150, showed only one band migrating under nonreducing conditions with the molecular weight of the monomer (Fig. 11.3b, lanes 3 and 4). Moreover, no VPg dimers were observed in freshly expressing insect cells (Fig. 11.3b, lane 6, his-tagged VPg is somewhat retarded in comparison with untagged protein). It seems thus that the VPg dimer arises during VPg purification or manipulation, when atmospheric oxygen induces formation of the disulfide bond.

Appearance of dimers because of the formation of disulfide bridges may have an impact on VPg functionality. For example, cysteine oxidation prevents p53 dimerization and hence inhibits DNA binding (Delphin et al., 1994). In some redox-responsive proteins, the formation of disulfide bonds modulates protein–protein interactions; on oxidation, the *Escherichia coli* chaperone Hsp33 undergoes a large conformational transition to a state that can bind polypeptide substrates and rescue them from aggregation (Graumann et al., 2001). To explore whether VPg dimerization affects the interaction with its host partner, eIF4E, we studied the complex formation using monomeric and dimeric VPgs in parallel. No difference was observed in the interaction of DTT-treated VPg (monomeric form) and older VPg (dimeric form), with glutathione S-transferase (GST)-eIF4E immobilized on a GST column (results not shown). This suggests that the VPg dimerization regions are distinct from those implicated in its interaction with eIF4E. Note that the N-terminal part of TuMV and PVY VPgs seems to interact with host eIF4E (Léonard et al., 2000; Grzela et al., 2006), while the C-terminal part of Yellow Vein *Virus* (CYVV) VPg is involved in dimerization (Yambao et al., 2003), which is in agreement with our results for PVY VPg (Fig. 11.3b). Whether disulfide brige formation in potyvirus VPgs is related to the infection needs to be determined via a mutation of the sole cysteine of VPg in infectious cDNA and subsequent infection analysis.

11.2.3 Hydrodynamic Properties of Potyviral VPg

11.2.3.1 Analytical Ultracentrifugation (AUC) Sedimentation velocity experiments were performed to evaluate the shape and oligomeric state of the PVY VPg under nondenaturating conditions (Fig. 11.4). As described earlier, VPg has a tendency to oligomerize with prolonged storage. VPg thawed after storage at −20°C (VPgM + D sample) has a maximum mean $s_{20,w}$ of 2.5 S. Protein stored for two weeks at 4°C (VPg) and believed to be a dimer (VPgD sample) has a larger mean $s_{20,w}$ value of 3.0 S, while VPg purified in the presence of DTT (monomer, Fig. 11.4) has the smallest $s_{20,w}$ mean value of 2.0 S (VPgM). The three sets of data were also modeled as a mixture of noninteracting monomer and dimer species (Table 11.1). Following this approach, the molar mass is fixed, while the sedimentation coefficient for monomer and dimer, as well as the proportion of both species in different samples, are fitted. The results are compatible with the presence of about 90% monomer in the VPgM sample, about 90% dimer in the VPgD sample, and equal amounts of the monomer and dimer in the VPgM + D sample. The results from the monomer/dimer mixture analysis gives the $s_{20,w}$ values 1.7 and 3.0 S for VPg monomer and dimer, respectively, that are rather small compared

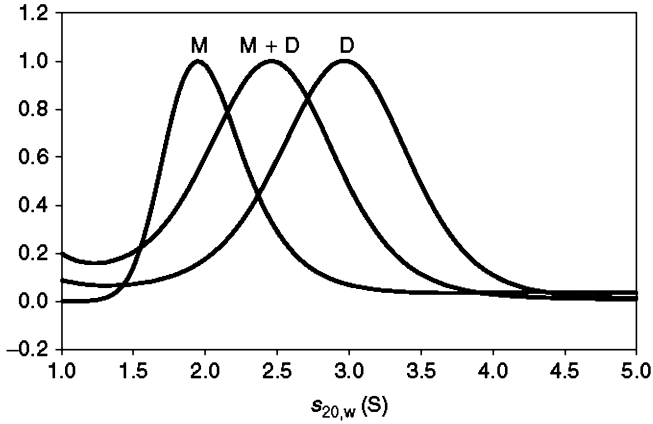


Figure 11.4 Analytical ultracentrifugation $c(s)$ analysis of PVY VPg. $c(s)$ curves renormalized up to maximum value of one. Samples are described in the text.

TABLE 11.1 Hydrodynamic Parameters of Potyviral VPgs Obtained by Analytical Ultracentrifugation (AUC) and Size Exclusion Chromatography (SEC)

	PVY VPg			LMV VPg			
Molecular Species	MM (kDa)	$s_{20,w}$ AUC	$s_{20,w}$ Glob Calc[a]	MM (kDa)	R_S (Å) SEC	$s_{20,w}$ Deduced	$s_{20,w}$ Glob Calc[a]
Monomer	22	1.7	2.1	26	27.9	2.25	2.66
Dimer	44	3	3.4	52	32.3	3.9	4.14

[a]Glob Calc stands for values calculated for globular proteins of identical molecular weight.

to values calculated for globular compact species of comparable molecular mass, namely, 2.1 and 3.4 S. These results thus imply an elongated structure for both the monomeric and the dimeric VPg. Other IDPs with extended or elongated shape have been described, such as C-terminal domain of the measles virus nucleoprotein or N-terminal domain of HPV16 E7 protein (Bourhis et al., 2005; García-Alai et al., 2007).

11.2.3.2 Size Exclusion Chromatography (SEC) In the mobile phase of a chromatographic system, macromolecules drag a "coat" of solute molecules in their motion. This "coat" added to the volume of the macromolecule itself can be assumed to be a hydration sphere. IDPs shape often diverge from the ideal sphere assumed for globular proteins. Consequently, the resulting hydration sphere is larger than what can be expected for globular proteins. The hydration volume of a protein can be deduced from the experimental determination of its hydration radius R_S (also called *Stokes radius*). Size exclusion chromatography (SEC) is a straightforward and reliable method to evaluate the R_S of proteins (Uversky, 2010). Potyviral VPgs are known to form dimers, but under reducing conditions, VPg is

expected to be monomeric (see previous section, Grzela et al., 2006); therefore, a purified recombinant form of the LMV VPg (theoretical mass 26,250 Da for a monomer) was first submitted to SEC in the presence of 10 mM DTT. Its elution volume suggested a 42,000-Da globular protein. The same test under nonreducing conditions revealed a molecular weight of 63,100 Da (Fig. 11.5a). Clearly, under both sets of conditions, the LMV VPg does not behave like a globular protein. A linear relationship exists between log (MM) and log (R_s), which, knowing the SEC-derived apparent molecular weight of a protein, allows the determination of its R_s. A method was developed to classify IPDs according to their Molecular Mass (MM) versus R_s characteristics (Uversky, 2002). From Figure 11.5b, it can be seen that the monomeric (26 kDa) form of the LMV VPg displays structural elements in between molten and premolten globule features. The dimeric (52 kDa) form seems more compact with mainly molten globule features. Although these results were obtained by different experimental approaches, we attempted to compare data obtained for the PVY VPg and LMV VPg. Since the sedimentation data were not available for the LMV VPg, $s_{20,w}$ values were deduced from SEC-determined R_s using Equation (11.1):

$$s_{20,w} = \frac{M(1 - \rho_{20,w}\upsilon^-)}{N_A 6\pi \eta_{20,w} R_S} \tag{11.1}$$

with $\rho_{20,w} = 0.998$ g/mL and $\eta_{20,w} = 1.022$ mPa/s being the density and viscosity of water at 20°C, respectively. The protein partial specific volume υ^- was estimated to be 0.724 mL/g from the LMV VPg amino acid composition using the freeware SEDNTERP (http:/www.jphilo.mailway.com), for both VPg forms. The hydrodynamic parameters of both VPgs are shown in Table 11.1. The $s_{20,w}$ values obtained were 2.25 and 3.9 S for the monomeric and dimeric forms, respectively. Using a linear fit of log (R_S) versus log (MM) over a population of globular proteins (Uversky, 2002), values of 2.66 and 4.14 S were deduced for compact globular proteins of the same molecular weight. In conclusion, the monomeric LMV VPg displays structural features reflected in a hydrodynamic behavior in between a native premolten and a molten globule. Analytical centrifugation (AUC) (PVY VPg) and SEC (LMV VPg) data seem to be in rather good agreement (Table 11.1). Within the limits of both approaches, the LMV VPg seems a little more compact than its PVY counterpart.

11.2.4 VPg Structure Analyzed by Circular Dichroism (CD) and Fluorescence Spectroscopy

IDPs possess distinctive far-UV CD spectra with a characteristic deep minimum in the vicinity of 200 nm and a relatively low ellipticity at 215 and 222 nm resulting from the low content of ordered secondary structure (Sreerama et al., 1999). The CD spectra of the PVY VPg are identical for the monomeric and dimeric forms (Fig. 11.6a) and exhibit some weak positive maxima around 190 nm, negative ellipticity near 205 nm, and a shoulder at 220 nm. These spectral features have been assigned to a group of β-sheet-rich proteins containing short segments of β-strands

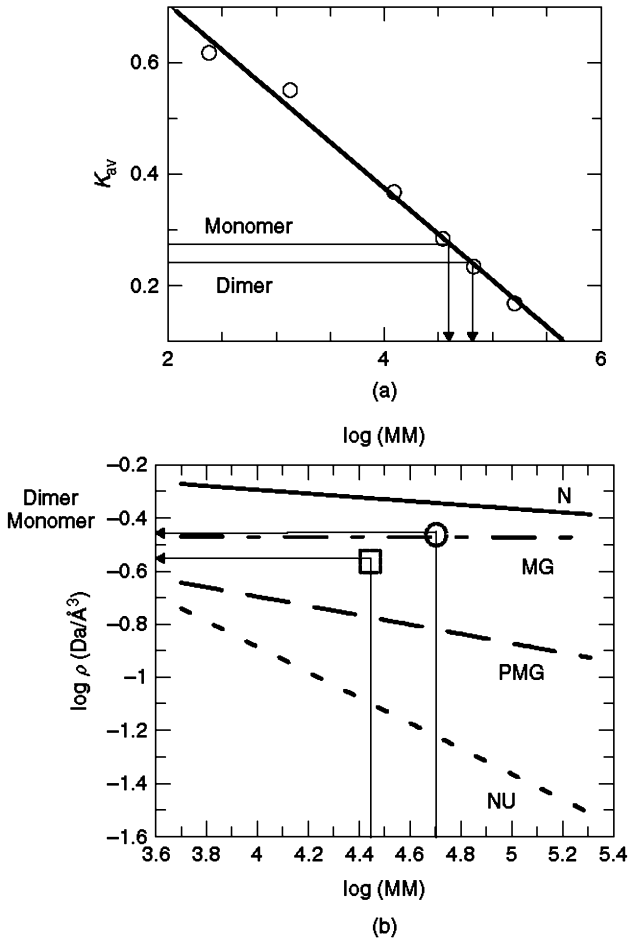


Figure 11.5 Hydrodynamic properties of the LMV VPg (SEC analysis). (a) A Superose-12 HR 10/30 column was calibrated with IgG (MM 160,000), BSA (MM 67,000), β-lactoglobulin (MM 35,000), cytochrome C (MM 12,400), vitamin B12 (MM 1355), and cytidine (MM 243). Recombinant VPg samples (0.3–0.5 mg/mL) were loaded onto the column in nonreducing conditions. Two molecular species were reproducibly eluted with K_{av} of 0.24 and 0.27, corresponding to a molecular weight (MM) of 41,686 and 63,095 Da, respectively. When the samples were reduced (10 mM DTT) and run under the same conditions, the 63-kDa species was no longer present. The hydrodynamic radii (R_S) were derived from these experimentally determined MM values according to Uversky (2002), assuming a behavior of globular proteins. (b) The apparent molecular density ρ of both species was calculated using the expression $\rho = M/(4/3pR_S^3)$, where M is the MM derived from amino acid composition and R_S the hydrodynamic radius. The log ρ values were reported to the log of these true MM on the x-axis. It is observed that both monomer and dimer species fall out of the domain of ordered globular proteins (N). The dimer form falls within the group of molten globules (MGs), while the less compact monomer falls in between molten and premolten globules (PMGs). Straight lines that define the different groups of conformational states were calculated from Uversky (2010). NU, native coil-like proteins.

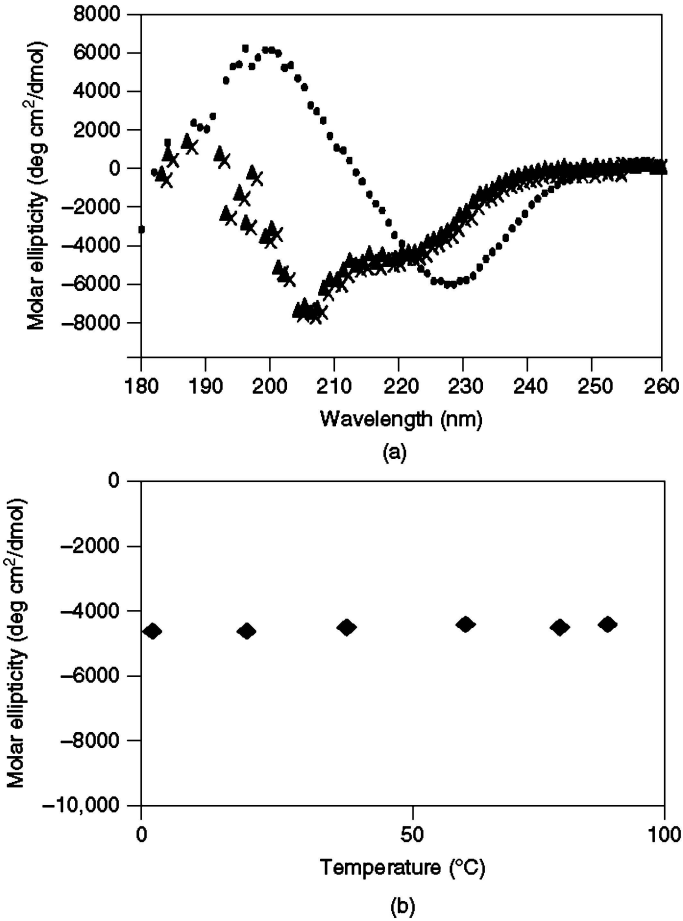


Figure 11.6 CD of PVY VPg. (a) Far-UV CD spectra were recorded at room temperature for the monomeric (triangles) and dimeric (crosses) forms as well as for the monomer in 20% TFE (dots). (b) Temperature dependence of the molar ellipticity of VPg followed at 222 nm.

(so called βII proteins) (Sreerama and Woody, 2003). Very similar spectral features were observed for the LMV VPg (Hébrard et al., 2009). 2,2,2-Trifluoroethanol (TFE) is known to promote formation of α-helical structures in proteins (Jasanoff and Fersht, 1994). The PVY VPg spectrum was not changed when recorded in the presence of 10% TFE but was significantly modified in the presence of 20% TFE (Fig. 11.6a, dots). This spectrum has a positive molar ellipticity peaking around 200 nm and a negative band at 228 nm, strikingly similar to spectra observed with βI-rich proteins (Sreerama and Woody, 2003). In contrast, the LMV VPg revealed a propensity for α-helix folding at TFE concentrations as low as 5% (Hébrard et al., 2009). It appears that VPg undergoes a structural reorganization in TFE,

which is known to favor local interactions through increasing helical propensity. It is known that the dielectric constant of TFE is approximately one-third of that of water, so charge interactions should be more important in TFE. Intramolecular hydrogen bonds are likely to be strengthened by the addition of TFE in an aqueous solution (Thomas and Dill, 1993, and the references therein).

When the molar ellipticity of PVY VPg was followed at 222 nm as a function of increasing temperature, no effect was observed (Fig. 11.6b). CD data for PVA VPg suggest 7–10% of α-helix, 20–27% of β-sheet, and 14–17% of turn, accompanied by about 50% of random coil (Rantalainen et al., 2008). As described earlier, T_m for certain changes in the structure of PVA VPg is 42°C. However, in common with other unfolded proteins (Sanchez-Puig et al., 2005), no cooperative transition of the molar ellipticity due to increasing temperature was observed. The tertiary structure of PVA VPg is disrupted at lower denaturant concentrations than the secondary structures, which is likely due to the formation of partially folded intermediates with pronounced secondary structures before complete denaturation. The shape of the unfolding curve does not support cooperative conformational transitions and is more indicative of significantly unstable overall structure of PVA VPg.

The presence of 12 Tyr residues in the PVA VPg (but only three in PVY VPg) opened up the possibility to study in more detail the compactness and degree of this protein folding, using the fluorescent spectroscopy analysis and applying the Stern–Volmer plot. Structural transitions of ordered proteins produce typical sigmoidal curve, whereas unfolding of molten globular forms is much less cooperative (Baily et al., 2001; Kuznetsova et al., 2007). 1-Anilino-8-naphthalene sulfonate (ANS)-binding data obtained for PVA VPg imply a partially folded molecule that contains a hydrophobic core domain (Rantalainen et al., 2008).

11.2.5 NMR Spectroscopy

The degree of folding is reflected by the degree of dispersion of the proton signals observed by 1D-NMR. Recently, a new NMR experimental procedure, called *HET-SOFAST*, has been proposed by Schanda et al. (2006), allowing the quantification of protein structural compactness through the ratio of average proton density λ_{NOE} of a reference and a saturated spectrum. While for compact, well-folded proteins, this ratio falls between 0.1 and 0.4, for totally unstructured proteins the ratio increases up to 0.8. HET-SOFAST experiments were applied to samples of PVY VPg in monomeric and dimeric states, at 10 and 30°C. The ^{1}H-NMR spectra of the VPg protein recorded under these different conditions were quite similar and showed a relatively poor dispersion of the amide protons in the 6- to 9-ppm region (Fig. 11.7a) characteristic of a rather unfolded protein. However, several aliphatic peaks were observed at around 0 ppm, located away from the random coil frequency of the main peak at around 0.7 ppm. These peaks most likely arise from regularly folded regions. The λ_{NOE} parameters measured for the protein in the monomeric state were around 0.5, independently of the conditions. These values of λ_{NOE} are characteristic of partially unfolded or degraded proteins (Schanda et al., 2006). As the VPg

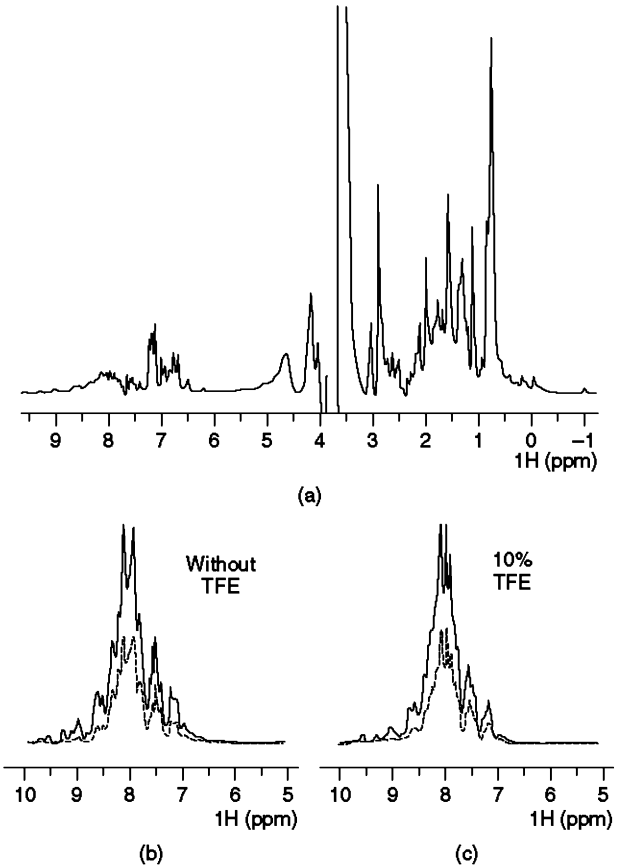


Figure 11.7 NMR spectroscopy of VPg in 10 mM phosphate buffer pH 7, containing 150 mM NaCl and 5 mM DTT. (a) 1D ^{1}H-NMR of the dimer of VPg recorded at 30°C. (b,c) Overlay of the reference (continuous line) and saturated spectra (dotted line) of the 1D HET-SOFAST experiment of the monomer of VPg recorded at 30°C without TFE (b) and in the presence of 10% TFE (c).

samples were found not to be degraded after the NMR experiment (data not shown), these data strongly substantiate the unfolded character of PVY VPg protein.

Both the NMR and CD spectra obtained in 10% TFE (Figs. 11.6a and 11.7b and 11.7c) did not indicate any significant change in PVY VPg structure on TFE addition. In contrast, CD spectra analysis performed in the presence of 20% TFE indicated a gross structural change (Fig. 11.6a). Therefore, we attempted to carry the NMR experiment at 20% TFE; however, a high TFE concentration resulted in VPg aggregation, which precluded analysis. It should be noted that NMR analysis requires approximately 10 times higher protein concentrations than CD analysis. VPg aggregates obtained in 20% TFE dissolved upon 3 min incubation at 100°C and reformed during cooling. This treatment could be repeated

several times, which again shows VPg insensitivity to denaturation. This unusual behavior can be explained by local interactions resulting, in the presence of TFE, in the acquisition of the secondary structure leading to VPg aggregation, while the weakening of local interactions with temperature increase induced protein resolubilization.

11.2.6 Proteolytic Analysis

Localization of disordered parts in dynamic structural environment of VPg is difficult due to the heterogeneity of the oligomerization status and constant movement of the disordered regions. Proteolytic digestion is one way to analyze which parts of the protein are exposed and thus available for cleavage. Chymotrypsin and trypsin cleavage sites on PVY VPg were located between N-terminal amino acids 11–21 and 41–61, regions that are predicted to be disordered (Grzela et al., 2008). In addition, a cut that is noncanonical for trypsin was observed at Gly119. On PVA VPg, the fastest cleaved sites were similarly located mostly on the N-terminus and, again noncanonically, at Asn115, leaving two more stable fragments after prolonged digestion. Asn115 in PVA VPg and Gly119 in PVY VPg both locate on the proposed central helix, which is shown to take part in interaction of LMV VPg with host eIF4E and viral HcPro (Roudet-Tavert et al., 2007). This region is poorly conserved, and it is conceivable that it takes part in host adaptation and specificity through a structural adaptation mechanism. This is supported by the fact that the amino acid residues 116 and 118 that vary between different isolates of PVA (Met/His on isolate B11 or Val/Tyr in isolate M) are determinants of systemic infection within different plant species (Rajamäki and Valkonen, 2002; Rajamäki and Valkonen, 1999). Localization of disordered regions in potyviral VPgs by proteolytic digestion is consistent with PONDR™-predicted regions of disorder as well as the predicted localization of more stable regions (Section 11.5).

11.3 INTERACTION OF PVA VPg WITH LIPIDS

PVA VPg was shown to bind anionic lipid vesicles (Rantalainen et al., 2009). The regions that are accessible to proteolysis—loosely folded nucleotide binding/NLS site at the N-terminus (amino acids 38–50) and Asn115—are shielded after lipid attachment. The N-terminal region has a high local positive charge; in fact, the entire N-terminus has a positive charge that raises the isoelectric point of the PVA VPg to ~8.9. It is very likely that the binding to anionic vesicles is driven by the electrostatic attraction between positively charged patches of VPg N-terminus and negatively charged vesicle surface. The fact that the first ~25 N-terminal amino acids are still trypsin accessible after vesicle interaction demonstrates that the lipid-protected region starts from around Thr30. It shields the putative central helix but leaves C-terminus accessible for trypsin. This suggests a clear division of exposed domains resulting in different structural states, which are conceivably involved in different VPg functions.

When the effect of TFE on VPg from PVY and LMV (genus *Potyvirus*), as well as on VPg of rice yellow mottle virus RYMV (genus *Sobemovirus*), was studied, a change in structure was observed, which was considered to indicate an increase in the α-helical content (Grzela et al., 2008; Hébrard et al., 2009). On the basis of changes observed in the far-UV region CD spectra, it appears that the interaction of potyviral VPg with anionic phospholipids is accompanied by similar structure stabilization. From the secondary structure predictions for PVA VPg, about 20% higher α-helix content could be expected than one actually calculated from the far-UV CD spectra. It is thus conceivable that α-helical stabilization pathway is

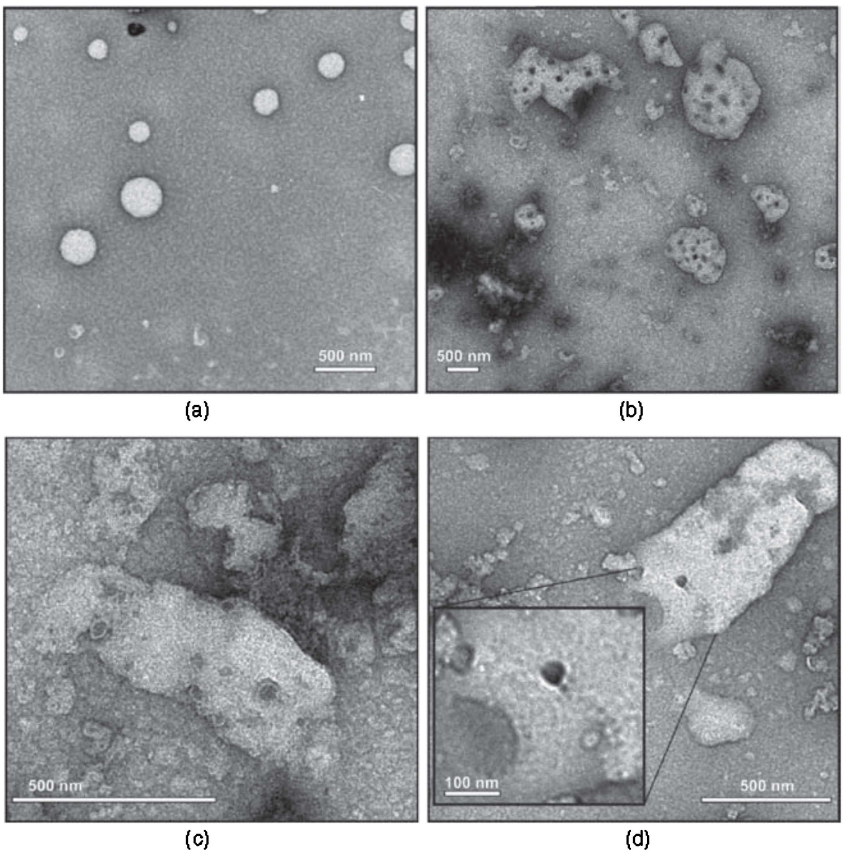


Figure 11.8 Electron microscopic images of VPg and DOPS vesicles. (a) DOPS vesicles in the absence of VPg appear as symmetrical spheres with an average size of ~100 nm. (b) On incubation of DOPS vesicles with VPg for 1 h at room temperature, electron-dense spots were formed, with an approximately fivefold increase in size compared to vesicles without VPg, which resulted in disruption of vesicle surface. (c,d) With closer inspection, the electron-dense spots reveal pore-like features. Additionally, micelle-like smaller vesicle structures appeared outside the swollen vesicles after VPg addition. Inset: digital magnification of a region with swollen and disrupted lipid structure. *Source*: From PhD thesis of K. Rantalainen (http://urn.fi/URN:ISBN:978-952-10-6014-4).

promoted by interaction with lipids. The mechanism of vesicle disruption seen by EM upon attachment of recombinant PVA VPg to synthetic vesicles (Fig. 11.8) could therefore be an α-helix penetration-driven event, which is a typical membrane lytic pathological mechanism (Shai et al., 1999). EM images also showed that the vesicle binding of VPg results in a formation of electron-dense protein spots on the vesicle surface. Taking into account studies with membrane lytic peptides, this phenomenon could be explained by formation of protein channels that disturb the bilayer and enable the leakage of molecules from a membrane-enclosed compartment. In a biological context this would evidently lead to the lysis of the organelle. A specific class of viral proteins termed *viroporins* is known to expand the membrane, enabling the passage of ions and small molecules (Gonzalez et al., 2003). Interestingly, the VPg precursor in tomato ringspot virus (TRV) (genus *Nepovirus*), called NTB-VPg protein, is proposed to have viroporin-type activity. It binds to the membrane surface by two different regions, C-terminal transmembrane helix and N-terminal amphipathic helix. The C-terminal transmembrane helix is able to traverse the membranes both *in vitro* and in infected plants, which results in a luminal orientation of the VPg domain. It is suggested that oligomerization of the amphipathic helix promotes its translocation in the membrane and leads to pore formation (Zhang et al., 2005). Other examples of viral proteins with similar membrane penetration and lysis activities include HIV-1 TAT, poliovirus 3AB protein, and the movement protein of barley yellow dwarf virus (genus *Luteovirus*) (Lama et al., 1996; Herce et al., 2007). It should be noted that membrane interactions and modifications are important for critical virus infection cycle events. Positive-strand RNA viruses create an environment for active replication complexes using different host organelles as a lipid source (Miller and Krijnse-Locker, 2008). For potyviruses, this involves vesicle formation from the endoplasmic reticulum (ER) and vesicles targeted to the chloroplasts (Schaad et al., 1997; Wei and Wang 2008; Wei et al., 2010). Replication-associated proteins, including VPg, localize to the vesicles, but the vesicle formation is induced by short 6K2 protein. Potyviral VPgs do not have a typical hydrophobic amino acid stretch that is capable of forming a transmembrane helix, but 6K2 does. It is conceivable that during viral life cycle, newly synthesized VPg enters vesicles while being in complex with 6K2 (in 6K2-NIa form) but is released by proteolysis in late phase of infection. Therefore, the lipid-associated binding and stabilization of VPg should help in the formation of peripheral membrane structures.

11.4 LMV VPg DISORDER-TO-ORDER TRANSITION UPON eIF4E BINDING

Upon interaction with protein partners, intrinsically disordered domains involved in the molecular recognition may undergo transitions toward more ordered structures (Dyson and Wright, 2002; Vacic et al., 2007; Oldfield et al., 2007). The only VPg structures available to date (Picornaviridae) were obtained either in the presence of a stabilizing agent (Schein, et al., 2006) or as a result of an association with the viral RNA-dependent RNA polymerase (RdRP), which could likely stabilize

the VPg folded state (Ferrer-Orta et al., 2006; Gruez et al., 2008). The folding propensity of the LMV VPg is underlined experimentally by helix formation induced at low TFE concentration. Secondary structure predictions indicate that helical folding may be adopted by the VPg central part. A fragment of this domain (residues 89–105) participates in interactions with the lettuce eIF4E and the viral HcPro (Michon et al., 2006; Roudet-Tavert et al., 2007). This domain is predicted to be disordered and could fold into a helix upon its interaction with eIF4E (Section 11.5). The interaction with $VPg_{(89-105)}$ is associated with perturbations in the environment of some tryptophan residues in eIF4E as shown by fluorescence emission spectra. The fluorescence decreases proportionally with the amount of eIF4E-$VPg_{(89-105)}$ complex formed (Fig. 11.9). One may expect that the complex formation between a natively folded protein such as eIF4E and a protein that undergoes disorder-to-order transitions on binding could be a two-step process (Oldfield et al., 2005; Cheng et al., 2007). The binding could be initiated by a "fly-casting" effect, whereby the disordered protein binds weakly to its target and folds as it approaches the cognate binding site (Shoemaker et al., 2000; Portman et al. 2000; Sugase et al., 2007). This might involve a fast association of eIF4E with $VPg_{(89-105)}$ followed by a slow change in conformation from the first association complex [eIF4E–$VPg_{(89-105)}$]* to produce the stable eIF4E–$VPg_{(89-105)}$ complex.

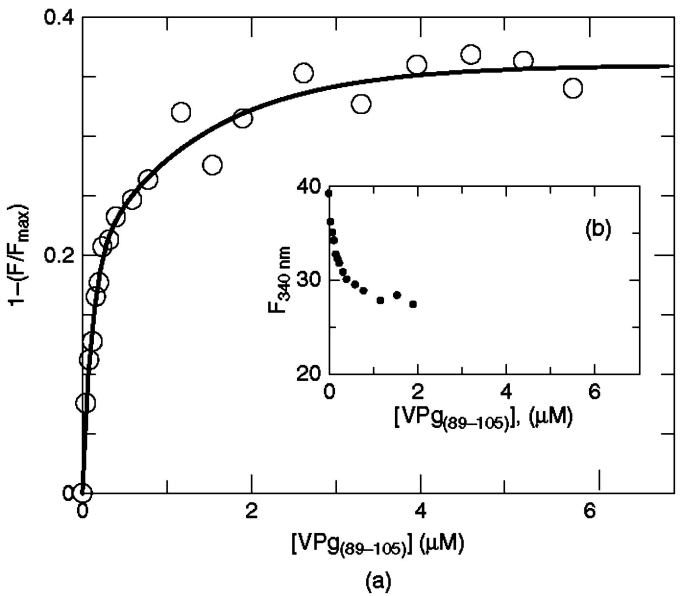


Figure 11.9 The interaction between a recombinant lettuce eIF4E and the LMV $VPg_{(89-105)}$ peptide was monitored using eIF4E fluorescence decay on addition of $VPg_{(89-105)}$, as previously described (Michon et al., 2006; Roudet-Tavert et al., 2007). (a) Binding isotherm of $VPg_{(89-105)}$ association with eIF4E. (b) Decrease of eIF4E fluorescence proportional to complex formation (raw data).

In order to probe this mechanism, fast kinetic measurements of the association between translation initiation factor eIF4E and the central domain $VPg_{(86-105)}$ were performed. The reaction can be schematized as follows:

$$\text{eIF4E} + \text{VPg}_{(89-105)} \overset{K_1}{\rightleftarrows} [\text{eIF4E}-\text{VPg}_{(89-105)}] * \overset{k_2}{\rightarrow} \text{eIF4E}-\text{VPg}_{(89-105)} \qquad (11.2)$$

According to this scheme, the binding rates are related to the concentration of $VPg_{(89-105)}$ as follows:

$$\frac{1}{k_{obs}} = \frac{1}{k_2} + \frac{K_1}{k_2[\text{VPg}_{(89-105)}]_{tot}} \qquad (11.3)$$

with k_{obs} being the experimental first-order rate constant; k_2, the forward rate constant for the second step; K_1, the association equilibrium constant for the first step; and $[VPg_{(89-105)}]_{tot}$, the total concentration of $VPg_{(89-105)}$. For an excess of $VPg_{(89-105)}$ with respect to eIF4E, the plot of $1/k_{obs}$ versus $1/[VPg_{(89-105)}]_{tot}$ is a

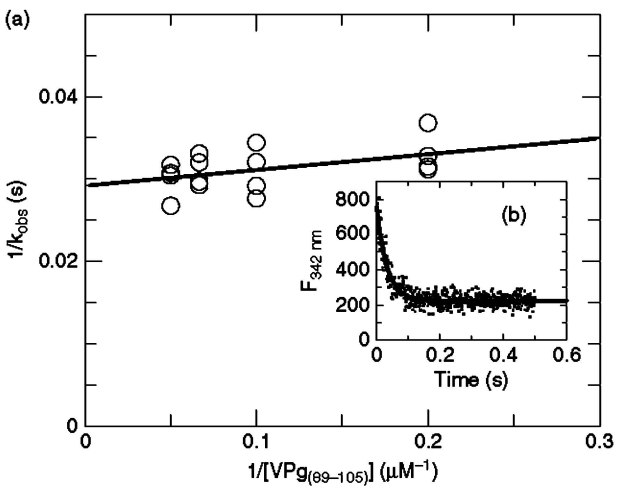


Figure 11.10 Kinetic analysis of the two-step process of eIF4E–$VPg_{(89-105)}$ complex formation. In a typical experiment, a solution of recombinant eIF4E (1 μM) was mixed with an equal volume of a 20-μM $VPg_{(89-105)}$ solution in 20 mM HEPES/NaOH pH 7.5, containing 200 mM NaCl, and 5 mM DTT at 20°C, using a stopped flow apparatus (dead time 5–10 ms). The final concentrations were 0.5 and 10 μM for eIF4E and $VPg_{(89-105)}$, respectively. A decrease in eIF4E fluorescence at 340 nm concomitant with the complex formation was observed. Raw data were fitted to a single exponential $F_t = F_0(1 - e^{-k_{obs}^t})$ and the exponential coefficient k_{obs} deduced. Fitting over a double exponential did not improve the fit. Several complex formations were monitored for varying $VPg_{(89-105)}$ concentrations, while keeping eIF4E concentration constant. (a) Plot of $1/k_{obs}$ versus $1/[VPg_{(89-105)}]$ gives a straight line expected for a two-step mechanism (see text). (b) Typical record of the transient kinetics of complex formation (eIF4E 0.5 μM, $VPg_{(89-105)}$ 10 μM).

straight line compatible with a two-step mechanism (Fig. 11.10). Similar data have been previously reported for the VPg from TEV (Khan et al., 2006, 2008; Miyoshi et al., 2006, 2008).

11.5 IS INTRINSIC DISORDER A COMMON FEATURE OF VPg? AN *IN SILICO* COMPARATIVE STUDY

Experimental data described above demonstrate that VPgs of PVY, PVA, and LMV are IDPs. Rather than repeating the experimentations for other potyviral VPgs, intrinsic disorder (ID) was assessed using *in silico* tools. This approach was a good opportunity to test several complementary disorder predictors on viral ID molten globule-like proteins. Moreover, two other phytoviral VPgs from SeMV and RYMV, genus *Sobemovirus*, that are not related to potyviruses, have been previously reported to be IDP (Chapter 2, Satheshkumar et al., 2005; Hébrard et al., 2009). Moreover, VPgs are not restricted to phytoviruses. For instance, the genus *Calicivirus* contains VPgs displaying intermediary lengths between those of potyviral and sobemoviral VPgs. *In silico* study of potyviral VPgs was then extended to sobemoviral and caliciviral VPgs to allow the comparison of disorder propensity at intra- and intergenera levels.

A collection of 16 VPg sequences representative of various virus species was analyzed using eight predictive tools. The sequence composition was analyzed to detect deviation from the composition of average globular proteins taken from Tompa et al. (2002), according to the method by Karlin et al. (2003). IDP has been shown to have a biased amino acid composition with depletion of "order-promoting" residues (W, C, F, I, Y, V, L, N) and an enrichment of "disorder-promoting" residues (A, R, G, Q, S, P, E, K) (Dunker et al., 2001). Version VLXT of PONDR is a neural network principally based on local amino acid composition, flexibility, and hydropathy (Romero et al., 2001; http://www.pondr.com). FoldIndex© is based on charge and hydropathy analyzed locally using a sliding window (Prilusky et al., 2005; http://bip.weizmann.ac.il/fldbin/findex). DISOPRED2 is also a neural network, but it incorporates information from multiple sequence alignments generated by PSI-BLAST (Ward et al., 2004; http://bioinf.cs.ucl.ac.uk/disopred). PONDR VSL2 has achieved higher accuracy and improved performance on short disordered regions, while maintaining high performance on long disordered regions (Obradovic et al., 2005; http://www.ist.temple.edu/disprot/predictorVSL2.php). IUPred uses a novel algorithm that evaluates the energy resulting from interresidue interactions (Dosztanyi et al., 2005; http://iupred.enzim.hu). PONDR VLXT and VSL2, as well as DISOPRED2, were all trained on data sets of disordered proteins, while FoldIndex and IUPred were not. Binary classifications of VPgs as ordered or disordered were performed using CDF and charge–hydropathy (CH) plot analyses. Cumulative distribution function (CDF) curves were generated for each data set using PONDR VLXT scores for each of the VPgs (Oldfield et al., 2005). CH plots were also analyzed using the method described by Uversky et al. (2000).

11.5.1 Prediction of Disordered Domains in Potyviral VPgs

The ID distribution was analyzed over the full-length PVY polyprotein using FoldIndex. The results clearly show that most of the VPg lies within one of the main unfolded regions of the polyprotein (Fig.11.12a). There are three other disorder "hot spots" in the potyviral polyprotein, which are P1, HcPro, and the N-terminal part of CP, shared by all potyviral polyproteins. The structural disorder propensity was evaluated for PVY, PVA, and LMV VPgs and three other potyviral VPgs, for which correlations between sequence and function are documented (Fig. 11.11b). Amino acid composition analysis of the six potyviral VPgs reveals a significant depletion in several order-promoting residues, such as tryptophan, cysteine, and valine, and a significant enrichment in disorder-promoting residues, such as arginine, glutamic acid, and lysine (Fig. 11.11a). However, the content of other residues is not

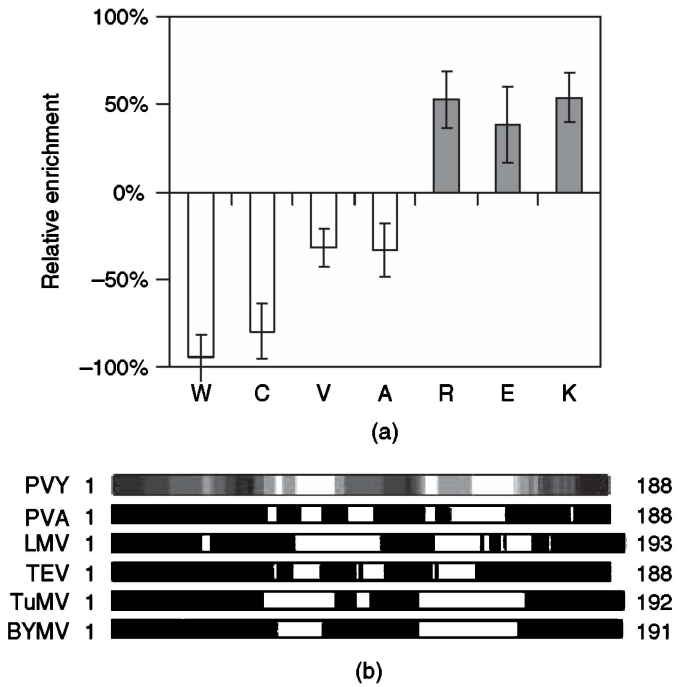


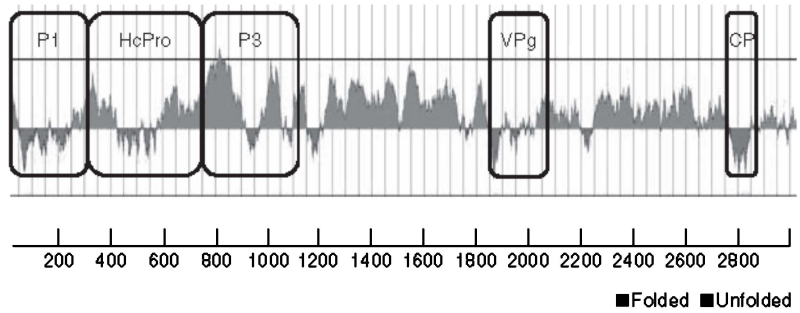
Figure 11.11 ID predictions for six potyviral VPgs. (a) Average deviation in amino acid composition from the average values for globular proteins taken from Tompa et al. (2002) and modified after Karlin et al. (2003). Only the residues for which compositional bias is significant and conserved in the six potyviral VPgs are shown. (b) Location of predicted disordered regions with five predictor software (PONDR® VLXT, FoldIndex©, DISOPRED2, VSL2, and IUPred). The intensity of gray in rectangles representing VPgs depends on the number of methods predicting ID. Numbering indicates the VPg length. PVY, potato virus Y; PVA, potato virus A; LMV, lettuce mosaic virus; TEV, tobacco etch virus; TuMV, turnip mosaic virus; BYMV, bean yellow mosaic virus.

significantly biased, not in favor of ID (such as alanine and serine), or not shared by all potyviral VPgs (data not shown).

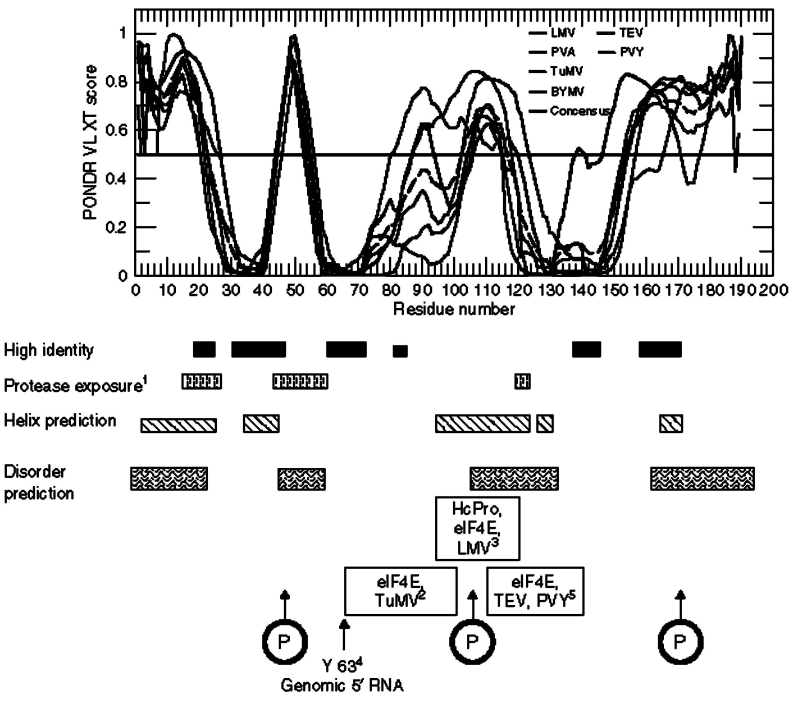
The sequence identity between these VPgs is between 42% and 54%, which is high in comparison to other viral families. Most of the conserved regions are located within domains predicted to be ordered (Hébrard et al., 2009). However, predicted disordered regions (PDRs) are detected in these six VPgs including those from PVA, PVY, and LMV, in which intrinsic disordered regions were experimentally detected (Figs. 11.11b and 11.12b). Although the length of the PDR varies among potyvirus VPgs and some discrepancies are observed between predictors, consensus PDRs can be mapped (Hébrard et al., 2009). The N- and C-terminal parts of all VPgs are predicted as disordered; however, prediction of disorder at the extremity of polypeptidic chains has to be considered cautiously. Indeed, disorder is often found at the N- and, to some extent, at the C-terminus of highly ordered proteins. N- and C-termini are not necessarily involved in biological functions, and examples of crystallographic structures obtained after N- or C-terminus removal are abundant. Three additional PDRs can be identified in potyviral VPgs around residues 40–60, 100–120, and 150–170. To illustrate this diversity, PONDR VLXT scores were plotted together with a derived consensus sequence (Fig. 11.12b). It is noteworthy that this central region of LMV VPg (89–105), which is involved in the binding of at least two different partners, the viral HcPro and the eIF4E host factor (Roudet-Tavert et al., 2007), is less conserved among potyviruses. Clearly, this region is permissive to mutations, which is interesting, as it also contains both major sites implicated in potyvirus virulence and binding domains for several functional factors (Charron et al., 2008). This suggests a tuning of the virus fitness to conditions imposed by the host biology, while keeping its vital specific network of interactions. Mutations in the disordered central part of potyviral VPgs may have a lower destabilizing effect than in more structured regions, which could constitute a net advantage for the viral adaptive evolution (Tokuriki et al., 2009). This observation is consistent with the hypothesis of protein plasticity whereby a single domain may modulate its surface overlap to accommodate various partners.

11.5.2 Prediction of Disordered Domains in Sobemoviral VPgs

The disorder propensity of six sobemoviral VPgs (77-residue long), including RYMV and SeMV for which ID has been reported, was evaluated. Amino acid composition analysis does not reveal a significant depletion in most order-promoting residues, except cysteine and asparagine, and a significant enrichment in disorder-promoting residues, except glutamic acid (Fig. 11.13a). The content of tryptophan and glutamine rather indicates propensity to order. However, because of the large diversity between the sobemoviral VPg sequences (only 20% identity between SeMV and RYMV VPgs), two viral groups, based on hydrophobic cluster analysis (HCA), secondary structure predictions (Hébrard *et al.*, 2009), and phylogenetic relationships (Hull and Fargette, 2005), were analyzed separately. One group contains the VPgs from RYMV/CoMV/RGMoV and the other the VPgs



(a)



(b)

Figure 11.12 Potyviral VPgs. (a) FoldIndex© prediction for whole PVY polyprotein (upper panel) (NCBI database). VPg is located between amino acids 1844–2031. (b) Comparative mapping of potyviral VPg disordered domains as predicted by PONDR VLXT. The medium horizontal line, set at score 0.5, delimits ordered residues in the lower part from the disordered regions, in upper part. Some experimentally determined interacting domains (VPg-eIF4E and VPg-HcPro) are located along the sequence. Somewhat different from the TuMV VPg-eIF4E interacting domain, interaction domains of the LMV, TEV, and PVY VPgs with their partners are located in the central part of the protein, which is predicted as disordered. Some putative phosphorylation sites are also marked. Following an alignment of the six VPg sequences, a mean score was calculated to build a consensus profile of PDR (in black). 1 Grzela et al. (2008), 2 Léonard et al. (2000), 3 Roudet-Tavert, et al. (2007), 4 Murphy et al. (1996), 5 Charron et al. (2008). (*See insert for color representation of the figure.*)

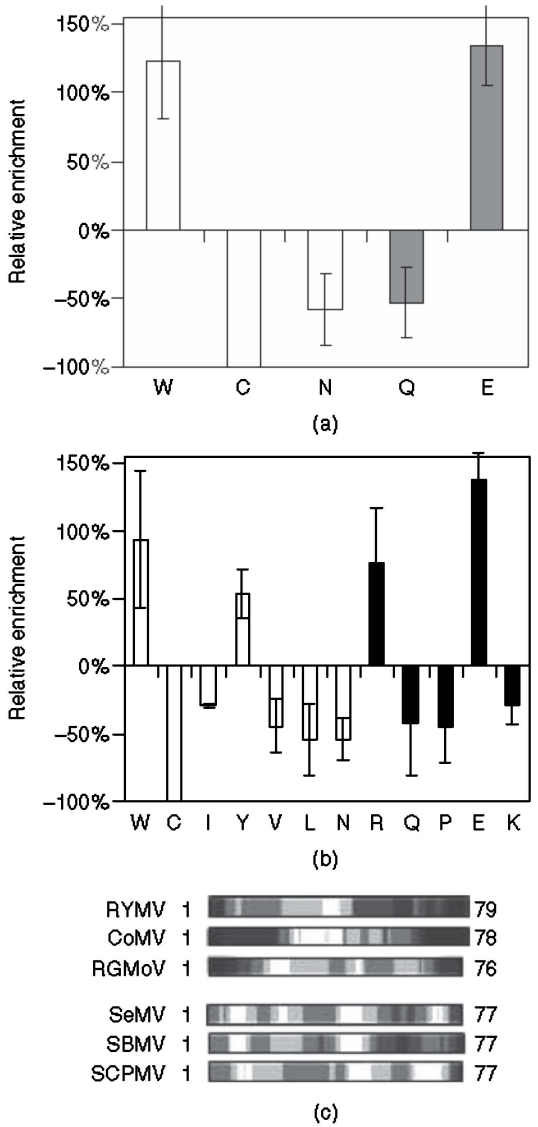


Figure 11.13 ID predictions for six sobemoviral VPgs. (a) Average deviation in amino acid composition from the average values for globular proteins taken from Tompa et al. (2002) and modified after Karlin et al. (2003). Only the residues for which compositional bias is significant and conserved are marked. (b) Average deviation in amino acid composition of RYMV, CofMV, and RGMoV VPgs. (c) Location of predicted disordered regions with five predictor software (PONDR® VLXT, FoldIndex©, DISOPRED2, VSL2, and IUPred). The intensity of gray in rectangles representing VPgs depends on the number of methods predicting ID. Numbering indicates the VPg length. RYMV, rice yellow mottle virus; CoMV, cocksfoot mottle virus; RGMoV, ryegrass mottle virus; SeMV, *Sesbania* mottle virus; SBMV, southern bean mosaic virus; SCPMV, southern cowpea mosaic virus.

from SeMV/SBMV/SCPMV. In the RYMV subgroup, depletion in order-promoting isoleucine, valine, leucine, and asparagine and, to a lower extent, enrichment in disorder-promoting proline and lysine have been noted (Fig. 11.13b). Regarding the PDR distribution within the sequence, the same groups of VPgs were confirmed (Hébrard *et al.,* 2009) (Fig. 11.13c). The N- and C-terminal ends are predicted as PDRs in the RYMV group, whereas PDRs are generally shorter and the consensus is less obvious in the SeMV group. Located in the C-terminal region, residues 48 and 52 associated with RYMV virulence (Pinel-Galzi et al., 2007) have been proposed to participate in the interaction between two antiparallel helices of the eIF(iso)4G central domain containing E309 and E321, two residues involved in rice resistance (Hébrard et al., 2008). Despite discrepancies within the species and a low correlation between the results of the five predictors, a structural disorder propensity is visible in all sobemoviral VPgs. This includes the SeMV and RYMV VPgs, for which intrinsic disorder was experimentally demonstrated (Chapter 2).

11.5.3 Prediction of Disordered Domains in Caliciviral VPgs

The Caliciviridae comprises four genera of human and animal viruses. These viruses possess intermediate-size VPgs between sobemovirus and potyvirus VPgs (Sadowy et al., 2001; Milner et al., 2001). Amino acid composition analysis of the VPg of a representative member of each genus reveals a significant depletion in cysteine and, to a lower extent, some other order-promoting residues, such as phenylalanine, isoleucine, valine, and leucine, and a significant enrichment in one disorder-promoting residue, arginine (Fig. 11.14a). The Norwalk virus (NV) VPg, the largest calicivirus VPg, is seen by most of the predictors as a completely disordered protein (Fig. 11.14b). The PDR distribution in three other calicivirus VPgs is conserved in spite of a low sequence identity ranging from 25% to 36%. In addition to several internal domains, the N- and C-termini are again predicted as disordered. Tyrosine residue involved in the uridylylation (position 20–30 depending on the virus) (Machin *et al.,* 2001) is not located in a PDR.

11.5.4 α-Helix-Forming Molecular Recognition Feature (α-MoRF) Predictions

On interaction with a partner, intrinsically disordered regions involved in molecular recognition may undergo a conformational transition toward an ordered structure (Garner et al., 1999; Wright and Dyson 1999; Dunker et al., 2001; Dyson and Wright 2002, 2005; Cheng *et al.*, 2007; Oldfield, et al. 2008). A correlation was established between some signatures of the plot generated by the PONDR VLXT algorithm and the ability for a given short disordered region to operate a disorder-to-order transition on binding (Garner et al., 1999). On the basis of these specific features, α-molecular recognition feature (α-MoRF) predictor was developed (Garner et al., 1999; Oldfield et al., 2005, 2008; Mohan et al., 2006; Cheng et al., 2007;

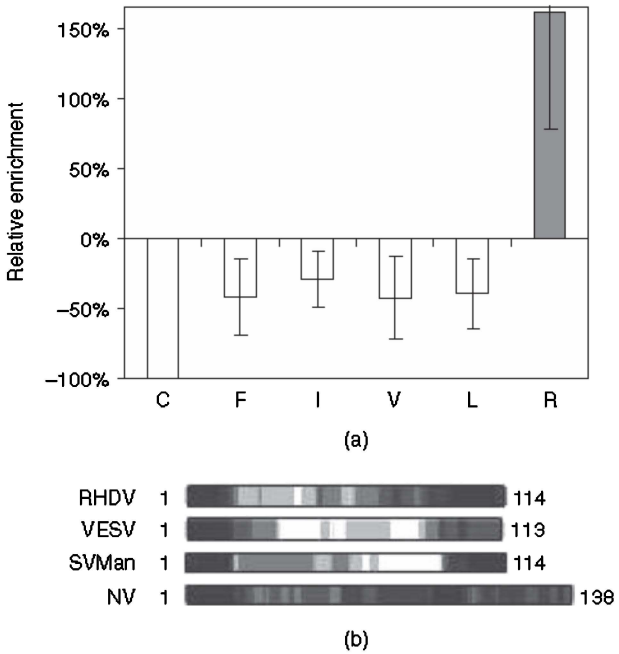


Figure 11.14 ID predictions for four caliciviral VPgs. (a) Average deviation in amino acid composition from the average values for globular proteins taken from Tompa et al. (2002)and modified after Karlin et al. (2003). Only the residues for which compositional bias is significant and conserved are shown. (b) Location of predicted disordered regions with five predictor software (PONDR® VLXT, FoldIndex©, DISOPRED2, VSL2, and IUPred). The intensity of gray in rectangles representing VPgs depends on the number of methods predicting ID. Numbering indicates the VPg length. RHDV, rabbit hemorrhagic disease virus; VESV, vesicular exanthema of swine virus; SV Man, Sapporo virus Manchester virus; NV, Norwalk virus.

Vacic et al., 2007). α-MoRF prediction indicates the presence of a relatively short (20 residues), loosely structured helical region within a largely disordered sequence that is stabilized on a disorder-to-order transition induced by binding to a partner. It utilizes a stacked architecture where PONDR VLXT is used to identify short predictions of order within long predictions of disorder. Then a second level predictor determines whether the order prediction is likely to be a binding site based on attributes of both the predicted ordered region and the predicted flanking disordered region. The α-MoRF predictor revealed 15 α-MoRFs in 12 of the 16 VPgs analyzed, with two α-MoRFs found at position 24–26 and 41–43 in all potyviral VPgs. Interestingly, there is no conserved α-MoRF in sobemoviral VPgs or caliciviral VPgs. This may reflect the weak sequence conservation in these proteins. However, other types of MoRF, such as strands or loops, (which may be stabilized on phosphorylation (Zhang et al., 2007; Stec et al. 2007) may be induced by some interactions (Mohan et al., 2006; Vacic et al., 2007). Three putative phosphorylation sites were predicted in the PDR of potyviral VPgs (Fig. 11.12b).

11.5.5 CDF and CH Plot Analyses

Sequences of VPgs from the various viral genera were in addition analyzed using two binary predictors, the CH plot (Uversky et al., 2000; Oldfield et al., 2005) and the CDF analysis (Oldfield et al., 2005). The two predictors classify the whole protein as ordered or disordered as opposed to the other predictors used in this study, which estimate the propensity to disorder for each position in the amino acid sequence. This analysis showed that a large majority of potyviral VPgs are predicted to belong to the class of native molten globules (results not shown). This is in accordance with experimental results on PVY and LMV VPgs (Section 11.2). All caliciviral VPgs are also classified as native molten globules, as well as the sobemoviral RYMV VPg. However, half of the sobemoviral VPgs are found in the group of ordered proteins. Moreover, PVA VPgs is predicted as ordered, which shows the limits of utility of some predictors and the necessity of confronting predictions with the experimental data.

11.6 CONCLUSIONS

Potyvirus VPgs do not have any homologs among nonviral proteins, and their atomic structure is unknown. As described earlier, the attempts at obtaining VPg crystals have been unsuccessful, which can be attributed at least in part to the intrinsically disordered nature of the protein. Taken together, the physicochemical data indicate that VPg has an elongated tertiary structure dominated by many disordered regions. Potyviral VPgs seem to contain some regions that are more stable and others that are rather loosely folded and capable of adopting several structures. Typical for disordered proteins, the loosely folded regions of VPgs also to be less conserved. This creates a structural environment with exposed, loosely folded, and interaction-prone regions accompanied by more stable and more conserved regions. Such a flexible nature of plant virus VPgs enables them to be involved in several different functions during the virus infection cycle since the less conserved regions predicted as disordered contain binding domains for several functional factors. This observation is consistent with the hypothesis of protein plasticity whereby a single domain is able to modulate its surface to accommodate various partners. Importantly, intrinsic disorder in VPgs yields a powerful strategy in which mutations would have less destabilizing effect in ID than in more structured regions. This could constitute a net advantage for the viral adaptive evolution.

ACKNOWLEDGMENTS

We are indebted to our collaborators:

J. C. to Renata Grzela, Ewa Szolajska, Christine Ebel, Dominique Madern, Adrien Favier, Jean-Pierre Andrieu, Izabela Wojtal, Bernard Dublet, and Wlodzimierz Zagorski. Financial support from the French National Agency for Research is acknowledged.

T. M. and E. H. to Nathalie Declerck, Sonia Longhi, Vladimir Uversky, Yannick Bessin, Jocelyne Walter, Denis Fargette, and Bénédicte Doublet. We thank the region Aquitaine (EPR ref. 20.000307004), the Genoplante program (Newvir, Gen56), and the French National Agency for Research (Poty4E, ANR-05-Blan-0302-01) for their financial support.

K. M. and K. R. to Vladimir Uversky, Perttu Permi, Nisse Kalkkinen, Keith Dunker, Peter Christensen, Anders Hafren, and Daniel Ozen. Financial support from the Academy of Finland (grants 115922 and 121622 to K. M.) is gratefully acknowledged.

LIST OF VIRUSES

BYDV	barley yellow dwarf virus
BYMV	bean yellow mosaic virus
ClYVV	clover yellow vein virus
CoMV	cocksfoot mottle virus
HIV	human immunodeficiency virus
HPV	human papilloma virus
LMV	lettuce mosaic virus
NV	Norwalk virus
PPV	plum pox virus
PSbMV	pea seed-borne mosaic virus
PVA	potato virus A
PVY	potato virus Y
RYMV	rice yellow mottle virus
RGMoV	ryegrass mottle virus
RHDV	rabbit hemorrhagic disease virus
SBMV	southern bean mosaic virus
SCPMV	southern cowpea mosaic virus
SeMV	*Sesbania* mosaic virus
SV Man	Sapporo virus Manchester virus
TEV	tobacco etch virus
TRV	tomato ringspot virus
TuMV	turnip mosaic virus
TVMV	tobacco vein mottling virus
VESV	vesicular exanthema of swine virus

ABBREVIATIONS

AF4	asymmetric flow field-flow fractionation
ANS	1-anilinonaphthalene-8-sulfonic acid
AUC	analytical ultracentrifugation
BSA	bovine serum albumin
CBB	Coomassie brilliant blue

CD	circular dichroism
CDF	cumulative distribution function analysis
CH	plots charge–hydropathy distributions
CP	coat protein
DLS	dynamic light scattering
DOPS	dioleylphosphoserine
eIF4E	eukaryotic initiation factor 4E
EM	electron microscopy
ER	endoplasmic reticulum
HcPro	helper-component protease
HET-SOFAST	fast detection of structural compactness and heterogeneity along polypeptide chains
ID	intrinsic disorder
IDPs	intrinsically disordered proteins
K_d	dissociation constant
λ_{NOE}	measure of average proton density
MoRF	molecular recognition feature
NIa	nuclear inclusion a
NIb	nuclear inclusion b
NLS	nuclear localization signal
ORF	open reading frame
PDRs	predicted disordered regions
PIPO	pretty interesting Potyviridae ORF
Pro	protease
PONDR	predictors of naturally disordered regions
RdRP	RNA-dependent RNA polymerase
SEC	size exclusion chromatography
VPg	viral protein genome linked

REFERENCES

Alonso LG, García-Alai MM, Nadra AD, Lapeña AN, Almeida FL, Gualfetti P, Prat-Gay GD. High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. Biochemistry 2002;41:10510–10518.

Anindya R, Chittori S, Savithri HS. Tyrosine 66 of pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. Virology 2005;336:154–162.

Bailey RW, Dunker AK, Brown CJ, Garner EC, Griswold MD. Clusterin, a binding protein with a molten globule-like region. Biochemistry 2001;40:11828–11840.

Beauchemin C, Boutet N, Laliberte JF. Visualization of the interaction between the precursors of VPg, the viral protein linked to the genome of turnip mosaic virus, and the translation eukaryotic initiation factor iso 4E in Planta. J Virol 2007;81:775–782.

Belliot G, Sosnovtsev SV, Chang KO, McPhie P, Green KY. Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. Virology 2008;374(1):33–49.

Bokros CL, Hugdahl JD, Kim HH, Hanesworth VR, van Heerden A, Browning KS, Morejohn LC. Function of the p86 subunit of eukaryotic initiation factor (iso)4F as a microtubule-associated protein in plant cells. Proc Natl Acad Sci U S A 1995;92:7120–7124.

Borgstrøm B, Johansen IE. Mutations in pea seedborne mosaic virus genome-linked protein VPg after pathotype-specific virulence in Pisum sativum. Mol Plant-Microbe Interact 2001;14:707–714.

Bourhis JM, Receveur-Bréchot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, Canard B, Finet S, Longhi S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci 2005;14:1975–1992.

Carrington JC, Freed DD, Leinicke AJ. Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. Plant Cell 1991;3:953–962.

Charron C, Nicolai M, et al. Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J 2008;54:56–68.

Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK, Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 2007;46:13468–13477.

Chung BY, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the Potyviridae. Proc Natl Acad Sci USA 2008;105:5897–5902.

Cotton S, Dufresne PJ, Thivierge K, Ide C, Fortin MG. The VPgPro protein of turnip mosaic virus: In vitro inhibition of translation from a ribonuclease activity. Virology 2006;351:92–100.

Decroocq V, Sicard O, Alamillo JM, Lansac M, Eyquard JP, García JA, Candresse T, Le Gall O, Revers F. Multiple resistance traits control Plum pox virus infection in Arabidopsis thaliana. Mol Plant-Microbe Interact 2006;19:541–549.

Delphin C, Cahen P, Lawrence JJ, Baudier J. Characterization of baculovirus recombinant wild-type p53. Dimerization of p53 is required for high-affinity DNA binding and cysteine oxidation inhibits p53 DNA binding. Eur J Biochem 1994;223:683–689.

Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21:3433–3434.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59. Review.

Dunoyer P, Thomas C, Harrison S, Revers F, Maule A. A cysteine-rich plant protein potentiates Potyvirus movement through an interaction with the virus genome-linked protein VPg. J Virol 2004;78:2301–2309.

Duprat A, Caranta C, Revers F, Menand B, Browning KS, Robaglia C. The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. Plant J 2002;32:927–934.

Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.

Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.

Fellers J, Wan J, Hong Y, Collins GB, Hunt AG. In vitro interactions between a potyvirus-encoded, genome-linked protein and RNA-dependent RNA polymerase. J Gen Virol 1998;79:2043–2049.

Ferrer-Orta C, Arias A, et al. The structure of a protein primer-polymerase complex in the initiation of genome replication. EMBO J 2006;25:880–888.

Gao Z, Johansen E, Eyers S, Thomas CL, Noel Ellis TH, Maule AJ. The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. Plant J 2004;40:376–385.

García-Alai MM, Alonso LG, de Prat-Gay G. The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry 2007;46:10405–10412.

Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting Binding Regions within Disordered Proteins. Genome Inform Ser Workshop Genome Inform 1999;10:41–50.

Gonzalez ME, Carrasco L. Viroporins. FEBS Lett 2003;552:28–34.

Graumann J, Lilie H, Tang X, Tucker KA, Hoffmann JH, Vijayalakshmi J, Saper M, Bardwell JC, Jakob U. Activation of the redox-regulated molecular chaperone Hsp33–a two-step mechanism. Structure 2001;9:377–387.

Gruez A, Selisko B, Roberts M, Bricogne G, Bussetta C, Jabafi I, Coutard B, De Palma AM, Neyts J, Canard B. The crystal structure of coxsackievirus B3 RNA-dependent RNA polymerase in complex with its protein primer VPg confirms the existence of a second VPg binding site on Picornaviridae polymerases. J Virol 2008;82:9577–9590.

Grzela R, Szolajska E, Ebel C, Madern D, Favier A, Wojtal I, Zagorski W, Chroboczek J. Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. J Biol Chem 2008;283:213–221.

Grzela R, Strokovska L, Andrieu JP, Dublet B, Zagorski W, Chroboczek J. Potyvirus terminal protein VPg, effector of host eukaryotic initiation factor eIF4E. Biochimie 2006;88:887–896.

Hafren A, Mäkinen K. Purification of viral genome-linked protein VPg from Potato virus A infected plants reveals several posttranslationally modified forms of the protein. J Gen Virol 2008;89:1509–1518.

Hébrard E, Pinel-Galzi A, Fargette D. Virulence domain of the RYMV genome-linked viral protein VPg towards rice rymv1-2-mediated resistance. Arch Virol 2008;153:1161–1164.

Hébrard E, Bessin Y, Michon T, Longhi S, Uversky VN, Delalande F, Van Dorsselaer A, Romero P, Walter J, Declerk N, Fargette D. Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. Virol J 2009;6:23–36.

Herbert TP, Brierley I, Brown TD. Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. J Gen Virol 1997;78:1033–10340.

Herce HD, Garcia AE. Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. Proc Natl Acad Sci USA 2007;104:20805–20810.

Hong Y, Levay K, Murphy JF, Klein PG, Shaw JG, Hunt AG. A potyvirus polymerase interacts with the viral coat protein and VPg in yeast cells. Virology 1995;214: 159–166.

Huang TS, Wei T, Laliberté JF, Wang A. A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. Plant Physiol 2010;152:255–266.

Hull R, Fargette D. Sobemovirus. Fauquet C, Mayo MA, Maniloff J, Desselberger U, Ball LA, editors. In: Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier Academic Press; 2005. pp. 885–890.

Ivanov KI, Puustinen P, Merits A, Saarma M, Mäkinen K. Phosphorylation down-regulates the RNA binding function of the coat protein of potato virus A. J Biol Chem 2001;276:13530–13540.

Jasanoff A, Fersht AR. Quantitative determination of helical propensities from trifluoroethanol titration curves. Biochemistry 1994;33:2129–2135.

Karlin S, Mrázek J, Gentles AJ. Genome comparisons and analysis. Curr. Opin. Struct. Biol. 2003;13:344–352.

Keller KE, Johansen IE, Martin RR, Hampton RO. Potyvirus genome-linked protein (VPg) determines pea seed-borne mosaic virus pathotype-specific virulence in Pisum sativum. Mol Plant-Microbe Interact 1998;11:124–130.

Khan MA, Miyoshi H, Gallie DR, Goss DJ. Potyvirus genome-linked protein, VPg, directly affects wheat germ In vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. J Biol Chem 2008;283:1340–1349.

Khan MA, Miyoshi H, Ray S, Natsuaki T, Suehiro N, Goss DJ. Interaction of genome-linked protein (VPg) of turnip mosaic virus with wheat germ translation initiation factors eIFiso4E and eIFiso4F. J Biol Chem 2006;281:28002–28010.

Kuznetsova IM, Turoverov KK, Dunker AK, Uversky VN. Analysis of folded, partially folded and misfolded proteins with fluorescent dyes. In Methods in Protein Structure and Stability Analysis: Luminescence Spectroscopy and Circular Dichroism. Uversky VN, Permyakov EA. editors. Hauppauge, NY, USA: Nova Science Publishers, Inc.; 2007. pp. 73–104.

Lama J, Carrasco L. Screening for membrane-permeabilizing mutants of the poliovirus protein 3AB. J Gen Virol 1996;77:2109–2119.

Lellis AD, Kasschau KD, Whitham SA, Carrington JC. Loss-of-susceptibility mutants of Arabidopsis thaliana reveal an essential role for eIF(iso)4E during potyvirus infection. Curr Biol 2002;12:1046–1051.

Léonard S, Viel C, Beauchemin C, Daigneault N, Fortin MG, Laliberté JF. Interaction of VPg-Pro of turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta. J Gen Virol 2004;85:1055–1063.

Léonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberté JF. Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. J Virol 2000;74:7730–7737.

Li XH, Valdez P, Olvera RE, Carrington JC. Functions of the tobacco etch virus RNA polymerase (NIb): subcellular transport and protein-protein interaction with VPg/proteinase (NIa). J Virol 1997;71:1598–1607.

Machin A, Martin Alonso JM, Parra F. Identification of the amino acid residue involved in rabbit hemorrhagic disease virus VPg uridylylation. J Biol Chem 2001;276:27787–27792.

Merits A, Rajamäki ML, Lindholm P, Runeberg-Roos P, Kekarainen T, Puustinen P, Mäkeläinen K, Valkonen JP, Saarma M. Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. J Gen Virol 2002;83:1211–1221.

Miller S, Krijnse-Locker J. Modification of intracellular membrane structures for virus replication. Nat Rev Microbiol 2008;6:363–374.

Michon T, Estevez Y, Walter J, German-Retana S, Le Gall O. The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. FEBS J 2006;273:1312–1322.

Mohan A, Oldfield CJ, Radivojac p, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362:1043–1059.

Miyoshi H, Suehiro N, Tomoo K, Muto S, Takahashi T, Tsukamoto T, Ohmori T, Natsuaki T. Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. Biochimie 2006;88(3–4):329–340.

Miyoshi H, Okade H, Muto S, Suehiro N, Nakashima H, Tomoo K, Natsuaki T. Turnip mosaic virus VPg interacts with arabidopsis thaliana eIF(iso)4E and inhibits In vitro translation. Biochimie 2008;90:1427–1434.

Murphy JF, Klein PG, Hunt AG, Shaw JG. Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. Virology 1996;220:535–538.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2005;61 (Suppl 7):176–182.

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genom 2008;9(Suppl 1): S1.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44:12454–12470.

Oruetxebarria I, Guo D, Merits A, Mäkinen K, Saarma M, Valkonen JPT. Identification of the genome-linked protein in virions of Potato virus A, with comparison to other members in genus Potyvirus. Virus Res 2001;73:103–112.

Paul AV, van Boom JH, Filippov D, Wimmer E. Protein-primed RNA synthesis by purified poliovirus RNA polymerase. Nature 1998;393:280–284.

Pinel-Galzi A, Rakotomalala M, Sangu E, Sorho F, Kanyeka Z, Traoré O, Sérémé D, Poulicard N, Rabenantoandro Y, Séré Y, Konaté G, Ghesquiére A, Hébrard E, Fargette D. Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. PLoS Pathog 2007;3(11):e180.

Prilusky J, Zeev-Ben-Mordehai T, Rydberg E, Felder C, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 2005;21:3435–3438.

Puustinen P, Mäkinen K. Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. J Biol Chem 2004;279: 38103–38110.

Puustinen P, Rajamäki ML, Ivanov KI, Valkonen JP, Mäkinen K. Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. J Virol 2002;76:12703–12711.

Rajamäki ML, Valkonen JP. Control of nuclear and nucleolar localization of nuclear inclusion protein a of picorna-like Potato virus A in Nicotiana species. Plant Cell 2009;21:2485–2502.

Rajamäki ML, Valkonen JPT. The 6K2 protein and the VPg of potato virus A are determinants of systemic infection in Nicandra physaloides. Mol Plant-Microbe Interact 1999;12:1074–1081.

Rajamäki ML, Valkonen JPT. Viral genome-linked protein (VPg) controls accumulation and phloem-loading of a potyvirus in inoculated potato leaves. Mol Plant-Microbe Interact 2002;15:138–149.

Rantalainen KI, Uversky VN, Permi P, Kalkkinen N, Dunker AK, Mäkinen K. Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. Virology 2008;377:280–288.

Rantalainen KI, Christensen PA, Hafrén A, Otzen DE, Kalkkinen N, Mäkinen K. Interaction of a potyviral VPg with synthetic anionic vesicles. Virology 2009;395:114–120.

Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Proteins 2006;62:24–45.

Restrepo MA, Freed DD, Carrington JC. Nuclear transport of plant potyviral proteins. Plant Cell 1990;2:987–998.

Restrepo-Hartwig MA, Carrington JC. The tobacco etch potyvirus 6-kilodalton protein is membrane associated and involved in viral replication. J Virol 1994;68:2388–2397.

Reyes JL, Campos F, Wei H, Arora R, Yang Y, Karlson DT, Covarrubias AA. Functional dissection of hydrophilins during In vitro freeze protection. Plant Cell Environ 2008;31:1781–1790.

Robaglia C, Caranta C. Translation initiation factors: A weak link in plant RNA virus infection. Trends Plant Sci 2006;11:40–45.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.

Roudet-Tavert G, Michon T, Walter J, Delaunay T, Redondo E, Le Gall O. Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPro. J Gen Virol 2007;88:1029–1033.

Ruffel S, Dussault MH, Palloix A, Moury B, Bendahmane A, Robaglia C, Caranta C. A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). Plant J 2002;32:1067–1075.

Sadowy E, Milner M, Haenni AL. Proteins attached to viral genomes are multifunctional. Adv Virus Res 2001;57:185–262.

Sanchez-Puig N, Veprintsev DB, Fersht AR. Human full-length Securin is a natively unfolded protein. Protein Sci 2005;14:1410–1418.

Satheshkumar PS, Gayathri P, Prasad K, Savithri HS. Natively unfolded. VPg is essential for Sesbania mosaic virus serine protease activity. J Biol Chem 2005;280: 30291–30300.

Sato M, Nakahara K, Yoshii M, Ishikawa M, Uyeda I. Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of arabidopsis thaliana by potyviruses. FEBS Lett 2005;579:1167–1171.

Schaad MC, Haldeman-Cahill R, Cronin S, Carrington JC. Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. J Virol 1996;70:7039–7048.

Schaad MC, Lellis AD, Carrington JC. VPg of tobacco etch potyvirus is a host genotype-specific determinant for long-distance movement. J Virol 1997;71:8624–8631.

Schanda P, Forge V, Brutscher B. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. Magn Reson Chem 2006;44: S177–S184.

Schein CH, Oezguen N, Volk DE, Garimella R, Paul A, Braun W. NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. Peptides 2006;27:1676–1684.

Shai Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. Biochim Biophys Acta 1999;1462:55–70.

Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 2000;97: 8868–8873.

Shukla DD, Ward CW, Bunt AA. The Potyviridae. Wallingford, UK: CAB International; 1994. p. 516.

Sigalov A, Aivazian D, Stern L. Homooligomerization of the cytoplasmic domain of the T cell receptor zeta chain and of other proteins containing the immunoreceptor tyrosine-based activation motif. Biochemistry 2004;43:2049–2061.

Sreerama N, Venyaminov SY, Woody RW. Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. Protein Sci 1999;8:370–380.

Sreerama N, Woody RW. Structural composition of betaI- and betaII-proteins. Protein Sci 2003;12:384–388.

Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007;447:1021–1025.

Thomas PD, Dill KA. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. Protein Sci 1993;2:2050–2065.

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? Trends Biochem Sci 2009;34:53–59.

Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.

Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579:3346–3354.

Urcuqui-Inchima S, Haenni AL, Bernardi F. Potyvirus proteins: a wealth of functions. Virus Res 2001;74:157–175. Review.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.

Uversky VN, Longhi S. Instrumental analysis of intrinsically disordered proteins. In: Uversky VN, editor. Protein and peptide science. Hoboken: Wiley; 2010.

Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK. Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 2007;6:2351–2366.

van de Ven FJ, Lycksell PO, van Kammen A, Hilbers CW. Computer-aided assignment of the 1H-NMR spectrum of the viral-protein-genome-linked polypeptide from cowpea mosaic virus. Eur J Biochem 1990;190:583–591.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.

Wei T, Wang A. Biogenesis of cytoplasmic membranous vesicles for plant potyvirus replication occurs at endoplasmic reticulum exit sites in a COPI- and COPII-dependent manner. J Virol 2008;82:12252–12264.

Wei T, Huang TS, McNeil J, Laliberté JF, Hong J, Nelson RS, Wang A. Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant potyvirus replication. J Virol 2010;84:799–809.

Wittmann S, Chatel H, Fortin MG, Laliberté JF. Interaction of the viral protein genome linked of turnip mosaic potyvirus with the translational eukaryotic initiation factor (iso) 4E of Arabidopsis thaliana using the yeast two-hybrid system. Virology 1997;234:84–92.

Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.

Yambao ML, Masuta C, Nakahara K, Uyeda I. The central and C-terminal domains of VPg of Clover yellow vein virus are important for VPg-HCPro and VPg-VPg interactions. J Gen Virol 2003;84:2861–2869.

Zhang SC, Zhang G, Yang L, Chisholm J, Sanfacon H. Evidence that insertion of Tomato ringspot nepovirus NTB-VPg protein in endoplasmic reticulum membranes is directed by two domains: a C-terminal transmembrane helix and an N-terminal amphipathic helix. J Virol 2005;79:11752–11765.

Zhang Y, Stec B, Godzik A. Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins. Structure 2007;15:1141–1147.

12

INTRINSIC DISORDER IN THE HUMAN PAPILLOMAVIRUS E7 PROTEIN

LUCÍA B. CHEMES, IGNACIO E. SÁNCHEZ, LEONARDO G. ALONSO, AND GONZALO DE PRAT-GAY

12.1 INTRODUCTION

12.1.1 Structure and Replication of Papillomaviruses

Papillomaviruses (PVs) are a group of small, nonenveloped, double-stranded DNA viruses, which infect the skin and mucous epithelia of reptiles, birds, and mammals, and have recently been classified as a separate virus family, the Papillomaviridae (de Villiers et al., 2004). To date, over 200 types have been identified, including more than 100 types that infect humans (Bernard et al., 2010). PVs have a narrow host range, and many virus types can infect a single host species. The virion particles are 50–55 nm in diameter and have T7 icosahedral symmetry (Baker et al., 1991; Wolf et al., 2010) (Fig. 12.1a).

12.1.1.1 Genomic Organization All PVs have a similar genomic organization, consisting of an ~8000-bp double-stranded circular genome, containing between 7 and 10 open reading frames (ORFs) (Howley and Lowy, 2007) (Fig. 12.1b). The ORFs are classified as early (E) or late (L) based on their location in the genome. All genes are encoded in the same DNA strand and are transcribed unidirectionally. The position, size, and function of each ORF are generally well conserved across different PV types. A 1000-bp regulatory region, called *long control region* or LCR is located between the L1 and E6/E7 ORFs (Fig. 12.1b). This region contains

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

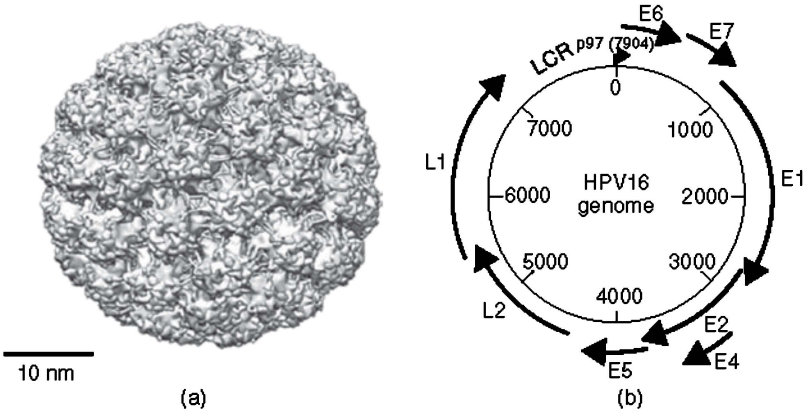


Figure 12.1 Papillomavirus structure and genome. (a) Structure of the papillomavirus capsid, based on a cryo-EM (electron microscopy) reconstruction of bovine papillomavirus particles. *Source:* Adapted from Wolf et al. (2010). (b) HPV16 genome organization showing the location of early (E) and late (L) open reading frames and of the long control region (LCR) containing the p97 promoter. ORFs are displayed in color: E6 and E7 in orange, E1 and E2 in light blue, E4 and E5 in dark blue, and L1 and L2 in green. (*See insert for color representation of the figure.*)

the origin of replication and transcriptional control elements regulating viral gene expression (Howley and Lowy, 2007). The early ORFs code for nonstructural proteins: E1 and E2 are required for viral DNA replication. E1 can bind the origin of replication and has DNA helicase and ATPase activity; E2 acts as a viral transcription factor and cooperates with E1 in viral DNA replication. The E6 and E7 proteins mediate S-phase reentry and are required for plasmid maintenance, among other functions. The functions of E4 and E5 are less well defined; E4 regulates several viral processes such as viral genome amplification, virus maturation, and release through changes in cellular structures such as keratin intermediate filaments, ND10 domains, and mitochondria, and E5 is required for cell growth and has transforming properties through mechanisms that include c-Src activation. The late genes code for the L1 and L2 structural proteins that make up the viral capsid.

*12.1.1.2 **Viral Replication*** With a small genome and only one enzyme, PVs strongly depend on the host cell replication machinery in order to amplify their genome. The virus life cycle is intimately linked to the epithelium differentiation program. PVs infect basal, undifferentiated cells of skin and mucous epithelia, where the genome is initially amplified and maintained at a low copy number. Subsequent viral replication takes place in differentiating cells, which have withdrawn from the cell cycle. The E7 protein is essential for stimulating S-phase reentry in differentiating keratinocytes, producing an uncoupling of proliferation and differentiation in these cells (Banerjee et al., 2006). This allows for the expression of the host DNA replication machinery, which is required for viral DNA amplification

(Chow and Broker, 2006). In the final stages of the virus life cycle, L1 and L2 proteins are produced, and mature virions shed from the epithelial surface.

12.1.1.3 DNA Transcription PV viral transcription is a complex process, given the presence of several promoters, multiple alternative splice sites, and tissue-specific differences in viral mRNA expression patterns (Zheng and Baker, 2006). This complexity expands the repertoire of viral proteins through the expression of protein fragments and proteins derived from the fusion of two or more ORFs. Regarding the E6 and E7 ORFs, there is evidence for the existence of several functional transcripts. In HPV16, the E6 and E7 proteins are transcribed from bi- or multicistronic E6/E7 transcripts synthesized from the P97 promoter (Fig. 12.1b). Several transcripts are produced for E6/E7 leading to the translation of two truncated forms of E6 (E6*I and E6*II) and of an E6/E7 fusion transcript containing the N-terminal region of E6 fused to the C-terminal region of E7 (Ordonez et al., 2004; Zheng and Baker, 2006). Transcription of the viral genome has a stringent cell-type specificity and is limited to keratinocytes and cervical cancer tissues (Thierry, 2009). Transcription is positively regulated by cellular core and tissue-specific factors such as Sp1, AP1, and NF1 among others (Apt et al., 1993; Thierry, 2009), and negatively regulated by the viral E2 protein, which can bind to the P97 promoter displacing Sp1 and TFIID, a mechanism that maintains low E6/E7 expression levels in HPV-infected cells (Tan et al., 1994b).

The HPV DNA is frequently found integrated into the host genome of cervical cancer-derived cell lines (Baker et al., 1987; Shirasawa et al., 1987) and cervical cancer biopsies (Choo et al., 1987b), with a loss of the region coding for the E2 (Choo et al., 1987a; Collins et al., 2009) or the E2/E1 (Shirasawa et al., 1987; Romanczuk and Howley, 1992; Cricca et al., 2009) proteins. Therefore, upon genome integration, transcription of the E6/E7 oncogenes is under positive cell–host factor control through the HMGI/Y factor among others, which is thought to lead to the higher expression levels observed in cervical cancer cell lines (Stoler et al., 1992; Bouallaga et al., 2000; Thierry, 2009). It has also been observed that HPV genome integration is associated with an increase in E6/E7 mRNA stability (Jeon and Lambert, 1995), which would further increase E6/E7 expression levels. The involvement of cell-type-specific host cell factors in E6/E7 transcriptional control suggests that E6/E7 expression levels may be differentially regulated *in vivo* depending on the infection niche, the differentiation state, or the transformation state of the infected epithelial cells.

12.1.2 Papillomaviruses and Human Cancer

HPVs are the etiological agent for cervical cancer, the second most frequent cancer in women (zur Hausen, 1994). Although most HPV exposures are transient and are cleared spontaneously by the host, a small fraction of infections persist and, over the years, may lead to the development of cancer. Histological studies have shown that over 95% of cervical cancer tissue samples are HPV positive, and epidemiological studies indicate that infection with high-risk HPV types is associated with both a

higher persistence rate of infection and a higher risk of developing cervical cancer (Howley and Lowy, 2007; zur Hausen, 2009). Of the more than 35 HPV types found in the genital tract, a few (HPV 16, 18, 45, 31, and 33) account for over 80% of the total cases of cervical cancer (Bosch et al., 2006). Infection by HPV is also linked, although to a lesser degree, to anogenital cancers such as cancer of the vulva, vagina, and penis; head and neck cancers; and nonmelanoma skin cancer (zur Hausen, 2009). The E6 and E7 proteins have been defined as the two main viral oncoproteins (Section 12.2). Functional and biochemical differences between high-risk and low-risk E6 and E7 proteins are thought to account, at least in part, for the different transforming potential of these HPV types (Munger et al., 1991; Moody and Laimins, 2010).

12.1.3 Functions Shared by DNA Tumor Viruses

The DNATVs comprise the PVs, the adenoviruses, and the polyomaviruses, which share a small double-stranded circular DNA genome. In the 1930s, the study of papillomas in wild rabbits led to the discovery of the first DNATV, cottontail rabbit PV (Shope and Hurst, 1933), and it was later shown that adenoviruses were able to cause tumors in hamsters and that the SV40 polyomavirus caused tumors in primates (Eddy et al., 1962; Trentin et al., 1962). The DNATVs have been highly studied and have served as fundamental tools to elucidate basic molecular mechanisms such as DNA transcription and replication and to uncover some of the major cellular pathways perturbed in human cancer (Howley and Livingston, 2009). DNATVs provided the first identification of tumor suppressors such as p53 and retinoblastoma (Rb) (Javier and Butel, 2008). DNATVs code for multifunctional oncoproteins capable of inducing cell proliferation and immortalization (Howley et al., 1991). The transforming properties of the adenovirus E1A (AdE1A), simian virus 40 large T antigen (SV40 LT) and human papillomavirus E7 (HPV E7) oncoproteins are related to their ability to bind to and inactivate Rb (DeCaprio et al., 1988; Whyte et al., 1988; Dyson et al., 1989), while the AdE1b, SV40-LT, and HPV E6 proteins are able to interfere with the function of p53 (Sarnow et al., 1982; Werness et al., 1990).

12.2 THE PAPILLOMAVIRUS E7 ONCOPROTEIN

In the 1980s, reverse genetics studies in rodent cells revealed that the main transforming activity of high risk HPV is localized to the early region of the genome, which codes for the E6 and E7 ORFs (Bedell et al., 1987) (Fig. 12.1b). Expression of high-risk HPV E6 and E7 proteins leads to immortalization of primary human keratinocytes *in vitro* (Munger et al., 1989), and to the development of cervical tumors in estrogen-treated transgenic mice (Arbeit et al., 1996). In both models, HPV E7 shows stronger transformation potential when compared to E6. In addition, the E6 and E7 genes are highly transcribed in cervical cancer tissues and cervical cancer-derived cell lines such as CaSki and HeLa (Smotkin and Wettstein,

1986; Bouallaga et al., 2000), which led to the classification of E6 and E7 as the main oncoproteins of HPV. It has been shown that E7 is required both for the induction and maintenance of the transformed phenotype in human cells (Goodwin and DiMaio, 2000). The transforming activity of E7 may be potentiated as a consequence of its overexpression upon HPV genome integration (Section 12.1.1).

12.2.1 E7 Domains and Structure

E7 is a small acidic Zn^{2+}-binding protein, comprising ~100 amino acids. On the basis of structural and biophysical evidence, HPV16 E7 is composed of two domains: an N-terminal domain (E7N) spanning residues 1–40 and a C-terminal domain (E7C) spanning residues 40–98 (Fig. 12.2a). The E7N domain has been classified as an intrinsically disordered domain (IDD) (Section 12.5), while the E7C domain has a globular conformation (Liu et al., 2006; Ohlenschlager et al., 2006) (Fig. 12.7b). Three conserved regions have been identified in E7 and have been named CR1, CR2, and CR3 for "conserved regions" 1–3. CR1 and CR2 are located within E7N, while CR3 is located within E7C (Fig. 12.2a). The three conserved regions are required for E7 transforming activity (Phelps et al., 1988; Brokaw et al., 1994). CR1 is also required for E7 and Rb degradation and mediates

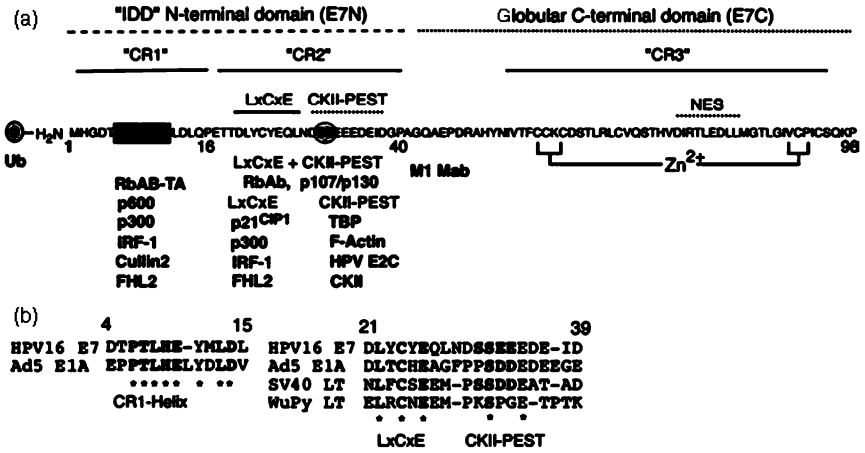


Figure 12.2 The human papillomavirus E7 protein. (a) HPV16 E7 sequence and domains. The E7N and E7C domains are marked above the sequence. CR1–3, conserved homology regions 1, 2, and 3; LxCxE, Rb-binding motif; CKII-PEST, Casein kinase II and PEST region; NES, nuclear export signal; Ub, ubiquitynation site. The main targets of the E7N domain and their interaction sites are displayed below the sequence. Violet box, E2F-like helix motif; green circle, serine-phosphorylation site. (b) Sequence conservation among DNATVs in the CR1 (left panel) and CR2 (right panel) regions. Numbers above the sequences denote the HPV16 E7 amino acid numbering. Asterisks denote conserved positions. HPV16 E7, human papillomavirus type 16 E7 protein; Ad5 E1A, adenovirus type 5 E1A protein; SV40 LT, Simian virus 40 large T antigen; WuPy LT, human Wu polyomavirus large T antigen. (*See insert for color representation of the figure.*)

binding to several cellular proteins (Section 12.2.2 and Fig. 12.2a). CR2 contains both the conserved LxCxE motif and the CKII-PEST region. The LxCxE motif mediates high-affinity binding to Rb, while the CKII-PEST region is a conserved acidic region containing one or two consensus (S/TxxD/E) sites for casein kinase II (CKII) recognition. The CR3 region corresponds to the globular C-terminal domain and mediates Zn^{2+} binding and dimerization. CR1 is conserved in PV E7 proteins and in adenovirus E1A proteins (Dyson et al., 1992), while CR2 is conserved in all three prototypical DNATV proteins (AdE1A, SV40-LT, and HPV E7) as well as in polyomavirus large T antigens (Dyson et al., 1990; Chellappan et al., 1992; Feng et al., 2008) (Fig. 12.2b). Additionally, the LxCxE motif is also found in several other viral proteins (Section 12.7.4).

12.2.1.1 Functional Role of CKII Phosphorylation In HPV16 and HPV18 E7 proteins, two serine residues (marked with a circle in the sequence of Fig. 12.2a) are phosphorylated both *in vitro* and *in vivo* by CKII (Barbosa et al., 1990). Phosphorylation is required for E7 transforming activity (Firzlaff et al., 1991) and for induction of S-phase reentry in infected keratinocytes (Banerjee et al., 2006; Genovese et al., 2008) and has been recently found to modulate Rb-binding affinity (Chemes et al., 2010) (Section 12.6.3). Phosphorylation of the conserved CKII sites also regulates transforming properties and Rb-binding affinity of AdE1A (Whalen et al., 1996), and nuclear–cytoplasmic shuttling of SV40-LT (Rihs et al., 1991), indicating that phosphorylation plays a functional role in all DNATV proteins.

12.2.1.2 E7 Degradation E7 is a short-lived protein, with a fast turnover in cells (Smotkin and Wettstein, 1987; Selvey et al., 1994) and whose proteasomal degradation involves ubiquitination at the N-terminus (Reinstein et al., 2000) (Fig. 12.2a). Additionally, the acidic region contained within CR2 of HPV16 E7 scores as a PEST protein-degradation motif (Rechsteiner and Rogers, 1996). PEST motifs have been shown to regulate turnover of short-lived cellular proteins (Lin et al., 1996) and of viral proteins such as the bovine papillomavirus (BPV) E2 protein (Penrose et al., 2004; Garcia-Alai et al., 2006). As will be further discussed in Section 12.5.3, biophysical evidence suggests that, in addition to the already described proteasomal degradation pathways, the PEST motif of E7 may also be involved in the regulation of E7 turnover.

12.2.2 E7 Targets

E7 has been reported to associate with over 100 cellular targets, a very large number taking into account the small size of this protein. These targets play diverse functional roles, and the significance of many if not most E7 interactions is currently unknown. For a recent review, see Moody and Laimins (2010). The primary, and best-characterized cellular target of the E7 protein is the Rb tumor suppressor. The Rb, p107, and p130 proteins comprise the "pocket protein" family, which has important roles in cell cycle regulation. Mutant E7 proteins incapable of binding Rb, p107, and p130 loose their transforming properties in cell culture (Phelps et al.,

1992) and the ability to induce S-phase reentry in differentiating keratinocytes in raft models (Banerjee et al., 2006). The E7–Rb interaction leads to disruption of Rb–E2F complexes, and in addition, high-risk HPV E7 proteins induce Rb degradation, thereby stimulating cell cycle progression (Boyer et al., 1996). The LxCxE motif of E7 constitutes the high-affinity Rb-binding site (Fig. 12.2a), but at least two other regions of E7 (one in CR1 and another in E7C) have been shown to contribute to Rb binding with lower affinity (Patrick et al., 1994; Chemes et al., 2010;) (Fig. 12.2a).

E7 has several Rb-independent functions, mediated through interactions with a wide array of cellular proteins that function as cell cycle regulators ($p21^{CIP1}$, $p27^{KIP1}$), transcriptional regulators (AP1, MPP2, TBP, and TAF-110), and chromatin structure modifiers (pCAF, Mi2β) (Moody and Laimins, 2010; Pim and Banks, 2010). E7 has also been found to interact with the PV E2 protein (Gammoh et al., 2006; Smal et al., 2009). Except for very few targets such as Rb, $p21^{CIP1}$, and HPV E2 (Lee et al., 1998; Singh et al., 2005; Ohlenschlager et al., 2006; Smal et al., 2009; Chemes et al., 2010), the affinities and specific binding sites of most E7 interactions are not known. The high number of interactions described for E7 suggests that this protein is able to interact in a networklike manner, similar to related viral oncoproteins such as AdE1A (Ferrari et al., 2008). A high number of E7 targets bind to sites located in the intrinsically disordered E7N domain including the CR1 region, the LxCxE motif, and the CKII-PEST region (Fig. 12.2a). Owing to its small size, one can hardly imagine such a wide range of interactions with many structurally unrelated proteins as being "specific." The promiscuity of these interactions must necessarily be related to distinctive structural and conformational properties, which are discussed throughout this chapter.

12.3 HPV16 E7: CONFORMATIONAL EQUILIBRIA AND STRUCTURE

The first reports of recombinant expression and purification of HPV16 E7 showed, through affinity pull-down assays, that the protein could bind underphosphorylated Rb and dissociate the Rb–E2F complex, and the far-ultraviolet (UV) circular dichroism (CD) spectrum of E7 indicated some α-helix content (Imai et al., 1991; Pahel et al., 1993). The first indications that E7 might contain nonglobular regions came from early studies of translation products and from the identification of the HPV16 E7 polypeptide in model cancer-derived cell lines. In these studies, it was found that E7 migrated as an ~20-kDa polypeptide in a reducing (sodium dodecyl sulfate) (SDS)-polyacrylamide gel electrophoresis (PAGE), even though its expected molecular weight is 11 kDa (Smotkin and Wettstein, 1987).

12.3.1 Hydrodynamic Properties

The above reported anomalous electrophoretic behavior of E7 is mirrored by an equally anomalous hydrodynamic behavior under native conditions in solution because E7 has an elution profile corresponding to that of a 44-kDa globular

protein in size exclusion chromatography (SEC). Techniques such as multiangle light scattering, which provide a measure of the molecular weight that is independent of the hydrodynamic radius, yield a molecular weight of 22 kDa matching that expected for an E7 dimer (Alonso et al., 2002). Sedimentation equilibrium experiments show that HPV16 E7 displays a monomer–dimer equilibrium with a dissociation constant in the order of 1 μM (Clements et al., 2000). Finally, dynamic light scattering of E7 shows a hydrodynamic diameter of about 6.8 nm, twice the diameter expected for a globular dimer of the same molecular weight as E7 (Smal et al., 2009). These experiments indicate that E7 can associate to form a nonglobular, extended, and rather weak dimer in solution, which has hydrodynamic properties comparable to those of a 44-kDa globular protein (Fig. 12.3).

12.3.2 Secondary and Tertiary Structure

The first evidence of the complex structural organization of the E7 protein came from the analysis of its CD spectrum. The far-UV CD spectrum of E7 shows a minimum at 205 nm, with a negative band at 220 nm and a maximum at 190 nm (Fig. 12.5a). The band at 220 nm is indicative of α-helix content, and the minimum at 205 nm can be interpreted as a mixture of α-helix (208 nm) and disordered (200 nm) structure. Evidences such as the presence of a strongly bound and constitutive Zn^{2+} atom per monomer, a high secondary structure content, the ability to dimerize, the observation of a cooperative chemical unfolding transition, and the ability to bind ANS (1-anilino-8-naphtalene-sulfonate) (Fig. 12.3) clearly indicate the presence of tertiary structure. Nevertheless, the CD spectrum is in agreement with the low dispersion of chemical shifts observed in ^1H nuclear magnetic resonance (NMR) spectra, indicating the presence of disordered structure (Alonso et al.,

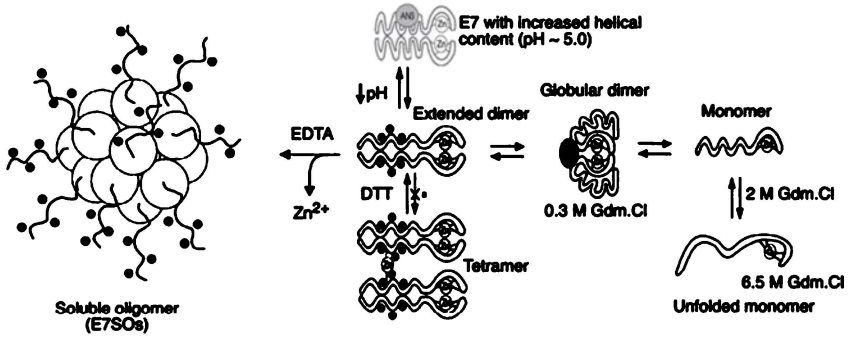


Figure 12.3 Conformational transitions of the HPV16 E7 protein. The hydrodynamic diameters are 6.8 nm for the extended dimer and 50 nm for the soluble oligomers (see text for details). Zn^{2+}, zinc ion; ANS, 1-anilinonaphthalene-8-sulfonic acid; Gdm.Cl, guanidinium chloride; DTT, dithiothreitol, EDTA: ethylenediamine tetraacetate, (−): negative charges. Double arrows, reversible processes; single arrows, irreversible processes. *Source*: Adapted from (Alonso et al., 2002).

2002). Furthermore, E7 has an extremely high thermal stability and a low overall cooperativity of unfolding. These biophysical properties, together with the high net charge/hydrophobicity ratio, include E7 in the list of intrinsically disordered proteins (IDPs) (Alonso et al., 2002; Uversky et al., 2006). Given the lack of structural information, E7 was initially proposed to have a conformation whose nature was intermediate between globular protein and IDP, or alternatively, to be formed by two domains, each presenting one of these structural properties (Alonso et al., 2002).

12.3.3 Three-Dimensional Structures

Structural information of E7 became available almost 20 years after the first description of E7 as a viral oncogene (Phelps et al., 1988), and 15 years after the first report on the purification of recombinant E7 protein (Imai et al., 1991), which pointed out the unusual behavior of this protein in solution. Structures of the isolated C-terminal domain from two HPV E7 proteins (types 1A and 45) have recently been determined by X-ray crystallography (Liu et al., 2006) and NMR (Ohlenschlager et al., 2006), respectively. Both structures reveal a fold consisting of monomers with a $\beta1\beta2\alpha1\beta3\alpha2$ topology (Fig. 12.7 c–d). In each monomer, one Zn^{2+} atom is coordinated by four cysteine residues. Two monomers associate to form a symmetric and flattened dimer with a 1820-Å^2 interface made up of residues that also participate in each monomer's inner core, thus creating a contiguous hydrophobic core (Fig. 12.7c). The structure of HPV45 E7 was solved by NMR using the isolated E7 C-terminal domain (Ohlenschlager et al., 2006). This study reported that the NMR spectrum of full-length HPV45 E7 showed additional signals in a very narrow chemical shift range when compared to the spectrum of the isolated C-terminal domain and that signals arising from the C terminus remained unchanged (Ohlenschlager et al., 2006). This low chemical shift dispersion was indicative of a disordered and flexible conformation of the N-terminal domain, in accordance with previously described results for HPV16 E7 (Alonso et al., 2002). The coincidence in the signals arising from the C-terminal domain in both constructs indicated that the isolated C-terminal domain could fold independently to a structure very similar to the one adopted in the context of the full-length protein (Ohlenschlager et al., 2006). The fact that the N-terminal domain could neither be crystallized nor assigned by NMR is a further indication that this region of E7 is responsible for its anomalous hydrodynamic properties (Alonso et al., 2002).

12.3.4 Conformational Equilibria

12.3.4.1 Structural Transitions of the E7 Dimer Under mild denaturing conditions or upon variation of the solution conditions, the extended conformation of HPV16 E7 is altered. In 0.3*M* guanidinium chloride (Gdm.Cl), E7 remains as a dimer as indicated by light scattering measurements and shows a reversible increase in canonical α-helix secondary structure content and a decrease in its hydrodynamic radius, indicating a shift to a more globular conformation (Fig. 12.3). This form

of E7 has an increased ANS-binding capacity, indicative of exposure of hydrophobic surfaces to the solvent (Alonso et al., 2002). This structural rearrangement is thought to involve interactions between the E7N and E7C domains. Moreover, E7 undergoes a reversible pH-dependent conformational transition from pH 7.0 to 5.0, which involves the exposure of those hydrophobic surfaces to the solvent, and an increase in its α-helical content (Fig. 12.3). At pH 2.5, the far-UV CD spectrum also shows an increase in α-helical content (Alonso et al., 2002). The effect of pH on E7 conformation is probably caused by the protonation of the numerous acidic side chains present within E7, which breaks the charge repulsion that holds an extended conformation, allowing for the formation of α-helix structure. Experimental evidence from Gdm.Cl unfolding and refolding experiments (Alonso et al., 2002) and from equilibrium sedimentation (Clements et al., 2000) support the existence of monomeric and tetrameric forms of E7, adding to the conformational plasticity of this protein (Fig. 12.3).

12.3.4.2 Oligomerization Early biochemical and more recent structural data indicate that the zinc atom bound by the C-terminal domain of each monomer is essential for the folding of this domain (McIntyre et al., 1993; Ohlenschlager et al., 2006). When Zn^{2+} is removed by a chelating agent, HPV16 E7 readily self-assembles into spherical particles with an average molecular weight of 790 kDa and a diameter of 50 nm (Alonso et al., 2004) (Fig. 12.3). The E7 oligomers, named E7SOs (for *E7 S*pherical *O*ligomers) are soluble and conformationally homogeneous. CD analysis reveals that E7SOs formation is likely to involve a coil to β-sheet transition in E7, and E7SOs bind congo red and thioflavin T, indicating the presence of amyloidlike, repetitive β-sheet structure (Klunk et al., 1999; LeVine, 1999). However, the E7 oligomers are highly soluble and thermo-stable and show no traces of insoluble fibrillar material. The features of oligomerization reveal a limited and ordered process, clearly distinct from amorphous aggregation. Several lines of evidence indicate that the E7C domain mediates E7 oligomerization. First, Zn^{2+} removal from E7C is a requirement for E7SOs formation. Second, the isolated E7N domain does not oligomerize upon EDTA (ethylenediamine tetraacetate) addition. Third, E7SOs assembly is accompanied by the burial of side chain 93 located in the E7 C-terminal domain, as revealed by the I93W E7 mutant(Alonso et al., 2004). Taken together, these results provide strong evidence that oligomerization requires the C-terminal domain of E7, although no structural information could be obtained to date on E7SOs. Last, E7SOs undergo spontaneous oxidation by forming a network of disulfide bridges and do not dissociate or unfold upon Gdm.Cl addition, indicating that this structure is highly stable and irreversible (Alonso et al., 2004).

12.4 CONFORMATIONAL DIVERSITY OF THE E7 PROTEIN *IN VIVO*

Studies of E7SOs showed that this species displayed *in vitro* chaperone-holdase activity for model substrate proteins (Section 12.6.1) (Alonso et al., 2006). Given

the ability of E7 to form oligomers with amyloid characteristics and chaperone-holdase activity, we set out to investigate the possible existence and localization of such conformers *in vivo* in established prototypic and transiently transfected human cell lines. Conformation-specific antibodies located endogenous oligomeric E7 species in the cytosol of CaSki, TC1, and E7-transfected U-2OS cell lines, showing a strong colocalization with amyloid structures (Dantur et al., 2009). These cytosolic oligomers appear as the most abundant species in all cell systems tested, and were found to be in equilibrium with dimeric–monomeric forms that localized to the nucleus (Fig. 12.4). Nuclear E7 levels are replenished dynamically from the cytosolic pool and do not result from protein synthesis.

Therefore, similar to what is observed *in vitro*, *in vivo* studies demonstrate a variety of E7 conformers that exist in cells and which could be related to the transforming activity of E7 (Fig. 12.4). Long-term events related to derepression of E7 expression, for example upon HPV genome integration, would cause accumulation of oligomeric E7 species, a possibility further supported by the detection of E7 oligomers in the cytosol of cancerous cells from tissue biopsies (Dantur

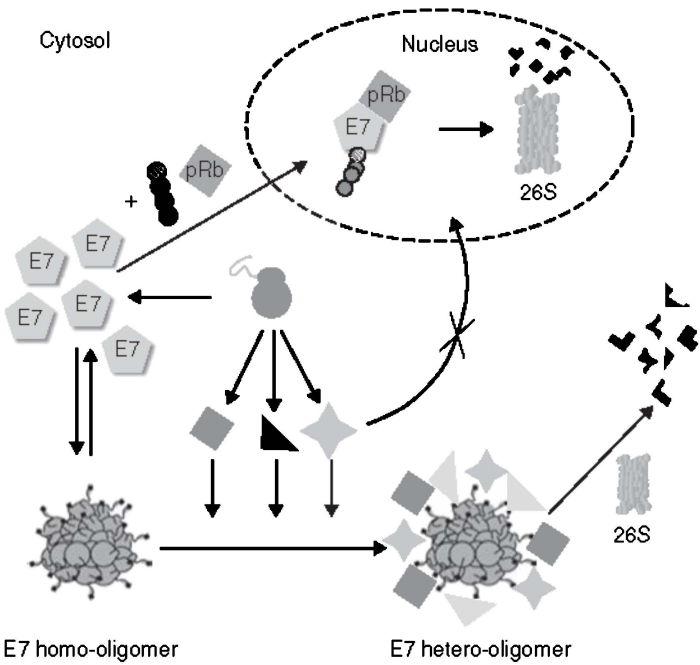


Figure 12.4 Distribution of E7 conformers in cells. Scheme depicting the localization of E7 conformers within mammalian cells. Grey shapes represent cellular E7 targets, which can be bound by E7 oligomers and degraded by the 26S proteasome (indicated as 26S). Both E7 (pentagon) and Rb (diamond) can be degraded in the nucleus. *Source:* Reproduced from Dantur et al. (2009), Copyright © [2009], copyright owner as specified in the journal.

et al., 2009). E6 and E7 share a common origin of their "treble-cleft" fold with other viral and cellular RING domain proteins (de Souza et al., 2009) (Section 12.7.1). RING domains mediate oligomerization in other proteins such as the *Arenavirus* Z protein (Kentsis et al., 2002b) and cellular PML protein (Kentsis et al., 2002a). Interestingly, the supramolecular assemblies formed by these proteins both *in vitro* and *in vivo* have morphologies similar to those of the E7 oligomers (Kentsis et al., 2002a).

12.5 E7N IS A BONA FIDE IDD

The intrinsically disordered nature of full-length HPV16 E7 (Alonso et al., 2002) and the compact tertiary structure of the E7 C-terminal domain (Liu et al., 2006) suggested that the N-terminal domain of E7 was the domain responsible for the extended and nonglobular properties of the full-length protein. As mentioned, E7N contains conserved regions CR1 and CR2, which present high sequence conservation among different PV types (Section 12.7.2). Even though CR1 and CR2 are frequently referred to as *domains*, because of their small size it is more accurate to refer to them as *linear motifs* (Gould et al., 2010). These include the conserved motif located within CR1 and the LxCxE Rb-binding motif and the acidic CKII-PEST region containing the CKII phosphorylation site located within CR2 (Fig. 12.2a). Despite the highly modular characteristics of the E7 protein, it remained unclear whether E7N qualified as an independent structural domain, which led us to investigate the biophysical properties of the isolated E7N fragment.

12.5.1 Noncanonical Secondary Structure within E7N

The far-UV CD spectrum of the N-terminal 40 amino acids of HPV16 E7 at neutral pH is characteristic of a disordered polypeptide, being devoid of any evidence of persistent canonical secondary structure and presenting a minimum at 198 nm (Fig. 12.5a). Subtraction of the E7N spectrum from that of the full-length protein yields a spectrum characteristic of α-helical conformation, with two minima at 208 and 222 nm, which is in accordance with the CD spectrum of the isolated E7C domain (Fig. 12.5a). This result suggests that E7N contains the disordered regions of E7 and that if there are any interactions between domains, they do not have a major effect on structure. The hydrodynamic radius of E7N is much larger than that expected for a random-coil peptide of the same number of residues. This has been assessed by SEC and pulse field gradient NMR, where E7N yielded a Stokes radius of 20.9 Å (SEC) and 21.16 Å (NMR) (Garcia-Alai et al., 2007), a result which confirms its extended nature (Fig. 12.5b).

A close examination of the CD spectrum of E7N at neutral pH shows a positive band at 218 nm and a negative band at 228 nm, which disappear at low pH (Fig. 12.5c). This behavior is characteristic of polyproline type II (PII) conformation and was previously investigated in poly glutamic and poly lysine model peptides (Tiffany and Krimm, 1968). In E7N, the pH-induced conformational transition can

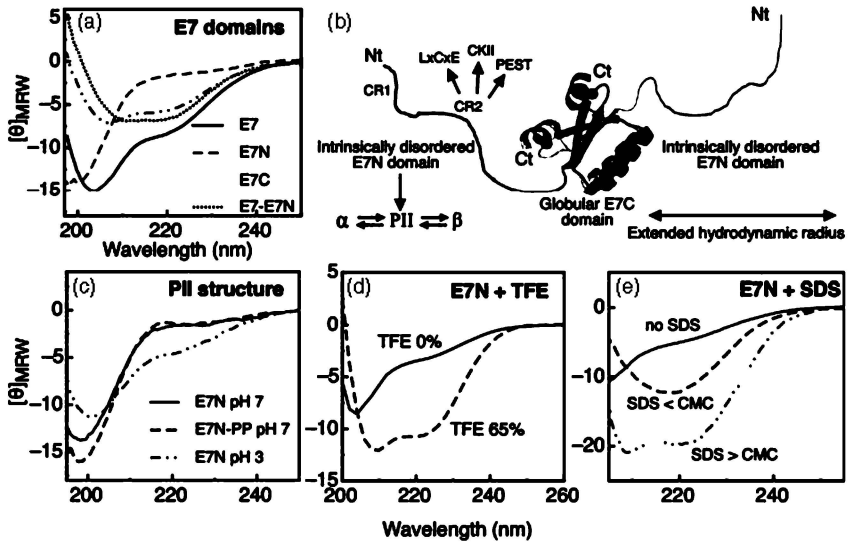


Figure 12.5 The E7N IDD. (a) Circular dichroism (CD) spectrum of full-length E7 (full black line), E7N (broken green line), E7C (broken and dotted red line), and E7–E7N difference spectrum (dotted black line). (b) Schematic diagram of the extended E7 dimer depicting conserved regions (CR1 and CR2) and motifs (LxCxE, CKII, PEST). Monomers are shown in red and yellow, respectively; gray spheres, Zn^{2+} atom; Nt, N-terminus; Ct, C-terminus. The structure of the E7C domain is from PDB ID: 2F8B. (c) E7N PII structure. CD spectrum of E7N at neutral pH (full green line) and acidic pH (broken and dotted red line), and CD spectrum of phosphorylated E7N (broken black line). (d) Effect of TFE on E7N conformation. CD spectrum of E7N at 0% v/v TFE (full green line) and 65% v/v TFE (broken black line) at pH 4.0. (e) Effect of SDS on E7N conformation. CD spectrum of E7N in the absence of SDS (full green line) and in the presence of SDS at submicellar (1 mM) (broken black line) and supramicellar (10 mM) (broken and dotted red line) concentrations. *Source:* Adapted from Garcia-Alai et al. (2007). (*See insert for color representation of the figure.*)

be explained by the protonation of the negatively charged residues present within the CKII-PEST region, which disrupts PII structure. This spectroscopic signature of PII conformation has also been confirmed by correlation to NMR measurements (Garcia-Alai et al., 2006). PII structure is stabilized at low temperature (Shi et al., 2002), an effect which is also observed in E7N. The loss of extended PII structure is also evidenced by a decrease in the hydrodynamic radius of E7N at pH 4.0. The conformational change observed in E7N at low pH can also involve an increase in α-helix content caused by neutralization of repulsive negative charges that stabilize an extended conformation, as observed in the full-length E7 protein (Section 12.3.4) (Alonso et al., 2002) (Fig. 12.3).

E7N also exhibits α-helical propensity. Preexisting or residual α-helical populations of E7N can be calculated from trifluoroethanol (TFE) stabilization experiments (Jasanoff and Fersht, 1994), which show that the α-helical population at 60% v/v [TFE] fluctuates from 14% at pH 7.5 to 29% at pH 4.0. This is an

indication that charge neutralization at low pH disrupts PII structure, allowing for a higher content of α-helix evidenced upon TFE addition. The residues involved in helix formation appear to be the same at high and low pH, as suggested by the fact that the stability (ΔG^{H_2O}) and the *m*-value for α-helix formation are the same (Garcia-Alai et al., 2007). These results also suggest that the residues involved in PII structure are not involved in the nucleation of the α-helix.

Addition of detergents such as SDS can stabilize both α-helix and β-sheet structures depending on its concentration and the nature of the peptides (Wu et al., 1981; Zhong and Johnson, 1992). A behavior similar to that observed at low pH is observed by addition of the anionic detergent SDS to E7N. Whereas no α-helical structure is induced at pH 7.5 at concentrations above the critical micellar concentration (CMC) as expected for this detergent, a drastic increase in α-helix is observed at pH 4.0 above CMC values (Fig. 12.5d) and a transition to β-sheet structure is observed at submicellar concentrations (Fig. 12.5e). The fact that the formation of both α-helix and β-sheet structures takes place at pH 4.0 but not at pH 7.5 is an indication that when PII structure is disrupted, E7N can undergo different structural transitions. This structural plasticity of E7N is remarkably similar to that observed in the IDP α-synuclein, which exhibits both α-helix and β-sheet transitions depending on the chemical environment (Woods et al., 2007).

Detailed residue-by-residue NMR analysis of E7N remains an essential but technically challenging task (Daughdrill et al., 2005). Nevertheless, spectroscopic analyses clearly reveal the presence of regions presenting PII, α-helix, and β-sheet structures in equilibrium with disordered (coil) regions. Furthermore, E7N has the ability to fine-tune these different conformations depending on the chemical environment (Garcia-Alai et al., 2007). In α-synuclein, α-helix and β-sheet conformational transitions correlate with specific target protein-binding activities (Ferreon et al., 2009a), acting as a molecular switch. Similarly to α-synuclein, the conformational plasticity of E7N and its acidic nature provide plausible molecular explanations for the broad protein-binding specificity of the E7 protein, allowing for differential target protein binding of E7N conformers and for nonspecific electrostatic interactions with targets presenting positively charged interaction surfaces.

12.5.2 E7N and Oligomerization

It was shown that E7N is exposed to the solvent in the E7SOs, while the oligomerization interface is buried from the solvent (Alonso et al., 2006) (Fig. 12.3). The varied functionality that E7 can exert on diverse cellular proteins involves direct protein–protein interactions, which in most cases require the E7N domain (Section 12.2.2). This suggests that the E7SOs species retain the ability to interact with many E7 targets, which has been shown to be the case for the E7SOs–Rb interaction (Alonso et al., 2006). Whereas the full-length E7 protein forms stable and soluble oligomers on Zn^{2+} removal (Section 12.3.4), the oligomers formed by the isolated E7C domain progress readily to insoluble aggregates (Smal, 2010). Therefore, besides providing with a molecular scaffold for protein–protein interactions,

the E7N IDD provides a hydrophilic module that maintains protein solubility of the E7SOs.

12.5.3 Polyproline II Structure in E7N and its Modulation by Phosphorylation

HPV16 E7 is phosphorylated at serine residues S31 and S32, which are located within the CKII-PEST region of E7 (Fig. 12.2a). As discussed in Section 12.5.1, this region presents dynamic combinations of coil, PII, α-helix, and β-sheet structures and confers E7N with its extended and intrinsically disordered conformation. Phosphorylation of both the E7N domain (Garcia-Alai et al., 2007) and the $E7_{16-40}$ (CR2) fragment (Chemes et al., 2010) causes an increase in PII content (Fig. 12.5c). A similar type of PII structure was observed in a disordered hinge region of the BPV E2 protein that contains a CKII-PEST degradation motif. CKII phosphorylation within this region disrupts PII structure and leads to unstable polypeptides susceptible to intracellular degradation, indicating that phosphorylation causes a loss in local thermodynamic stability (Penrose et al., 2004; Garcia-Alai et al., 2006). As native proteins need to be at least partly unfolded for proteolytic degradation, local, or global thermodynamic stabilization opposes degradation (Johnston et al., 1995; Lee et al., 2001). Given that PII structures are predominant in the unfolded state or in exposed regions of proteins (Tompa and Csermely, 2004), the marginal stability of these regions ensures a fast and effective way to control intracellular protein levels, and thus, protein activity through the regulation of PII content (Garcia-Alai et al., 2006).

The E7N domain plays a role in E7 turnover, because ubiquitination of the N terminus is required for proteasome-mediated degradation (Reinstein et al., 2000). The structural features of the E7N IDD resemble those of degradation initiation sites (Prakash et al., 2004). Therefore, similar to what is observed in BPV E2, E7 phosphorylation may act as a conformational switch, regulating E7 protein stability and turnover. Furthermore, E7 conformers with different susceptibility to phosphorylation may have different turnover rates, as may be the case for E7SOs, which cannot be phosphorylated *in vitro* by CKII (Alonso et al., 2004). The fact that E7SOs are not phosphorylated implies that the site is not exposed enough or cannot be recognized by the modifying enzyme. This result is surprising given that the E7N domain faces the solvent in the E7SOs, potentially exposing both the CKII phosphorylation and PEST sites (Alonso et al., 2004) (Fig. 12.3), and indicates an accessibility hindrance of yet unknown nature. Nevertheless, the *in vitro* phosphorylated dimeric full-length E7 can form E7SOs (Alonso et al., 2004), in line with evidence suggesting that oligomerization is mediated by the E7C domain (Section 12.3.4).

12.6 INTERACTION MECHANISMS OF HPV16 E7

The only enzymatic activity coded for in the HPV genome is the helicase activity of the E1 protein, which participates together with E2 in origin recognition and

subsequent unwinding for replication initiation. So far, There is no other chemical activity described for HPV proteins, and E7 is no exception. As mentioned in Section 14.2.2, E7 is a scaffold protein whose action is mediated through its interaction with numerous cellular proteins. Of the many interaction targets described for E7 (Section 14.2.2), only a few mechanisms have been analyzed in detail in solution using purified components, and a single structure has been published, that of the HPV16 E7 nine-residue peptide containing the LxCxE motif (DLYCYEQLN) bound to the RbAB domain (Lee et al., 1998). The few interactions analyzed involve binding sites located within the E7N IDD (Section 14.2.2 and Fig. 12.2a).

12.6.1 HPV16 E7 Chaperone-Holdase Activity

In vitro studies of the E7SOs species showed that it is able to prevent incorrect folding and aggregation of two model chaperone substrates, citrate synthase (CS) and luciferase, and that this activity takes place at substoichiometric concentrations (Alonso et al., 2006) (Fig. 12.6a). E7SOs are able to bind these model polypeptides after partial thermal denaturation or in the late stages of renaturation, holding them in a folding-competent conformational state with significant secondary and tertiary structure. E7SOs do not require ATP for this activity, and are unable to bind the fully folded native proteins (Alonso et al., 2006), which defines E7SOs as a *chaperone holdase*. The E7 dimer does not present chaperone activity.

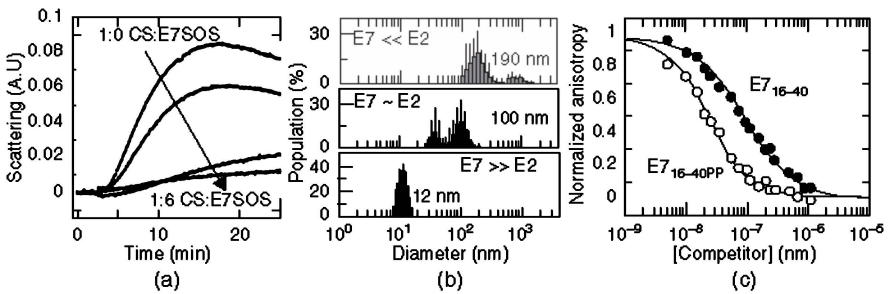


Figure 12.6 HPV16 E7 protein–protein interactions. (a) E7SOs-induced inhibition of citrate synthase (CS) thermal aggregation followed by light scattering. The E7SOs:CS molar ratio is expressed considering the concentration of monomeric species. CS concentration was 150 nM. *Source*: Adapted from Alonso et al. (2006) (b) Size distribution of HPV E7–E2 complexes measured by dynamic light scattering. The size distribution was measured at different relative E7:E2C concentrations, at 2 μM E2C. The mean size of the particle population is shown in each graph. *Source*: Adapted from Smal et al. (2009) (c) Increase in RbAB binding affinity caused by phosphorylation of the E7 CKII-PEST region. A stoichiometric complex of 25 nM RbAB and FITC-$E7_{16-31}$ was competed with unphosphorylated CR2 peptide ($E7_{16-40}$, filled circles) or with phosphorylated CR2 peptide ($E7_{16-40}$PP, open circles). *Source*: Adapted from Chemes et al. (2010).

The chaperone activity of E7SOs may play a role in Rb degradation. Consistent with this possibility, the oligomeric E7 chaperone binds strongly to full-length native Rb. This is not unexpected, given the fact that the high-affinity LxCxE Rb-binding motif is located in E7N IDD and that this domain is exposed to the solvent in E7SOs, being accessible for protein–protein interactions (Alonso et al., 2006) (Fig. 12.3). The ~70 molecules of E7 that form E7SOs(Alonso et al., 2004) would present a high concentration of accessible Rb-binding motifs, providing with a high-avidity molecular assembly that could sequester Rb from the Rb–E2F complexes or target the tumor suppressor for proteasomal degradation (Fig. 12.4). Interestingly, the E7N IDD is able to compete with E7SOs for Rb binding but does not interfere with E7SOs chaperone-holdase activity (Alonso et al., 2006).

The polyomavirus SV40-LT protein requires both the LxCxE motif and its J domain for displacement of the Rb–E2F complex and for proteasomal targeting of Rb (Sullivan et al., 2000). An interaction between the SV40-LT J domain and the hsc70 cellular chaperone is required for this activity. Therefore, for both viral proteins, a chaperone activity (coded directly by E7 and coded in a separate protein in the case of SV40-LT) seems to be required for Rb degradation. In line with this evidence, the AdE1A protein, which has no known associations to chaperone activities, is able to bind Rb and displace E2F but does not cause Rb degradation.

12.6.2 Interaction of E7 with the E2 Master Regulator

As mentioned in Section 12.1, the HPV E2 protein controls gene transcription and viral genome replication activity, essential for the virus life cycle (Hamid et al., 2009; Thierry, 2009). E2 acts as a loading factor for the viral E1 helicase at the HPV origin of replication. E2 can also repress transcription of the E6/E7 ORFs (Section 12.1.1). The gene-silencing effect of E2 depends on interaction with bromodomain 4 (Brd4) and other cellular proteins. The E2 proteins are about 400-amino acid polypeptides consisting of an N-terminal transactivation domain (about 200 amino acids) and a C-terminal DNA-binding and dimerization domain (E2C, about 85 amino acids), separated by a nonconserved "hinge" domain. E2C domains from all strains analyzed so far share a particular folding topology, the dimeric β-barrel (Hegde, 2002).

Integration of the viral DNA genome into the host cell genome plays a role in HPV-mediated carcinogenesis, and there is a correlation between the progress of the cancerous lesion and DNA integration (Kalantari et al., 1998). The E2 ORF is disrupted when integration takes place (Section 12.1.1), and the consequent loss of E2-mediated repression of E6 and E7 transcription increases E6 and E7 expression levels with the consequent increase in proliferation and genetic instability (Section 12.1.1). Interaction between E2 and E7 has been reported *in vivo*, where E2 was capable of inhibiting E7–ras cooperation in cell transformation and had a direct effect on E7 stability and localization (Gammoh et al., 2006, 2009).

Interaction of the HPV16 E7 protein with the DNA-binding domain of the HPV16 E2 protein leads to different soluble (12, 40, and 115 nm diameters, respectively) and insoluble complexes depending on the relative concentration of each

protein in solution (Smal et al., 2009) (Fig. 12.6b). The interaction is strongly dependent on ionic strength, which suggested that the acidic region within E7N was involved. Indeed, the N-terminal domain of E7 interacts with E2C with a K_D of 0.1 μM, and the stretch of residues 25–40 of E7, encompassing both the PEST motif and the two phosphorylation sites (S31 and S32), was mapped as the primary binding site (Fig. 12.2a) (Smal et al., 2009). Displacement of the soluble E7–E2C complex by an E2 site DNA duplex and site-directed mutagenesis indicated that the protein–protein interface involves the DNA-binding helix of E2 (Smal et al., 2009). These results suggest a finely tuned mechanism for regulating the relative availability of the E2 and E7 proteins. Through this mechanism, E7 could block its own repression through interaction with the E2 repressor, an interaction that involves the E7N IDD. This implies that the imbalance in the relative protein levels that would result from the loss of E2 on integration of the viral genome to the host cell would potentiate the increase of both E7 and E6 oncoproteins. In addition, E7 could sequester E2 in oligomeric species in the cytosol before integration, thus increasing E7 expression levels in cells bearing episomal viral DNA.

12.6.3 Interaction of E7 with the Retinoblastoma Tumor Suppressor

As described in Section 12.1, Rb was the first target of HPV E7 to be uncovered. Despite early semiquantitative analysis of this interaction (Jones et al., 1990; Munger et al., 1991; Dong et al., 2001), structural information is limited to the structure of the RbAB domain bound to a nine-residue peptide containing the Rb-binding motif (DLYCYEQLN) of HPV16 E7 (Lee et al., 1998). The structure shows that the peptide adopts an extended conformation, which is stabilized by hydrogen bonds involving both backbone and side chain residues and by hydrophobic interactions involving the L22, C24, E26, and L28 side chains of the peptide (Lee et al., 1998). Both the acidic nature of the E7 peptide and the presence of a lysine patch surrounding the binding pocket suggest that this interaction is further stabilized by electrostatic interactions (Dick and Dyson, 2002).

A detailed thermodynamic study on the E7–RbAB interaction was performed using a set of fragments spanning the E7N IDD, the E7C domain, and the full-length E7 protein, which allowed the dissection of all E7 regions involved in RbAB binding and of their energy of interaction (Chemes et al., 2010). E7 can bind RbAB as a monomer with a 1:1 stoichiometry, and 90% of the total binding energy is provided by the LxCxE motif (Fig. 12.2a), with this motif also being the main determinant of binding for the HPV11 and HPV18 E7 proteins. The free energy of binding for full-length E7 is 1.0 kcal/ mol higher than that of the E7N domain, indicating the presence of an additional binding determinant located in the E7 C-terminal domain. This dual-binding mode of the E7 monomer is conserved among prototypic high risk and low risk E7 proteins (Chemes et al., 2010).

Several regions within the E7N IDD participate in the RbAB interaction. A low-affinity ($K_D = 20$ μM) site was described in the E7 CR1 region (Fig. 12.2a), and this isolated region was shown to present high α-helix propensity *in vitro* (Chemes

et al., 2010). The CR1 site of AdE1A also binds to RbAB with similar micromolar affinity ($K_D = 1$ μM), and the structure of the complex reveals that the binding surface involves the E2F binding site of RbAB. Furthermore, the AdE1A CR1-binding motif presents an α-helix conformation (Liu and Marmorstein, 2007). These results highlight the structural and functional similarity between the E7 and AdE1A proteins (Ferreon et al., 2009b). Despite the fact that the CR1 site does not contribute to binding of the E7 monomer (Chemes et al., 2010), because of the short distance separating the CR1 and LxCxE motifs, reports on E7–RbAB complexes with 2:1 stoichiometry (Clements et al., 2000) support the possibility that the CR1 region participates in binding of the E7 dimer to Rb. Serine phosphorylation at the conserved CKII sites adjacent to the LxCxE motif increases RbAB binding affinity (Fig. 12.6c) (Chemes et al., 2010). As phosphorylation increases the PII content of the isolated E7 CR2 region (Section 12.5.3), it is possible that this posttranslational modification favors an extended conformation that facilitates the E7–RbAB interaction (Chemes et al., 2010). Structural studies of the SV40-LT/RbAB complex show high-temperature factors for the residues of the CKII-PEST region, suggesting that it remains flexible on complex formation (Kim et al., 2001). This suggests the possibility that phosphorylation induces a structural change which brings the CKII-PEST region into contact with the RbAb surface (Kim et al., 2001).

The properties of the E7–RbAB interaction highlight the structural modularity of E7 and of its interaction with target proteins, bringing together the IDP nature of the E7N domain and its function. The E7N IDD contains several tightly packed short motifs (CR1, LxCxE, and CKII-PEST) that present structural plasticity and participate in RbAB binding (Garcia-Alai et al., 2007; Chemes et al., 2010), similar to what is observed in the AdE1A protein (Ferreon et al., 2009b) (Section 12.7.3) and in other IDP proteins such as α-synuclein (Ferreon et al., 2009a). The wide range of affinities found for the E7–RbAB interaction and the similarity to the affinities reported for cellular Rb targets (Chemes et al., 2010) suggest that although E7 is optimized for high-affinity binding to Rb through the LxCxE motif, the effect of E7 on the host cell protein–protein interaction network will include Rb as well as many of the large number of interaction targets reported (Moody and Laimins, 2010) and will depend strongly on E7 expression levels.

12.7 EVOLUTION OF THE E7 PAPILLOMAVIRUS PROTEIN

12.7.1 Origin and Diversification of the E6 and E7 Papillomavirus Genes

A plausible genomic history of PVs can be reconstructed from several lines of evidence (Garcia-Vallve et al., 2005). It was proposed that the common ancestor of the present PVs had a genome that coded for just the early replication proteins E1 and E2 and the late structural proteins L1 and L2 (Garcia-Vallve et al., 2005). The incorporation of an early transforming gene to the PV genome was a crucial event. Not being an essential part of the replication machinery of the virus, the new

gene was able to evolve new functions quickly and play an adaptive role in the infection of new hosts and tissues (Garcia-Vallve et al., 2005). This led to drastic changes in sequence and structure during PV evolution. Before the diversification of PVs, the primordial transforming gene underwent duplication, leading to what would become the extant E6 and E7 ORFs (Cole and Danos, 1987; de Souza et al., 2009). The common origin of E6 and E7 is reflected in both sequence and structure similarities (Cole and Danos, 1987; de Souza et al., 2009), which also indicate that their ancestor was probably a host "treble-clef fold" domain with two zinc-binding sites, each composed of two CxxC motifs (de Souza et al., 2009). Early on, the E6 protein lost one of the zinc-binding sites and acquired two extra helices (de Souza et al., 2009). After the divergence between avian/reptilian and mammalian PVs, the E6 gene itself underwent an internal duplication event and attained its present, longer form (Van Doorslaer et al., 2009).

E7 genetics is best discussed in terms of the E7N and E7C domains (Fig. 12.2a). During PV evolution, E7C gained the ability to dimerize and lost one of the zinc-binding sites and the N-terminal strand (de Souza et al., 2009). E7N shows a high degree of conservation across most PVs, but in reptilian (Herbst et al., 2009), avian (Terai et al., 2002), and some artiodactyl (Tomita et al., 2007) PVs, E7N is substituted by a domain with no sequence similarity to canonical E7N sequences. This genetic information supports the view, derived from the nearly independent conformations of E7N and E7C, that they are two separate, bona fide domains. The presence of diversifying evolutionary pressures in the PV transforming genes also manifests itself in the fast rate of evolution of E6 and E7 compared to the core PV genes (Garcia-Vallve et al., 2005) and the disappearance of either the E7 (Van Bressem et al., 2007) or the E6 (Chen et al., 2007; Tomita et al., 2007; Van Doorslaer et al., 2009) ORFs in some lineages. Once the E6 and E7 genes of a PV type have adapted to a given evolutionary niche, most codons are under strong purifying selection to retain the acquired traits. This leads to low sequence variability and less evidence for intratype adaptive selection compared to intertype adaptive selection (Chen et al., 2005; Chen et al., 2009).

12.7.2 Sequence Conservation in the E7 Papillomavirus Protein

Function of the PV E7 transforming protein is tightly coupled to its fine structural properties, most prominently for the intrinsically disordered E7N domain. We now review the interplay between the structure and function of E7 and its evolution.

12.7.2.1 Sequence Conservation in E7C Figure 12.7a and 12.7b displays sequence conservation in the E7N and E7C domains in the form of sequence logos. The two domains show similar degrees of sequence conservation across PVs, but as a consequence of evolutionary pressures of different nature. The E7C domain shows a conservation pattern that is readily explained by its main functions of zinc binding and dimerization (Fig. 12.7b) (Liu et al., 2006; Ohlenschlager et al., 2006). Most highly conserved positions constitute the core of the dimeric globular structure. Six conserved amino acids (displayed in blue) form the core

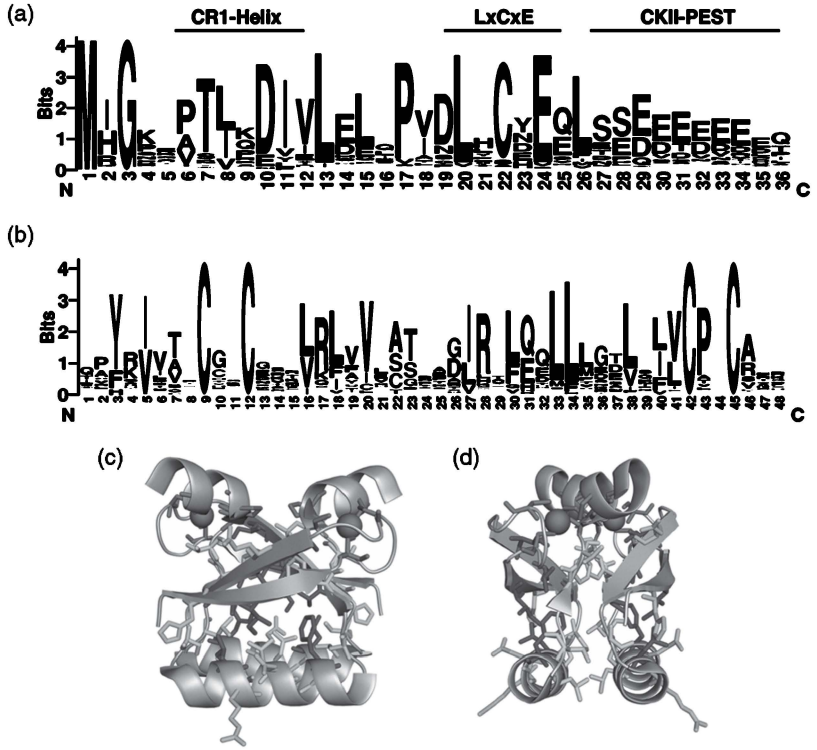


Figure 12.7 Sequence conservation of the Papillomavirus E7N and E7C domains. (a) Sequence logo of the E7N domain obtained from the alignment of 184 PV E7 sequences. CR1-helix, CR1 helix motif; LxCxE, LxCxE motif; CKII-PEST, CKII-PEST motif. (b) Sequence logo of the E7C domain obtained from the alignment of 196 PV E7 sequences. (c) Front view of the E7C domain; (d) side view of the E7C domain (PDB ID: 2F8B). Conserved residues are colored and shown in stick representation. Blue, conserved residues from each monomer's hydrophobic core; cyan, conserved residues from the dimer interface; red, conserved cysteine residues involved in Zn^{2+} coordination, Zn^{2+} atoms, and neighboring proline; green, conserved residues involved in protein–protein interactions. In both sequence logos, positions in the alignment with more than 25% gaps were removed. (*See insert for color representation of the figure.*)

of each monomer, while six different conserved amino acids (displayed in cyan) are mainly responsible for stabilizing the dimerization interface. The four cysteine residues responsible for zinc binding (displayed in red) are nearly invariable, and a neighboring proline (displayed in red) important for the structural integrity of this functional site is also highly conserved. Four other sequence positions on the surface of the domain (displayed in green) also show high conservation. Of these, positions 3, 28, and 31 in the sequence logo (Fig. 12.7b) may be involved in binding of the host target proteins Rb and p21CIP1 (Liu et al., 2006; Ohlenschlager et al., 2006), while the role of position 17 is currently unknown.

12.7.2.2 Sequence Conservation in E7N E7N is an IDD without a stable globular structure. It harbors several tightly packed linear motifs, such as the LxCxE motif, the CKII phosphorylation site, the acidic stretch, a helix-forming sequence located in CR1, and a ubiquitynation site at the N-terminus, resulting in a high functional density (Figs. 12.2a and 12.7a). A reason behind this compact arrangement may be the need to keep the viral genome small. The sequence and spatial proximity of motifs could also allow them to operate in a coordinated manner as in AND or OR molecular switches, thus increasing the E7 functional repertoire. E7N shows a high degree of conservation along its whole length, without any long stretches of low conservation (Fig. 12.7a). This is unusual for an IDD and can be explained only in part by the known motifs. The conserved residues lying between the known motifs may be motifs of yet unassigned function. Another possibility is that they provide a physical connection between the known motifs, thus allowing the propagation of conformational changes to distant functional sites. This is compatible with the observed coupling between the helix–coil transition in CR1 and the PII–coil transition in CR2 (Garcia-Alai et al., 2007) (Section 12.5). A third hypothesis, not mutually exclusive with the other two, is that conservation of these additional residues is necessary to fine-tune the details of the extended E7N structure, which is highly dynamic but by no means random (Garcia-Alai et al., 2007). In any case, the continuity of conservation throughout the whole E7N further confirms that it is a bona fide domain and not a mere joining of two independent regions, CR1 and CR2 (Garcia-Alai et al., 2007).

12.7.3 Role of Intrinsic Disorder in the Evolution of the E7 Protein

Intrinsically disordered regions have distinct evolutionary properties. Their sequence is not under the pressure to retain a specific tertiary structure and may change at a faster rate than the sequence of a globular domain. The IDD of E7 harbors several short functional motifs related to the transforming properties and oncogenic potential of PVs, such as the E2F-like helix, the LxCxE motif, the CKII phosphorylation site, and the acidic stretch. Most of these motifs consist of less than five function-determining residues and can thus appear or disappear altogether with only a few point mutations, a process facilitated by a flexible structural context. Accordingly, we observe a diversity of motif arrangements within canonical E7N domains in the PV family (Bernard et al., 2010). The LxCxE motif is missing in the whole *Nupapillomavirus* and *Zetapapillomavirus* genera (Hirt et al., 1991; Ghim et al., 2004). The CKII site is absent from the whole *Mupapillomavirus* and *Iotapapillomavirus* genera and in several PV types, from the *Betapapillomavirus, Kappapapillomavirus, Lambdapapillomavirus*, and *Xipapillomavirus* genera (Danos et al., 1982; Tan et al., 1994a; Christensen et al., 2000; Rector et al., 2007; Hatama et al., 2008; Chouhy et al., 2010). The acidic stretch is present in all known canonical E7N domains, although the amount of acidic residues is variable. Both the CKII site and the LxCxE motifs are lacking in several PV types from the Alpha and Gamma genera (Ekstrom et al., 2010; Kohler et al., 2010).

Remarkably, the changes in the E7N functional motifs do not follow the phylogenetic relationships between PV types (Bernard et al., 2010), indicating that these changes have occurred independently, several times in the recent past. This is direct evidence that structural disorder has facilitated adaptive evolution of short functional motifs, thereby modulating the diverse functional repertoire of the PV E7 protein (Chemes et al., 2010). Remarkably, structural disorder also allows for relocation of a short functional motif within the PV proteome. The Rhesus PV E6 protein does not contain a functional PDZ-binding motif, but the E7 protein from the same virus does contain one (Tomaic et al., 2009). In a similar manner, the phosphorylation/PEST sequence is not present in the E7 protein from BPV type 1 (Narechania et al., 2004), but it seems to have migrated to the E2 protein (Garcia-Alai et al., 2006). Last, it is worth mentioning that a recent study found an increased amount of disorder in E7 proteins from oncogenic genital HPVs, compared with their nononcogenic counterparts (Uversky et al., 2006). This indicates that the increased functional adaptability imparted by structural disorder may also have physiological consequences that go beyond the life cycle of the virus.

12.7.4 Sequence Similarity to Other Viral Proteins

The PV E7 transforming protein shares features with several transforming proteins from other human viruses. It has long been recognized that the canonical PV E7N domain and the adenovirus E1A protein share two disordered regions of significant sequence similarity, CR1 and CR2 (Dyson et al., 1992). The CR2 region is also present in the large T antigen from simian as well as human polyomaviruses (Dyson et al., 1990; Chellappan et al., 1992; Feng et al., 2008) (Section 12.2.1) (Fig. 12.2b). In later years, functional examples of the LxCxE Rb-binding motif have been characterized in proteins from other human viruses, such as rubella virus (Forng and Atreya, 1999), cytomegalovirus (Kalejta et al., 2003), human T-cell leukemia virus (Kehn et al., 2005), hepatitis C virus (Munakata et al., 2005), and molluscum contagiosum virus (Mohr et al., 2008). CKII-PEST motifs in noncanonical E7N domains from reptilian and avian PVs have also been reported (Terai et al., 2002; Herbst et al., 2009). The similarities among viruses also extend to the globular treble-cleft fold domain present in E6 and E7. A treble-cleft fold domain is present in the *Arenavirus* protein Z (Kentsis et al., 2002b). It is interesting to note that the treble-cleft fold domains from PV E6, E7 and *Arenavirus* protein Z share the ability to form large oligomeric assemblies, an ability which is also present in certain host domains derived from the treble-cleft fold (Section 12.4).

The origin of these similarities between viral transforming proteins is still not understood. Some similarities may be because of convergent evolution. This is a plausible explanation in the case of short functional motifs within disordered domains. For example, the CR2 region containing the LxCxE and CKII-PEST motifs may be a mimic of similar regions found in about a quarter cellular targets of the Rb protein (Chemes, 2010). Alternatively, some regions of sequence similarity may have a common ancestor, that is, be homologous. Such regions may have been transferred through recombination between PV genomes of different types

(Narechania et al., 2005; Hu et al., 2009) and between PV and polyomavirus genomes (Woolford et al., 2007; Bennett et al., 2008).

12.8 CONCLUDING REMARKS

The HPV E7 oncoprotein is a prototypic viral oncoprotein that shares sequence and functional similarities with oncoproteins from other DNATVs. It was through the investigation of E7 and other DNATV proteins that the function of the Rb protein as an essential cell cycle regulator was first uncovered. The E7–Rb interaction stimulates cell cycle progression, but it cannot explain by itself the oncogenic nature of the E7 protein, given that it takes place in all HPV-infected cells. Cancer development is a rare and exceptional event, which probably requires the participation many other target–protein interactions.

We argue that the multitarget nature of E7 has an impact on the cell transformation process because of its intrinsically disordered nature, which endows E7 with the ability to adopt a number of different conformations by combining multiple equilibria that depend on the chemical environment. This property resides in the E7 N-terminal module, which we can identify as a bona fide domain, irrespective of the fact that it does not bear a compact tertiary fold.

The modular nature of E7 provides with two domains that can contact protein partners, including the dimer interface, which is exposed on dissociation. A vast number of important interacting sites are located at the E7N domain, while others reside in the E7C domain, as identified by in-cell or pull-down methods. However, the interaction with cellular binding partners appears to depend on a combination of contact sites that contribute differentially to the binding free energy.

E7N is a plastic domain, which can accommodate a large number of interaction partners including an enzyme such as CKII. Features such as the presence of several linear motifs within an intrinsically disordered region, the presence of additional interaction sites in the compact C-terminal domain, and the highly acidic and therefore ionic nature of E7 form the structural and molecular basis for the large number of interacting partners described. Thus, E7 qualifies as a scaffold or hub protein, similar to its main target Rb.

So far, the only high-affinity and high-specificity E7 target analyzed in detail is the Rb protein. Because of its high affinity, the E7–Rb interaction is expected to take place at low E7 levels typical of early stages of infection. As E7 levels build up on viral life cycle progression, oligomerization in the cytosol and the establishment of low-affinity interactions through the E7N domain may become significant. This would generate a network of interactions that far from being productive to cell physiology could interfere and eventually lead to transformation through many of the different routes that have been described.

The marked statistical presence of IDP in cell signaling and cancer-related protein interaction networks (Uversky et al., 2009), which are highly related to each other, argues for the requirement of multiple low-specificity interactions mediated by these domains. The active participation of E7 in HPV biology, its small size,

the large number of viral sequences that allow bioinformatics approaches, and finally, the availability of a large number of genetic and cell biology tools make E7 an excellent model for investigating the molecular basis for the role of IDPs in biological processes.

ACKNOWLEDGMENTS

We thank Dr Gilles Travé for careful reading and helpful comments on the manuscript and Ms Juliana Glavina for careful proofreading of the manuscript. L. B. C. is supported by a CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) postdoctoral fellowship, and L. G. A., I. E. S., and G. de P.-G. are career investigators from CONICET.

ABBREVIATIONS

PV	papillomavirus
HPV	human Papillomavirus
BPV	bovine Papillomavirus
DNATV	DNA tumor viruses
ORF	open reading frame
E7N	E7 N-terminal domain
E7C	E7 C-terminal domain
CR1	E7 conserved region 1
CR2	E7 conserved region 2
CR3	E7 conserved region 3
Rb	retinoblastoma protein
PEST	P, E, S, T-rich degradation motif
CKII	casein kinase II
IDP	intrinsically disordered protein
IDD	intrinsically disordered domain
E7SOs	E7 soluble oligomers
CD	circular dichroism
PII	polyproline type-II structure
Gdm.Cl	guanidinium chloride
TFE	1,1,1-trifluoro ethanol
SDS	sodium dodecyl sulfate
ANS	1-anilino-8-naphtalene sulfonate

REFERENCES

Alonso LG, Garcia-Alai MM, Nadra AD, Lapena AN, Almeida FL, Gualfetti P, Prat-Gay GD. High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. Biochemistry 2002;41(33):10510–10518.

Alonso LG, Garcia-Alai MM, Smal C, Centeno JM, Iacono R, Castano E, Gualfetti P, de Prat-Gay G. The HPV16 E7 viral oncoprotein self-assembles into defined spherical oligomers. Biochemistry 2004;43(12):3310–3317.

Alonso LG, Smal C, Garcia-Alai MM, Chemes L, Salame M, de Prat-Gay G. Chaperone holdase activity of human papillomavirus E7 oncoprotein. Biochemistry 2006;45(3):657–667.

Apt D, Chong T, Liu Y, Bernard HU. Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. J Virol 1993;67(8):4455–4463.

Arbeit JM, Howley PM, Hanahan D. Chronic estrogen-induced cervical and vaginal squamous carcinogenesis in human papillomavirus type 16 transgenic mice. Proc Natl Acad Sci USA 1996;93(7):2930–2935.

Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley PM. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J Virol 1987;61(4):962–971.

Baker TS, Newcomb WW, Olson NH, Cowsert LM, Olson C, Brown JC. Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophys J 1991;60(6):1445–1456.

Banerjee NS, Genovese NJ, Noya F, Chien WM, Broker TR, Chow LT. Conditionally activated E7 proteins of high-risk and low-risk human papillomaviruses induce S phase in postmitotic, differentiated human keratinocytes. J Virol 2006;80(13):6517–6524.

Barbosa MS, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. EMBO J 1990;9(1):153–160.

Bedell MA, Jones KH, Laimins LA. The E6-E7 region of human papillomavirus type 18 is sufficient for transformation of NIH 3T3 and rat–1 cells. J Virol 1987;61(11):3635–3640.

Bennett MD, Woolford L, Stevens H, Van Ranst M, Oldfield T, Slaven M, O'Hara AJ, Warren KS, Nicholls PK. Genomic characterization of a novel virus found in papillomatous lesions from a southern brown bandicoot (Isoodon obesulus) in Western Australia. Virology 2008;376(1):173–182.

Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 2010;401(1):70–79.

Bosch FX, Sanjose S, Castellagué X, Moreno V, Muñoz N. Epidemiology of human papillomavirus infections and associations with cervical cancer: new opportunities for prevention. In: Saveria Campo M, editor. Papillomavirus research: from natural history to vaccines and beyond. Norfolk, UK: Caister Academic Press; 2006. pp.19–39.

Bouallaga I, Massicard S, Yaniv M, Thierry F. An enhanceosome containing the Jun B/Fra–2 heterodimer and the HMG-I(Y) architectural protein controls HPV 18 transcription. EMBO Rep 2000;1(5):422–427.

Boyer SN, Wazer DE, Band V. E7 protein of human papilloma virus–16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Res 1996;56(20):4620–4624.

Brokaw JL, Yee CL, Munger K. A mutational analysis of the amino terminal domain of the human papillomavirus type 16 E7 oncoprotein. Virology 1994;205(2):603–607.

Chellappan S, Kraus VB, Kroger B, Munger K, Howley PM, Phelps WC, Nevins JR. Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. Proc Natl Acad Sci USA 1992;89(10):4549–4553.

Chemes, LB. PhD Thesis. La proteina supresora de tumores Retinoblastoma: caracterización de su dominio AB y mecanismo de interacción con la oncoproteína E7 del papilomavirus humano. Univerisity of Buenos Aires, Argentina; 2010.

Chemes LB, Sanchez IE, Smal C, de Prat-Gay G. Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. Structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7. FEBS J 2010;277(4):973–988.

Chen Z, DeSalle R, Schiffman M, Herrero R, Burk RD. Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. J Virol 2009;83(3):1443–1455.

Chen Z, Schiffman M, Herrero R, Desalle R, Burk RD. Human papillomavirus (HPV) types 101 and 103 isolated from cervicovaginal cells lack an E6 open reading frame (ORF) and are related to gamma-papillomaviruses. Virology 2007;360(2):447–453.

Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J Virol 2005;79(11):7014–7023.

Choo KB, Pan CC, Han SH. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. Virology 1987a;161(1):259–261.

Choo KB, Pan CC, Liu MS, Ng HT, Chen CP, Lee YN, Chao CF, Meng CL, Yeh MY, Han SH. Presence of episomal and integrated human papillomavirus DNA sequences in cervical carcinoma. J Med Virol 1987b;21(2):101–107.

Chouhy D, Gorosito M, Sanchez A, Serra EC, Bergero A, Fernandez Bussy R, Giri AA. New generic primer system targeting mucosal/genital and cutaneous human papillomaviruses leads to the characterization of HPV 115, a novel Beta-papillomavirus species 3. Virology 2010;397(1):205–216.

Chow LT, Broker, TR. Mechanisms and regulation of papillomavirus DNA replication. In: Campo MS, editor. Papillomavirus research. Norfolk, UK: Caister Academic Press; 2006 pp.53–62.

Christensen ND, Cladel NM, Reed CA, Han R. Rabbit oral papillomavirus complete genome sequence and immunity following genital infection. Virology 2000;269(2):451–461.

Clements A, Johnston K, Mazzarelli JM, Ricciardi RP, Marmorstein R. Oligomerization properties of the viral oncoproteins adenovirus E1A and human papillomavirus E7 and their complexes with the retinoblastoma protein. Biochemistry 2000;39(51):16033–16045.

Cole ST, Danos O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. J Mol Biol 1987;193(4):599–608.

Collins SI, Constandinou-Williams C, Wen K, Young LS, Roberts S, Murray PG, Woodman CB. Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. Cancer Res 2009;69(9):3828–3832.

Cricca M, Venturoli S, Leo E, Costa S, Musiani M, Zerbini M. Disruption of HPV 16 E1 and E2 genes in precancerous cervical lesions. J Virol Methods 2009;158(1–2):180–183.

Danos O, Katinka M, Yaniv M. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. EMBO J 1982;1(2):231–236.

Dantur K, Alonso L, Castano E, Morelli L, Centeno-Crowley JM, Vighi S, de Prat-Gay G. Cytosolic accumulation of HPV16 E7 oligomers supports different transformation routes for the prototypic viral oncoprotein: the amyloid-cancer connection. Int J Cancer 2009;125(8):1902–1911.

Daughdrill, GW, Pielak, GJ, Uversky, VN, Cortese, MS, Dunker AK. Natively disordered proteins. Volume 2, In: J Buchner, T Kiefhaber, editor. Protein folding handbook. Weinheim, Germany: Wiley-VCH; 2005. pp.275–357.

de Souza RF, Iyer LM, Aravind L. Diversity and evolution of chromatin proteins encoded by DNA viruses. Biochim Biophys Acta 2009;1799(3–4):302–318.

de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology 2004;324(1):17–27.

DeCaprio JA, Ludlow JW, Figge J, Shew JY, Huang CM, Lee WH, Marsilio E, Paucha E, Livingston DM. SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. Cell 1988;54(2):275–283.

Dick FA, Dyson NJ. Three regions of the pRB pocket domain affect its inactivation by human papillomavirus E7 proteins. J Virol 2002;76(12):6224–6234.

Dong WL, Caldeira S, Sehr P, Pawlita M, Tommasino M. Determination of the binding affinity of different human papillomavirus E7 proteins for the tumour suppressor pRb by a plate-binding assay. J Virol Methods 2001;98(1):91–98.

Dyson N, Bernards R, Friend SH, Gooding LR, Hassell JA, Major EO, Pipas JM, Vandyke T, Harlow E. Large T antigens of many polyomaviruses are able to form complexes with the retinoblastoma protein. J Virol 1990;64(3):1353–1356.

Dyson N, Guida P, Munger K, Harlow E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. J Virol 1992;66(12):6893–6902.

Dyson N, Howley PM, Munger K, Harlow E. The human papilloma virus–16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 1989;243(4893):934–937.

Eddy BE, Borman GS, Grubbs GE, Young RD. Identification of the oncogenic substance in rhesus monkey kidney cell culture as simian virus 40. Virology 1962;1765–75.

Ekstrom J, Forslund O, Dillner J. Three novel papillomaviruses (HPV109, HPV112 and HPV114) and their presence in cutaneous and mucosal samples. Virology 2010;397(2):331–336.

Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008;319(5866):1096–1100.

Ferrari R, Pellegrini M, Horwitz GA, Xie W, Berk AJ, Kurdistani SK. Epigenetic reprogramming by adenovirus e1a. Science 2008;321(5892):1086–1088.

Ferreon AC, Gambin Y, Lemke EA, Deniz AA. Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. Proc Natl Acad Sci USA 2009a;106(14):5645–5650.

Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. Proc Natl Acad Sci USA 2009b;106(32):13260–13265.

Firzlaff JM, Luscher B, Eisenman RN. Negative charge at the casein kinase II phosphorylation site is important for transformation but not for Rb protein binding by the E7 protein of human papillomavirus type 16. Proc Natl Acad Sci USA 1991;88(12):5187–5191.

Forng RY, Atreya CD. Mutations in the retinoblastoma protein-binding LXCXE motif of rubella virus putative replicase affect virus replication. J Gen Virol 1999;80(Pt 2):327–332.

Gammoh N, Grm HS, Massimi P, Banks L. Regulation of human papillomavirus type 16 E7 activity through direct protein interaction with the E2 transcriptional activator. J Virol 2006;80(4):1787–1797.

Gammoh N, Isaacson E, Tomaic V, Jackson DJ, Doorbar J, Banks L. Inhibition of HPV–16 E7 oncogenic activity by HPV–16 E2. Oncogene 2009;28(23):2299–2304.

Garcia-Alai MM, Alonso LG, de Prat-Gay G. The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry 2007;46(37):10405–10412.

Garcia-Alai MM, Gallo M, Salame M, Wetzler DE, McBride AA, Paci M, Cicero DO, de Prat-Gay G. Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. Structure 2006;14(2):309–319.

Garcia-Vallve S, Alonso A, Bravo IG. Papillomaviruses: different genes have different histories. Trends Microbiol 2005;13(11):514–521.

Genovese NJ, Banerjee NS, Broker TR, Chow LT. Casein kinase II motif-dependent phosphorylation of human papillomavirus E7 protein promotes p130 degradation and S-phase induction in differentiated human keratinocytes. J Virol 2008;82(10):4862–4873.

Ghim SJ, Rector A, Delius H, Sundberg JP, Jenson AB, Van Ranst M. Equine papillomavirus type 1: complete nucleotide sequence and characterization of recombinant virus-like particles composed of the EcPV–1 L1 major capsid protein. Biochem Biophys Res Commun 2004;324(3):1108–1115.

Goodwin EC, DiMaio D. Repression of human papillomavirus oncogenes in HeLa cervical carcinoma cells causes the orderly reactivation of dormant tumor suppressor pathways. Proc Natl Acad Sci USA 2000;97(23):12513–12518.

Gould CM, Diella F, Via A, Puntervoll P, Gemund C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C, Seiler M, Davey NE, Haslam N, Weatheritt RJ, Budd A, Hughes T, Pas J, Rychlewski L, Trave G, Aasland R, Helmer-Citterich M, Linding R, Gibson TJ. ELM: the status of the 2010 eukaryotic linear motif resource. Nucleic Acids Res 2010;38(Database issue):D167–D180.

Hamid NA, Brown C, Gaston K. The regulation of cell proliferation by the papillomavirus early proteins. Cell Mol Life Sci 2009;66(10):1700–1717.

Hatama S, Nobumoto K, Kanno T. Genomic and phylogenetic analysis of two novel bovine papillomaviruses, BPV–9 and BPV–10. J Gen Virol 2008;89(Pt 1):158–163.

Hegde RS. The papillomavirus E2 proteins: structure, function, and biology. Annu Rev Biophys Biomol Struct 2002;31343–360.

Herbst LH, Lenz J, Van Doorslaer K, Chen Z, Stacy BA, Wellehan JF, Manire CA, Burk RD Jr. Genomic characterization of two novel reptilian papillomaviruses, Chelonia mydas papillomavirus 1 and Caretta caretta papillomavirus 1. Virology 2009;383(1):131–135.

Hirt L, Hirsch-Behnam A, de Villiers EM. Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type without a typical E2 binding site consensus sequence. Virus Res 1991;18(2–3):179–189.

Howley PM, Livingston DM. Small DNA tumor viruses: large contributors to biomedical sciences. Virology 2009;384(2):256–259.

Howley PM, Munger K, Romanczuk H, Scheffner M, Huibregtse JM. Cellular targets of the oncoproteins encoded by the cancer associated human papillomaviruses. Princess Takamatsu Symp 1991;22:239–248.

Howley, PM, Lowy DR. Papillomaviruses. Volume 2. In: DM Knipe, Howley PM, editors. Fields virology. Philadelphia, PA: Lippincott, Williams and Wilkins; 2007 pp.2299–2354.

Hu J, Cladel NM, Budgeon L, Balogh KK, Christensen ND. Papillomavirus DNA complementation in vivo. Virus Res 2009;144(1–2):117–122.

Imai Y, Matsushima Y, Sugimura T, Terada M. Purification and characterization of human papillomavirus type 16 E7 protein with preferential binding capacity to the underphosphorylated form of retinoblastoma gene product. J Virol 1991;65(9):4966–4972.

Jasanoff A, Fersht AR. Quantitative determination of helical propensities from trifluoroethanol titration curves. Biochemistry 1994;33(8):2129–2135.

Javier RT, Butel JS. The history of tumor virology. Cancer Res 2008;68(19):7693–7706.

Jeon S, Lambert PF. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci USA 1995;92(5):1654–1658.

Johnston JA, Johnson ES, Waller PR, Varshavsky A. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. J Biol Chem 1995;270(14):8172–8178.

Jones RE, Wegrzyn RJ, Patrick DR, Balishin NL, Vuocolo GA, Riemen MW, Defeo-Jones D, Garsky VM, Heimbrook DC, Oliff A. Identification of HPV–16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. J Biol Chem 1990;265(22):12782–12785.

Kalantari M, Karlsen F, Kristensen G, Holm R, Hagmar B, Johansson B. Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. Int J Gynecol Pathol 1998;17(2):146–153.

Kalejta RF, Bechtel JT, Shenk T. Human cytomegalovirus pp71 stimulates cell cycle progression by inducing the proteasome-dependent degradation of the retinoblastoma family of tumor suppressors. Mol Cell Biol 2003;23(6):1885–1895.

Kehn K, Fuente Cde L, Strouss K, Berro R, Jiang H, Brady J, Mahieux R, Pumfery A, Bottazzi ME, Kashanchi F. The HTLV-I Tax oncoprotein targets the retinoblastoma protein for proteasomal degradation. Oncogene 2005;24(4):525–540.

Kentsis A, Gordon RE, Borden KL. Control of biochemical reactions through supramolecular RING domain self-assembly. Proc Natl Acad Sci USA 2002a;99(24):15404–15409.

Kentsis A, Gordon RE, Borden KL. Self-assembly properties of a model RING domain. Proc Natl Acad Sci USA 2002b;99(2):667–672.

Kim HY, Ahn BY, Cho Y. Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. EMBO J 2001;20(1–2):295–304.

Klunk WE, Jacob RF, Mason RP. Quantifying amyloid by congo red spectral shift assay. Methods Enzymol 1999;309285–305.

Kohler A, Gottschling M, Forster J, Rowert-Huber J, Stockfleth E, Nindl I. Genomic characterization of a novel human papillomavirus (HPV–117) with a high viral load in a persisting wart. Virology 2010;399(1):129–133.

Lee C, Schwartz MP, Prakash S, Iwakura M, Matouschek A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. Mol Cell 2001;7(3):627–637.

Lee JO, Russo AA, Pavletich NP. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. Nature 1998;391(6670):859–865.

LeVine H 3rd Quantification of beta-sheet amyloid fibril structures with thioflavin T. Methods Enzymol 1999;309:274–284.

Lin R, Beauparlant P, Makris C, Meloche S, Hiscott J. Phosphorylation of IkappaBalpha in the C-terminal PEST domain by casein kinase II affects intrinsic protein stability. Mol Cell Biol 1996;16(4):1401–1409.

Liu X, Clements A, Zhao K, Marmorstein R. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. J Biol Chem 2006;281(1):578–586.

Liu X, Marmorstein R. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. Genes Dev 2007;21(21):2711–2716.

McIntyre MC, Frattini MG, Grossman SR, Laimins LA. Human papillomavirus type 18 E7 protein requires intact Cys-X-X-Cys motifs for zinc binding, dimerization, and transformation but not for Rb binding. J Virol 1993;67(6):3142–3150.

Mohr S, Grandemange S, Massimi P, Darai G, Banks L, Martinou JC, Zeier M, Muranyi W. Targeting the retinoblastoma protein by MC007L, gene product of the molluscum contagiosum virus: detection of a novel virus-cell interaction by a member of the poxviruses. J Virol 2008;82(21):10625–10633.

Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer 2010;10(8):550–560.

Munakata T, Nakamura M, Liang Y, Li K, Lemon SM. Down-regulation of the retinoblastoma tumor suppressor by the hepatitis C virus NS5B RNA-dependent RNA polymerase. Proc Natl Acad Sci USA 2005;102(50):18159–18164.

Munger K, Phelps WC, Bubb V, Howley PM, Schlegel R. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. J Virol 1989;63(10):4417–4421.

Munger K, Yee CL, Phelps WC, Pietenpol JA, Moses HL, Howley PM. Biochemical and biological differences between E7 oncoproteins of the high- and low-risk human papillomavirus types are determined by amino-terminal sequences. J Virol 1991;65(7):3943–3948.

Narechania A, Chen Z, DeSalle R, Burk RD. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. J Virol 2005;79(24):15503–15510.

Narechania A, Terai M, Chen Z, DeSalle R, Burk RD. Lack of the canonical pRB-binding domain in the E7 ORF of artiodactyl papillomaviruses is associated with the development of fibropapillomas. J Gen Virol 2004;85(Pt 5):1243–1250.

Ohlenschlager O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, Baum M, Korbas M, Meyer-Klaucke W, Durst M, Gorlach M. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene 2006;25(44):5953–5959.

Ordonez RM, Espinosa AM, Sanchez-Gonzalez DJ, Armendariz-Borunda J, Berumen J. Enhanced oncogenicity of Asian-American human papillomavirus 16 is associated with impaired E2 repression of E6/E7 oncogene transcription. J Gen Virol 2004;85(Pt 6):1433–1444.

Pahel G, Aulabaugh A, Short SA, Barnes JA, Painter GR, Ray P, Phelps WC. Structural and functional characterization of the HPV16 E7 protein expressed in bacteria. J Biol Chem 1993;268(34):26018–26025.

Patrick DR, Oliff A, Heimbrook DC. Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein. J Biol Chem 1994;269(9):6842–6850.

Penrose KJ, Garcia-Alai M, de Prat-Gay G, McBride AA. Casein Kinase II phosphorylation-induced conformational switch triggers degradation of the papillomavirus E2 protein. J Biol Chem 2004;279(21):22430–22439.

Phelps WC, Munger K, Yee CL, Barnes JA, Howley PM. Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. J Virol 1992;66(4):2418–2427.

Phelps WC, Yee CL, Munger K, Howley PM. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. Cell 1988;53(4):539–547.

Pim D, Banks L. Interaction of viral oncoproteins with cellular target molecules: infection with high-risk vs low-risk human papillomaviruses. APMIS 2010;118(6–7):471–493.

Prakash S, Tian L, Ratliff KS, Lehotzky RE, Matouschek A. An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat Struct Mol Biol 2004;11(9):830–837.

Rechsteiner M, Rogers SW. PEST sequences and regulation by proteolysis. Trends Biochem Sci 1996;21(7):267–271.

Rector A, Lemey P, Tachezy R, Mostmans S, Ghim SJ, Van Doorslaer K, Roelke M, Bush M, Montali RJ, Joslin J, Burk RD, Jenson AB, Sundberg JP, Shapiro B, Van Ranst M. Ancient papillomavirus-host co-speciation in Felidae. Genome Biol 2007;8(4):R57.

Reinstein E, Scheffner M, Oren M, Ciechanover A, Schwartz A. Degradation of the E7 human papillomavirus oncoprotein by the ubiquitin-proteasome system: targeting via ubiquitination of the N-terminal residue. Oncogene 2000;19(51):5944–5950.

Rihs HP, Jans DA, Fan H, Peters R. The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40T-antigen. EMBO J 1991;10(3):633–639.

Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci USA 1992;89(7):3159–3163.

Sarnow P, Ho YS, Williams J, Levine AJ. Adenovirus E1b–58kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54kd cellular protein in transformed cells. Cell 1982;28(2):387–394.

Selvey LA, Dunn LA, Tindle RW, Park DS, Frazer IH. Human papillomavirus (HPV) type 18 E7 protein is a short-lived steroid-inducible phosphoprotein in HPV-transformed cell lines. J Gen Virol 1994;75(Pt 7):1647–1653.

Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99(14): 9190–9195.

Shirasawa H, Tomita Y, Sekiya S, Takamizawa H, Simizu B. Integration and transcription of human papillomavirus type 16 and 18 sequences in cell lines derived from cervical carcinomas. J Gen Virol 1987;68(Pt 2):583–591.

Shope RE, Hurst EW. Infectious Papillomatosis of Rabbits: With a Note on the Histopathology. J Exp Med 1933;58(5):607–624.

Singh M, Krajewski M, Mikolajka A, Holak TA. Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. J Biol Chem 2005;280(45):37868–37876.

Smal, C. PhD Thesis. Oligomerización de la oncoproteína E7 del papilomavirus humano y su interacción con el regulador de la transcripción y replicación viral E2. Univarsidad de Buenos Aires, Argentina ; 2010.

Smal C, Wetzler DE, Dantur KI, Chemes LB, Garcia-Alai MM, Dellarole M, Alonso LG, Gaston K, de Prat-Gay G. The human papillomavirus E7-E2 interaction mechanism in vitro reveals a finely tuned system for modulating available E7 and E2 proteins. Biochemistry 2009;48(50):11939–11949.

Smotkin D, Wettstein FO. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. Proc Natl Acad Sci USA 1986;83(13):4680–4684.

Smotkin D, Wettstein FO. The major human papillomavirus protein in cervical cancers is a cytoplasmic phosphoprotein. J Virol 1987;61(5):1686–1689.

Stoler MH, Rhodes CR, Whitbeck A, Wolinsky SM, Chow LT, Broker TR. Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. Hum Pathol 1992;23(2):117–128.

Sullivan CS, Cantalupo P, Pipas JM. The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism. Mol Cell Biol 2000;20(17):6233–6243.

Tan CH, Tachezy R, Van Ranst M, Chan SY, Bernard HU, Burk RD. The Mastomys natalensis papillomavirus: nucleotide sequence, genome organization, and phylogenetic relationship of a rodent papillomavirus involved in tumorigenesis of cutaneous epithelia. Virology 1994a;198(2):534–541.

Tan SH, Leong LE, Walker PA, Bernard HU. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. J Virol 1994b;68(10):6411–6420.

Terai M, DeSalle R, Burk RD. Lack of canonical E6 and E7 open reading frames in bird papillomaviruses: Fringilla coelebs papillomavirus and Psittacus erithacus timneh papillomavirus. J Virol 2002;76(19):10020–10023.

Thierry F. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. Virology 2009;384(2):375–379.

Tiffany ML, Krimm S. New chain conformations of poly(glutamic acid) and polylysine. Biopolymers 1968;6(9):1379–1382.

Tomaic V, Gardiol D, Massimi P, Ozbun M, Myers M, Banks L. Human and primate tumour viruses use PDZ binding as an evolutionarily conserved mechanism of targeting cell polarity regulators. Oncogene 2009;28(1):1–8.

Tomita Y, Ogawa T, Jin Z, Shirasawa H. Genus specific features of bovine papillomavirus E6, E7, E5 and E8 proteins. Virus Res 2007;124(1–2):231–236.

Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. FASEB J 2004;18(11):1169–1175.

Trentin JJ, Yabe Y, Taylor G. The quest for human cancer viruses. Science 1962;137: 835–841.

Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 2009;10(Suppl 1):S7.

Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res 2006;5(8):1829–1842.

Van Bressem MF, Cassonnet P, Rector A, Desaintes C, Van Waerebeek K, Alfaro-Shigueto J, Van Ranst M, Orth G. Genital warts in Burmeister's porpoises: characterization of Phocoena spinipinnis papillomavirus type 1 (PsPV–1) and evidence for a second, distantly related PsPV. J Gen Virol 2007;88(Pt 7):1928–1933.

Van Doorslaer K, Sidi AO, Zanier K, Rybin V, Deryckere F, Rector A, Burk RD, Lienau EK, van Ranst M, Trave G. Identification of unusual E6 and E7 proteins within avian papillomaviruses: cellular localization, biophysical characterization, and phylogenetic analysis. J Virol 2009;83(17):8759–8770.

Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science 1990;248(4951):76–79.

Whalen SG, Marcellus RC, Barbeau D, Branton PE. Importance of the Ser–132 phosphorylation site in cell transformation and apoptosis induced by the adenovirus type 5 E1A protein. J Virol 1996;70(8):5373–5383.

Whyte P, Buchkovich KJ, Horowitz JM, Friend SH, Raybuck M, Weinberg RA, Harlow E. Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. Nature 1988;334(6178):124–129.

Wolf M, Garcea RL, Grigorieff N, Harrison SC. Subunit interactions in bovine papillomavirus. Proc Natl Acad Sci USA 2010;107(14):6298–6303.

Woods WS, Boettcher JM, Zhou DH, Kloepper KD, Hartman KL, Ladror DT, Qi Z, Rienstra CM, George JM. Conformation-specific binding of alpha-synuclein to novel protein partners detected by phage display and NMR spectroscopy. J Biol Chem 2007;282(47):34555–34567.

Woolford L, Rector A, Van Ranst M, Ducki A, Bennett MD, Nicholls PK, Warren KS, Swan RA, Wilcox GE, O'Hara AJ. A novel virus detected in papillomas and carcinomas of the endangered western barred bandicoot (Perameles bougainville) exhibits genomic features of both the Papillomaviridae and Polyomaviridae. J Virol 2007; 81(24):13280–13290.

Wu CS, Ikeda K, Yang JT. Ordered conformation of polypeptides and proteins in acidic dodecyl sulfate solution. Biochemistry 1981;20(3):566–570.

Zheng ZM, Baker CC. Papillomavirus genome structure, expression, and post-transcriptional regulation. Front Biosci 2006;112286–2302.

Zhong L, Johnson WC Jr. Environment affects amino acid preference for secondary structure. Proc Natl Acad Sci USA 1992;89(10):4462–4465.

zur Hausen H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. Curr Top Microbiol Immunol 1994;186:131–156.

zur Hausen H. Papillomaviruses in the causation of human cancers - a brief historical account. Virology 2009;384(2):260–265.

13

THE SEMLIKI FOREST VIRUS CAPSID PROTEASE IS DISORDERED AND YET DISPLAYS CATALYTIC ACTIVITY

MANUEL MORILLAS, HEIKE EBERL, FRÉDERIC H.-T. ALLAIN, RUDI GLOCKSHUBER, AND EVA KUENNEMANN

13.1 INTRODUCTION

Togaviruses are enveloped plus-strand RNA viruses with an icosahedral nucleocapsid and a spherical virion morphology. The Togaviridae family is composed of two genera: *Alphavirus* and *Rubivirus*. The *Alphavirus* genus consists of more than 40 recognized members, while the rubella virus is the only member of the *Rubivirus* genus (Matthews, 1979; Porterfield, 1986). Genome organization and nucleotide homology make the difference between the two genera. Alphaviruses are about 11.7 kb in length, while rubella virus is 2 kb shorter (Dominguez et al., 1990; Strauss et al., 1984). Alphaviruses are transmitted in vertebrates and insects, while rubella virus is restricted to humans. The Semliki Forest virus (SFV), for instance, infects birds and rodents via mosquito bites, causing rash and arthritis.

The genome of SFV is a nonsegmented single-stranded plus-strand RNA, 13 kb in length. It has an open reading frame (ORF) in the 5′ region that encodes nonstructural viral proteins and a second ORF in the 3′ region that encodes structural proteins (Fig. 13.1). The genome carries a methylated nucleotide cap at the 5′-terminus cap and a polydeanylated tail at the 3′-terminus, resembling an mRNA. Alphavirus-infected cells produce three species of RNAs: genomic plus-strand

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

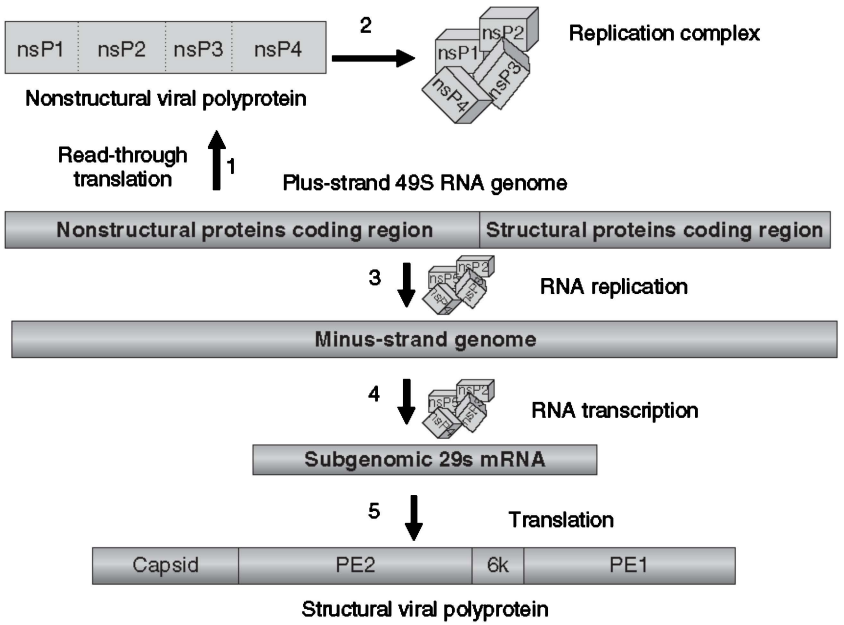


Figure 13.1 Replication and processing of the genome of an alphavirus. Alphaviruses have (+)ss RNA, and the replication and translation occurs in two cycles. First cycle (step 1 and 2) produces nonstructural polyproteins required for RNA replication and transcription. The second cycle (step 3 and 4) produces a (−)ss RNA, subgenomic mRNA, and the viral structural protein. *Source:* Adapted from Kuhn (2006).

RNA, complementary minus-strand RNA, and subgenomic mRNA (Fig. 13.2). Translation of the (+)ss RNA occurs in two cycles. The first cycle produces a polyprotein encoded in the 5′ region of the genome forming four nonstructural proteins responsible for gene replication and expression (Fig. 13.1, steps 1 and 2). The second cycle of translation produces a negative sense fragment that acts as the template for further synthesis of the positive-sense RNA and the viral polyprotein (Fig. 13.1, steps 3–5; Kuhn, 2006). The polyprotein translated in this second cycle carries four structural proteins: the capsid protein C, enveloped glycoproteins E1 and E2, and leader peptide proteins E3 and 6K (Raju and Huang, 1991).

The capsid protein constitutes the N-terminal segment of the polyprotein (residues 1–267) and is released from the viral polyprotein by autoproteolysis (Fig. 13.3) The new viral polyprotein devoid of the capsid protein now bears a signal sequence for translocation across the ER where it is glycosylated and translocated to the cell membrane. Proteins E1, E2, and 6K are transmembrane proteins (Sefton, 1977; de Curtis and Simons, 1988; Garoff et al., 1990; Presley et al., 1991).

Viral proliferation is followed by interaction of the capsid protein associated with the ribosome and viral RNA, thus triggering viral capsid formation. Interaction of the capsid and the glycoproteins E1 and E2 promotes viral budding (Miller and

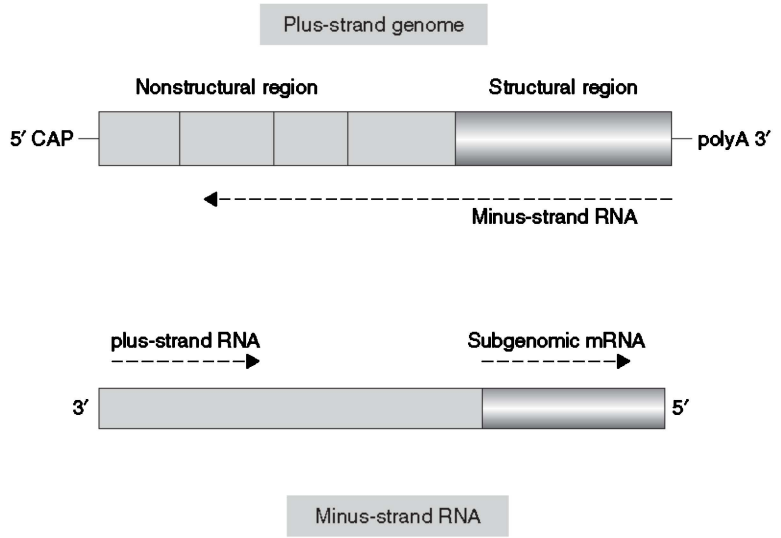


Figure 13.2 Structure of the alphavirus genome regulation of RNA synthesis. Alphaviruses produce three species of RNAs: genomic plus-strand RNA, complementary minus-strand RNA, and subgenomic mRNA. The synthesis of these three species is tightly regulated by specific nonstructural viral proteins. Synthesis is asymmetric with minus-strand synthesis to about 5% of the level of plus-strand genome RNA. RNA replication begins with the synthesis of minus-strand synthesis (top panel). As minus-strand synthesis continues, nonstructural proteins continue to be translated and concentrations of the protease precursors increase. A change in the composition of the replication complex shifts to plus-strand synthesis and production of subgenomic mRNA.

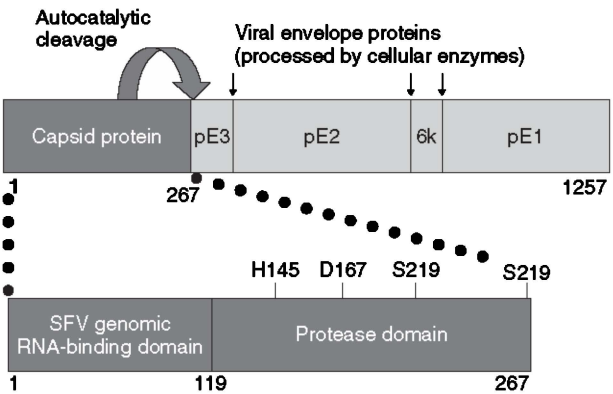


Figure 13.3 Processing and structure of the viral polyprotein. The viral polyprotein is processed by intramolecular and intermolecular proteolysis. The SFV capsid protein autocleaves the peptide bond between the residues Trp267 and Ser268 cotranslationally. The new polyprotein is processed by cellular proteases. The mature capsid protein is composed of an unfolded N-terminal segment and the protease C-terminal segment that remains inactive. The residues of the catalytic triad are indicated.

Brown, 1992; Roman and Garoff, 1986; von Bonsdorff and Harrison, 1978). The mature envelope virion is 70 nm in diameter with a molecular mass of 52,000 kDa and a density of 1.22 g/cc. The mature virion is composed of an outer host-derived lipid layer and repeating units of the E1 and E2 transmembrane glycoproteins, and an inner capsid formed by 240 copies of the capsid protein arranged in $T = 4$ icosahedral lattice encapsulating the single molecule of the genomic RNA (Harrison et al., 1971; von Bonsdorff and Harrison, 1975, 1978; Harrison et al., 1992; Fuller et al., 1995; Fig. 13.4).

The capsid protein consists of an unstructured region (residues 1–118) and a folded domain with serine protease activity (residues 119–267). The capsid protein displays several functions in the life cycle of SFV: (i) cotranslational folding of the C-terminal protease domain catalyzes the cleavage of the Trp267-Ser268 bond, thus generating the mature capsid protein (SFVP) (Melancon and Garoff, 1987; Skoging and Liljestrom, 1998; Nicola et al., 1999); (ii) viral capsid assembly is initiated by an interaction between the positively charged N-terminal segment of the SFVP, the ribosome, and the negatively charged viral RNA (Weiss et al., 1989; Geigenmuller-Gnirke et al., 1993; Forsell et al., 1995; Owen and Kuhn, 1996); (iii) interaction between the viral capsid and the glycoproteins E1 and E2 promotes viral budding (von Bonsdorff and Harrison, 1978; Roman and Garoff, 1986; Miller and Brown, 1992); the interaction between SFVP and the transmembrane glycoprotein is needed for effective capsid assembly and viral release (Zhao et al., 1994; Lee et al., 1996; Owen and Kuhn, 1997); and (iv) SFVP interacts with ribosomes in the host cell to trigger the release of the genomic RNA in the

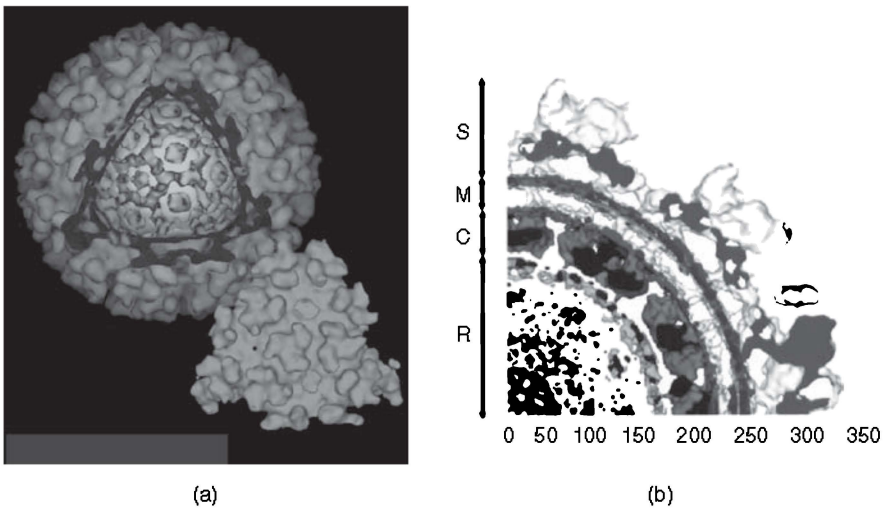


Figure 13.4 Structure of alphavirus determined using cryo-EM. (a,b) The trimeric petal-shaped spikes and the inner capsid core are visible. S, spikes; M, membrane; C, capsid; R, RNA. *Source:* Adapted from http://www.embl.de/ExternalInfo/fuller/EMBL_Virus_Structure.html. (*See insert for color representation of the figure.*)

cell cytoplasm (Singh and Helenius, 1992). These activities are shared among other homologous capsid proteins (63–92% sequence homology) of the genus *Alphavirus* (Strauss and Strauss, 1994). The capsid protein of the Sindbis virus, togavirin (E.C. 3.4.21.90), identifies the serine protease activity displayed in all alphavirus capsid proteins.

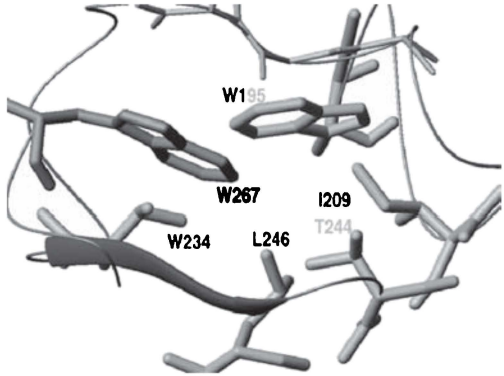
We have concentrated our studies on the C-terminal protease domain of SFVP (residues 119–267). Its structure and folding properties are well characterized (Choi et al., 1996, 1997; Sanchez et al., 2004), but to the best of our knowledge nobody has attempted to reactivate the protease for biotechnological purposes. The mature SFVP and its C-terminal protease domain in its mature form are inactive. The protein performs a single proteolytic reaction in the viral polyprotein, and then its serine protease activity remains inactive. The three-dimensional structure of SFVP shows two β-barrel domains (residues 119–182 and 183–267) and a catalytic serine protease triad at the domain interface (His145, Asp167, and Ser219) (Fig. 13.5a). This fold is similar to the structure of chymotrypsin (Choi et al., 1997). In the mature capsid protein, $N^{\varepsilon 2}$ of His145 and O^{γ} of Ser219 form hydrogen bonds with the free carboxylate of the C-terminal Trp267, thus inhibiting further proteolytic activity. Thus, SFVP performs a single enzymatic reaction before it is assembled in the viral capsid shell. It is worth remarking the 100% conservancy of Trp267 among the *Alphavirus* genus members and its relevance in the viral cycle (Boege et al., 1981; Kinney et al., 1986; Chang and Trent, 1987; Faragher et al., 1988; Hahn et al., 1988; Levinson et al., 1990; Shirako et al., 1991; Rumenapf et al., 1995). Trp267 is buried in a hydrophobic pocket, and single mutations of Trp267 have shown poor viral infectivity (Skoging and Liljestrom, 1998) (Fig. 13.5b).

We have investigated if reactivation of SFVP is possible by deletion of the autoinhibitory C-terminal Trp267. Our objective is to obtain a novel enzyme derived from SFVP to test if the new enzyme is highly specific for tryptophan substrates, which can be of significant biotechnological relevance. In the following sections, we describe the catalytic and structural properties of the C-terminal truncated variants of SFVP and their relationships with other natively unfolded proteins. We have obtained six variants of SFVP to test their catalytic and structural properties (Table 13.1).

13.2 ENZYMATIC ACTIVITY AND STEADY-STATE KINETICS OF THE SFV C-TERMINAL TRUNCATED VARIANTS

We have tested the catalytic properties of the SFVP variants with esters and amide compounds of aromatic and nonaromatic amino acids. The four C-terminal SFVP deletion variants E266 (−1 aa), E265 (−2 aa), S264 (−3 aa), and T260 (−7 aa) and the variant W267A catalyze the hydrolysis of ester derivatives of aromatic amino acids. The hydrolysis of *N*-acetyl-tryptophan-4-nitrophenyl ester (Ac-Trp-ONp) and *N-tert*-butoxycarbonyl-tyrosine-4-nitrophenyl ester (Boc-Tyr-ONp) is accelerated 5×10^4-fold in the presence of nanomolar amounts of the SFVP C-terminal deletion variants. Increasing amounts of the wild-type (wt) protein SFVP

(a)



(b)

Figure 13.5 Three-dimensional X-ray structure of wild-type SFVP. (a) Ribbon representation of the X-ray structure of wild-type SFVP (residues 119–267). The residues that form the catalytic triad (S219, H145, and D167) are in the center of the figure. The three tryptophan residues are located in the C-terminal β-barrel domain. The free carboxylic group of Trp267 makes hydrogen bonds (dotted lines) with $N^{\varepsilon 2}$ of His145 (distance 2.75 Å) and O^{γ} of Ser219 (distance 2.85 Å). (b) Hydrophobic pocket where the C-terminal Trp267 is buried in the structure of wild-type SFVP. The pocket formed is by the hydrophobic groups of the labeled residues, the polypeptide chain of a β-sheet strand (residues Gly236-Ala238, bottom of the figure), and the polypeptide chain of residues Ala213-Lys215 (top of the figure).

TABLE 13.1 List of Protein Constructs of SFVP Used in the Study

Name of Protein Construct	C-Terminal Sequence of SFVP Proteins
SFVP (wt)[a]	x-259-T260-P261-E262-G263-S264-E265-E266-W267
SFVP (−1 aa)	x-259-T260-P261-E262-G263-S264-E265-E266
SFVP (−2 aa)	x-259-T260-P261-E262-G263-S264-E265
SFVP (−3 aa)	x-259-T260-P261-E262-G263-S264
SFVP (−7 aa)	x-259-T260
SFVP (−3 aa) S219G	x-259-T260-P261-E262-G263-S264

[a]The wild-type SFVP has a catalytic triad with sequence D... -H... S that is maintained in all constructs with the exception of SFVP (−3 aa) S219G.

or the variant SFVP Ser219Ala (−3 aa) do not have any effect on catalysis of these compounds (Fig. 13.6).

The steady-state kinetic constants k_{cat}, K_M, and k_{cat}/K_M were calculated from the saturation curves of the enzyme with increasing concentration of substrate (Table 13.2). We measured a value for k_{cat}/K_M of 5×10^5/s/M for the variant with the substrate Ac-Trp-ONp close to some of the highest values of $k_{cat}/K_M \approx 2 \times 10^6$/s/M reported for serine proteases (Lottenberg et al., 1981; Odake et al., 1991; Powers and Kam, 1995). The variant SFVP (−7 aa) showed a threefold decrease in k_{cat} and k_{cat}/K_M relative to the variant SFVP (−1 aa). K_M remained equal in all C-terminal deleted variants within the experimental error.

We measured a 10-fold decrease in k_{cat}/K_M and a 20-fold decrease in k_{cat} with the substrate Boc-Tyr-ONp relative to the substrate Ac-Trp-ONp (Table 13.2). We

TABLE 13.2 Steady-State Kinetic Parameters of C-Terminal Deleted Variants of SFVP

Protein Constructs	Substrates					
	Ac-Trp-ONp			Boc-Tyr-ONp		
	k_{cat}/K_M ($s^{-1}M^{-1}$)	k_{cat}) (s^{-1})	K_M (M)	k_{cat}/K_M (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_M (M)
SFVP (−1 aa)	5×10^5	15 ± 0.8	30 ± 6	—	—	—
SFVP (−2 aa)	1.7×10^5	14.9	89 ± 14	6×10^4	0.76 ± 0.01	13 ± 0.8
SFVP (−3 aa)	2×10^5	12 ± 0.6	60 ± 8	3×10^4	0.70 ± 0.04	24 ± 4
SFVP (−7 aa)	1.6×10^5	5.9 ± 0.4	36 ± 7	1.0×10^4	0.54 ± 0.05	52 ± 9
SFVP (wt)	n.d.			n.d.		
SFVP (−3 aa) S219G	n.d.			n.d.		
Saneous Rate. No Enzyme.	3.7×10^{-4}/s			1.5×10^{-5}/s		

Protein samples (10–90 nM) were added to solutions containing the substrate in 10 mM MOPS-NaOH pH 7.0 at 25°C. The enzymatic activity of SFVP W267G is identical to SFVP (−1 aa), and the catalytic parameters of SFVP (−2 aa) with *p*-nitrophenyl acetate are $k_{cat} = 0.0012 \pm 0.0002$ s^{-1} and $K_M = 213 \pm 66.3$ μM under identical experimental conditions.
Abbreviation: n.d. = not determined.

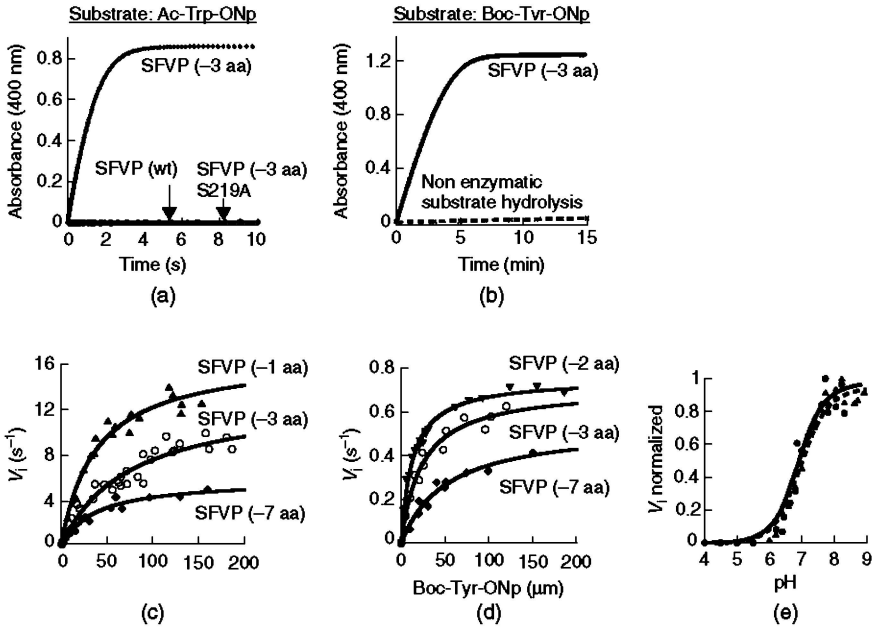


Figure 13.6 Enzymatic activity of SFVP C-terminal deletion variants. (a,b) Examples of full-progress curves of SFVP (−3 aa) with Ac-Trp-ONp (a) and Boc-Tyr-ONp (b). (a) Substrate = 110 μM and enzyme = 5.7 μM. (b) Substrate = 158 μM and enzyme = 5 μM. Reactions measured in 10 mM MOPS-NaOH pH 7.0, $T = 25°C$. SFVP wild type and SFVP S219A (−3 aa) cannot catalyze the hydrolysis of either Ac-Trp-ONp or Boc-Tyr-ONp. (c,d) Effect of increasing the substrate concentration on the enzymatic hydrolysis of Ac-Trp-ONp (c) and Boc-Tyr-ONp (d). The initial velocity of the reaction (micromolar product s^{-1})/(micromolar enzyme) is displayed against substrate concentration. Enzyme concentration = 10–90 nM. Reactions were measured in 10 mM MOPS-NaOH pH 7.0, $T = 25°C$. ▲, SFVP (−1 aa); ▼, SFVP (−2 aa); ○, SFVP (−3 aa); and ♦, SFVP (−7 aa). (e) pH-dependence of the enzymatic activity of SFVP C-terminal deletion (−3 aa) variant with Ac-Trp-ONp and Boc-Tyr-ONp. The initial velocities of full progress curves were measured at 2.5 μM substrate concentration and 60–120 nM enzyme concentration. The enzymatic activity of SFVP(−3 aa) exhibits a $pK1 = 6.97 \pm 0.09$ and 6.98 ± 0.09 with Ac-Trp-ONp (•) and Boc-Tyr-ONp (▲), respectively.

also observed a small acceleration in the hydrolysis of the ester *p*-nitrophenyl acetate in the presence of the SFVP C-terminal deletion variants. The value of k_{cat}/K_M is only 5.7/s/M. This represents a 10,000-fold difference in enzymatic efficiency relative to specific substrates. We did not detect any enzymatic catalysis with other esters of aromatic acids such as *N*-acetyl-leucine *p*-nitrophenyl ester and *N*-acetyl-glycine *p*-nitrophenyl ester. These results suggest substrate specificity for tryptophan over tyrosine and basically nonspecificity over any other amino acid. These results agree with *in vivo* mutagenesis studies on Trp267 (Skoging and Liljestrom, 1998) and the intramolecular protease activity that we observed *in vivo* for SFVP fusion proteins. A model protein (p53) was fused C-terminal to residue

267 of either wt SFVP (W267) or SFVP W267Y and expressed in the cytoplasm of *Escherichia coli*. We observed that 100% of the model fusion proteins were cleaved off the SFVP if the residue at position 267 was tryptophan (wt). However, only 40–50% of the model fusion proteins was cleaved off if the residue at position 267 was tyrosine (data not shown).

The enzymatic activity exhibited a pH dependence with a p*K1* of 6.97, a value similar to the ionization of histidine in proteins as it was expected from an active catalytic triad. We can conclude that deletion of the C-terminal tryptophan is sufficient to reactivate the hydrolytic activity of SFVP. We can also conclude that Ser219 is an essential nucleophilic residue in the reactivated variants. These data suggest that the catalytic triad (His145, Asp167, and Ser219) (Choi et al., 1996, 1997) is responsible for enzymatic catalysis of the C-terminal truncated SFVP variants.

Tryptophan 267 is fully conserved among capsid proteins of alphaviruses, and its side chain is located in a buried hydrophobic pocket surrounded by hydrophobic and noncharged residues (Boege et al., 1981; Chang and Trent, 1987; Shirako et al., 1991; Choi et al., 1997). We, however, could not detect any enzymatic activity with esters of small aliphatic groups and other amino acids, indicating that SFV variants are enzymatically active only with tryptophan substrates and less efficiently with tyrosine substrates. This came as no surprise since previous studies showed the pivotal role of Trp267 in viral assembly and viral viability (Skoging and Liljestrom, 1998). Skoging and Liljeström showed that mutation of Trp267 by phenylalanine, alanine, or arginine decreased the count of virus proliferation to five orders of magnitude. These Trp267 mutations were shown to be responsible for an increase in capsid aggregation and a reduction in autoprotease activity and nucleocapsid formation. Our enzymatic results with C-terminal truncated variants of SFVP support the essential role of Trp267 in the efficient catalytic activity of the capsid protein and hence in processing of the capsid assembly and viral proliferation.

13.3 PRE-STEADY-STATE KINETICS OF THE SFV C-TERMINAL TRUNCATED VARIANTS

We have studied the mechanism of action of SFV C-terminal truncated variants by determining individual rate constants. Hydrolases and proteases are well known to accelerate catalysis by stabilization of a tetrahedral intermediate and formation of an acyl enzyme in the enzymatic reaction (Polgar, 1989; Taylor, 1993; Roa et al., 1996; Morillas et al., 1999; Turk, 1999). We have strong evidence that the same catalytic triad responsible for catalysis on SFVP wt is also responsible for the catalysis observed with esters of aromatic amino acids. We followed in the enzymatic reaction of SFVP C-terminal deletion variants with Boc-Tyr-ONp and Ac-Trp-ONp. The enzymatic hydrolysis of the two substrates progresses in two distinctive phases (Fig. 13.7). The amplitude of the first phase is proportional to the enzyme concentration, and the rate of the reaction increases with increasing concentrations of substrate. The second phase corresponds to the steady state of the enzymatic reaction that was used to calculate the steady-state parameters.

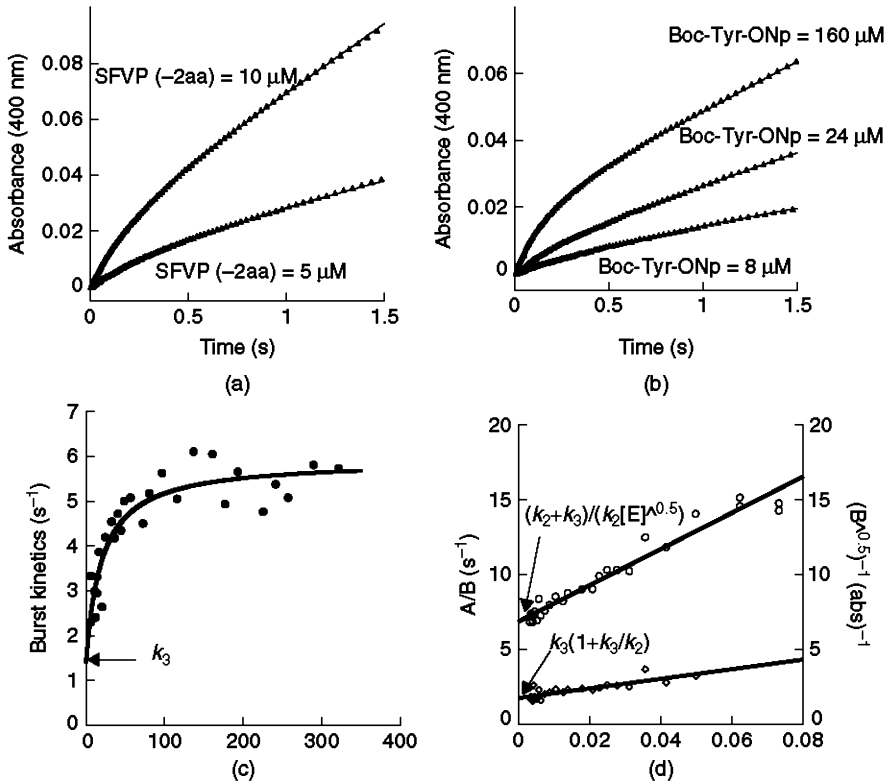


Figure 13.7 Burst kinetics of SFVP C-terminal deletion variants and pre-steady-state analysis. (a) Stopped-flow measurements of the enzymatic hydrolysis of SFVP (−2 aa) with Boc-Tyr-ONp at two enzyme concentrations. Two phases were detected at the beginning of the enzymatic reaction. The amplitude of the first phase increased with increasing enzyme concentrations. The slope of the second phase increased with enzyme concentration. Concentration of substrate = 150 μM. (b) Effect of substrate concentration on the enzymatic hydrolysis of SFVP (−2 aa). The rate of the first phase and the slope of the second phase increased with substrate concentration. (c) The rate of the burst phase is plotted against substrate concentration. The rate of the burst phase saturates with substrate concentrations. (d) Effect of $(\text{substrate})^{-1}$ on the slope/amplitude (A/B) of the kinetic traces (o). The abscissa intercept was used to obtain individual rate constants. Effect of $(\text{substrate})^{-1}$ on the inverse of square root of the burst amplitude $1/\sqrt{B}$(◇). The abscissa intercept was used to calculate the total amount of active sites.

We interpret the biphasic release of the product as the transient accumulation of an acyl intermediate (EA) with a rate k_2, followed by its steady-state hydrolysis to free enzyme with a rate k_3 (Polgar, 1989; Taylor, 1993; Morillas et al., 1999; Turk, 1999). See the following model.

$$E + S \overset{K_S}{\longleftrightarrow} ES \overset{k_2}{\longrightarrow} EA + P_1 \overset{k_3}{\longrightarrow} E + P_2$$

TABLE 13.3 Rate Constants and Parameters Obtained from Burst Kinetics

	Rate Constants and Parameters				
	k_2	k_3	K_M^{acyl}	k_{cat}	[EA][a]
Boc-Tyr-ONp	$4.4 \pm 0.6\ s^{-1}$	$1.6 \pm 0.5\ s^{-1}$	$22 \pm 9\ \mu M$	0.76 ± 0.05	95%
Ac-Trp-ONp[b]	$21 \pm 5.2\ s^{-1}$	$6.3 \pm 3.5\ s^{-1}$	$44 \pm 32\ \mu M$	14 ± 0.8	94%

[a]Concentration of active enzyme molecules.
[b]The stopped-flow data with Ac-Trp-ONp are more scattered because of the higher spontaneous hydrolysis of this ester.

The saturation value of $k_2 + k_3$ as a function of substrate concentration is consistent with $k_2 \geq k_3$, k_{cat} is not rate limited, and $K_M \approx K_M^{acyl}$. Individual rate constants of SFV variants and the concentration of active sites are shown in Table 13.3. The concentration of active sites relative to the enzyme concentration indicates that 95% of the enzyme molecules are active and have one active site per molecule, supporting this enzymatic model.

The value of k_2 is slightly bigger than the value of k_3 in SFV variants, $k_2/k_3 = 2.8$ and 3.3 with Boc-Tyr-ONp and Ac-Trp-ONp respectively. The values of k_2/k_3 are bigger in other proteases and hydrolases with ester substrates (Mao et al., 1992; Roa et al., 1996; Morillas et al., 1999); for instance, $k_2/k_3 > 600$ in chymotrypsin (Mao et al., 1992). The close values of k_2 and k_3 in SFV variants indicate that the acyl enzyme intermediate accumulates to 55% of the total enzyme concentration during the pre-steady-state, but it accumulates to nearly 100% of the total enzyme concentration in chymotrypsin.

Given the lower thermodynamic stability of the SFV variants (Section 13.5), we tested if the small ratio of $k_2 : k_3$ may be the consequence of high energy barriers from a rate-dependent protein-folding reaction. We have studied the effect of temperature on the enzymatic reaction of the C-terminal truncated variants of SFVP and determined individual rate constants and steady-state constants with the substrate Boc-Tyr-ONp.The initial velocity of the enzymatic reaction was determined between 5 and 42°C. The Arrhenius plots show a deviation from linearity above 25°C. Similar deviation of linearity in the Arrhenius plots has been observed in other proteases and hydrolases above 22 and 25°C, which has been interpreted either as evidence of enzyme isomerization or change in the rate limiting step of the enzymatic reaction (Rajender et al., 1970; Adams and Swart, 1977; Wang et al., 1981; Nohara et al., 1989; Martin et al., 1990; Morillas et al., 1999; Truhlar and Kohen, 2001). To simplify our analysis, we have focused our thermodynamic studies at temperatures between 5 and 25°C. The activation energies for the individual rate constants k_2 and k_3 and the steady-state rate constant k_{cat} vary between 30 and 37 kJ/mol, and this corresponds to a decrease in the activation energy of 23 kJ/mol relative to the uncatalyzed reaction (Fig. 13.8 and Table 13.4). The values for the activation energy of SFVP variants fall within the values of activation energies measured with other enzymes (14–49 kJ/mol (Sanyal and Maren, 1981; Valley and Fitzpatrick, 2004; Yu and Li, 2006) and 13.4 ± 1 kJ/mol (Morillas et al., 1999) for

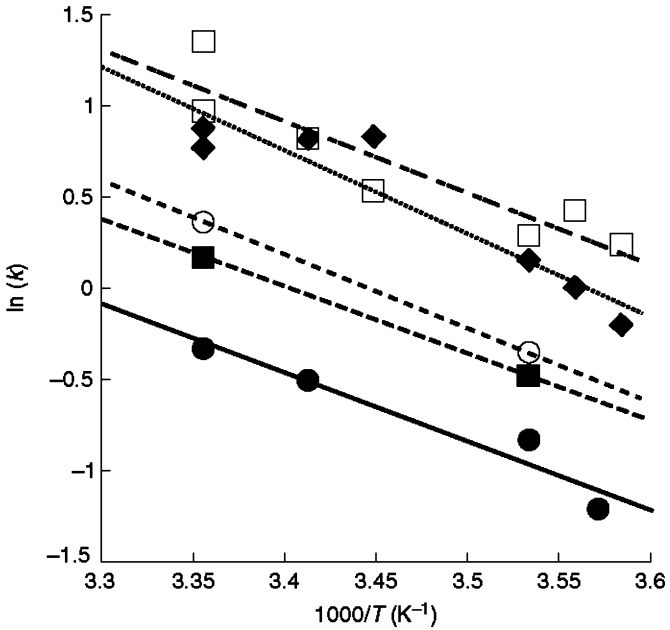


Figure 13.8 Arrhenius plots for the kinetic parameters of SFVP (−3 **aa**). Arrhenius plots for the kinetic parameters of SFVP (−3 aa). The natural logarithms of the kinetic parameters V_i(○), k_2(○), k_3(■), $k_2 + k_3$(□), and slope of V_i/burst amplitude (♦) are plotted against the inverse of temperature. Enzymatic reactions under steady-state conditions contained 100 μM of the substrate Boc-Tyr-ONp and 190 nM of SFVP (−3 aa) in 10 mM MOPS-NaOH pH 7.0. Enzymatic reactions under pre-steady-state conditions contained 100 μM of the substrate Boc-Tyr-ONp and 3.7 μM of SFVP (−3 aa).

TABLE 13.4 Activation Energies from Rate Constants and Parameters

Kinetic Parameter	E_{act} (kJ/mol)
k_{cat} (steady state)	30.9 ± 5.9
k_2	33.3
k_3	30.1
$k_2 + k_3$	32.5 ± 6.2
Slope (V_i)/burst amplitude	37.5 ± 6.3

Activation energies for steady-state and individual rate constants of the C-terminal truncated variant of SFVP (−3 aa) at 25°C.

k_2 of penicillin G acylase). However, higher activation energy values have been reported for other enzymatic reactions; for instance, the activation energies for the chymotrypsin- and trypsin-catalyzed reactions vary between 46–53 kJ/mol (Snoke and Neurath, 1950; Baggott and Klapper, 1976) and 59–93 kJ/mol (Blazyk et al., 1981; Toth et al., 2006), respectively, depending on the reaction. Experimental

values for the activation energy of protein-folding reactions also vary with different proteins (16–155 kJ/mol) (Oliveberg et al., 1995; Su et al., 1996; Zerovnik et al., 1998, 1999; Perl et al., 2002; Verheyden et al., 2004) because it involves a big change on, U, whose associated free energy is temperature dependent (Scalley and Baker, 1997; Perl et al., 2002), and it also depends on the value of the diffusion coefficient (Naganathan et al., 2007). We conclude that the activation energy for the enzymatic reaction of the C-terminal truncated variants of SFVP falls within the observed values for protein-folding reactions, and we cannot exclude conformational changes with small energy barriers (Oliveberg et al., 1995; Perl et al., 2002).

13.4 SUBSTRATE SEARCH AND INHIBITION STUDIES

We have not detected any enzymatic activity of SFV variants with the compounds *N*-acetyl-tryptophan *p*-nitroanilide, *N*-glutaryl-phenylalanine *p*-nitroanilide, 6-nitro-3-phenylacetamide benzoic acid, Suc-Ala-Ala-Ala-Pro-Phe-*p*-nitroanilide, and *N*-acetyl-glutamate-glutamate-tryptophan *p*-nitroanilide (Ac-EEW-pNa), a compound that mimics the last three residues deleted in the variant. Reasons for the lack of activity may reside in the low stability of the SFV variants (Section 13.5). It is well known that the transition-state energy barrier for amide compounds is about 11–19 kcal/mol higher than for esters compounds (Chong et al., 2003). Chong et al., showed that an engineered catalytic antibody could hydrolyze ester substrates, but it could not surmount the energy barrier and catalyze the hydrolysis of amide compounds. We interpret that we could not detect enzymatic activity of SFV variants with amide compounds for similar reasons, and the novelty resides in the efficient catalysis of an inactive protease with ester substrates that was restricted to a single intramolecular turnover.

Known covalent inhibitors of serine proteases such as phenylmethylsulfonyl fluoride and choromethylketones such as *N*-benzyloxycarbonyl-leucine-tyrosine-chloromethylketone and *N*-benzyloxycarbonyl-phenylalanine-chloromethylketone did not inhibit the C-terminal variants of SFVP, even though they are well-known inhibitors of serine proteases (Sidorowicz et al., 1980; Eguchi and Kuriyama, 1985; Jung et al., 1995; Johnson and Moore, 2000, 2002). The inhibition constant of SFV variants with free tryptophan ($K_I = 5–7$ mM) is a very high value to consider free tryptophan as an inhibitor (Ohno et al., 1976; Sato et al., 1977). *N*-acetyl-glutamate-glutamate-tryptophan *p*-nitroanilide that mimics the last three residues deleted in the variant did not show any inhibition to the catalysis of ester substrates. The enzymology studies do not provide a definitive answer to the lack of inhibition by these compounds, but structural studies of natively unfolded proteins help to interpret these data (Section 13.5).

We conclude that SFV variants cannot stabilize the high energy required for buildup, stabilization, and decomposition of an amide transition-state intermediate, and this may be explained by the particular 3D structure of the SFV variants (see next section).

13.5 STRUCTURAL STUDIES OF THE C-TERMINAL TRUNCATED SFV CAPSID PROTEINS

We have shown in the preceding sections that SFV variants are specific Michaelis–Menten enzymes for esters of aromatic amino acids. We have reached our original objective to reactivate the enzymatic activity of SFVP. In this section, we explain the structural changes and conformational stability observed in the enzymatically active SFV variants.

The first indication that structural differences might occur between SFV variants and SFVP wt was observed during the expression and purification stages of the recombinant SFV variant. SFVP wt is fully expressed and purified as a soluble protein from the cytoplasm of *E. coli*. SFV variants are expressed (more than 95% of the expressed protein) in the insoluble fraction of *E. coli*. Attempts to recover the soluble fraction failed, and we attempted to refold the SFV variants from the insoluble fraction. The methodology to purify and recover the SFVP variants has been described in detailed elsewhere (Morillas et al., 2008).

The structure and biophysical properties of the SFVP C-terminal deletion variants were studied using spectroscopic methods and equilibrium sedimentation experiments. We determined the mass of the protein, independent of its shape in solution, by equilibrium sedimentation experiments. These experiments performed at different protein concentrations showed that the C-terminal deletion variants are monomeric, thus eliminating the possibility of protein oligomerization in our studies (Fig. 13.9a and Table 13.5).

The 1D ^{1}H NMR spectrum of the wt SFVP shows a chemical shift dispersion expected for a well-folded protein with a chemical shift range from 11 to −0.5 ppm. However, the C-terminal truncated variant SFVP (−2 aa) shows broader and less dispersed spectra indicative of a natively unfolded protein (Fig. 9b,c). The spectra were recorded at pH 6.0 to reduce the amide exchange with water, and similar results were obtained at pH 7.0, where the C-terminal deletion variants are more active. We attempted to repeat the same 1D ^{1}H NMR experiments in the presence of the substrates Boc-Tyr-ONp and Ac-Trp-ONp and detect protein conformations, but the high enzyme concentration needed for these experiments (0–2 mM, 3.2 mg/mL) hydrolyzes the substrate very fast before any measurement may be recorded. Owing to the high absorbance and fluorescence of the substrates, we could not follow protein conformational changes in a stopped flow.

In an attempt to explain how nonnative polypeptides could efficiently hydrolyze esters of aromatic acids, we performed 2D total correlation spectroscopy (TOCSY) experiments on SFVP (−2 aa) in the absence and presence of the poor substrate *p*-nitrophenyl acetate. The value of k_{cat}/K_M with this compound is only 5.7/s/M. This represents a 10,000-fold difference in enzymatic efficiency relative to the specific substrates Boc-Tyr-ONp and Ac-Trp-ONp. The reaction is so slow that we could measure over a 24-h period the enzymatic hydrolysis of the ester. Additionally, we could not observe burst kinetics with the compound *p*-nitrophenyl acetate compatible with either $k_3 > k_2$ or $K_M^{acyl} > K^M$. We measured the enzymatic reaction of SFVP (−2 aa) with *p*-nitrophenyl acetate by 2D TOCSY NMR. We chose

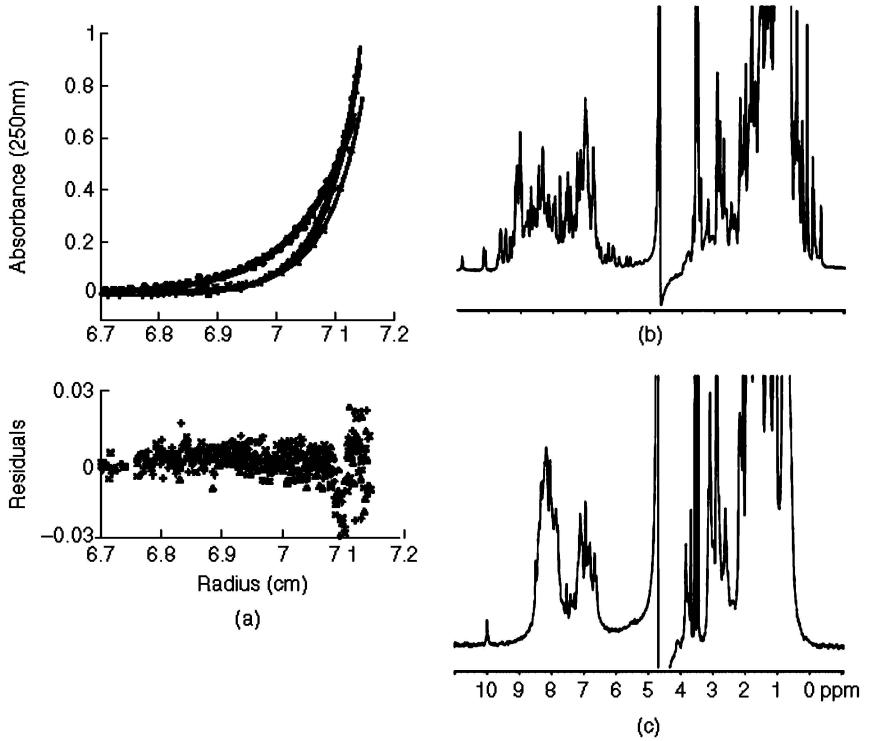


Figure 13.9 Equilibrium sedimentation analysis and 1D ^{1}H spectra of wild-type and C-terminal truncated variant of SFVP. (a) Equilibrium sedimentation analysis of SFVP was performed in a Beckman Optima XL-I analytical ultracentrifuge at pH 7.0 and 20°C. Experiments were run at 27,000 and 35,000 rpm in an An-50 Ti Beckman rotor. SFVP (wild type) at 27,000 rpm (◇) and 35,000 rpm (△). SFVP (−2 aa) at 27,000 rpm (□) and 35,000 rpm (+). SFVP (−3 aa) at 27,000 rpm ○ and 35,000 rpm (×). The data shown here were obtained at a protein concentration of 40 μM. Continuous lines represent fits according to a single-component system analysis, and the residuals of this analysis are shown in the bottom panel. Identical results were obtained at a protein concentration of 10 μM. (b,c) One-dimensional ^{1}H spectra of the free SFVP wild type (b) and free SFVP (−2 aa) (c). Protein solutions of 0.29 mM for SFVP wild type and 0.2 mM for SFVP (−2 aa) were prepared in 50 mM acetate(D^2)-HCl pH 6.0, $T = 25$°C.

conditions such that the reaction developed with a $t_{1/2} \approx 4$ h. The 2D TOCSY spectrum of SFVP (−2 aa) in the absence of substrate shows a disordered protein structure compared with the 2D spectrum of the well-folded SFVP wt (Fig. 13.10). Successive spectra recorded every 2 h showed that although the reaction is efficient as indicated by disappearance of the cross-peaks from the substrate and appearance of cross-peak of the product (Fig. 13.10c–e, circled), we did not observe any chemical shift changes in the protein signals (Fig. 13.10c–e), suggesting that the protein remains mostly unfolded, although the reaction proceeds.

TABLE 13.5 Equilibrium Sedimentation Analysis of SFVP wt and Calculated Mass

Protein	M_r, Equilibrium Sedimentation Analysis	M_r, Calculated Mass
SFVP wt	15844 ± 109 Da	16209 Da
SFVP (−2 aa)	16586 ± 510 Da	15894 Da
SFVP (−3 aa)	16996 ± 1380 Da	15765 Da

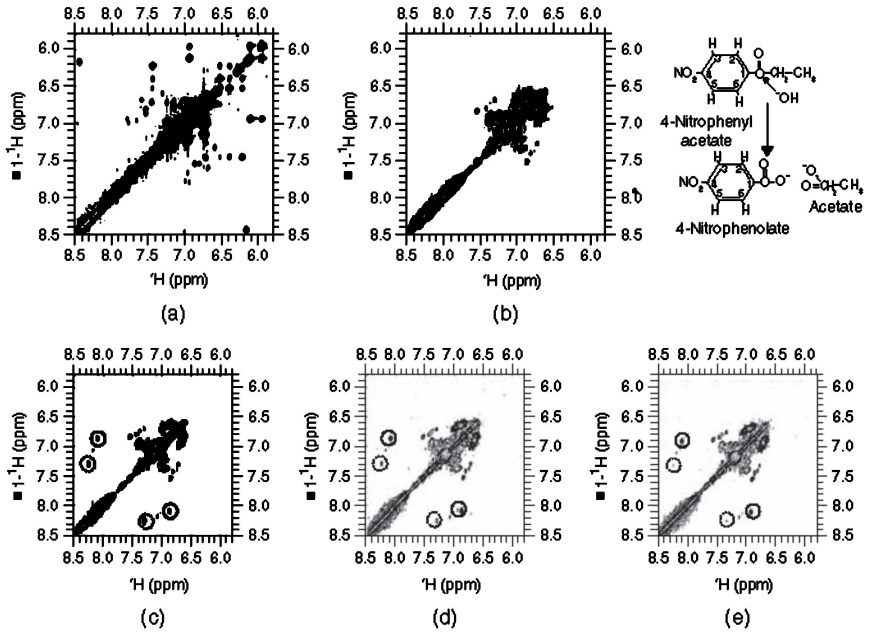


Figure 13.10 Two-dimensional TOCSY NMR experiments of SFVP proteins. (a) SFVP wild type [protein] = 0.29 mM. (b) SFVP (−2 aa) in the absence of substrate [protein] = 0.2 mM. (c–e) 0.2 mM SFVP (−2 aa) +0.2 mM *p*-nitrophenyl acetate. The spectra were recorded between 0 and 2 h (c), 11 and 13 h (d), and 19 and 21 h (e) after the start of the reaction. The chemical structure of the substrate 4-nitrophenyl acetate and the products 4-nitropenolate and acetate are shown. The 2D TOCSY spectra shown in (c–e) also show correlations between the H2,H6 and H3,H5 aromatic protons within the substrate (dotted line) and the product (solid line). Buffer = 50 mM acetate(D^2)-HCl pH 6.0, $T = 25°C$.

The far-UV and near-UV CD spectra of the C-terminal deletion variants also differ significantly from those of the wt SFVP spectra. The spectra of the C-terminal deletion variants show lack of tertiary structure and incomplete secondary structure (Fig. 13.11a). The far-UV CD spectra of the SFV variants show a minimum of around 200 nm observed in unfolded polypeptides (Uversky et al., 1999), and a positive band at 230 nm that disappears under mild denaturing conditions (Fig. 13.11b). The near-UV CD spectra of the SFV variants under native conditions

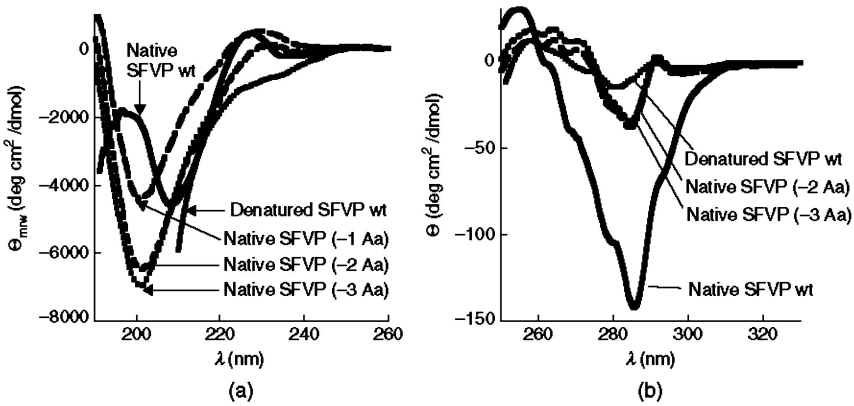


Figure 13.11 Spectroscopic characterization and urea-induced unfolding and refolding of wild-type and C-terminal truncated variants of SFVP transition by CD. (a) Far-UV CD spectra of SFVP under native and denaturing conditions. Spectra of protein samples (22–50 μM) of SFVP wild-type, SFVP variants (−1 aa), SFVP (−2 aa), and SFVP (−3 aa) under native conditions are labeled. The spectra of all the variants and wild-type SFVP are identical under denaturing conditions, and only one is shown. Native conditions = 10 mM MOPS-NaOH pH 7.0. Denaturing conditions = 6*M* urea, 10 mM MOPS-NaOH pH 7.0. The spectra recorded in 0.2-mm path length quartz cuvettes. (b) Near-UV CD spectra of SFVP under native and denaturing conditions. Protein samples of 110 μM SFVP (wild type and C-terminal deletion variants) were used to measure these spectra in 10-mm path length quartz cuvettes. The spectra of SFVP wild type, SFVP variants (−2 aa), and SFVP (−3 aa) under native conditions are labeled. The spectra of all the SFVP variants and wild-type SFVP are identical under denaturing conditions, and only one is shown. Native conditions = 10 mM MOPS-NaOH pH 7.0. Denaturing conditions = 6*M* urea, 10 mM MOPS-NaOH pH 7.0. (c) Esterase activity under mild denaturing conditions and urea-induced unfolding and refolding of the C-terminal deletion SFVP variants monitored by CD. Protein samples were incubated in urea for 16 h before monitoring the enzymatic activity with 100 μM Ac-Trp-ONp or recording the protein CD signal at λ = 230 nm. The initial velocity (V_i) of the enzymatic reaction of SFVP (−3 aa) (black squares) was recorded and normalized to the initial velocity in the absence of urea. The urea-induced unfolding reaction of SFVP C-terminal deletion variants measured by CD at λ = 230 nm is reversible but does not show a clear pretransition region, and the signal decreases from the native state at 0 *M* urea to the unfolded state at 2.5 *M* urea. The urea-induced unfolding reaction of SFVP wild type measured by CD at λ = 227 nm is, however, a reversible and cooperative transition between two states, and the results were fitted to the two-state model (the dotted line is the fitted curve). CD signal of SFVP (−3 aa, ○), SFVP (−2 aa, △), SFVP (−1 aa, □).

are similar to the spectrum of the unfolded protein under denaturing conditions. This suggests a loss in long-range interactions responsible for the tertiary structure and indicates that the aromatic residues are flexible and can rotate freely. These results show that the SFV variants have lost most of the tertiary structure that was present in the wt and retain residual secondary structure typical of natively unfolded proteins.

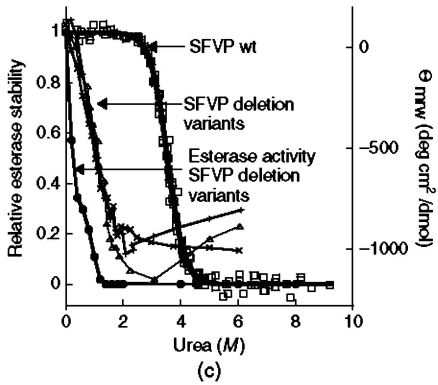


Figure 13.11 (*Continued*)

The structural studies of the SFV variants were completed with intrinsic protein fluorescence measurements (Fig. 13.12). The wt SFVP contains three conserved tryptophan residues (Trp195, Trp251, and Trp267) located in the C-terminal β-barrel domain, but the C-terminal deletion variants contain only two tryptophan residues (Trp195 and Trp251). The fluorescence spectrum of wt SFVP shows a λ_{max} of 333 nm under native conditions, indicative of an apolar environment, and is consistent with the X-ray crystallographic data showing each one of the three tryptophan residues buried in the protein interior (Fig. 13.5b). We observed different fluorescence intensities of SFV variants under native conditions, but this may be related to the different local environment, λ_{max}, quantum yield, and the tryptophan residues in each variant (Verheyden et al., 2003). All SFV variants have identical fluorescence spectra under denaturing conditions, and the intensity is two-third of the fluorescence intensity of the SFVP wt under identical denaturing conditions. There is, however, a big difference in the fluorescence λ_{max} of the SFV variants relative to the SFVP WT. The λ_{max} of SFV variants is shifted toward higher wavelengths indicative of more exposure to the solvent. The $\Delta\lambda_{max}$ is 3 nm for SFVP (−1 aa), 6 nm for SFVP (−2 aa), 8 nm for SFVP (−3 aa), and 9 nm for SFVP (−7 aa) (Fig. 13.12a). This suggests that the tryptophan residues of the C-terminal truncated SFV variants are more exposed to the solvent under native conditions than the wt indicating a lack of long-range interaction responsible for the tertiary structure.

Binding of the dye 1-anilinonaphthalene-8-sulfonate (ANS) to the protein (Fig. 13.12c) also indicates exposure of the C-terminal deletion variants to the solvent. The fluorescence spectrum of ANS shows a λ_{max} blueshift of 30 nm and a 420% increase in fluorescence intensity on binding to the SFVP truncated variants, but only a 6-nm blueshift and a 20% fold increase in fluorescence intensity in the presence of wt SFVP. These results confirm the presence of hydrophobic patches accessible to the solvent in the C-terminal deletion variants.

We have followed intrinsic protein fluorescence and far-UV CD protein signal under native and denaturing conditions to study the thermodynamic stability

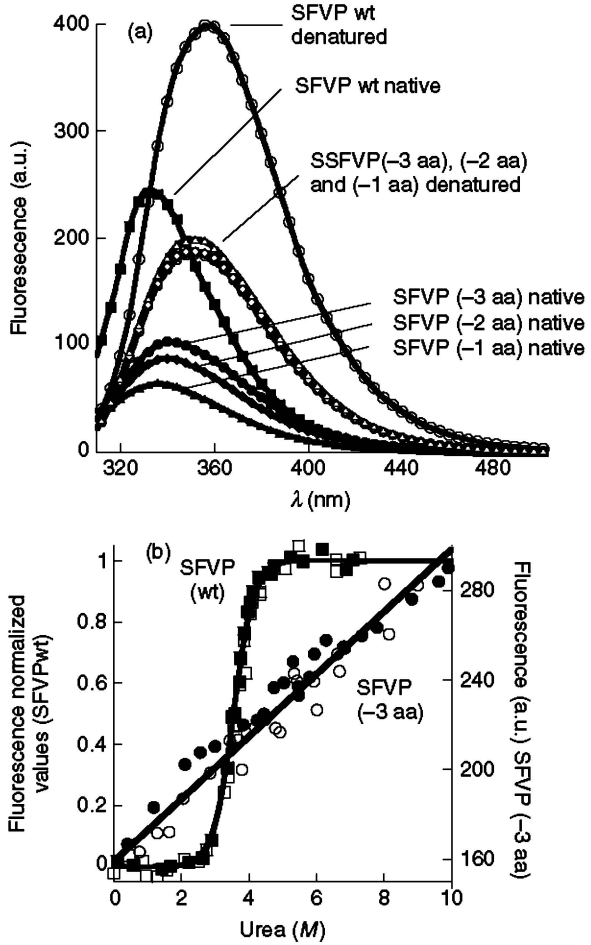


Figure 13.12 Spectroscopic characterization and urea-induced unfolding and refolding of wild-type and C-terminal truncated variants of SFVP transition by intrinsic protein fluorescence. (a) Intrinsic protein fluorescence spectra of SFVP C-terminal deletion variants and wild-type SFVP. SFVP wt (■, □), SFVP (−1 aa) (▲, △), SFVP (−2 aa) (♦, ◇), and SFVP (−3 aa) (●, ○). λ_{exc} = 295 nm. Closed symbols: native conditions in 10 mM MOPS-NaOH pH 7.0. Protein concentration = 2 μM. Open symbols: denaturing conditions in 6*M* urea, 10 mM MOPS-NaOH pH 7.0. (b) Urea-induced unfolding–refolding transitions measured by intrinsic protein fluorescence. The SFVP wild type shows a reversible cooperative transition that was normalized according to a two-state model. The unfolding–refolding reaction of all the SFVP C-terminal deletion variants shows a linear dependence with urea. Protein samples (1–3 μM) were incubated at different concentrations of urea for 16 h, and the fluorescence signal at 360 nm was recorded. SFVP wild type normalized values (■, □). SFVP (−3 aa) (●, ○) values at 360 nm, λ_{exc} = 295 nm. Closed symbols, unfolding reaction. Open symbols, refolding reaction. (c) Fluorescence spectra of 1-anilinonaphthalene-8-sulfonate (ANS) in free solution and bound to SFVP wild type and C-terminal truncated variants. A solution containing 34 μM ANS and 2 μM SFVP was preincubated for 10 min in 10 mM MOPS-NaOH pH 7.0. The fluorescence spectra of the free dye and the dye bound to the protein sample are shown. λ_{exc} = 360 nm.

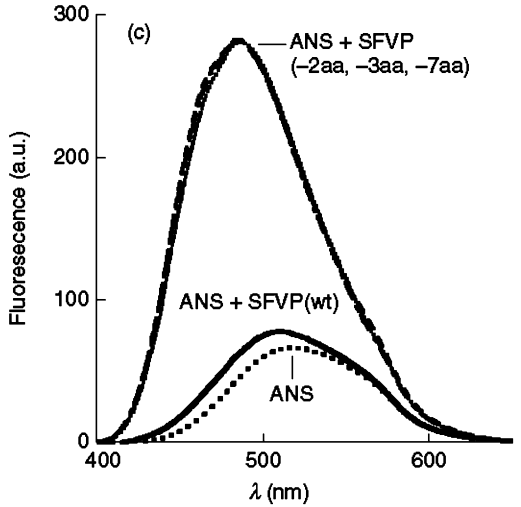


Figure 13.12 (*Continued*)

of SFV variants. We have previously shown that wt SFVP unfolds and refolds in a cooperative and reversible manner with a ΔG of 30 kJ/mol (Sanchez et al., 2004); however, very different results are obtained with the C-terminal truncated variants compatible with natively unfolded proteins (Anil et al., 2006; Khaymina et al., 2007; Neira et al., 2010). The fluorescence signal of the C-terminal deletion variants changes linearly with urea concentration between the native state and the unfolded state, and there is no indication of a cooperative transition (Fig. 13.12b). The CD signal in the urea-induced unfolding reaction decreases between 0 and 2.5 *M* urea in the C-terminal deletion variants measured; it does not show a pretransition baseline and remains constant at higher concentrations of the denaturant (Fig. 13.11c). Moreover, the decrease in enzymatic activity of the C-terminal deletion variants under mild denaturing conditions does not coincide with the change in CD signal. The enzymatic activity at 0.5 *M* urea is 50% of its value under native conditions, and there is no enzymatic hydrolysis at concentrations higher than 1 *M* urea. These results show evidence of very small energy barriers between the native and denatured states, which are expected in natively unfolded proteins and undetectable with the current experimental techniques.

Overall, we interpret the efficient catalysis of the C-terminal truncated variants as a rapid equilibrium of natively unfolded polypeptides and the active native state (Fink, 2005; Receveur-Brechot et al., 2006). The equilibrium is displaced to the ensemble of natively unfolded proteins in the absence of substrate and shifted toward the native state in the presence of the substrate. By making favorable interactions with the hydrophobic pocket, the substrate might stabilize the native structure by fitting in the aromatic moiety in the empty hydrophobic pocket where the side chain of tryptophan 267 is buried in the wt. This model explains that nearly 100% of the enzyme molecules are active but that covalent inhibitors of

serine proteases do not inhibit SFVP since they cannot bind to the ensemble of natively unfolded proteins or the enzyme–substrate complex.

In this model, folding to the active state precedes catalysis, and since protein folding is generally slower than substrate binding, we would expect protein folding to rate limiting catalysis by high energy barriers. We have calculated the activation energies of the individual rate constants. and we conclude that the activation energy for the enzymatic reaction of the C-terminal truncated variants of SFVP falls within the observed values for protein-folding reactions (Sanyal and Maren, 1981; Morillas et al., 1999; Valley and Fitzpatrick, 2004; Yu and Li, 2006). These data do not allow us to state that protein folding limits the enzymatic reaction of SFVP C-terminal truncated variants, but we propose that if in a protein conformational change occurs along with enzymatic catalysis, this may have a small activation energy (at least not bigger than the observed activation energy of the enzymatic reaction), otherwise we would have observed bigger slopes in the Arrhenius plots.

Examples of natively unfolded viral proteins are found in this book, but to the best of our knowledge, only two active enzymes with nonnative structural characteristics have been reported (MacBeath et al., 1998; Li and Jing, 2000; Vamvaca et al., 2004). Genetic selection and structure-based design were used to obtain a monomeric active form of the dimeric enzyme chorismate mutase (MacBeath et al., 1998). The active monomer enzyme is, however, a molten globule polypeptide. A transition-state analog bound to the active site produced a conformational change in the structure of the enzyme. This conformational change shows an increase in protein stability and tertiary structure (MacBeath et al., 1998; Vamvaca et al., 2004). Another example of natively unfolded enzyme is the double mutant of staphylococcal nuclease. In this case, binding of the substrate resulted in a reorganization of the enzyme required for efficient catalysis (Li and Jing, 2000).

In summary, we show that we can obtain polypeptides derived from the SFV capsid protein with high enzymatic activity toward esters of aromatic amino acids. Although we have not observed a peptidase activity, the esterase activity is, to the best of our knowledge, the highest enzymatic activity achieved by a natively unfolded enzyme and opens new perspectives in biotechnology and biochemistry. We predict that other capsid proteins of the *Alphavirus* genus can be converted to active Michaelis–Menten enzymes by deletion of the highly conserved C-terminal tryptophan.

ACKNOWLEDGMENTS

This work was supported by the Schweizerische Nationalfonds and the ETH Zurich within the framework of the NCCR Structural Biology program.

ABBREVIATIONS

SFVP	capsid protein of the Semliki Forest virus
Ac-Trp-ONp	*N*-acetyl-tryptophan-4-nitrophenyl ester

Boc-Tyr-ONp	*N*-*tert*-butoxycarbonyl-tyrosine-4-nitrophenyl ester
ANS	1-anilinonaphtalene-8-sulfonate
Ac-EEW-pNA	*N*-acetyl-glutamate-glutamate-tryptophan *p*-nitroanilide

REFERENCES

Adams PA, Swart ER. The effect of temperature on the individual stages of the hydrolysis of non-specific-p-nitrophenol esters by alpha-chymotrypsin. Biochem J 1977; 161(1):83–92.

Anil B, Li Y, Cho JH, Raleigh DP. The unfolded state of NTL9 is compact in the absence of denaturant. Biochemistry 2006;45(33):10110–10116.

Baggott JE, Klapper MH. Rate enhancement specificity with alpha-chymotrypsin: temperature dependence of deacylation. Biochemistry 1976;15(7):1473–1481.

Blazyk JF, Blazyk JM, Kline CM. Temperature-dependent changes in the activity and tryptic susceptibility of membrane-bound transhydrogenase. J Biol Chem 1981;256(2): 691–694.

Boege U, Wengler G, Wittmann-Liebold B. Primary structures of the core proteins of the alphaviruses Semliki Forest virus and Sindbis virus. Virology 1981;113(1):293–303.

Chang GJ, Trent DW. Nucleotide sequence of the genome region encoding the 26S mRNA of eastern equine encephalomyelitis virus and the deduced amino acid sequence of the viral structural proteins. J Gen Virol 1987;68(Pt 8):2129–2142.

Choi HK, Lee S, Zhang YP, McKinney BR, Wengler G, Rossmann MG, Kuhn RJ. Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis. J Mol Biol 1996;262(2):151–167.

Choi HK, Lu G, Lee S, Wengler G, Rossmann MG. Structure of Semliki Forest virus core protein. Proteins 1997;27(3):345–359.

Chong LT, Bandyopadhyay P, Scanlan TS, Kuntz ID, Kollman PA. Direct hydroxide attack is a plausible mechanism for amidase antibody 43C9. J Comput Chem 2003;24(12):1371–1377.

de Curtis I, Simons K. Dissection of Semliki Forest virus glycoprotein delivery from the trans-Golgi network to the cell surface in permeabilized BHK cells. Proc Natl Acad Sci U S A 1988;85(21):8052–8056.

Dominguez G, Wang CY, Frey TK. Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution. Virology 1990;177(1): 225–238.

Eguchi M, Kuriyama K. Purification and characterization of membrane-bound alkaline proteases from midgut tissue of the silkworm, *Bombyx mori*. J Biochem (Tokyo) 1985;97(5):1437–1445.

Faragher SG, Meek AD, Rice CM, Dalgarno L. Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. Virology 1988;163(2):509–526.

Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15(1):35–41.

Forsell K, Suomalainen M, Garoff H. Structure-function relation of the NH2-terminal domain of the Semliki Forest virus capsid protein. J Virol 1995;69(3):1556–1563.

Fuller SD, Berriman JA, Butcher SJ, Gowen BE. Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. Cell 1995; 81(5):715–725.

Garoff H, Huylebroeck D, Robinson A, Tillman U, Liljestrom P. The signal sequence of the p62 protein of Semliki Forest virus is involved in initiation but not in completing chain translocation. J Cell Biol 1990;111(3):867–876.

Geigenmuller-Gnirke U, Nitschko H, Schlesinger S. Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. J Virol 1993;67(3):1620–1626.

Hahn CS, Lustig S, Strauss EG, Strauss JH. Western equine encephalitis virus is a recombinant virus. Proc Natl Acad Sci U S A 1988;85(16):5997–6001.

Harrison SC, David A, Jumblatt J, Darnell JE. Lipid and protein organization in Sindbis virus. J Mol Biol 1971;60(3):533–538.

Harrison SC, Strong RK, Schlesinger S, Schlesinger MJ. Crystallization of Sindbis virus and its nucleocapsid. J Mol Biol 1992;226(1):277–280.

Johnson G, Moore SW. Cholinesterase-like catalytic antibodies: reaction with substrates and inhibitors. Mol Immunol 2000;37(12–13):707–719.

Johnson G, Moore SW. Catalytic antibodies with acetylcholinesterase activity. J Immunol Methods 2002;269(1–2):13–28.

Jung G, Ueno H, Hayashi R, Liao TH. Identification of the catalytic histidine residue participating in the charge-relay system of carboxypeptidase Y. Protein Sci 1995;4(11): 2433–2435.

Khaymina SS, Kenney JM, Schroeter MM, Chalovich JM. Fesselin is a natively unfolded protein. J Proteome Res 2007;6(9):3648–3654.

Kinney RM, Johnson BJ, Brown VL, Trent DW. Nucleotide sequence of the 26S mRNA of the virulent Trinidad donkey strain of Venezuelan equine encephalitis virus and deduced sequence of the encoded structural proteins. Virology 1986;152(2):400–413.

Kuhn, RJ. Togaviridae: the viruses and their replication. In Knipe DM, Howley PM, editors. Fields virology. Philadelphia, PA: Lippincott Williams & Wilkins; 2006.

Lee S, Owen KE, Choi HK, Lee H, Lu G, Wengler G, Brown DT, Rossmann MG, Kuhn RJ. Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. Structure 1996;4(5):531–541.

Levinson RS, Strauss JH, Strauss EG. Complete sequence of the genomic RNA of O'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. Virology 1990;175(1):110–123.

Li Y, Jing G. Double point mutant F34W/W140F of staphylococcal nuclease is in a molten globule state but highly competent to fold into a functional conformation. J Biochem (Tokyo) 2000;128(5):739–744.

Lottenberg R, Christensen U, Jackson CM, Coleman PL. Assay of coagulation proteases using peptide chromogenic and fluorogenic substrates. Methods Enzymol 1981;80(Pt C):341–361.

MacBeath G, Kast P, Hilvert D. Redesigning enzyme topology by directed evolution. Science 1998;279(5358):1958–1961.

Mao Q, Walde P, Luisi PL. Kinetic behaviour of alpha-chymotrypsin in reverse micelles. A stopped-flow study. Eur J Biochem 1992;208(1):165–170.

Martin J, Prieto I, Barbero JL, Perez-Gil J, Mancheno JM, Arche R. Thermodynamic profiles of penicillin G hydrolysis catalyzed by wild-type and Met—Ala168 mutant penicillin acylases from *Kluyvera citrophila*. Biochim Biophys Acta 1990;1037(2): 133–139.

Matthews RE. The classification and nomenclature of viruses. Summary of results of meetings of the International Committee on Taxonomy of Viruses in The Hague, September 1978. Intervirology 1979;11(3):133–135.

Melancon P, Garoff H. Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. J Virol 1987;61(5):1301–1309.

Miller ML, Brown DT. Morphogenesis of Sindbis virus in three subclones of *Aedes albopictus* (mosquito) cells. J Virol 1992;66(7):4180–4190.

Morillas M, Eberl H, Allain FH, Glockshuber R, Kuennemann E. Novel enzymatic activity derived from the Semliki Forest virus capsid protein. J Mol Biol 2008;376(3):721–735.

Morillas M, Goble ML, Virden R. The kinetics of acylation and deacylation of penicillin acylase from *Escherichia coli* ATCC 11105: evidence for lowered pKa values of groups near the catalytic centre. Biochem J 1999;338(Pt 1):235–239.

Naganathan AN, Doshi U, Munoz V. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. J Am Chem Soc 2007;129(17):5673–5682.

Neira JL, Contreras LM, de los Panos OR, Sanchez-Hidalgo M, Martinez-Bueno M, Maqueda M, Rico M. Mechanisms of protease action unfolded enterocin EJ97. Protein Eng Des 2010;23(7):507–518.

Nicola AV, Chen W, Helenius A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. Nat Cell Biol 1999;1(6):341–345.

Nohara D, Wakamatsu M, Goto M, Sakai T. Kinetic anomalies in chymotryptic hydrolyses of *p*-nitrophenyl acetate and N-benzoyl-L-alanine methyl ester. Chem Pharm Bull (Tokyo) 1989;37(7):1685–1690.

Odake S, Kam CM, Narasimhan L, Poe M, Blake JT, Krahenbuhl O, Tschopp J, Powers JC. Human and murine cytotoxic T lymphocyte serine proteases: subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. Biochemistry 1991;30(8):2217–2227.

Ohno M, Sato S, Karasaki Y, Tsukamoto S. Kinetic study of alpha-chymotrypsin catalysis with regard to the interaction between the specificity-determining site and the aromatic side chain of substrates. J Biochem (Tokyo) 1976;80(2):239–251.

Oliveberg M, Tan YJ, Fersht AR. Negative activation enthalpies in the kinetics of protein folding. Proc Natl Acad Sci U S A 1995;92(19):8926–8929.

Owen KE, Kuhn RJ. Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation. J Virol 1996;70(5):2757–2763.

Owen KE, Kuhn RJ. Alphavirus budding is dependent on the interaction between the nucleocapsid and hydrophobic amino acids on the cytoplasmic domain of the E2 envelope glycoprotein. Virology 1997;230(2):187–196.

Perl D, Jacob M, Bano M, Stupak M, Antalik M, Schmid FX. Thermodynamics of a diffusional protein folding reaction. Biophys Chem 2002;96(2–3):173–190.

Polgar L. Mechanisms of protease action. Boca Raton, FL: CRC-Press; 1989.

Porterfield JS, editors. Comparative and historical aspects of the Togaviridae and Flaviridae. New York: Plenum Press; 1986.

Powers JC, Kam CM. Peptide thioester substrates for serine peptidases and metalloendopeptidases. Methods Enzymol 1995;248:3–18.

Presley JF, Polo JM, Johnston RE, Brown DT. Proteolytic processing of the Sindbis virus membrane protein precursor PE2 is nonessential for growth in vertebrate cells but is required for efficient growth in invertebrate cells. J Virol 1991;65(4):1905–1909.

Rajender S, Han M, Lumry R. Studies of the chymotrypsinogen family of proteins. IX. Steady-state kinetics of the chymotryptic hydrolysis of N-acetyl-L-tryptophan ethyl ester at pH 8.0. J Am Chem Soc 1970;92(5):1379–1385.

Raju R, Huang HV. Analysis of Sindbis virus promoter recognition in vivo, using novel vectors with two subgenomic mRNA promoters. J Virol 1991;65(5):2501–2510.

Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Proteins 2006;62(1):24–45.

Roa A, Goble ML, Garcia JL, Acebal C, Virden R. Rapid burst kinetics in the hydrolysis of 4-nitrophenyl acetate by penicillin G acylase from Kluyvera citrophila. Effects of mutation F360V on rate constants for acylation andde-acylation. Biochem J 1996;316(Pt 2):409–412.

Roman LM, Garoff H. Alteration of the cytoplasmic domain of the membrane-spanning glycoprotein p62 of Semliki Forest virus does not affect its polar distribution in established lines of Madin-Darby canine kidney cells. J Cell Biol 1986;103(6Pt 2):2607–2618.

Rumenapf T, Strauss EG, Strauss JH. Aura virus is a New World representative of Sindbis-like viruses. Virology 1995;208(2):621–633.

Sanchez IE, Morillas M, Zobeley E, Kiefhaber T, Glockshuber R. Fast folding of the two-domain semliki forest virus capsid protein explains co-translational proteolytic activity. J Mol Biol 2004;338(1):159–167.

Sanyal G, Maren TH. Thermodynamics of carbonic anhydrase catalysis. A comparison between human isoenzymes B and C. J Biol Chem 1981;256(2):608–612.

Sato S, Karasaki Y, Ohno M. Interactions of alpha-chymotrypsin with peptides containing tryptophan or its derivatives at the C-terminus. J Biochem (Tokyo) 1977;82(1): 231–237.

Scalley ML, Baker D. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. Proc Natl Acad Sci U S A 1997;94(20):10636–10640.

Sefton BM. Immediate glycosylation of Sindbis virus membrane proteins. Cell 1977; 10(4):659–668.

Shirako Y, Niklasson B, Dalrymple JM, Strauss EG, Strauss JH. Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. Virology 1991;182(2): 753–764.

Sidorowicz W, Jackson GC, Behal FJ. Multiple molecular forms of human pancreas alanine aminopeptidase. Clin Chim Acta 1980;104(2):169–179.

Singh I, Helenius A. Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. J Virol 1992;66(12):7049–7058.

Skoging U, Liljestrom P. Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. J Mol Biol 1998;279(4):865–872.

Snoke, JE and Neurath, H. The effect of temperature on the esterase activity of chymotrypsin. J Biol Chem 1950;182:577–584.

Strauss EG, Rice CM, Strauss JH. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology 1984;13392–110.

Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. Microbiol Rev 1994;58(3):491–562.

Su ZD, Arooz MT, Chen HM, Gross CJ, Tsong TY. Least activation path for protein folding: investigation of staphylococcal nuclease folding by stopped-flow circular dichroism. Proc Natl Acad Sci U S A 1996;93(6):2539–2544.

Taylor A. Aminopeptidases: towards a mechanism of action. Trends Biochem Sci 1993; 18(5):167–171.

Toth J, Gombos L, Simon Z, Medveczky P, Szilagyi L, Graf L, Malnasi-Csizmadia A. Thermodynamic analysis reveals structural rearrangement during the acylation step in human trypsin 4 on 4-methylumbelliferyl 4-guanidinobenzoate substrate analogue. J Biol Chem 2006;281(18):12596–12602.

Truhlar D, Kohen A. Convex Arrhenius plots and their interpretation. Proc Natl Acad Sci U S A 2001;98(3):848–851.

Turk, V, editor. Proteases: new perspectives. Molecular and cell biology updates. Basel, Switzerland: Birkhäuser Verlag; 1999.

Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. Biochemistry 1999;38(45):15009–15016.

Valley MP, Fitzpatrick PF. Comparison of enzymatic and non-enzymatic nitroethane anion formation: thermodynamics and contribution of tunneling. J Am Chem Soc 2004;126(20):6244–6245.

Vamvaca K, Vogeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: efficient coupling of folding and catalysis. Proc Natl Acad Sci U S A 2004;101(35): 12860–12864.

Verheyden G, Matrai J, Volckaert G, Engelborghs Y. A fluorescence stopped-flow kinetic study of the conformational activation of alpha-chymotrypsin and several mutants. Protein Sci 2004;13(9):2533–2540.

Verheyden S, Sillen A, Gils A, Declerck PJ, Engelborghs Y. Tryptophan properties in fluorescence and functional stability of plasminogen activator inhibitor 1. Biophys J 2003;85(1):501–510.

von Bonsdorff CH, Harrison SC. Sindbis virus glycoproteins form a regular icosahedral surface lattice. J Virol 1975;16(1):141–145.

von Bonsdorff CH, Harrison SC. Hexagonal glycoprotein arrays from Sindbis virus membranes. J Virol 1978;28(2):578–583.

Wang CL, Calvo KC, Klapper MH. A lpha-chymotrypsin deacylation: temperature dependence of hydrolysis and transesterification reactions. Biochemistry 1981;20(5): 1401–1408.

Weiss B, Nitschko H, Ghattas I, Wright R, Schlesinger S. Evidence for specificity in the encapsidation of Sindbis virus RNAs. J Virol 1989;63(12):5310–5318.

Yu XW, Li YQ. Kinetics and thermodynamics of synthesis of propyl gallate by mycelium-bound tannase from *Aspergillus niger* in organic solvent. J Mol Catal B: Enzym 2006;40(1–2):44–50.

Zerovnik E, Janjic V, Francky A, Mozetic-Francky B. Equilibrium and transient intermediates in folding of human macrophage migration inhibitory factor. Eur J Biochem 1999;260(3):609–618.

Zerovnik E, Jerala R, Virden R, Kroon Zitko L, Turk V, Waltho JP. On the mechanism of human stefin B folding: II. Folding from GuHCl unfolded, TFE denatured, acid denatured, and acid intermediate states. Proteins 1998;32(3):304–313.

Zhao H, Lindqvist B, Garoff H, von Bonsdorff CH, Liljestrom P. A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. Embo J 1994;13(18):4204–4211.

14

CORE-LATIONS BETWEEN INTRINSIC DISORDER AND MULTIFACETED ACTIVITIES IN HEPATITIS C VIRUS AND RELATED VIRUSES

ROLAND IVANYI-NAGY, EVE-ISABELLE PÉCHEUR, AND JEAN-LUC DARLIX

14.1 THE FLAVIVIRIDAE FAMILY OF RNA VIRUSES

Members of the Flaviviridae family are RNA viruses with no DNA intermediate during the viral replication cycle. They have in common a nonsegmented, linear, single-stranded RNA genome of positive polarity, between 9.6 and 12.3 kb in length. Virus particles are enveloped and spherical, about 40–60 nm in diameter. All three genera of Flaviviridae—*Hepacivirus, Flavivirus*, and *Pestivirus*—contain pathogens of great medical and/or economic significance (Lindenbach et al., 2007).

The sole members of *Hepacivirus* are hepatitis C virus (HCV) and the closely related GB virus B (GBV-B), causing liver disease in humans and New World monkeys, respectively. Chronic HCV infection, with an estimated worldwide prevalence of 2–3%, entails substantial morbidity and mortality (Lavanchy, 2009). Long-term sequelae of infection, including liver cirrhosis and hepatocellular carcinoma, as well as various metabolic diseases, constitute a major healthcare and economic burden. At present, no protective vaccine is available against HCV, and standard treatment, consisting of a combination of pegylated interferon-α and ribavirin, is plagued by relatively poor efficacy and tolerability (Shimakami et al., 2009). However, a large number of small molecule antiviral compounds targeting viral enzymatic

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

activities are currently under clinical evaluation, with the promise of a specifically targeted antiviral therapy for hepatitis C (STAT-C) in the near future (Pereira and Jacobson, 2009). *Flaviviruses* comprise a large number of arthropod-borne human pathogens, including the West Nile virus (WNV), yellow fever virus (YFV), dengue viruses (DENV 1–4), and tick-borne encephalitis virus (TBEV). Owing to a complex interplay of factors, including climate change, increased international travel and urbanization, and reduced vector control coupled with the spread of vector resistance, numerous mosquito-borne viruses previously endemic only to tropical or subtropical regions are emerging or re-emerging in industrialized countries (Roehrig et al., 2002; Gould and Solomon, 2008; CDC, 2010). The members of the third genus, *Pestivirus*, are animal pathogens infecting economically important livestock (e.g., bovine viral diarrhea virus (BVDV) of cattle and classical swine fever virus (CSFV) of pigs) and wild ungulates.

Viruses in the Flaviviridae family infect a wide range of host organisms, target a variety of cells and tissues, and establish various types of infection patterns (acute vs chronic, from lethal to subclinical infections). Despite this variability, they share a similar genome organization, virion morphology, and basic replication strategy (Fig. 14.1; for recent reviews and references for the following section, refer to Bartenschlager et al. (2004), Lindenbach et al. (2007), and Moradpour et al. (2007)). The replication cycle starts with the attachment of viral particles to the host cell membrane through the recognition of specific receptors by the envelope protein(s). This step is followed by internalization of the receptor–virus complex via clathrin-dependent endocytosis. Owing to acidification in the endosome lumen, viral envelope proteins undergo structural rearrangements enabling viral membrane fusion. Subsequently, the nucleocapsid is released in the cytosol, and the genomic RNA is uncoated. Cap-dependent (flaviviruses) or cap-independent/IRES (internal ribosome entry site)-driven (hepaci- and pestiviruses) RNA translation results in the synthesis of a viral polyprotein precursor, which is co- and posttranslationally processed by cellular and viral proteases to yield the viral structural and nonstructural proteins. Similar to all positive-strand RNA viruses, the accumulating protein products remodel internal host cell membranes to serve as a scaffold for the viral replication complex. Altered endoplasmic reticulum (ER) membranes, termed *membranous web* for the HCV replication complex, possibly together with a subverted autophagy machinery, provide a protected environment for RNA replication that proceeds by a semiconservative and asymmetric manner through a negative-strand intermediate (replicative form (RF); Fig. 14.1). Virions are thought to form by budding into the ER and are released through the cellular secretory pathway. In the case of flaviviruses, immature and mature virions are released, together with slowly sedimenting haemagglutinins (SHAs) devoid of nucleocapsid. In contrast, only fully mature virions seem to be released in case of HCV.

14.2 THE CORE PROTEIN OF HEPATITIS C VIRUS

HCV core protein is released from the viral polyprotein precursor by two consecutive processing reactions carried out by ER-associated proteases. Cleavage by

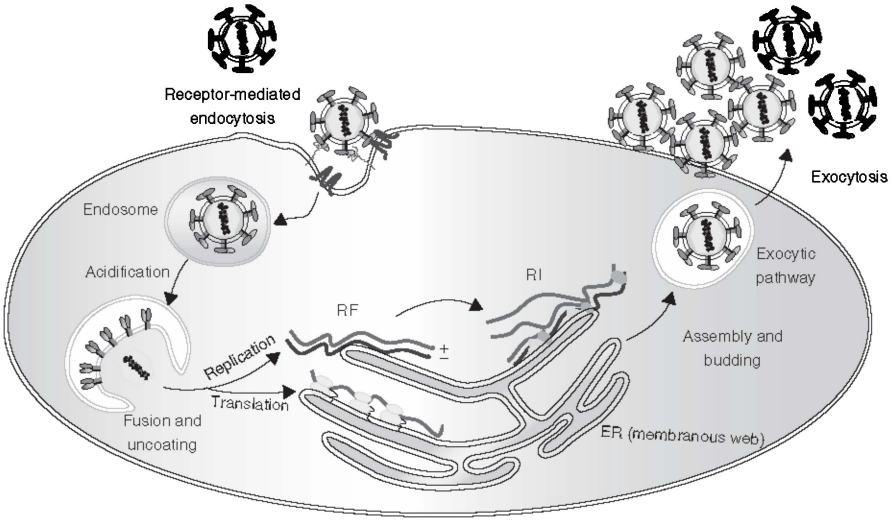


Figure 14.1 Schematic representation of the replicative cycle shared by viruses in the Flaviviridae family. For explanation, please refer to Section 14.1 of the main text. RF corresponds to the double-stranded replicative form of genomic RNA, while RI is the replicative intermediate generated during plus-strand RNA synthesis.

signal peptidase (SP) first releases a 191-residue form of core anchored to the ER membrane by its signal peptide (Hijikata et al., 1991; Santolini et al., 1994; Yasui et al., 1998), followed by a second maturation event at or near residue 177 (Okamoto et al., 2008) by the intramembrane-cleaving signal peptide peptidase (SPP), a presenilin-type aspartic protease (Lemberg and Martoglio, 2002; McLauchlan et al., 2002).

Mature HCV core protein consists of two functional domains (named D1 and D2), characterized by distinct amino acid contents and hydrophobicity profiles (Hope and McLauchlan, 2000; Boulant et al., 2005; Fig. 14.2). The N-terminal domain D1 (located approximately between residues 1 and 117) contains three highly basic amino acid clusters (BD1–BD3, Fig. 14.2). These basic regions mediate RNA binding (Santolini et al., 1994), promote RNA structural rearrangements (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006), and contain all the information necessary for nucleocapsid-like particle (NLP) formation in the presence of RNA (Kunkel et al., 2001; Majeau et al., 2004; Section 2.4). Besides RNA binding, the majority of mapped protein interaction sites are also localized to the N-terminal region (McLauchlan, 2000), indicating that D1 is a major organizer of the HCV infection network (Section 2.5). The C-terminal D2 domain, of hydrophobic character, serves as a membrane-binding module, mediating core protein attachment to ER membranes and, later in the replicative cycle, its recruitment to the putative assembly platform for viral particle formation, the surface of cellular lipid droplets (LDs) (Hope and McLauchlan, 2000; Miyanari et al., 2007; Shavinskaya et al., 2007; Section 2.2).

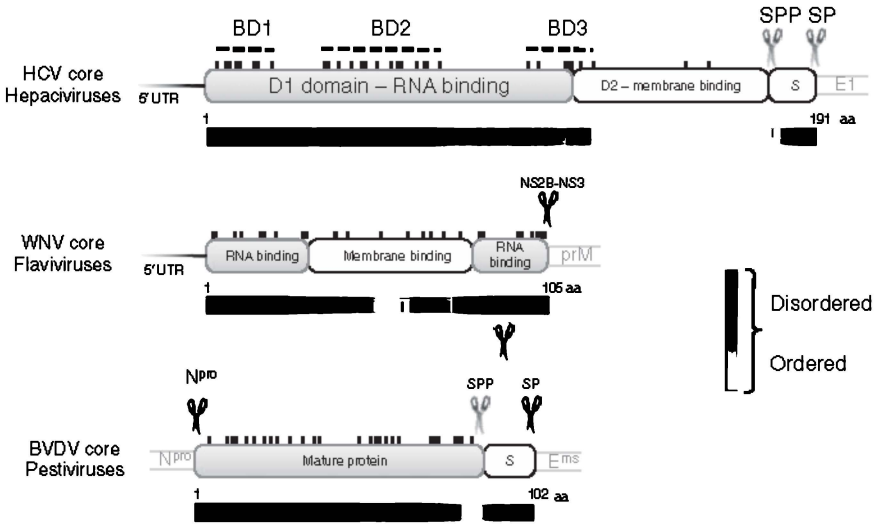


Figure 14.2 Domain organization and intrinsic disorder in Flaviviridae core proteins. Core proteins are released from the viral polyprotein precursor as a result of processing by viral and cellular proteases (SP; SPP; N^{pro}, N-terminal autoprotease; NS2B-NS3, nonstructural protein 2B-3). The mature hepacivirus (HCV) and flavivirus (WNV) core proteins contain RNA- and lipid-binding regions (shown in gray and white segments, respectively) in distinct arrangements, while the domain organization of pestivirus (BVDV) core proteins is currently unknown. Below the protein domains, predicted disorder is illustrated by a color-scale where dark gray/black segments correspond to disordered regions and light gray/white segments to ordered domains. Computer prediction of disordered regions was carried out using the DisProt VL3-H predictor (Obradovic et al., 2003). The position of basic amino acids is illustrated by black notches on top of the domain outlines. In HCV, they are clustered in three highly basic regions (BD1–BD3). As a general characteristic of Flaviviridae core proteins, RNA-binding regions are highly flexible and are enriched in basic residues.

14.2.1 HCV Core Protein as an Intrinsically Disordered Protein—Biophysical Characterization

The biophysical features of the N-terminal D1 domain of HCV core and its shorter segments have been extensively characterized by a variety of methods, demonstrating the lack of any stable secondary or tightly folded tertiary structure (reviewed in Ivanyi-Nagy and Darlix (2010)). D1 is hypersensitive to proteolytic digestion by either trypsin or chymotrypsin (Kunkel and Watowich, 2002; Duvignaud et al., 2009) and shows aberrant electrophoretic mobility on SDS gels (Duvignaud et al., 2009), features characteristic of intrinsically disordered proteins (IDPs) (Receveur-Brechot et al., 2006). In agreement with this, far-UV CD spectra of the N-terminal 124 (C124), 117 (C117), or 82 (C82) amino acids of core suggest a random coil-like conformation, with a pronounced ellipticity minimum in the spectrum observed at ~200 nm (Kunkel and Watowich, 2004; Boulant

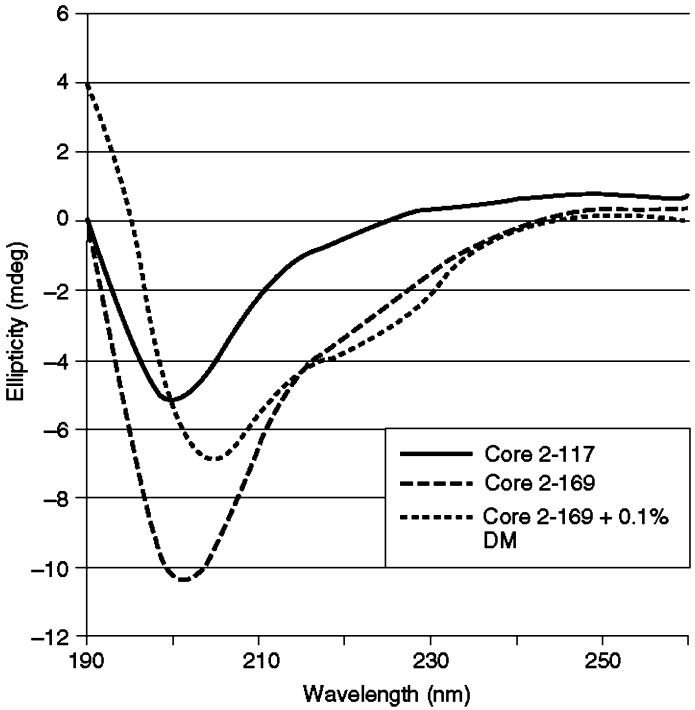


Figure 14.3 Far-UV CD spectra of HCV core protein. In the absence of detergents, both the D1 domain (core 2–117) and the full-length protein (core 2–169) are disordered, as indicated by the single ellipticity minimum at ~200 nm. In the presence of detergents, core protein adopts a partially α-helical conformation (core 2–169 + 0.1% DM). *Source:* Adapted from Ivanyi-Nagy et al. (2008).

et al., 2005; Ivanyi-Nagy et al., 2008; Duvignaud et al., 2009) (Fig. 14.3). NMR spectroscopy and chemical shift indexing further support the highly disordered nature of C82 (Duvignaud et al., 2009).

The structure of the D2 domain and how its presence influences the folding of the D1 region are still a matter of controversy. Secondary structures rich in β-sheets (Kunkel and Watowich, 2004), α-helices (Boulant et al., 2005), and a mostly unstructured conformation (Ivanyi-Nagy et al., 2008) have all been suggested for the full-length core protein (C179 or C169) based on far-UV CD spectroscopy. One confounding factor in these measurements is the formation of heterogeneous, stable homo-oligomeric aggregates of core protein (Kunkel and Watowich, 2004; Boulant et al., 2005). Indeed, equilibrium sedimentation analyses show that C179 forms large complexes with an average molecular mass of 500 kDa (~24–26 subunits) (Kunkel and Watowich, 2004), while C169 sediments mostly between 600 and 700 kDa (~30–35 subunits) (Boulant et al., 2005). Efficient solubilization of full-length core protein is dependent on the presence of mild detergents, such as DM (*n*-dodecyl β-D-maltoside), during its purification and subsequent handling,

suggesting that the proper folding (and function) of core protein necessitates membrane association (Boulant et al., 2005; Section 2.2; Fig. 14.3). Differences in the core protein constructs, their solubilization, the formation of unspecific aggregates, and varied buffer compositions used in these studies might all contribute to the contradictory CD results; thus, further research is required to clarify the structure of the full-length core protein.

Nevertheless, based on the available experimental data, a general model can be suggested, where an originally unstructured core protein undergoes a series of induced folding events upon its successive interactions with cellular membranes (Section 2.2), the viral RNA (Section 2.3), the viral envelope glycoproteins, and a variety of cellular proteins (Section 2.5). These individual interactions result only in partial structure formation (Boulant et al., 2005; Carmona and Molina, 2010), leading to the formation of "fuzzy complexes" (Tompa and Fuxreiter, 2008), where parts of core protein remain highly flexible and exposed, prone to binding further partners. The combination of dynamically changing structured and unstructured segments might permit the formation of a staggering number of various protein–protein, protein–RNA, and protein–lipid complexes, allowing a tight temporal and spatial regulation of the viral replicative cycle and host cell metabolism.

14.2.2 Membrane Binding by HCV Core

From its synthesis on the ER membrane to viral budding in concert with LDs, the HCV core protein is continuously associated with various intracellular membrane structures. Indeed, core protein trafficking, virion morphogenesis, and changes in the host cell lipid metabolism (ultimately leading to pathogenesis) are intimately interrelated (see Section 2.4 and recent reviews by McLauchlan (2009) and Piver et al. (2010)). Lipid association of core is mediated by its C-terminal D2 domain through the formation of two putative amphipathic α-helices between amino acids 119–136 and 148–164 interacting in plane with the LD phospholipid monolayer and connected by a flexible hydrophobic linker (Hope and McLauchlan, 2000; Boulant et al., 2006). However, conformational changes upon lipid binding most probably extend beyond the D2 region and lead to partial α-helical folding also in D1 (Boulant et al., 2005). Indeed, far-UV CD spectra of full-length core show an ~50% helical content for the protein in the presence of mild detergents mimicking a membrane environment (Boulant et al., 2005; Fig. 14.3). In agreement with an important role for lipid-induced folding, core protein mutants incapable of membrane association, which probably stay intrinsically disordered, are rapidly degraded by the proteasome (Boulant et al., 2006).

Core protein folding and membrane association might be further influenced by palmitoylation, a posttranslational modification that enhances surface hydrophobicity and membrane affinity of modified proteins (Greaves et al., 2009). Palmitoylation of core at cysteine 172 was found to be essential for infectious virion production (Majeau et al., 2009), although this finding may not hold true for all HCV sequences (Kopp et al., 2010).

14.2.3 HCV Core as an RNA-Binding and Chaperoning Protein

14.2.3.1 The "RNA Structural Code" of HCV Highly error-prone replication, caused by the lack of proof-reading activity in viral RNA-dependent RNA polymerases (RdRPs), puts RNA viruses under a constant selection pressure to keep their genome size small. Indeed, with close to one mutation/genome/replication cycle (Drake and Holland, 1999), longer RNA viruses are threatened by "error catastrophe", a point where mutational load reaches a threshold level, eventually leading to population extinction (Eigen, 1971; Biebricher and Eigen, 2005). An opposing selection pressure, demanding high variability and coding capacity, operates at the same time, endowing viruses with the flexibility to infect and replicate in host target cells and resist elimination by the host immune system or antiviral treatments.

Increased coding and regulatory capacity in the compact genome of HCV is achieved by two strategies: encoding protein products in multiple reading frames (Branch et al., 2005) and by the presence in the genome of a number of cis-acting RNA elements (CREs), RNA structures regulating various aspects of the viral life cycle. CREs in RNA viruses can either correspond to relatively stable RNA structures involved in viral or cellular protein binding (e.g., the signal for VPg uridylylation in poliovirus (Steil and Barton, 2009) or the encapsidation signal in hepatitis B virus (Kramvis and Kew, 1998)) or function as metastable structures, with two (or more) conformations possessing similar thermodynamic stability, thus providing a new layer of regulatory potential by switching between the alternative structures (e.g., RNA sequences regulating flavivirus genome cyclization (Villordo and Gamarnik, 2009) or retroviral RNA dimerization (Darlix et al., 1995)). In most cases, structural rearrangements are facilitated by RNA chaperone proteins, which destabilize RNA structures and thus help overcome the kinetic barrier hindering the interconversion of stable, alternative conformations (Herschlag, 1995). As a well-characterized example, the small nucleocapsid protein (NCp7) of HIV-1, which is generated by the viral protease-mediated cleavage of the Gag structural polyprotein, is endowed with nucleic-acid-binding and chaperoning properties. Binding of the nucleocapsid protein to viral RNA and DNA molecules has important consequences. First, it drives the structural rearrangement of viral nucleic acids into their most stable conformation in an energy landscape (Tompa and Csermely, 2004; Ivanyi-Nagy et al., 2005). Second, when two essential viral nucleic acid sequences with complementary regions, such as the TAR RNA-cTAR DNA sequences, are within the nucleocapsid-nucleic-acid complex, nucleocapsid drives their annealing allowing proviral DNA synthesis to continue (Cristofari and Darlix, 2002; Levin et al., 2005; Godet et al., 2006).

Similar to retroviruses, complex RNA–RNA interactions, RNA structural rearrangements, and protein binding to distinct RNA structures regulate various aspects of the HCV replicative cycle, from translation of viral proteins throughout the replication of the genomic RNA to (probably) encapsidation. The extent, structure, and regulation of this RNA structural code is as yet poorly explored, but recent advances have identified numerous CREs and long-range RNA–RNA interactions, providing the first glimpses to a highly sophisticated regulatory system.

14.2.3.1.1 RNA Structures in the Untranslated Regions of the HCV Genome The single, long open reading frame (ORF) of HCV coding for the viral structural and nonstructural proteins is flanked by highly structured and well-conserved untranslated regions (UTRs). The 5′ UTR folds into a complex secondary and tertiary structure (Kieft et al., 2002; Lukavsky et al., 2003), consisting of four stem loops (SLI–SLIV) and a pseudoknot (Fig. 14.4). This region contains overlapping signals for RNA replication and translation (Tsukiyama-Kohara et al., 1992; Friebe et al., 2001). While SLI and SLII are essential for replication (Friebe et al., 2001), the region between SLII and SLIV, together with the 5′ proximal part of the core-coding region, forms an IRES, mediating cap-independent translation for the synthesis of the viral polyprotein (Tsukiyama-Kohara et al., 1992; Fraser and Doudna, 2007). The 43S ribosomal complex binds directly to the IRES, without the requirement for the canonical translation initiation factors (Pestova et al., 1998). Direct positioning of the ribosomal P site to the AUG initiation codon leads to efficient ribosome recruitment without prior scanning of the 5′ highly structured UTR (Spahn et al., 2001; Otto and Puglisi, 2004).

The 3′ UTR has a tripartite organization, consisting of a ~40 nt long variable region, a poly(U/UC) tract, and a terminal highly conserved RNA structure of 98 nt, termed *X RNA* (Tanaka et al., 1995, 1996; Kolykhalov et al., 1996; Yamada

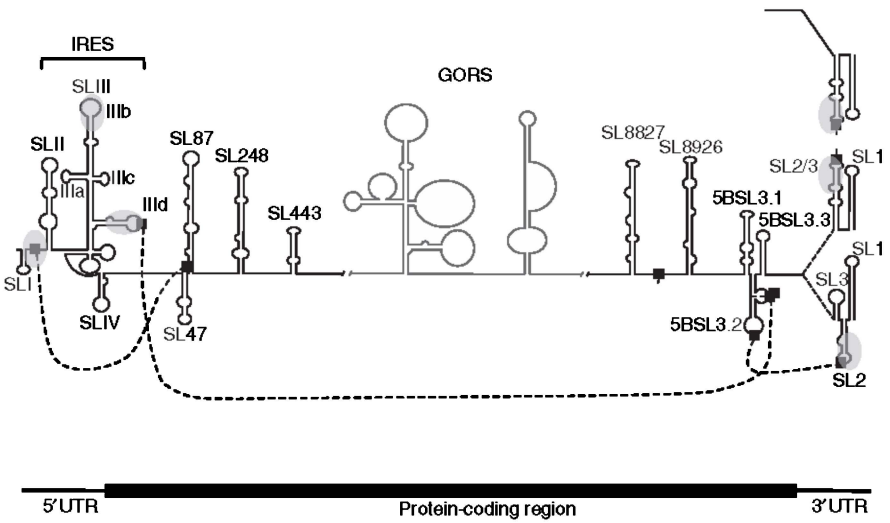


Figure 14.4 RNA structures and long-range interactions in the HCV genome. Schematic representation of RNA secondary structures in the 5′ UTR (SLI–SLIV), core-coding region (SL47–SL443), NS5B-coding region (SL8827–5BSL3.3), and 3′ UTR (SL3–SL1) of the HCV genomic RNA. Two alternative conformations, consisting of two or three stem-loop structures, are drawn for the 3′ UTR. In addition, the 3′ UTR might also dimerize through the X RNA region. Putative long-range RNA–RNA interactions are illustrated by dashed lines connecting the interaction sites (squares). High affinity core-protein-binding sites are indicated by shaded gray ovals.

et al., 1996). Both the linker pyrimidine-rich region and the X RNA sequences have been shown to be essential for HCV replication in cell culture and in chimpanzees (Yanagi et al., 1999; Kolykhalov et al., 2000; Friebe and Bartenschlager, 2002; Friebe et al., 2005; Yi and Lemon, 2003a,b). Although the mechanistic details of (−) strand RNA synthesis are poorly understood, it is believed that specific interaction of the RdRp NS5B with a promoter sequence in the X RNA region is involved in the initiation process (Cheng et al., 1999; Oh et al., 2000; Kim et al., 2002). Following a number of initially conflicting results (Fang and Moyer, 2000; Friebe and Bartenschlager, 2002; Kong and Sarnow, 2002; Imbert et al., 2003; Yi and Lemon, 2003a), an increasing consensus is pointing to a role for the 3′ UTR—and specifically the terminal SL1 structure—in stimulating the IRES-driven translation of the HCV genome (Ito et al., 1998; Ito and Lai, 1999; Michel et al., 2001; McCaffrey et al., 2002; Song et al., 2006; Lourenco et al., 2008; Weinlich et al., 2009; Bung et al., 2010). Enzymatic and chemical probings of the X RNA structure suggest the existence of a stable, long, terminal hairpin loop (SL1, Fig. 14.4) preceded by a region that may exist in multiple conformations, adopting either a single stem-loop or a double stem-loop structure (SL2 and SL3 or SL2/3 in Fig. 14.4) (Blight and Rice, 1997; Ito and Lai, 1997; Smith et al., 2002; Dutkiewicz and Ciesiolka, 2005; Ivanyi-Nagy et al., 2006).

14.2.3.1.2 RNA Structures in the Protein-Coding Region of the HCV Genome
In addition to the canonical RNA signals located in the UTRs, a number of conserved RNA structures have been identified in the core- and NS5B-encoding regions by a combination of bioinformatic analyses and structural probing (Smith and Simmonds, 1997; Tuplin et al., 2002, 2004). Of the four stem-loop structures, SL47, SL87, SL248, and SL443 (Fig. 14.4), located in the core-coding sequence, disruption of base pairing in SL47 and SL87 proved to be detrimental for HCV replication, probably owing to a reduced translation efficiency in their absence (McMullan et al., 2007; Vassilaki et al., 2008). In addition, the integrity of the 5BSL3.2 structure in the NS5B-coding region correlated with replication efficiency (You et al., 2004; Friebe et al., 2005; Diviney et al., 2008), while substitutions in SL8926 led to milder defects, resulting in reduced colony formation in a subgenomic replicon model system (Lee et al., 2004).

The mechanism by which these RNA structures regulate various aspects of the HCV replicative cycle is incompletely understood. Recruitment of viral or cellular proteins necessary for translation and/or replication might be one function associated with stable CREs. On the other hand, a number of long-range RNA–RNA interactions have been suggested to take place during HCV replication. Base pairing between SL87 in the core-coding region and an unpaired region in the 5′ UTR (between SLI and SLII, Fig. 14.4) might be involved in translation regulation (Honda et al., 1999; Kim et al., 2003; Beguiristain et al., 2005; Diaz-Toledano et al., 2009). Interestingly, the same UTR region serves as the binding site for microRNA-122, a cellular cofactor required for infectious HCV production (Jopling et al., 2005, 2008; Jangra et al., 2010), suggesting that RNA structural rearrangements upon miRNA binding might be involved in the regulation of HCV translation and/or

replication (McMullan et al., 2007; Diaz-Toledano et al., 2009). The 5BSL3.2 CRE, situated in the NS5B-coding region, serves as another metastable structural element, potentially orchestrating a complex regulatory system by base pairing with the IIId loop of the HCV IRES (Romero-Lopez and Berzal-Herranz, 2009), with an unstructured segment in the NS5B region (Diviney et al., 2008), or with the SL2 stem loop in the 3′ UTR (Friebe et al., 2005; You and Rice, 2008) (Fig. 14.4).

Besides these relatively stable RNA elements, the HCV genome is characterized by the presence of extensive base pairing throughout the polyprotein coding region (Simmonds et al., 2004). The existence of this genome-scale ordered RNA structure (GORS) is supported by large-scale thermodynamic predictions, by the relative inaccessibility of HCV genomic RNA to hybridization with antisense probes, and by the compact spheroid shape of the RNA molecule when visualized by atomic force microscopy (Simmonds et al., 2004; Davis et al., 2008). In contrast to the stable RNA structural elements involved in translation and genome replication, GORS was proposed to be highly dynamic, being based on transient, low stability interactions and thus easily accommodating the unwinding and exposure of desired regions during the viral replicative cycle (Simmonds et al., 2004). The significance of GORS in HCV remains elusive. A possible function in viral persistence, by contributing to the evasion of the intracellular defense mechanisms triggered by double-stranded RNA molecules, awaits experimental verification (Simmonds et al., 2004; Davis et al., 2008).

14.2.3.2 RNA Binding and Chaperoning by HCV Core Protein As a viral nucleocapsid protein, HCV core binds to, protects, and encapsidates the viral genomic RNA during virion morphogenesis. At present, the RNA signal(s) responsible for the specific packaging of the positive-strand genomic RNA are unknown. Indeed, the core protein is able to bind *in vitro* to a variety of RNA, single-stranded DNA, and double-stranded DNA substrates (Santolini et al., 1994; Cristofari et al., 2004), with a marked preference for G-C-rich sequences (Tanaka et al., 2000). Owing to the importance of the RNA structures present in the viral UTRs, most studies to date concentrated on the interaction of core protein with these regions. Core protein binding to the 5′ UTR, with variable specificity against competitor RNAs, was monitored *in vitro* by Northwestern blot analysis (Hwang et al., 1995), gel mobility shift assays (Fan et al., 1999; Wang et al., 2000; Li et al., 2003), and surface plasmon resonance (SPR) (Tanaka et al., 2000); and *ex vivo* by immunoprecipitation of RNA–core complexes (Shimoike et al., 1999). Binding to short oligonucleotides corresponding to segments of the HCV IRES identified loop IIId, SLI, and a single-stranded region between SLI and SLII as the highest affinity binding sites for HCV core (Tanaka et al., 2000) (Fig. 14.4). In addition, core also binds with high affinity to the X RNA region of the 3′ UTR *in vitro* (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006; Shetty et al., 2010). In agreement with these studies, a reporter gene assay system (Yu et al., 2009) identified SLIIIb and SLIIId in the 5′ UTR and SL2 in the 3′ UTR as specific interaction sites for core (Fig. 14.4). It is important to emphasize that, in most cases, study designs focused on the UTRs, precluding the identification of high affinity sites in the polyprotein

coding region, including (probably) the elusive encapsidation signal of the HCV genomic RNA.

Core protein binding to RNA not only leads to genome encapsidation but also facilitates structural rearrangements in the target structure (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006, 2008; Sharma et al., 2010; Shetty et al., 2010). During their folding process, RNA molecules are prone to adopt (meta)stable, but non-functional, conformations and can easily become trapped in these local minima of the free-energy landscape (Herschlag, 1995; Pan et al., 1997). Alternatively, two (or more) metastable conformations of the same RNA region, separated by a high kinetic barrier, might serve as a molecular switch regulating various aspects of cellular or viral RNA function. Protein cofactors that help RNA molecules reach their most stable conformation, by either resolving kinetically trapped RNA structures or facilitating structural switches, are termed *RNA chaperones* (Herschlag et al., 1994; Herschlag, 1995; reviewed in Cristofari and Darlix (2002) and Schroeder et al. (2004)). RNA chaperone proteins bind nucleic acids with a broad sequence specificity and assist in their folding without ATP consumption (Cristofari and Darlix, 2002). They are ubiquitous in living organisms, orchestrating cellular RNA metabolism from translation through RNA transport and localization to RNA degradation (Cristofari and Darlix, 2002). RNA viruses from diverse families also take advantage of the versatility provided by RNA chaperones in translating their "RNA structural code" into functional, adaptive replication strategies (reviewed in Ivanyi-Nagy et al. (2005) and Zuniga et al. (2009)). The best characterized example is the small nucleocapsid protein of human immunodeficiency virus type 1 (HIV-1) and of related retroviruses and retrotransposons, which facilitates a number of essential RNA–RNA interaction reactions during viral replication, including the annealing of the primer tRNA onto the primer-binding site, strand-transfer reactions during the reverse transcription reaction, and dimerization of retroviral RNA genomes upon encapsidation (reviewed in Darlix et al. (1995), Bampi et al. (2004), and Darlix et al. (2007)). Nucleocapsid proteins in hantaviruses (Bunyaviridae family) and coronaviruses (Coronaviridae family) also play essential roles in chaperoning viral replication, facilitating genome circularization (Mir and Panganiban, 2006) and template switching during RNA transcription (Zuniga et al., 2010), respectively.

A potent RNA chaperone activity for HCV core protein has been demonstrated by a variety of well-established *in vitro* assays (reviewed in Cristofari and Darlix (2002) and Rajkowitsch et al. (2005)). Core protein binding augments DNA- and RNA-annealing kinetics by at least three orders of magnitude (Cristofari et al., 2004; Sharma et al., 2010), facilitates strand-exchange reactions (Cristofari et al., 2004), and promotes the catalytic cleavage of substrates by hammerhead ribozymes (Ivanyi-Nagy et al., 2008), demonstrating bona fide chaperone activity. Interestingly, at equal protein/nucleic acid molar ratios, core was found to be even more active than the prototypic RNA chaperone NCp7 of HIV (Sharma et al., 2010).

What are the functional targets of core RNA chaperone activity in the HCV genomic RNA? In agreement with the specific association of core protein with the 5′ UTR of the HCV genome, several reports described an effect of core on IRES-mediated translation. Transient (over)expression of core protein in various cell lines

has been shown to downregulate IRES-driven reporter construct expression using either mono- or bicistronic RNA constructs (Shimoike et al., 1999, 2006; Zhang et al., 2002; Li et al., 2003). In contrast, low levels of core protein, either added in trans *in vitro* or expressed intracellularly, were found to stimulate IRES activity four- to fivefold, with a reduction in stimulation at higher core ratios (Boni et al., 2005; Lourenco et al., 2008). These seemingly contradictory findings might be reconciled by the existence of a dual, dose-dependent mechanism of action for IRES regulation. At low core protein levels, RNA chaperoning may facilitate the correct folding of the IRES structure or modulate its long-range interaction(s) with other regulatory regions in the HCV genome (Kim et al., 2003; Beguiristain et al., 2005; Romero-Lopez and Berzal-Herranz, 2009). This, in turn, could lead to high translational efficiency and rapid accumulation of the viral nonstructural proteins required for replication. Later in the replicative cycle, high levels of newly synthesized core might compete with the 40S ribosomal subunit for binding to the SLIIId loop of the IRES (Zhang et al., 2002) and/or modulate long-range RNA–RNA interactions, resulting in the downregulation of translation and redirecting the RNA genome for replication and encapsidation. Indeed, although the genomic RNA serves as a common template for translation and replication in positive-strand RNA viruses, the two processes cannot occur simultaneously (Gamarnik and Andino, 1998), necessitating fine-tuned regulation and timing. Another possible advantage of curbing translation might be to keep the viral load in infected cells at a relatively low level, facilitating the establishment of persistent infection (Boni et al., 2005; Shimoike et al., 2006).

Besides the IRES region, another high affinity core-protein-binding site in the genomic RNA was identified in the SL2 region of the 3′ UTR (Yu et al., 2009; Fig. 14.4). Core protein binding to a 16-nt long palindromic sequence (named *DLS* (dimer linkage sequence)) in SL2 triggered the dimerization of the 3′ UTR *in vitro* (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006; Shetty et al., 2010), similar to the action of the retroviral RNA chaperone NCp7 on the HIV-1 genome (Darlix et al., 1995). HCV RNA dimerization, if it occurs *in vivo*, could constitute a mechanism regulating (−) strand HCV RNA synthesis via conformational changes in the promoter region or could trigger a switch between viral RNA replication/translation or translation/packaging.

14.2.3.3 Intrinsic Disorder and RNA Chaperoning RNA chaperoning by HCV core does not require a folded protein before its contact with the RNA. The well-characterized, intrinsically disordered D1 domain and its shorter fragments containing two or three of the basic amino acid clusters were found to be as competent in *in vitro* chaperone assays as the full-length core protein (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006). In addition, boiling (heat denaturation) of core before nucleic acid binding does not significantly diminish its activity (Ivanyi-Nagy et al., 2008), a peculiar behavior characteristic of IDPs (Weinreb et al., 1996; Kim et al., 2000; Receveur-Brechot et al., 2006).

The presence of long intrinsically unstructured protein regions might be a hallmark of RNA chaperones (Tompa and Csermely, 2004; Ivanyi-Nagy et al., 2005).

Indeed, among all the functional protein classes examined so far, RNA chaperones contain the highest average level of intrinsic disorder (Tompa and Csermely, 2004). On the basis of bioinformatic predictions on a well-characterized subset of RNA chaperones, Tompa and Csermely found that more than half of their residues fall into disordered regions, a ratio not attained even by regulatory or signaling-associated proteins (Tompa and Csermely, 2004). A number of RNA chaperone proteins exert their function in the absence of a well-defined 3D structure. Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) promotes RNA and DNA annealing via its disordered C-terminal glycine-rich region (Kumar and Wilson, 1990; Pontius and Berg, 1990), while its N-terminal folded RNA-binding domain provides sequence specificity and binds RNA in a non-cooperative manner (Buvoli et al., 1990; Casas-Finet et al., 1993). Similarly, chaperone activity is associated with the disordered N-terminal region in the cellular prion protein PrP^C (Gabus et al., 2001a, 2001b). Disorder prediction was also successfully employed to identify the active chaperone region in the Gag-like structural polyprotein of the Gypsy retroelement (Gabus et al., 2006).

On the basis of the "entropy exchange model" of chaperone function suggested by Tompa and Csermely (Tompa and Csermely, 2004), the high flexibility of chaperone proteins would provide a number of advantages in facilitating broad-specificity RNA binding and disrupting non-native bonds. Disordered regions may provide a platform for reversible interactions with a variety of structurally diverse substrates (Kriwacki et al., 1996). Most importantly, a reciprocal entropy transfer between the chaperone protein and its target was suggested (Tompa and Csermely, 2004). According to this model, RNA chaperones first contact RNA in a partly unstructured state. Induced folding (disorder-to-order transition) of the protein is accompanied by partial melting (order-to-disorder transition) of the substrate RNA in an entropy exchange process, followed by the dissociation of the complex. At each successive cycle of entropy transfer, the RNA resumes a conformational search, eventually achieving its thermodynamically most stable conformation (Tompa and Csermely, 2004).

Although the entropy exchange model provides a plausible and appealing theoretical framework, the proposed mechanism of action remains to be experimentally verified. Results obtained on structural changes in HCV core protein and its target RNA support a "mutual induced folding" scenario, in agreement with the model's predictions. Fourier transform infrared (FTIR) spectroscopy indicates that the binding of the intrinsically unstructured D1 domain of core protein to its specific target (the SLIIId subdomain of the HCV IRES) leads to β-sheet formation (~22%) in core, although the majority of the protein remains unstructured (Carmona and Molina, 2010). Interpretation of these results is complicated by the fact that binding to structured RNA molecules is coupled to core protein oligomerization (Kunkel et al., 2001; Kunkel and Watowich, 2002), which in itself is accompanied by limited β-sheet formation (Rodriguez-Casado et al., 2007). Nevertheless, homo-oligomerization in the absence of RNA affects only ~10% of the protein structure, suggesting that RNA binding induces further disorder-to-order transitions in core protein (Rodriguez-Casado et al., 2007; Carmona and Molina, 2010).

14.2.4 HCV Core in Particle Assembly—Structure of the Viral Nucleocapsid and Virion

14.2.4.1 Structure of the HCV Nucleocapsid There is currently no high resolution structure of the HCV nucleocapsid available. However, several groups have performed *in vitro* reconstitution experiments with various efficacies, using recombinant HCV core and cell-free systems (Klein et al., 2004), or core mixed with structured RNA (Kunkel et al., 2001; Carmona and Molina, 2010). These systems led to the assembly of NLPs, forming electron-dense objects with a size between 35 and 60 nm in diameter (Kunkel et al., 2001; Klein et al., 2004; Carmona and Molina, 2010). Interestingly, the presence of the N-terminal disordered D1 domain of core was found to be important for proper assembly of these NLPs (Kunkel et al., 2001; Klein et al., 2004, 2005; Majeau et al., 2004), as it was for proper RNA chaperoning activity (Cristofari et al., 2004). This further emphasizes the major role played by intrinsically disordered regions of the HCV core protein in viral particle formation and RNA structural rearrangements during viral replication.

14.2.4.2 Structure of the HCV Virion HCV infects only humans, and its transmission does not involve insect vectors, which sets this virus apart from most other members of the Flaviviridae family. Another peculiarity of HCV is the exceptionally low density of the virus particles, resulting from the association of the virus with serum lipoproteins (Thomssen et al., 1992). Indeed, the majority of HCV circulating in blood was found to be associated with β-lipoproteins, VLDLs (very low density lipoproteins), and LDLs (low density lipoproteins) (Thomssen et al., 1992; Andre et al., 2005; Nielsen et al., 2006). A final peculiarity of HCV is that it targets the liver, and among liver cells, HCV infects hepatocytes, the polarized cells where the virus replicates. Recent data obtained in a chimeric mouse model with humanized liver suggest that HCV infection would not solely rely on the presence of human hepatocytes per se, but would also depend on sufficient levels of human lipoproteins (Steenbergen et al., 2010). This, together with previous observations (Nielsen et al., 2006), would support the notion that HCV association with lipoproteins could occur in hepatocytes as part of the viral assembly/release process (for a recent review see Jones and McLauchlan (2010)).

The presence of HCV RNA has been observed in other human tissues and cells, including B and T lymphocytes, monocytes, and dendritic cells. In spite of these observations, no HCV replication could be detected in these cells. It is thus very likely that blood cells do not support the full HCV infection cycle (Marukian et al., 2008). It has been recently suggested that B cells could constitute a reservoir for HCV, where antiviral immune responses would not be triggered (Ito et al., 2010); this in turn would help the virus to persist and be transmitted to the liver (Stamataki et al., 2009).

The structure of the virion has been investigated using HCV isolated from the serum of infected patients, or more recently using HCV grown in cell culture (HCVcc). The virus has been visualized by conventional techniques in transmission electron microscopy, revealing viral particles of approximately 60 nm in diameter (Kaito et al., 1994; Li et al., 1995; Prince et al., 1996; Wakita et al., 2005) and the

presence of envelope glycoproteins on their surface (Wakita et al., 2005; Kaito et al., 2006). Obviously, the morphology and biophysical properties of HCV particles are modified and eventually determined by their propensity to associate with cellular lipids and lipoproteins.

14.2.4.2.1 Serum HCV Fractionation of HCV-positive serum upon isopycnic density gradients revealed three major pools of low ($d < 1.10$ g/mL), intermediate ($1.10 < d < 1.14$ g/mL), and high ($d > 1.15$ g/mL) density particles. Interestingly, it was shown that viral infectivity inversely correlated with density. Low density viral particles contain HCV RNA and are consistently associated to triglyceride-rich lipids; are positive for apolipoproteins B, C1, C2, C3, and E; and also for HCV structural proteins core, E1 and E2 (Andre et al., 2002; Nielsen et al., 2006). They display an irregular and globular shape, and their size is ~100 nm (Andre et al., 2002) (Fig. 14.5). The overall structure of these virions, called *lipo-viro-particles* (LVPs), is therefore very reminiscent of that of serum lipoproteins of low and very low densities (Andre et al., 2005). Delipidation results in structures resembling nucleocapsids, recognized by specific anti-core antibodies (Takahashi et al., 1992; Andre et al., 2002). Particles of high densities displayed an electron-dense and compact structure of ~35 nm and were found to be positive for HCV core and RNA (Takahashi et al., 1992; Kanto et al., 1995). Consequently these particles, present in the serum of infected patients, were considered as naked HCV nucleocapsids (Maillard et al., 2001) (Fig. 14.5).

14.2.4.2.2 Cell Culture Grown HCV (HCVcc) Density fractionation of HCVcc obtained from cell culture supernatants revealed a similar tripartite profile, with low density fractions displaying the highest specific infectivity (Lindenbach et al., 2005; Haid et al., 2009). Chisari and coworkers recently visualized the structures present in each density pool and revealed a very heterogeneous pattern between and within fractions (Gastaminza et al., 2010). Indeed, low density viral particles displayed at least three types of morphologies: enveloped by a lipid bilayer, enveloped by a lipid monolayer, or exosome-like structures. All these particles were positive for HCV E2 and RNA, as well as for apolipoproteins E and C1 (Chang et al., 2007; Meunier et al., 2008). Fractions of intermediate densities were mainly composed of bilayer-enveloped particles of 60 nm, while high density particles appeared like nonenveloped structures of 45 nm in size, very reminiscent of nucleocapsids. These structures contained HCV RNA but were not infectious. This ultrastructural study was performed using cryo-electron microscopy, enabling high resolution imaging. The 45-nm non-enveloped structures identified by Chisari and coworkers did not display any discernible spatial symmetry, and when these structures were observed delimited by a lipid envelope in low or intermediate HCVcc density fractions, they were not tightly associated to the viral envelope (Gastaminza et al., 2010).

A common point shared by all HCVcc structures was the absence of visible projections or spikes at the virion surface, conferring an overall smooth appearance to the virions (Gastaminza et al., 2010). This seems to be an intrinsic property of HCV envelope glycoproteins since we also reported a similar "bald" aspect of

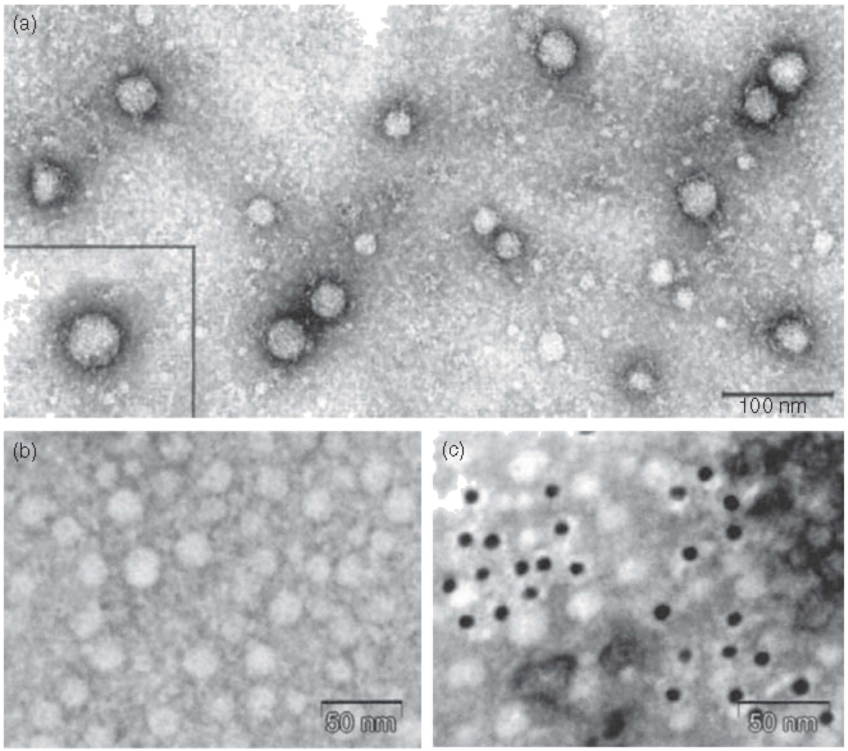


Figure 14.5 Electron microscopic images of HCV particles. (a) Naked nucleocapsids from HCV-infected serum. *Source:* Reprint from Maillard et al. (2001). Direct staining of virus particles isolated from serum by affinity chromatography, with 1% uranyl acetate. Inset, viral particle, 54 nm in diameter. *Source:* Reprint from Maillard et al. (2001). (b,c) Immunodetection of purified and delipidated lipo-viro-particles. Immunoelectron microscopy of Tween-80-treated LVP with a control antibody (b) or an anti-HCV core protein monoclonal antibody (c) and a 10-nm gold-labeled secondary antibody. The binding of core-protein-specific antibody on the particles can be seen. *Source:* Reprint from Andre et al. (2002).

HCV pseudotyped particles, harboring HCV E1 and E2 at their surface (Bonnafous et al., 2010). Also, these viral particles were consistently devoid of apolipoprotein B (Huang et al., 2007; Owen et al., 2009), suggesting that human hepatoma Huh-7 cells, used for the production of HCVcc, display defects in lipid metabolism that could be prejudicial to proper or complete HCV assembly (Icard et al., 2009; Pecheur et al., 2010).

14.2.4.3 Similarities and Differences between HCV and Other Flaviviridae Virion Structures Flavivirus virions analyzed to date by high resolution (cryo)-electron microscopy displayed a smooth particle surface, showing no membrane projections or spikes; this is the case, for example, for the dengue viruses and WNV (Mukhopadhyay et al., 2003; Zhang et al., 2003). Likewise, Gastaminza

et al. described HCVcc as ~60-nm particles with a spikeless outer membrane (Gastaminza et al., 2010). This feature seems to be a hallmark of the viral family and could be intrinsic to the peculiar folding of envelope glycoprotein(s) on the virion surface. However, as described in detail above, HCV also exists in other forms, enveloped or not by a lipid bilayer or a monolayer. Importantly, HCV is found associated to serum lipoproteins; models based on the latest studies of HCV assembly report the concomitant assembly of VLDL and viral particles to form the LVP in the ER lumen (Fig. 14.6). Such a behavior has not been reported for any other virus in the Flaviviridae family.

All Flaviviridae viruses analyzed to date displayed an icosahedral symmetry of the viral particle (Kuhn et al., 2002; Pokidysheva et al., 2006; Lindenbach et al., 2007). This could not be established concerning HCV, for which no symmetry could be attributed to any viral population at present (Gastaminza et al., 2010).

The HCV assembly pathway uses LDs as platforms where HCV core is enriched and serves as nucleation points for particle formation (McLauchlan et al., 2002; Miyanari et al., 2007; Shavinskaya et al., 2007) (Fig. 14.6). This peculiar behavior has not been shown in other Flaviviridae members, although a recent study reported that the capsid protein of dengue virus accumulates on the surface of LDs, suggesting a link between LD metabolism and dengue virus replication (Samsa et al., 2009).

Moreover, the process of HCV assembly and release readily gives rise to fully mature particles, whereas some Flaviviridae members are released as immature and mature particles (e.g., for dengue virus (Rodenhuis-Zybert et al., 2010)). In addition to mature virions, smaller, noninfectious particles can be released. These small (~14 nm in diameter) particles are termed slowly sedimenting haemagglutinin (SHA) because they can agglutinate red blood cells at low pH. Cells expressing the prM and E proteins alone can produce a related type of particle, the recombinant subviral particle (RSP, ~30-nm diameter) (Lindenbach et al., 2007). SHA and RSP are not observed in the life cycle of HCV.

14.2.5 Flexibility of HCV Core and the Viral Infection Network

Apart from coordinating genomic RNA packaging and virion formation, the HCV core protein plays pivotal roles within the host cell in the adaptation of the virus (possibly leading to persistent infection) and in the establishment of a variety of serious pathological manifestations. Indeed, expression of HCV core protein in transgenic mice leads to progressive steatosis, insulin resistance, and hepatocellular carcinoma, the hallmarks of chronic HCV infection in humans (Moriya et al., 1998; Lerat et al., 2002; Shintani et al., 2004). Core-mediated pathogenesis is the result of the perturbation of varied aspects of cellular homeostasis, including major changes in the regulation of gene transcription, signal transduction, apoptosis, and cellular lipid metabolism (reviewed in McLauchlan (2000) and Ray and Ray (2001)). This widespread effect is achieved by promiscuous interactions with a large number of cellular proteins. Systematic yeast two-hybrid screens, co-immunoprecipitation experiments, and extensive literature mining have identified close to 100 cellular

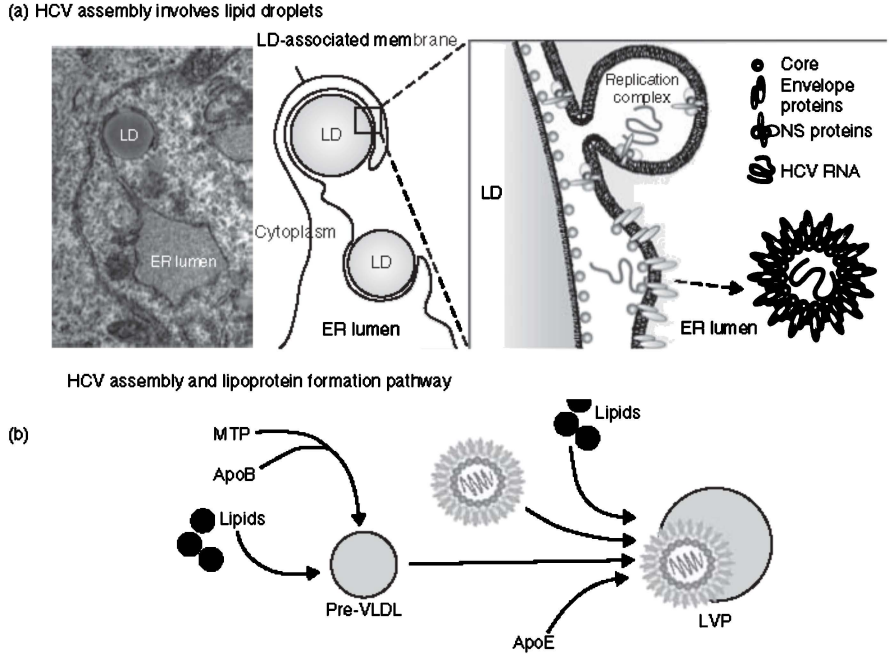


Figure 14.6 Model of HCV particle assembly. Among the Flaviviridae members, HCV has a peculiar assembly process. (a) HCV assembly involves lipid droplets (LDs). (A) Transmission electron micrograph of human hepatoma Huh-7 cells, showing an LD closely surrounded by endoplasmic reticulum (ER) (Pécheur EI, Jammart B, Durantel D, unpublished observation). (B) Schematic representation of a similar organization. (C) Model for the assembly of HCV viral particles: core is associated to the LD monolayer; HCV glycoproteins E1 and E2 and other proteins of the replication complex are set in the ER membrane; and viral RNA is brought to the site of assembly from its cytosolic location in the narrow space between LD and ER membranes. *Source:* Adapted with permission from Macmillan Publishers Ltd: Miyanari Y et al. The lipid droplet is an important organelle for hepatitis C virus production. Nat Cell Biol. Copyright (2007). (b) Proposed model for HCV assembly and lipoprotein formation pathway. In the ER lumen, the HCV virion is loaded onto an apo(lipoprotein)-B-positive pre-VLDL filled with lipids by the microsomal triglyceride transfer protein (MTP) to give rise to the lipo-viro-particle (LVP), which also contains the exchangeable apo(lipoprotein) E.

proteins directly interacting with HCV core (de Chassey et al., 2008; Navratil et al., 2009), among them important regulators of the JAK/STAT, insulin, and TGF-β signaling pathways; transcription factors; and effectors of apoptosis and cell cycle control. The majority of characterized interactions are mediated by the intrinsically unstructured D1 domain of the core protein (McLauchlan, 2000).

Intrinsic disorder and the ensuing large accessible surface area are advantageous to highly connected (hub) proteins by allowing high specificity/low affinity interactions with diverse, structurally heterogeneous substrates (Kriwacki et al., 1996; Dunker et al., 2002). Indeed, studies of eukaryotic interactomes show that hub

proteins, on average, contain higher levels of disorder than proteins with fewer interaction partners (Dunker et al., 2005; Dosztanyi et al., 2006; Haynes et al., 2006). Interestingly, the binding partners of core protein are also frequently unstructured (data not shown) and serve as hub proteins or as bottlenecks in cellular regulatory networks (de Chassey et al., 2008; Dyer et al., 2008). Thus, the flexible D1 domain of core may serve as a date hub, providing a platform to orchestrate a plethora of protein complexes dependent on the ever-changing requirements in the cellular environment.

14.3 STRUCTURE AND DISORDER IN THE CORE PROTEINS OF FLAVI- AND PESTIVIRUSES

Encoding a small, highly basic RNA-binding protein (core or capsid protein) at the N-terminal region of the viral polyprotein is a general feature in the three Flaviviridae genera, namely, *Hepacivirus, Flavivirus*, and *Pestivirus*. All core proteins bind RNA with a broad sequence specificity (Santolini et al., 1994; Cristofari et al., 2004; Ivanyi-Nagy et al., 2008; Murray et al., 2008), possess RNA chaperone activities *in vitro* (Cristofari et al., 2004; Ivanyi-Nagy et al., 2006, 2008), and are responsible for the condensation and packaging of the viral genomic RNA during virion morphogenesis. Despite these functional similarities, core proteins from the three Flaviviridae genera do not share significant sequence homology and exhibit genus-specific, diverse structural features and domain organization (recently reviewed in Ivanyi-Nagy and Darlix (2010), Fig. 14.2).

Flavivirus core proteins contain an internal hydrophobic region that is functionally analogous to the D2 domain of HCV core and is involved in homo-oligomerization (Wang et al., 2004), membrane binding (Markoff et al., 1997), and localization of the protein to LDs (Samsa et al., 2009). Flanking this hydrophobic domain, two short basic regions at the extremities of the protein are involved in RNA binding (Khromykh and Westaway, 1996), RNA chaperoning (Ivanyi-Nagy et al., 2008), and genome packaging (Patkar et al., 2007). In marked contrast to the HCV core, flavivirus core proteins exhibit extensive stable secondary and tertiary structures (Dokland et al., 2004; Ma et al., 2004), corresponding to a unique fold composed of four (or in some cases three) α-helices per monomer. The first 20 amino acids of the dengue virus core protein, corresponding roughly to the N-terminal RNA-binding region, were found to be disordered in solution (Ma et al., 2004), and phylogenetic comparisons further suggest that the first helix ($\alpha 1$) is also highly flexible (Dokland et al., 2004; Patkar et al., 2007). Interestingly, functional plasticity in flaviviral core proteins extends well beyond the structurally flexible regions. Large deletions in either the internal hydrophobic domain (Kofler et al., 2002, 2003; Schlick et al., 2009) or the RNA-binding regions (Patkar et al., 2007) are well tolerated, indicating that neither of the distinct structural elements are indispensable for RNA binding and infectious virion formation. In agreement with these data, WNV core protein retained its nucleic acid chaperone activity following heat denaturation (boiling for 5–10 min), suggesting that its activity is dependent on the

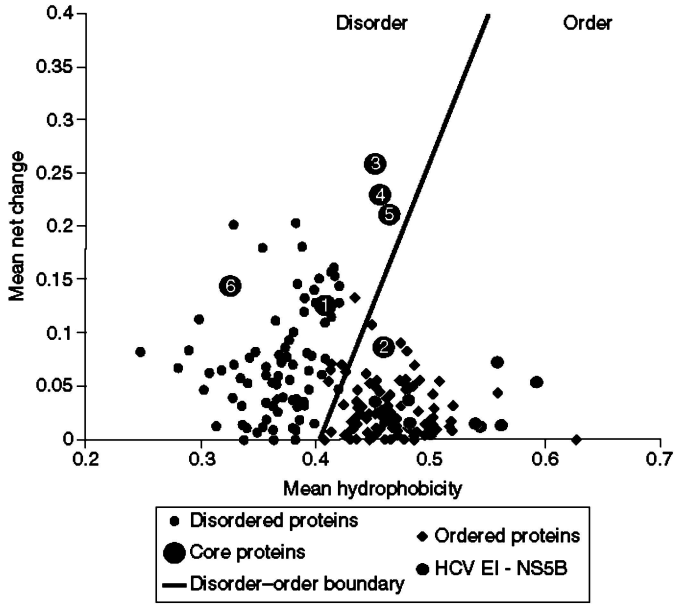


Figure 14.7 Charge–hydrophobicity (CH) plot of Flaviviridae core proteins. Low mean hydrophobicity combined with high net charge is frequently associated with intrinsic disorder in proteins (Uversky et al., 2000). Thus, ordered and disordered proteins are separated by a boundary line in the CH phase space (Uversky et al., 2000). The position of Flaviviridae core proteins (1, HCV; 2, GBV-B; 3, YFV; 4, TBEV; 5, WNV; and 6, BVDV) is shown on the CH plot, together with well-characterized data sets of ordered (black diamonds) and disordered (grey dots) proteins (Uversky et al., 2000). Flaviviridae core proteins are classified mostly as disordered, with flavivirus cores occupying a unique position on the plot. For comparison, the HCV envelope and structural proteins (large grey dots) are also included.

mass action of basic residues rather than on a well-folded RNA-binding domain (Ivanyi-Nagy et al., 2008). Interestingly, flavivirus core proteins are also classified among the disordered proteins based on their mean charge/mean hydrophobicity profiles (Fig. 14.7).

Pestivirus core proteins lack any recognizable domain structure and were suggested to contact RNA over their entire length (Ivanyi-Nagy et al., 2008; Murray et al., 2008). Intrinsic fluorescence spectroscopy and far-UV CD spectroscopy indicate that BVDV core protein is completely unstructured, supported by its sensitivity to proteinase digestion and by the heat resistance of the RNA chaperone activity of the protein (Ivanyi-Nagy et al., 2008; Murray et al., 2008).

In summary, despite the marked differences in their domain organization and overall structure (or lack of structure), core proteins from the three genera in Flaviviridae use IDP regions for RNA binding, chaperoning, and particle assembly (Ivanyi-Nagy et al., 2008; Ivanyi-Nagy and Darlix, 2010). Interfering with RNA

binding, or with the disorder-to-order transitions occurring upon core–RNA interactions, could constitute a general antiviral strategy effective against a wide range of human pathogenic viruses.

ACKNOWLEDGMENTS

Work supported (for J. L. D. and R. I. N) by ANRS, INSERM, and FINOVI. E. I. P. is supported by ANRS.

ABBREVIATIONS

BVDV	bovine viral diarrhea virus
CRE	cis-acting RNA element or cis-acting replication element
CSFV	classical swine fever virus
DENV	dengue virus
DLS	dimer linkage sequence
DM	*n*-dodecyl β-D-maltoside
ER	endoplasmic reticulum
far-UV CD	far ultraviolet circular dichroism
FTIR	Fourier transform infrared
GBV-B	GB virus B
GORS	genome-scale ordered RNA structure
HCV	hepatitis C virus
HCVcc	cell culture grown HCV
HIV-1	human immunodeficiency virus type 1
hnRNP A1	heterogeneous nuclear ribonucleoprotein A1
IDP	intrinsically disordered protein
IRES	internal ribosome entry site
LD	lipid droplet
LVP	lipo-viro-particle
NLP	nucleocapsid-like particle
ORF	open reading frame
PrP^C	cellular prion protein
RdRp	RNA-dependent RNA polymerase
RF	replicative form
RI	replicative intermediate
RSP	recombinant subviral particle
SHA	slowly sedimenting haemagglutinin
SP	signal peptidase
SPP	signal peptide peptidase
STAT-C	specifically targeted antiviral therapy for hepatitis C
TBEV	tick-borne encephalitis virus
UTR	untranslated region
WNV	West Nile virus
YFV	yellow fever virus

REFERENCES

Andre P, Komurian-Pradel F, Deforges S, Perret M, Berland JL, Sodoyer M, Pol S, Brechot C, Paranhos-Baccala G, Lotteau V. Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. J Virol 2002;76(14):6919–6928.

Andre P, Perlemuter G, Budkowska A, Brechot C, Lotteau V. Hepatitis C virus particles and lipoprotein metabolism. Semin Liver Dis 2005;25(1):93–104.

Bampi C, Jacquenet S, Lener D, Decimo D, Darlix JL. The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance. Int J Biochem Cell Biol 2004;36(9):1668–1686.

Bartenschlager R, Frese M, Pietschmann T. Novel insights into hepatitis C virus replication and persistence. Adv Virus Res 2004;63:71–180.

Beguiristain N, Robertson HD, Gomez J. RNase III cleavage demonstrates a long range RNA: RNA duplex element flanking the hepatitis C virus internal ribosome entry site. Nucleic Acids Res 2005;33(16):5250–5261.

Biebricher CK, Eigen M. The error threshold. Virus Res 2005;107(2):117–127.

Blight KJ, Rice CM. Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA. J Virol 1997;71(10):7345–7352.

Boni S, Lavergne JP, Boulant S, Cahour A. Hepatitis C virus core protein acts as a trans-modulating factor on internal translation initiation of the viral RNA. J Biol Chem 2005;280(18):17737–17748.

Bonnafous P, Perrault M, Le Bihan O, Bartosch B, Lavillette D, Penin F, Lambert O, Pecheur EI. Characterization of hepatitis C virus pseudoparticles by cryo-transmission electron microscopy using functionalized magnetic nanobeads. J Gen Virol 2010;91(Pt 8):1919–1930.

Boulant S, Montserret R, Hope RG, Ratinier M, Targett-Adams P, Lavergne JP, Penin F, McLauchlan J. Structural determinants that target the hepatitis C virus core protein to lipid droplets. J Biol Chem 2006;281(31):22236–22247.

Boulant S, Vanbelle C, Ebel C, Penin F, Lavergne JP. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. J Virol 2005;79(17):11353–11365.

Branch AD, Stump DD, Gutierrez JA, Eng F, Walewski JL. The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. Semin Liver Dis 2005;25(1):105–117.

Bung C, Bochkaeva Z, Terenin I, Zinovkin R, Shatsky IN, Niepmann M. Influence of the hepatitis C virus 3'-untranslated region on IRES-dependent and cap-dependent translation initiation. FEBS Lett 2010;584(4):837–842.

Buvoli M, Cobianchi F, Biamonti G, Riva S. Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. Nucleic Acids Res 1990;18(22):6595–6600.

Carmona P, Molina M. Interactions of protein and nucleic acid components of hepatitis C virus as revealed by Fourier transform infrared spectroscopy. Biochemistry 2010;49(23):4724–4731.

Casas-Finet JR, Smith JD, Kumar A, Kim JG, Wilson SH, Karpel RL Jr. Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains. Nucleic acid interaction, structural stability and self-association. J Mol Biol 1993;229(4):873–889.

CDC. Locally acquired dengue - Key West, Florida, 2009–2010. MMWR 2010;59(19):5.

Chang KS, Jiang J, Cai Z, Luo G. Human apolipoprotein e is required for infectivity and production of hepatitis C virus in cell culture. J Virol 2007;81(24):13783–13793.

Cheng JC, Chang MF, Chang SC. Specific interaction between the hepatitis C virus NS5B RNA polymerase and the 3′ end of the viral RNA. J Virol 1999;73(8):7044–7049.

Cristofari G, Darlix JL. The ubiquitous nature of RNA chaperone proteins. Prog Nucleic Acid Res Mol Biol 2002;72:223–268.

Cristofari G, Ivanyi-Nagy R, Gabus C, Boulant S, Lavergne JP, Penin F, Darlix JL. The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. Nucleic Acids Res 2004;32(8):2623–2631.

Darlix JL, Garrido JL, Morellet N, Mely Y, de Rocquigny H. Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. Adv Pharmacol 2007;55:299–346.

Darlix JL, Lapadat-Tapolsky M, de Rocquigny H, Roques BP. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. J Mol Biol 1995;254(4):523–537.

Davis M, Sagan SM, Pezacki JP, Evans DJ, Simmonds P. Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. J Virol 2008;82(23):11824–11836.

de Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugue S, Meiffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, Andre P, Rabourdin-Combe C, Lotteau V. Hepatitis C virus infection protein network. Mol Syst Biol 2008;4:230.

Diaz-Toledano R, Ariza-Mateos A, Birk A, Martinez-Garcia B, Gomez J. In vitro characterization of a miR-122-sensitive double-helical switch element in the 5′ region of hepatitis C virus RNA. Nucleic Acids Res 2009;37(16):5498–5510.

Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, Evans DJ. A hepatitis C virus cis-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B. J Virol 2008;82(18):9008–9022.

Dokland T, Walsh M, Mackenzie JM, Khromykh AA, Ee KH, Wang S. West Nile virus core protein; tetramer structure and ribbon formation. Structure 2004;12(7): 1157–1163.

Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. J Proteome Res 2006;5(11): 2985–2995.

Drake JW, Holland JJ. Mutation rates among RNA viruses. Proc Natl Acad Sci U S A 1999;96(24):13910–13913.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41(21):6573–6582.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 2005;272(20): 5129–5148.

Dutkiewicz M, Ciesiolka J. Structural characterization of the highly conserved 98-base sequence at the 3′ end of HCV RNA genome and the complementary sequence located at the 5′ end of the replicative viral strand. Nucleic Acids Res 2005;33(2):693–703.

Duvignaud JB, Savard C, Fromentin R, Majeau N, Leclerc D, Gagne SM. Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein. Biochem Biophys Res Commun 2009;378(1):27–31.

Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog 2008;4(2):e32.

Eigen M. Self organization of matter and the evolution of biological macromolecules. Naturwissenschaften 1971;58(10):465–523.

Fan Z, Yang QR, Twu JS, Sherker AH. Specific in vitro association between the hepatitis C viral genome and core protein. J Med Virol 1999;59(2):131–134.

Fang JW, Moyer RW. The effects of the conserved extreme 3′ end sequence of hepatitis C virus (HCV) RNA on the in vitro stabilization and translation of the HCV RNA genome. J Hepatol 2000;33(4):632–639.

Fraser CS, Doudna JA. Structural and mechanistic insights into hepatitis C viral translation initiation. Nat Rev Microbiol 2007;5(1):29–38.

Friebe P, Bartenschlager R. Genetic analysis of sequences in the 3′ nontranslated region of hepatitis C virus that are important for RNA replication. J Virol 2002;76(11): 5326–5338.

Friebe P, Boudet J, Simorre JP, Bartenschlager R. Kissing-loop interaction in the 3′ end of the hepatitis C virus genome essential for RNA replication. J Virol 2005;79(1):380–392.

Friebe P, Lohmann V, Krieger N, Bartenschlager R. Sequences in the 5′ nontranslated region of hepatitis C virus required for RNA replication. J Virol 2001;75(24): 12047–12057.

Gabus C, Auxilien S, Pechoux C, Dormont D, Swietnicki W, Morillas M, Surewicz W, Nandi P, Darlix JL. The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. J Mol Biol 2001a;307(4):1011–1021.

Gabus C, Derrington E, Leblanc P, Chnaiderman J, Dormont D, Swietnicki W, Morillas M, Surewicz WK, Marc D, Nandi P, Darlix JL. The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. J Biol Chem 2001b;276(22):19301–19309.

Gabus C, Ivanyi-Nagy R, Depollier J, Bucheton A, Pelisson A, Darlix JL. Characterization of a nucleocapsid-like region and of two distinct primer tRNALys,2 binding sites in the endogenous retrovirus Gypsy. Nucleic Acids Res 2006;34(20):5764–5777.

Gamarnik AV, Andino R. Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev 1998;12(15):2293–2304.

Gastaminza P, Dryden K, Boyd B, Wood M, Law M, Yeager M, Chisari FV. Ultrastructural and Biophysical Characterization of Hepatitis C Virus Particles Produced in Cell Culture. J Virol 2010;84(21):10999–11009.

Godet J, de Rocquigny H, Raja C, Glasser N, Ficheux D, Darlix JL, Mely Y. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. J Mol Biol 2006;356(5):1180–1192.

Gould EA, Solomon T. Pathogenic flaviviruses. Lancet 2008;371(9611):500–509.

Greaves J, Prescott GR, Gorleku OA, Chamberlain LH. The fat controller: roles of palmitoylation in intracellular protein trafficking and targeting to membrane microdomains (Review). Mol Membr Biol 2009;26(1):67–79.

Haid S, Pietschmann T, Pecheur EI. Low pH-dependent hepatitis C virus membrane fusion depends on E2 integrity, target lipid composition, and density of virus particles. J Biol Chem 2009;284(26):17657–17667.

Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2006;2(8):e100.

Herschlag D. RNA chaperones and the RNA folding problem. J Biol Chem 1995; 270(36):20871–20874.

Herschlag D, Khosla M, Tsuchihashi Z, Karpel RL. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. EMBO J 1994;13(12):2913–2924.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K. Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc Natl Acad Sci U S A 1991;88(13):5547–5551.

Honda M, Rijnbrand R, Abell G, Kim D, Lemon SM. Natural variation in translational activities of the 5′ nontranslated RNAs of hepatitis C virus genotypes 1a and 1b: evidence for a long-range RNA-RNA interaction outside of the internal ribosomal entry site. J Virol 1999;73(6):4941–4951.

Hope RG, McLauchlan J. Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein. J Gen Virol 2000;81(Pt 8):1913–1925.

Huang H, Sun F, Owen DM, Li W, Chen Y, Gale M, Ye J Jr. Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. Proc Natl Acad Sci U S A 2007;104(14):5848–5853.

Hwang SB, Lo SY, Ou JH, Lai MM. Detection of cellular proteins and viral core protein interacting with the 5′ untranslated region of hepatitis C virus RNA. J Biomed Sci 1995;2(3):227–236.

Icard V, Diaz O, Scholtes C, Perrin-Cocon L, Ramiere C, Bartenschlager R, Penin F, Lotteau V, Andre P. Secretion of hepatitis C virus envelope glycoproteins depends on assembly of apolipoprotein B positive lipoproteins. PLoS One 2009;4(1):e4233.

Imbert I, Dimitrova M, Kien F, Kieny MP, Schuster C. Hepatitis C virus IRES efficiency is unaffected by the genomic RNA 3'NTR even in the presence of viral structural or non-structural proteins. J Gen Virol 2003;84(Pt 6):1549–1557.

Ito M, Masumi A, Mochida K, Kukihara H, Moriishi K, Matsuura Y, Yamaguchi K, Mizuochi T. Peripheral B cells may serve as a reservoir for persistent hepatitis C virus infection. J Innate Immun 2010;2(6):607–617.

Ito T, Lai MM. Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. J Virol 1997; 71(11):8698–8706.

Ito T, Lai MM. An internal polypyrimidine-tract-binding protein-binding site in the hepatitis C virus RNA attenuates translation, which is relieved by the 3'-untranslated sequence. Virology 1999;254(2):288–296.

Ito T, Tahara SM, Lai MM. The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. J Virol 1998;72(11):8789–8796.

Ivanyi-Nagy R, Darlix JL. Intrinsic disorder in the core proteins of flaviviruses. Protein Pept Lett 2010;17(8):1019–1025.

Ivanyi-Nagy R, Davidovic L, Khandjian EW, Darlix JL. Disordered RNA chaperone proteins: from functions to disease. Cell Mol Life Sci 2005;62(13):1409–1417.

Ivanyi-Nagy R, Kanevsky I, Gabus C, Lavergne JP, Ficheux D, Penin F, Fosse P, Darlix JL. Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. Nucleic Acids Res 2006;34(9):2618–2633.

Ivanyi-Nagy R, Lavergne JP, Gabus C, Ficheux D, Darlix JL. RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. Nucleic Acids Res 2008;36(3): 712–725.

Jangra RK, Yi M, Lemon SM. Regulation of hepatitis C virus translation and infectious virus production by the microRNA miR–122. J Virol 2010;84(13):6615–6625.

Jones DM, McLauchlan J. Hepatitis C virus: assembly and release of virus particles. J Biol Chem 2010;285(30):22733–22739.

Jopling CL, Schutz S, Sarnow P. Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome. Cell Host Microbe 2008;4(1):77–85.

Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science 2005;309(5740):1577–1581.

Kaito M, Watanabe S, Tanaka H, Fujita N, Konishi M, Iwasa M, Kobayashi Y, Gabazza EC, Adachi Y, Tsukiyama-Kohara K, Kohara M. Morphological identification of hepatitis C virus E1 and E2 envelope glycoproteins on the virion surface using immunogold electron microscopy. Int J Mol Med 2006;18(4):673–678.

Kaito M, Watanabe S, Tsukiyama-Kohara K, Yamaguchi K, Kobayashi Y, Konishi M, Yokoi M, Ishida S, Suzuki S, Kohara M. Hepatitis C virus particle detected by immunoelectron microscopic study. J Gen Virol 1994;75(Pt7):1755–1760.

Kanto T, Hayashi N, Takehara T, Hagiwara H, Mita E, Naito M, Kasahara A, Fusamoto H, Kamada T. Density analysis of hepatitis C virus particle population in the circulation of infected hosts: implications for virus neutralization or persistence. J Hepatol 1995;22(4):440–448.

Khromykh AA, Westaway EG. RNA binding properties of core protein of the flavivirus Kunjin. Arch Virol 1996;141(3–4):685–699.

Kieft JS, Zhou K, Grech A, Jubin R, Doudna JA. Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. Nat Struct Biol 2002;9(5):370–374.

Kim M, Kim H, Cho SP, Min MK. Template requirements for de novo RNA synthesis by hepatitis C virus nonstructural protein 5B polymerase on the viral X RNA. J Virol 2002;76(14):6944–6956.

Kim TD, Ryu HJ, Cho HI, Yang CH, Kim J. Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes. Biochemistry 2000;39(48):14839–14846.

Kim YK, Lee SH, Kim CS, Seol SK, Jang SK. Long-range RNA-RNA interaction between the 5′ nontranslated region and the core-coding sequences of hepatitis C virus modulates the IRES-dependent translation. RNA 2003;9(5):599–606.

Klein KC, Dellos SR, Lingappa JR. Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system. J Virol 2005;79(11):6814–6826.

Klein KC, Polyak SJ, Lingappa JR. Unique features of hepatitis C virus capsid formation revealed by de novo cell-free assembly. J Virol 2004;78(17):9257–9269.

Kofler RM, Heinz FX, Mandl CW. Capsid protein C of tick-borne encephalitis virus tolerates large internal deletions and is a favorable target for attenuation of virulence. J Virol 2002;76(7):3534–3543.

Kofler RM, Leitner A, O'Riordain G, Heinz FX, Mandl CW. Spontaneous mutations restore the viability of tick-borne encephalitis virus mutants with large deletions in protein C. J Virol 2003;77(1):443–451.

Kolykhalov AA, Feinstone SM, Rice CM. Identification of a highly conserved sequence element at the 3′ terminus of hepatitis C virus genome RNA. J Virol 1996;70(6): 3363–3371.

Kolykhalov AA, Mihalik K, Feinstone SM, Rice CM. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3′ nontranslated region are essential for virus replication in vivo. J Virol 2000;74(4):2046–2051.

Kong LK, Sarnow P. Cytoplasmic expression of mRNAs containing the internal ribosome entry site and 3′ noncoding region of hepatitis C virus: effects of the 3′ leader on mRNA translation and mRNA stability. J Virol 2002;76(24):12457–12462.

Kopp M, Murray CL, Jones CT, Rice CM. Genetic analysis of the carboxy-terminal region of the hepatitis C virus core protein. J Virol 2010;84(4):1666–1673.

Kramvis A, Kew MC. Structure and function of the encapsidation signal of hepadnaviridae. J Viral Hepat 1998;5(6):357–367.

Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A 1996;93(21):11504–11509.

Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG, Baker TS, Strauss JH. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell 2002;108(5):717–725.

Kumar A, Wilson SH. Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. Biochemistry 1990;29(48):10717–10722.

Kunkel M, Lorinczi M, Rijnbrand R, Lemon SM, Watowich SJ. Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. J Virol 2001;75(5):2119–2129.

Kunkel M, Watowich SJ. Conformational changes accompanying self-assembly of the hepatitis C virus core protein. Virology 2002;294(2):239–245.

Kunkel M, Watowich SJ. Biophysical characterization of hepatitis C virus core protein: implications for interactions within the virus and host. FEBS Lett 2004;557(1–3): 174–180.

Lavanchy D. The global burden of hepatitis C. Liver Int 2009;29(Suppl 1):74–81.

Lee H, Shin H, Wimmer E, Paul AV. cis-acting RNA signals in the NS5B C-terminal coding sequence of the hepatitis C virus genome. J Virol 2004;78(20): 10865–10877.

Lemberg MK, Martoglio B. Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. Mol Cell 2002;10(4):735–744.

Lerat H, Honda M, Beard MR, Loesch K, Sun J, Yang Y, Okuda M, Gosert R, Xiao SY, Weinman SA, Lemon SM. Steatosis and liver cancer in transgenic mice expressing the structural and nonstructural proteins of hepatitis C virus. Gastroenterology 2002;122(2):352–365.

Levin JG, Guo J, Rouzina I, Musier-Forsyth K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. Prog Nucleic Acid Res Mol Biol 2005;80:217–286.

Li D, Takyar ST, Lott WB, Gowans EJ. Amino acids 1–20 of the hepatitis C virus (HCV) core protein specifically inhibit HCV IRES-dependent translation in HepG2 cells, and inhibit both HCV IRES- and cap-dependent translation in HuH7 and CV-1 cells. J Gen Virol 2003;84(Pt4):815–825.

Li X, Jeffers LJ, Shao L, Reddy KR, de Medina M, Scheffel J, Moore B, Schiff ER. Identification of hepatitis C virus by immunoelectron microscopy. J Viral Hepat 1995;2(5):227–234.

Lindenbach, B, Thiel, HJ, Rice, CM. Flaviviridae: the viruses and their replication. In Fields virology, Knipe DM, Howley PM, editors. Philadelphia: Lippincott-Raven Publishers; 2007, pp.1101–1152.

Lindenbach BD, Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, Rice CM. Complete replication of hepatitis C virus in cell culture. Science 2005;309(5734):623–626.

Lourenco S, Costa F, Debarges B, Andrieu T, Cahour A. Hepatitis C virus internal ribosome entry site-mediated translation is stimulated by cis-acting RNA elements and trans-acting viral factors. FEBS J 2008;275(16):4179–4197.

Lukavsky PJ, Kim I, Otto GA, Puglisi JD. Structure of HCV IRES domain II determined by NMR. Nat Struct Biol 2003;10(12):1033–1038.

Ma L, Jones CT, Groesch TD, Kuhn RJ, Post CB. Solution structure of dengue virus capsid protein reveals another fold. Proc Natl Acad Sci U S A 2004;101(10):3414–3419.

Maillard P, Krawczynski K, Nitkiewicz J, Bronnert C, Sidorkiewicz M, Gounon P, Dubuisson J, Faure G, Crainic R, Budkowska A. Nonenveloped nucleocapsids of hepatitis C virus in the serum of infected patients. J Virol 2001;75(17):8240–8250.

Majeau N, Fromentin R, Savard C, Duval M, Tremblay MJ, Leclerc D. Palmitoylation of hepatitis C virus core protein is important for virion production. J Biol Chem 2009;284(49):33915–33925.

Majeau N, Gagne V, Boivin A, Bolduc M, Majeau JA, Ouellet D, Leclerc D. The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. J Gen Virol 2004;85(Pt4):971–981.

Markoff L, Falgout B, Chang A. A conserved internal hydrophobic domain mediates the stable membrane integration of the dengue virus capsid protein. Virology 1997;233(1):105–117.

Marukian S, Jones CT, Andrus L, Evans MJ, Ritola KD, Charles ED, Rice CM, Dustin LB. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. Hepatology 2008;48(6):1843–1850.

McCaffrey AP, Ohashi K, Meuse L, Shen S, Lancaster AM, Lukavsky PJ, Sarnow P, Kay MA. Determinants of hepatitis C translational initiation in vitro, in cultured cells and mice. Mol Ther 2002;5(6):676–684.

McLauchlan J. Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. J Viral Hepat 2000;7(1):2–14.

McLauchlan J. Lipid droplets and hepatitis C virus infection. Biochim Biophys Acta 2009;1791(6):552–559.

McLauchlan J, Lemberg MK, Hope G, Martoglio B. Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. EMBO J 2002;21(15):3980–3988.

McMullan LK, Grakoui A, Evans MJ, Mihalik K, Puig M, Branch AD, Feinstone SM, Rice CM. Evidence for a functional RNA element in the hepatitis C virus core gene. Proc Natl Acad Sci U S A 2007;104(8):2879–2884.

Meunier JC, Russell RS, Engle RE, Faulk KN, Purcell RH, Emerson SU. Apolipoprotein c1 association with hepatitis C virus. J Virol 2008;82(19):9647–9656.

Michel YM, Borman AM, Paulous S, Kean KM. Eukaryotic initiation factor 4G-poly(A) binding protein interaction is required for poly(A) tail-mediated stimulation of picornavirus internal ribosome entry segment-driven translation but not for X-mediated stimulation of hepatitis C virus translation. Mol Cell Biol 2001;21(13):4097–4109.

Mir MA, Panganiban AT. The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. RNA 2006;12(2):272–282.

Miyanari Y, Atsuzawa K, Usuda N, Watashi K, Hishiki T, Zayas M, Bartenschlager R, Wakita T, Hijikata M, Shimotohno K. The lipid droplet is an important organelle for hepatitis C virus production. Nat Cell Biol 2007;9(9):1089–1097.

Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nat Rev Microbiol 2007;5(6):453–463.

Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, Matsuura Y, Kimura S, Miyamura T, Koike K. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nat Med 1998;4(9):1065–1067.

Mukhopadhyay S, Kim BS, Chipman PR, Rossmann MG, Kuhn RJ. Structure of West Nile virus. Science 2003;302(5643):248.

Mukhopadhyay S, Kuhn RJ, Rossmann MG. A structural perspective of the flavivirus life cycle. Nat Rev Microbiol 2005;3(1):13–22.

Murray CL, Marcotrigiano J, Rice CM. Bovine viral diarrhea virus core is an intrinsically disordered protein that binds RNA. J Virol 2008;82(3):1294–1304.

Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, Andre P, Lotteau V, Rabourdin-Combe C. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Res 2009;37(Database issue):D661–D668.

Nielsen SU, Bassendine MF, Burt AD, Martin C, Pumeechockchai W, Toms GL. Association between hepatitis C virus and very-low-density lipoprotein (VLDL)/LDL analyzed in iodixanol density gradients. J Virol 2006;80(5):2418–2428.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(Suppl 6):566–572.

Oh JW, Sheu GT, Lai MM. Template requirement and initiation site selection by hepatitis C virus polymerase on a minimal viral RNA template. J Biol Chem 2000;275(23):17710–17717.

Okamoto K, Mori Y, Komoda Y, Okamoto T, Okochi M, Takeda M, Suzuki T, Moriishi K, Matsuura Y. Intramembrane processing by signal peptide peptidase regulates the membrane localization of hepatitis C virus core protein and viral propagation. J Virol 2008;82(17):8349–8361.

Otto GA, Puglisi JD. The pathway of HCV IRES-mediated translation initiation. Cell 2004;119(3):369–380.

Owen DM, Huang H, Ye J, Gale M Jr. Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor. Virology 2009;394(1):99–108.

Pan J, Thirumalai D, Woodson SA. Folding of RNA involves parallel pathways. J Mol Biol 1997;273(1):7–13.

Patkar CG, Jones CT, Chang YH, Warrier R, Kuhn RJ. Functional requirements of the yellow fever virus capsid protein. J Virol 2007;81(12):6471–6481.

Pecheur EI, Diaz O, Molle J, Icard V, Bonnafous P, Lambert O, Andre P. Morphological characterization and fusion properties of triglyceride-rich lipoproteins obtained from cells transduced with hepatitis C virus glycoproteins. J Biol Chem 2010;285(33):25802–25811.

Pereira AA, Jacobson IM. New and experimental therapies for HCV. Nat Rev Gastroenterol Hepatol 2009;6(7):403–411.

Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. Genes Dev 1998;12(1):67–83.

Piver E, Roingeard P, Pages JC. The cell biology of hepatitis C virus (HCV) lipid addiction: molecular mechanisms and its potential importance in the clinic. Int J Biochem Cell Biol 2010;42(6):869–879.

Pokidysheva E, Zhang Y, Battisti AJ, Bator-Kelly CM, Chipman PR, Xiao C, Gregorio GG, Hendrickson WA, Kuhn RJ, Rossmann MG. Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. Cell 2006;124(3):485–493.

Pontius BW, Berg P. Renaturation of complementary DNA strands mediated by purified mammalian heterogeneous nuclear ribonucleoprotein A1 protein: implications for a mechanism for rapid molecular assembly. Proc Natl Acad Sci U S A 1990;87(21): 8403–8407.

Prince AM, Huima-Byron T, Parker TS, Levine DM. Visualization of hepatitis C virions and putative defective interfering particles isolated from low-density lipoproteins. J Viral Hepat 1996;3(1):11–17.

Rajkowitsch L, Semrad K, Mayer O, Schroeder R. Assays for the RNA chaperone activity of proteins. Biochem Soc Trans 2005;33(Pt3):450–456.

Ray RB, Ray R. Hepatitis C virus core protein: intriguing properties and functional relevance. FEMS Microbiol Lett 2001;202(2):149–156.

Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Proteins 2006;62(1):24–45.

Rodenhuis-Zybert IA, van der Schaar HM, da Silva Voorham JM, van der Ende-Metselaar H, Lei HY, Wilschut J, Smit JM. Immature dengue virus: a veiled pathogen?. PLoS Pathog 2010;6(1):e1000718.

Rodriguez-Casado A, Molina M, Carmona P. Spectroscopic study of conformational changes accompanying self-assembly of HCV core protein. Proteins 2007;66(1): 110–117.

Roehrig JT, Layton M, Smith P, Campbell GL, Nasci R, Lanciotti RS. The emergence of West Nile virus in North America: ecology, epidemiology, and surveillance. Curr Top Microbiol Immunol 2002;267:223–240.

Romero-Lopez C, Berzal-Herranz A. A long-range RNA-RNA interaction between the 5′ and 3′ ends of the HCV genome. RNA 2009;15(9):1740–1752.

Samsa MM, Mondotte JA, Iglesias NG, Assuncao-Miranda I, Barbosa-Lima G, Da Poian AT, Bozza PT, Gamarnik AV. Dengue virus capsid protein usurps lipid droplets for viral particle formation. PLoS Pathog 2009;5(10):e1000632.

Santolini E, Migliaccio G, La Monica N. Biosynthesis and biochemical properties of the hepatitis C virus core protein. J Virol 1994;68(6):3631–3641.

Schlick P, Taucher C, Schittl B, Tran JL, Kofler RM, Schueler W, von Gabain A, Meinke A, Mandl CW. Helices alpha2 and alpha3 of West Nile virus capsid protein are dispensable for assembly of infectious virions. J Virol 2009;83(11):5581–5591.

Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol 2004;5(11):908–919.

Sharma K, Didier P, Darlix JL, de Rocquigny H, Bensikaddour H, Lavergne JP, Penin F, Lessinger JM, Mely Y. Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein. Nucleic Acids Res 2010;38(11):3632–3642.

Shavinskaya A, Boulant S, Penin F, McLauchlan J, Bartenschlager R. The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. J Biol Chem 2007;282(51):37158–37169.

Shetty S, Kim S, Shimakami T, Lemon SM, Mihailescu MR. Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. RNA 2010;16(5):913–925.

Shimakami T, Lanford RE, Lemon SM. Hepatitis C: recent successes and continuing challenges in the development of improved treatment modalities. Curr Opin Pharmacol 2009;9(5):537–544.

Shimoike T, Koyama C, Murakami K, Suzuki R, Matsuura Y, Miyamura T, Suzuki T. Down-regulation of the internal ribosome entry site (IRES)-mediated translation of the hepatitis C virus: critical role of binding of the stem-loop IIId domain of IRES and the viral core protein. Virology 2006;345(2):434–445.

Shimoike T, Mimori S, Tani H, Matsuura Y, Miyamura T. Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. J Virol 1999;73(12):9718–9725.

Shintani Y, Fujie H, Miyoshi H, Tsutsumi T, Tsukamoto K, Kimura S, Moriya K, Koike K. Hepatitis C virus infection and diabetes: direct involvement of the virus in the development of insulin resistance. Gastroenterology 2004;126(3):840–848.

Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. RNA 2004;10(9):1337–1351.

Smith DB, Simmonds P. Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. J Mol Evol 1997;45(3):238–246.

Smith RM, Walton CM, Wu CH, Wu GY. Secondary structure and hybridization accessibility of hepatitis C virus 3'-terminal sequences. J Virol 2002;76(19):9563–9574.

Song Y, Friebe P, Tzima E, Junemann C, Bartenschlager R, Niepmann M. The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. J Virol 2006;80(23):11579–11588.

Spahn CM, Kieft JS, Grassucci RA, Penczek PA, Zhou K, Doudna JA, Frank J. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. Science 2001;291(5510):1959–1962.

Stamataki Z, Shannon-Lowe C, Shaw J, Mutimer D, Rickinson AB, Gordon J, Adams DH, Balfe P, McKeating JA. Hepatitis C virus association with peripheral blood B lymphocytes potentiates viral infection of liver-derived hepatoma cells. Blood 2009;113(3):585–593.

Steenbergen RH, Joyce MA, Lund G, Lewis J, Chen R, Barsby N, Zhu LF, Tyrrell L, Kneteman NM. Lipoprotein profiles in SCID/uPA mice transplanted with human hepatocytes become human-like and correlate with HCV infection success. Am J Physiol Gastrointest Liver Physiol 2010;299(4):G844–G854.

Steil BP, Barton DJ. Cis-active RNA elements (CREs) and picornavirus RNA replication. Virus Res 2009;139(2):240–252.

Takahashi K, Kishimoto S, Yoshizawa H, Okamoto H, Yoshikawa A, Mishiro S. p26 protein and 33-nm particle associated with nucleocapsid of hepatitis C virus recovered from the circulation of infected hosts. Virology 1992;191(1):431–434.

Tanaka T, Kato N, Cho MJ, Shimotohno K. A novel sequence found at the 3′ terminus of hepatitis C virus genome. Biochem Biophys Res Commun 1995;215(2):744–749.

Tanaka T, Kato N, Cho MJ, Sugiyama K, Shimotohno K. Structure of the 3′ terminus of the hepatitis C virus genome. J Virol 1996;70(5):3307–3312.

Tanaka Y, Shimoike T, Ishii K, Suzuki R, Suzuki T, Ushijima H, Matsuura Y, Miyamura T. Selective binding of hepatitis C virus core protein to synthetic oligonucleotides corresponding to the 5′ untranslated region of the viral genome. Virology 2000;270(1):229–236.

Thomssen R, Bonk S, Propfe C, Heermann KH, Kochel HG, Uy A. Association of hepatitis C virus in human sera with beta-lipoprotein. Med Microbiol Immunol 1992;181(5):293–300.

Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. FASEB J 2004;18(11):1169–1175.

Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci 2008;33(1):2–8.

Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. Internal ribosome entry site within hepatitis C virus RNA. J Virol 1992;66(3):1476–1483.

Tuplin A, Evans DJ, Simmonds P. Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J Gen Virol 2004;85(Pt10):3037–3047.

Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. RNA 2002;8(6):824–841.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–427.

Vassilaki N, Friebe P, Meuleman P, Kallis S, Kaul A, Paranhos-Baccala G, Leroux-Roels G, Mavromara P, Bartenschlager R. Role of the hepatitis C virus core{+}1 open reading frame and core cis-acting RNA elements in viral RNA translation and replication. J Virol 2008;82(23):11503–11515.

Villordo SM, Gamarnik AV. Genome cyclization as strategy for flavivirus RNA replication. Virus Res 2009;139(2):230–239.

Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Krausslich HG, Mizokami M, Bartenschlager R, Liang TJ. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat Med 2005;11(7):791–796.

Wang SH, Syu WJ, Hu ST. Identification of the homotypic interaction domain of the core protein of dengue virus type 2. J Gen Virol 2004;85(Pt8):2307–2314.

Wang TH, Rijnbrand RC, Lemon SM. Core protein-coding sequence, but not core protein, modulates the efficiency of cap-independent translation directed by the internal ribosome entry site of hepatitis C virus. J Virol 2000;74(23):11347–11358.

Weinlich S, Huttelmaier S, Schierhorn A, Behrens SE, Ostareck-Lederer A, Ostareck DH. IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR. RNA 2009;15(8):1528–1542.

Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 1996;35(43):13709–13715.

Yamada N, Tanihara K, Takada A, Yorihuzi T, Tsutsumi M, Shimomura H, Tsuji T, Date T. Genetic organization and diversity of the 3′ noncoding region of the hepatitis C virus genome. Virology 1996;223(1):255–261.

Yanagi M, St Claire M, Emerson SU, Purcell RH, Bukh J. In vivo analysis of the 3′ untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. Proc Natl Acad Sci U S A 1999;96(5):2291–2295.

Yasui K, Wakita T, Tsukiyama-Kohara K, Funahashi SI, Ichikawa M, Kajita T, Moradpour D, Wands JR, Kohara M. The native form and maturation process of hepatitis C virus core protein. J Virol 1998;72(7):6048–6055.

Yi M, Lemon SM. 3′ nontranslated RNA signals required for replication of hepatitis C virus RNA. J Virol 2003a;77(6):3557–3568.

Yi M, Lemon SM. Structure-function analysis of the 3′ stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. RNA 2003b;9(3):331–345.

You S, Rice CM. 3′ RNA elements in hepatitis C virus replication: kissing partners and long poly(U). J Virol 2008;82(1):184–195.

You S, Stump DD, Branch AD, Rice CM. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. J Virol 2004;78(3):1352–1366.

Yu KL, Jang SI, You JC. Identification of in vivo interaction between Hepatitis C Virus core protein and 5′ and 3′ UTR RNA. Virus Res 2009;145(2):285–292.

Zhang J, Yamada O, Yoshida H, Iwai T, Araki H. Autogenous translational inhibition of core protein: implication for switch from translation to RNA replication in hepatitis C virus. Virology 2002;293(1):141–150.

Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, Baker TS, Strauss JH, Rossmann MG, Kuhn RJ. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. Nat Struct Biol 2003;10(11):907–912.

Zuniga S, Cruz JL, Sola I, Mateos-Gomez PA, Palacio L, Enjuanes L. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. J Virol 2010;84(4):2169–2175.

Zuniga S, Sola I, Cruz JL, Enjuanes L. Role of RNA chaperones in virus replication. Virus Res 2009;139(2):253–266.

15

THE NS5A DOMAIN II OF HCV: CONSERVATION OF INTRINSIC DISORDER IN SEVERAL GENOTYPES

XAVIER HANOULLE, ISABELLE HUVENT, ARNAUD LEROY, HONG YE, CONG BAO KANG, YU LIANG, CLAIRE ROSNOBLET, JEAN-MICHEL WIERUSZESKI, HO SUP YOON, AND GUY LIPPENS

15.1 THE NS5A DOMAIN II IN THE HCV REPLICATIVE CYCLE

With nearly 3% of the world population infected, of which a sizeable portion will evolve to liver cirrhosis and/or liver carcinoma, the hepatitis C virus (HCV) poses a serious health problem. The present standard care, pegylated interferon (IFN)-α combined with the nucleoside analogue ribavirin, provides both a general hostile cell environment for the virus (through the IFN) and an inhibition of the replication through the ribavirin (Fried et al., 2002). However, the treatment outcome is dependent on the genotype of the virus. Effective for over 80% of patients who carry genotypes 3 and 4, mainly found in the East, it leads to complete viral clearing in only 40–50% of patients who carry the genotype 1 virus, which is the dominant form in the Western countries (NIH, 2002). From a clinical point of view, the comparative study of different genotypes is hence of uttermost importance, both in the definition of the treatment and for the development of new antiviral entities.

Ribavirin targets the RNA-dependent RNA polymerase NS5B, and a wealth of structural data have turned this replicative enzyme into an important drug target. In addition to its active site, it features at least three known allosteric binding pockets that regulate RNA synthesis and are suitable for inhibitor design (Beaulieu, 2007). Other proteins with a well-defined three-dimensional structure have equally been

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

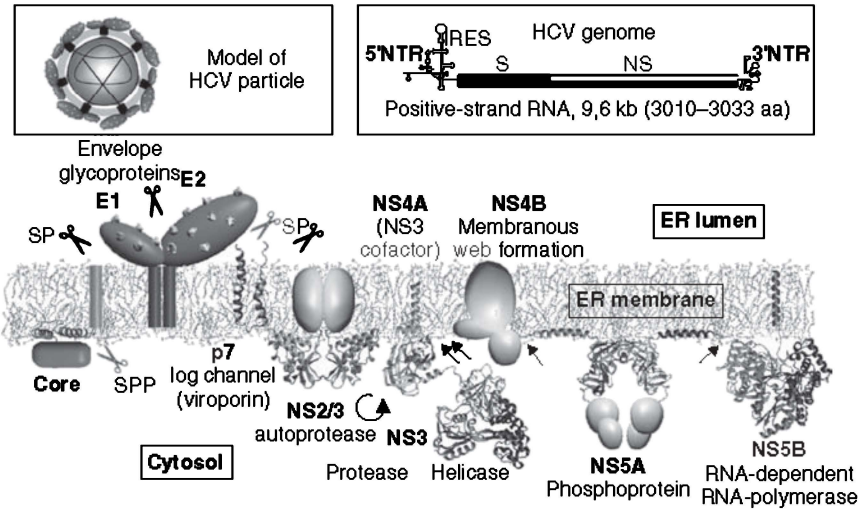


Figure 15.1 Review of the experimental structures of the HCV viral proteins. The left inset shows a schematic model of the HCV particle exhibiting E1 and E2 surface glycoproteins anchored to the viral membrane and surrounding the viral nucleocapsid, including core protein associated to viral RNA genome. A schematic view of this 9.6-kb RNA genome is shown in the right inset. NTR, N-terminal region. *Source*: Adapted from Moradpour et al. (2007).

the subject of pharmaceutical focus, the latest being a zinc-binding domain encoded by domain I of NS5A Tellinghuisen et al. (2005); Gao et al., 2010). However, when summarizing the structural knowledge of the viral proteins (Fig. 15.1) (Moradpour et al., 2007), we note that besides the membrane proteins that have currently escaped crystallization, some proteins are identified as lacking a well-defined structure. Less understood from a functional point of view than their structured counterparts, these proteins, however, do present an obvious pharmaceutical interest. The unstructured domain II of the NS5A protein, for example, has been shown to contain an interferon sensitivity-determining region (ISDR) (Enomoto et al., 1995). Although challenged over time, with an apparent geographical partitioning of the validity of this concept (Pascu et al., 2004), these findings suggest that a closer understanding of these natively unfolded proteins or protein domains might further advance our understanding of and hence capacity to interfere with the viral life cycle.

Because the comparative structural features of the NS5A domain II is the subject of this chapter, it is interesting to summarize its role in the life cycle of the virus (for a thorough review, see (Tellinghuisen et al., 2007)). Akin to that of other members of the Flaviviridae family, the life cycle of HCV starts when extracellular HCV virions interact via HCV glycoproteins (notably the E2 protein) with receptor molecules at the cell surface and undergo receptor-mediated endocytosis into a low-pH vesicle. After release of the viral RNA into the cytoplasm, it is translated into a single large polyprotein. This polyprotein contains part of the protease activity necessary for its self-processing into the final 10 mature HCV proteins. Some of

these will associate with an endoplasmic reticulum (ER)-like membrane structure termed *the membranous web*, thereby making up the replication process that will assure viral RNA replication via a minus-strand replicative intermediate to produce daughter RNA. A portion of this newly synthesized RNA becomes packaged into nucleocapsids and associates with the HCV glycoproteins, leading to budding into the ER. Virions follow the cellular secretory pathway and mature during this transit into productive particles. Mature virions are released from the cell, completing the life cycle. The third (unstructured) domain of NS5A (Hanoulle et al., 2009b) is involved in viral packaging (Appel et al., 2008; Tellinghuisen et al., 2008), but otherwise, NS5A is thought to play an important role mainly in the RNA replication process. Its second domain interacts with both the RNA component (Foster et al., 2010) and the RNA-dependent RNA polymerase NS5B itself (Shirota et al., 2002), but the list of its interaction partners is much longer (for a review, see Macdonald and Harris (2004)). Therefore, its role in the constitution of a multiprotein replication complex seems central. One of the best-studied interaction partners is the IFN-induced double-stranded RNA-activated protein kinase (PKR), which recognizes the internal ribosome entry site (IRES) present in the 5′-untranslated region (UTR) of HCV RNA and the ISDR region present in NS5A domain II equally (Gale et al., 1997). Other regulatory proteins such as the Lyn or Fyn kinase domain(s) would equally interact with the domain (Macdonald et al., 2004), and we and others recently showed that the prolyl cis/trans isomerase cyclophilin A (CypA) interacts with and isomerizes selected proline residues in this domain (Chatterji et al., 2009; Hanoulle et al., 2009a). The functional outcome of this structural posttranslational modification is not clear, especially in the context of an intrinsically unstructured protein (IUP) domain, but is actually pursued by several research groups.

Structural data on the NS5A domains II of two genotypes have been reported in the literature by our groups: the Japanese fulminant hepatitis (JFH-1) sequence is classified in genotype 2a (Hanoulle et al., 2009a, 2010), whereas the H77 sequence belongs to genotype 1a (Liang et al., 2006, 2007). We have recently obtained information on the Con1 NS5A domain II that belongs to genotype 1b. In this chapter, we focus on a comparative study of these three proteins, rather than on the particular aspects of one or the other. We thereby consider information contained in the primary sequences, both at the level of conservation and structural content. Second, we assess information obtained by macroscopic biochemical methods such as gel filtration and circular dichroism (CD) spectroscopy. Last, the comparison of the nuclear magnetic resonance (NMR) spectra of the three protein domains highlights that certain patterns contained in the primary sequence might translate into a conserved structural motif.

15.2 PRIMARY SEQUENCE INFORMATION OVER THE DIFFERENT GENOTYPES

When we align the primary sequences of these three sequences, we observe that nearly half of the residues are strictly conserved (Fig. 15.2a). This proportion

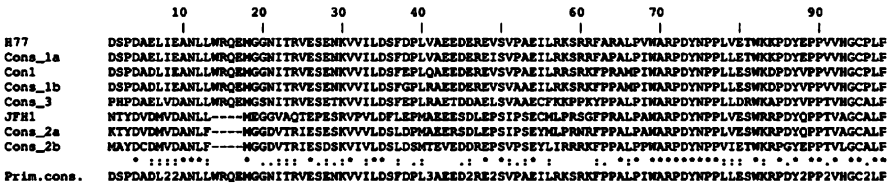


Figure 15.2 Sequence alignments of the three NS5A domain II sequences, from H77 (1a), Con1 (1b), and JFH-1 (2a) HCV strains, that we discuss in this chapter. Multiple sequence alignments were performed using ClustalW (Thompson et al., 1994) with default parameters. Star characters indicate strictly conserved amino acids, colon conservative mutations. In addition to the H77, Con1, and JFH-1 sequences, the consensus sequences of the NS5A domains II of genotypes 1a, 1b, 2a, 2b, and 3 are also shown. Consensus sequences were generated from the European HCV database (http://euhcvdb.ibcp.fr) by aligning the validated sequences of each genotype.

rises to 75% when we take into account the conservation of residues of related physicochemical nature. The two genotype 1 sequences, H77 and Con1, are even closer, with only 12 mutations out of 98 amino acids, of which only 4 are recognized as truly different amino acid substitutions. One obvious difference for the JFH-1 sequence is the deletion of the WRQE tetrapeptide in the N-terminal region of the NS5A domain II. In order to see whether this deletion is specific for this precise strain or for genotype 2 to which it belongs, we generated the consensus sequences for genotypes 1a, 1b, and 2a starting from the viral sequences deposited in the HCV database (Combet et al., 2007, http://euhcvdb.ibcp.fr/euHCVdb/). Caution should be taken regarding the great variability in the number of confirmed sequences that are deposited for the different genotypes: whereas 224 and 357 confirmed sequences have been deposited for genotypes 1a and 1b, respectively, only 19 and 26 sequences are deposited for the 2a and 2b genotypes, respectively, and only 30 sequences are deposited for genotype 3. Interestingly, the deletion is a defining factor for the genotype 2 (Fig. 15.2), and this despite previous observations that the JFH-1 strain fits into the cluster of genotype 2a with notable deviations in the 5′-UTR, core, NS3, and NS5A regions (Wakita and Kato, 2006). Because this deletion is in the putative ISDR region, we generated the consensus sequence of genotype 3 NS5A domain II as well, whereby patients carrying this viral strain are, in general, good responders to the interferon–ribavirin treatment. From the alignment, we find that the deletion is not a defining feature of the genotype 3 sequences, and thereby makes the genotype 3 more close to the genotype 1 than genotype 2.

15.3 DISORDER PREDICTIONS OVER THE DIFFERENT GENOTYPES

A large number of algorithms have been developed to estimate the disorder in a given protein sequence. Based solely on the primary sequence, these algorithms use parameters such as amino acid composition, sequence complexity, hydrophobicity, solvent accessibility, and so on, to estimate the presence or absence of secondary

structure. We show here the results of two commonly used programs, the PONDR VLXT algorithm (Romero et al., 1997, 2001; Li et al., 1999) and the metaPrDOS meta server (Ishida and Kinoshita, 2008), where the latter averages the results of 6 independent disorder predictions.

In agreement with the high sequence conservation of both sequences, the PONDR predictions for the H77 and Con1 NS5A domains II are nearly identical (Fig. 15.3a). Still, two very conservative mutations at positions 6 and 38 (D_6E and $D_{38}E$—numbering according to Fig. 15.2) do influence the predicted disorder for the intercalated segment to some extent. The resulting region predicted as disordered spans some 50 residues (17–69) in the center of both sequences. The N-terminal pentapeptide is equally predicted as disordered, whereas the C-terminal stretch would be rather ordered. However, one should take into account a possible border effect for these two parts. Indeed, when we apply the PONDR VLXT algorithm with the same parameters on the full-length NS5A protein sequences, we find less disorder in the N-terminal residues and more disorder in the C terminus.

For the JFH-1 domain II, the central region predicted as disordered is shorter, and spans only 40 residues. After this predicted disordered region, we obtain the

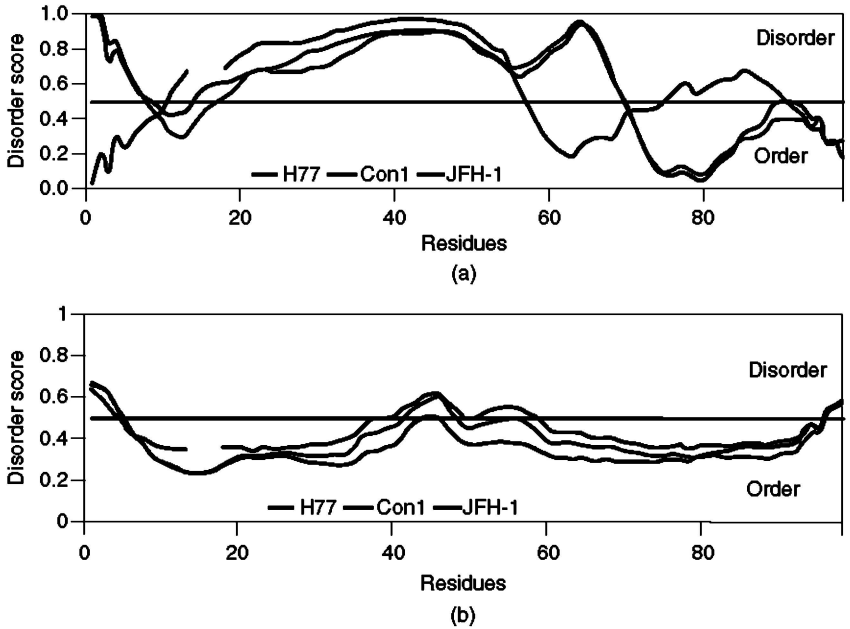


Figure 15.3 Disorder predictions on NS5A domain II sequences. (a). PONDR VLXT algorithm (Romero et al., 1997, 2001; Li et al., 1999) predictions of the three NS5A domain II sequences (dark gray: H77; medium gray: JFH-1; light gray: Con1). Sequence numbering is as in Figure 15.2. (b). Predictions of disorder of the same sequences with the metaPrDOS meta server (Ishida and Kinoshita, 2008). Scores above the threshold value of 0.5 are considered as disordered.

inversed profile compared to the Con1 and H77 sequences: first a short ordered region of 10 residues, followed by a disordered region that was predicted as ordered in the previous sequences. As the computer predictions allow ready evaluation of the effect of different point mutations (in contrast to the lengthy experimental approach!), we performed predictions with different hybrid sequences and traced back this order/disorder inversion to the presence of four positive charges ($R_{57}R/KSRK/R_{61}$) in the Con1/H77 sequences, whereas only one positive charge ($L_{53}PRSG_{57}$) characterizes the JFH-1 sequence.

When applying the metaPrDOS meta server to the three sequences, the resulting predictions are dramatically different. Indeed, when applying a threshold value of 0.5 for disorder prediction, only the central 20 residues for JFH-1 are predicted as disordered, whereas this portion drops to a mere 2 residues for the H77 sequence. Even more amazing, the Con1 sequence that differs only by a Q41V mutation in this central region adopts a profile closer to JFH-1 than H77.

Whereas we have not studied in depth the algorithmic subtleties that led to such large differences in output, we would have a tendency to conclude that experimental input is required when protein sequences do not lead to a clear consensus in predicted (dis)order.

15.4 MACROSCOPIC METHODS TO EVALUATE STRUCTURE: GEL FILTRATION AND CD

One simple method to evaluate the absence of order is gel filtration that was here performed on a Superdex 75 column. Indeed, when one establishes a molecular-weight scale by using structured proteins as markers, a disordered protein elutes consistently with a larger molecular mass as expected. When considering the elution profiles for the Con1 and JFH-1 domains (Fig. 15.4a), we indeed observe that they elute in the molecular-weight range of 30–35 kDa, although their calculated (and experimentally verified by mass spectroscopy) molecular weights are 12.4 and 11.6 kDa, respectively. Clearly, one could establish a molecular-weight scale with only disordered proteins as molecular markers, and this should allow recalibration of the column for this precise subclass of proteins.

A second method commonly used to evaluate the secondary structure of proteins is CD. Particularly sensitive to the presence of α-helices, which show up as a distinctive minimum at 222 nm, this spectroscopy has been extensively used to evidence the disordered character of polypeptides. We show here the comparison between the CD spectra of the Con1 and JFH-1 NS5A domains II, whereby both spectra were recorded on the same instrument, at the same concentration of 0.84 mg/ml, and at 4°C in order to stabilize potential secondary structure elements (Fig. 15.4b). Surprisingly, the spectra are not identical. Indeed, whereas the CD spectrum of the JFH-1 sequence is characteristic of a random coil polypeptide, with a deep minimum at 198 nm, this same minimum is less pronounced in the spectrum of the Con1 domain. This latter spectrum also shows a weak minimum at 222 nm that we previously observed in the H77 CD spectrum at 25°C (Liang

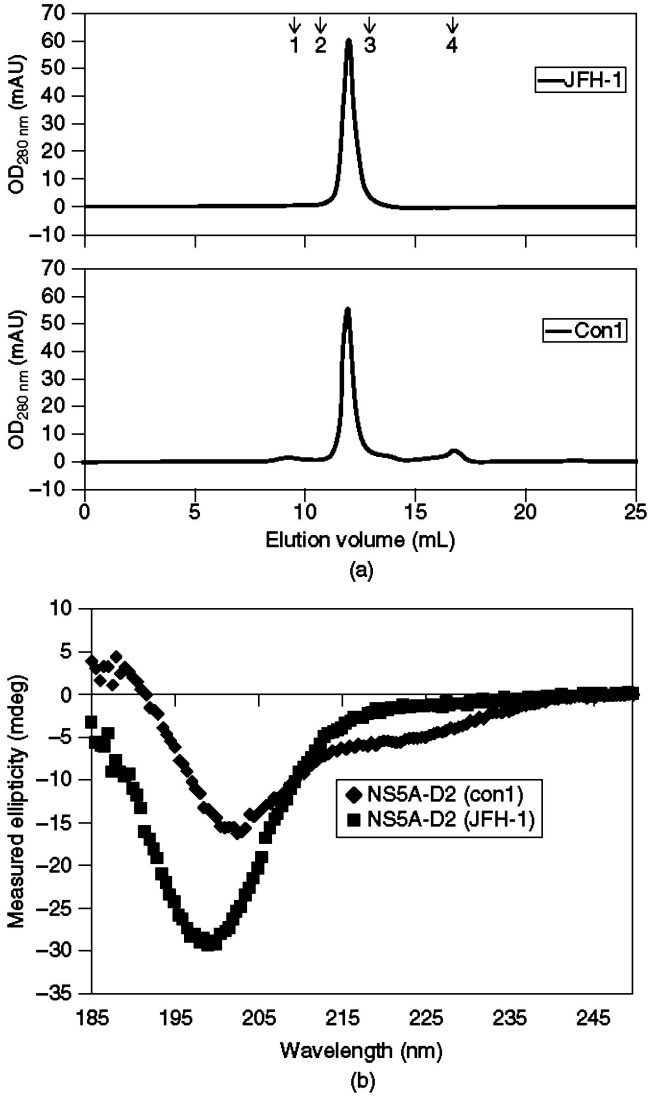


Figure 15.4 (a). Gel filtration analyses of the NS5A domains II from Con1 (bottom) and JFH-1 (top) HCV strains. Gel filtrations were performed on a Superdex 75 HR 10/30 column equilibrated in the following buffer: 50 mM NaPi pH6.8, 150 mM NaCl, 1 mM DTT. Arrows on top indicate the elution volumes of globular standards with the following molecular weight: 1, 67 kDa; 2, 43 kDa; 3, 25 kDa; 4, 13.7 kDa). (b). CD spectra of the JFH-1 and Con1 NS5A domains II. Spectra were obtained on a Model CD6 spectrometer (Jobin-Yvon-Spex, Longjumeau, France) at 4°C using a quartz sample cell with 0.1-mm path length. The ellipticity was scanned from 185 to 250 nm with an increment of 0.5 nm, an integration time of two seconds, and a constant band-pass of 2 nm. The concentration of the protein was adjusted to 0.84 mg/mL.

et al., 2007). These observations point to a higher degree of disorder in the JFH-1 sequence compared to those of the genotype 1 NS5A domains II, in good agreement with the predictions of the metaPrDOS server, where the JFH-1 disorder score was higher than the other two over the full length of the domain (Fig. 15.3b).

15.5 NMR SPECTROSCOPY OF THE NS5A D2 DOMAINS

NMR spectroscopy is particularly well suited to evaluate the structure of a polypeptide at the per-residue level. The first defining character of IUPs is the limited dispersion of the amide proton chemical shift. Indeed, when comparing the ^{1}H, ^{15}N heteronuclear single-quantum coherence spectroscopy (HSQC) spectra of the three NS5A domain II, we note that the majority of the residues have their amide proton chemical shift within the 8.6–7.6-ppm range (Fig. 15.5a). However, although the sequence identity between the three domains is high (Fig. 15.2), the comparison of the three spectra shows that this does not directly translate into a good spectral overlap. However, some common features can be deduced from the superposition of the spectra. We do, for example, note that a couple of residues fall outside the range of amide proton chemical shift values that are generally qualified as characteristic for disordered polypeptides. In particular, the Trp69 and Ala70 (numbering according to Fig. 15.2) resonate at amide proton values of 7.0–7.2 ppm and 7.4 ppm, respectively (Fig. 15.5a), and this in the three different NS5A domains II. Overlap is not perfect, but the unusual chemical shift is clearly conserved. Whereas the nitrogen chemical shift of the Trp cross peak is quite similar in the Con1 and H77 sequences, it adopts an even more extreme value of 113.5 ppm in the JFH-1 sequence. These two residues are strictly conserved between the three sequences, hinting to a possible structural motif that would escape both bioinformatic analysis and macroscopic biophysical characterization. We recently showed that at least in the JFH-1 sequence, this motif is recognized by the Cyp family of proline cis/trans isomerases (Hanoulle et al., 2009a). These enzymes are the target for nonimmunosuppressive cyclosporin A derivatives such as Debio025 (Paeshuyse et al., 2006; Flisiak et al., 2008; Coelmont et al., 2009; Landrieu et al., 2010), currently in clinical testing as alternative antiviral treatments.

We previously employed peptide mapping to assign the (complex) NMR spectrum of the 441-amino-acid-long Tau protein (Smet et al., 2004). Spectral identity for a given amino acid in the full-length protein and in a short peptide derived from the protein sequence and centered on the residue under study led to the conclusion that the residue samples the same conformational space in both contexts. Classical homonuclear NMR was then used to ascertain that the peptide is indeed devoid of secondary structure. Here, we extend this notion to the comparative NMR analysis of the sequences. When we consider, for example, the tripeptide V_{32}-I_{33}-L_{34} in the NMR spectrum of the H77 and Con1 sequence, the perfect superposition of the cross peaks (Fig. 15.5b) proves that these three residues sample the same conformational space in both polypeptides and therefore argues against the weight that one should give to subtle differences predicted by the different algorithms

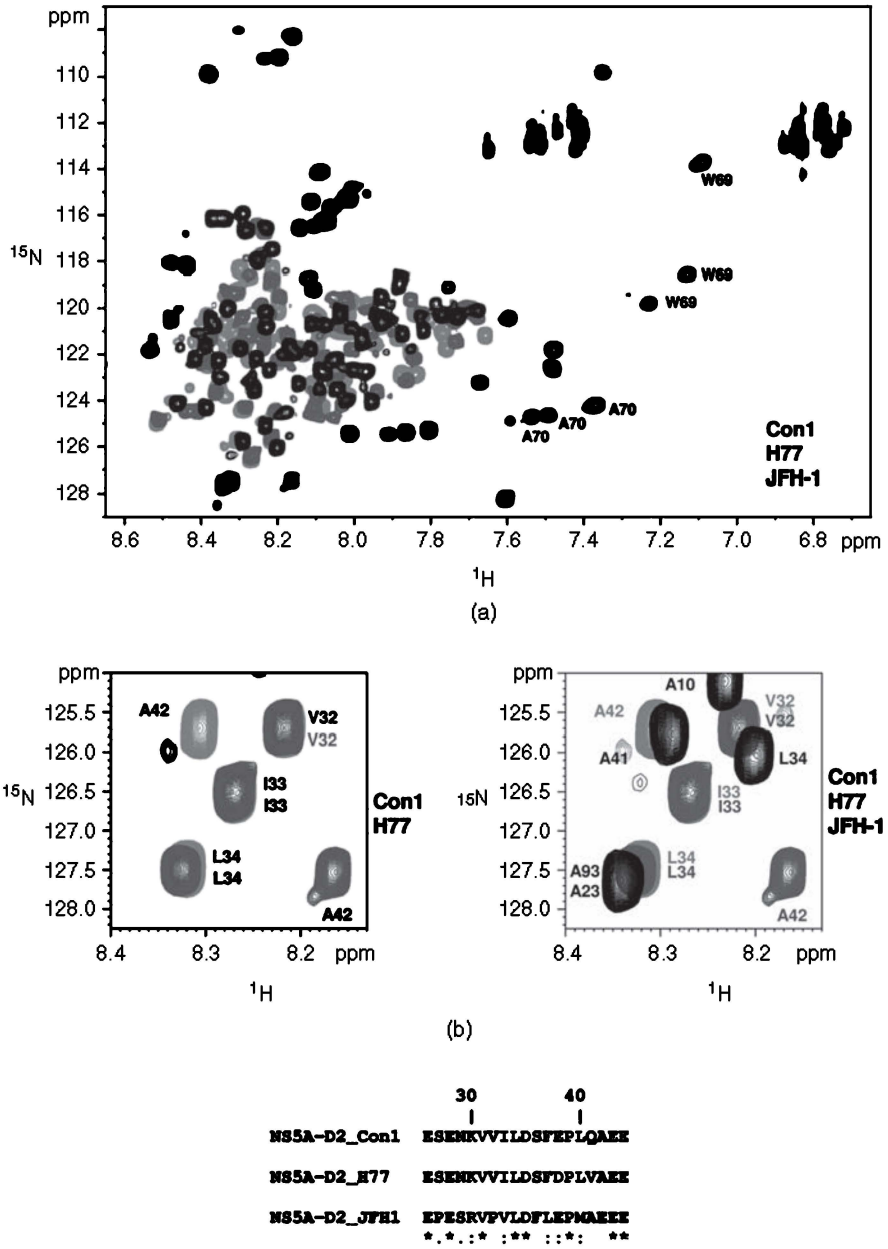


Figure 15.5 (a). Superposition of the $^{1}H,^{15}N$-HSQC NMR spectra of NS5A domain II from Con1 (red), H77 (blue) and JFH-1 (black) HCV strains. Spectra were acquired at 298 K on a 600-MHz spectrometer for Con1 and JFH-1 and at 700 MHz for H77. (b) Specific regions of these HSQC spectra showing the good overlap in the aligned sequence elements. (*See insert for color representation of the figure.*)

described above. When we do consider another residue such as A_{42}, although conserved between both sequences, the fact that it has a Q_{41} as neighbor in the Con1 sequence but a V_{41} in the H77 NS5A domain II leads to a shift that hence cannot be interpreted in terms of differential conformational sampling. The same argument is valid when we include it in the comparison of the $^{1}H,^{15}N$-HSQC spectrum of the JFH-1 domain II. L_{30} in JFH-1 can be aligned with L_{34} in the Con1/H77 sequences (Fig. 15.2) and has the same downstream aspartic acid, but a different upstream neighbor (V_{29} in JFH-1 versus I_{33} in Con1/H77). We find experimentally that both peaks do not resonate far from one another, but we cannot assign the spectral difference to a differential sampling of conformational space or to the electronic effect induced by a different neighbor.

Does this comparative analysis of closely related protein sequences constitute an experimental approach to evaluate disorder in closely related sequences? One manner to evaluate this is to apply it to two closely related proteins with well-defined secondary and tertiary structures. We have chosen the human CypA and CypB proteins, that were previously studied in our laboratory as potential interactors of the NS5A domains II (Hanoulle et al., 2009a). Sequence alignment of both proteins is high, with 64% sequence identity (107 identical residues among 166 overlapping residues). However, even when looking at peaks in identical stretches of the primary sequence, the tertiary fold does bring in different spatial neighbors that translate into differential chemical shift values (Fig. 15.6). When considering the identical FFI tripeptide, for example, located on the central strand of the β-sheet of the Cyps (Mikol et al., 1993, 1994), we note that even the F_{112} of CypA does not resonate at the same position as the corresponding F_{120} in CypB (Fig. 15.6). It should be noted that both residues are in the identical TNNGSQ$F_{112/120}$FI nonapeptide, with N_{106}D and C_{123}T being the closest substitutions. Because tertiary interactions induce chemical shift differences for residues that are in otherwise identical peptide sequences, spectral overlap can hence be considered as an experimental evidence of the absence of such interactions.

Whereas so far, we have only exploited the amide chemical shift values, another parameter observable is the carbon chemical shift values. Initially used to identify secondary structure elements in a method dubbed *chemical shift analysis* (Wishart and Sykes, 1994), comparison of the same chemical shifts to a database of random coil values can lead to the conclusion of the absence of secondary structure, which can then be interpreted in terms of disorder. A potential pitfall with the extension of the chemical shift index (CSI) method to the evaluation of intrinsic disorder is the choice of the random coil values. Different databases are indeed available in the literature and are based on either short peptides at different pH values or the chemical shift values derived from unstructured parts of full-length proteins. Whereas a deviation of 1 ppm is commonly found in bona fide secondary structure elements of folded proteins, deviations are generally smaller in IUPs and hence merit the careful consideration of the baseline. We previously reported the CSI values for the JFH-1 and H77 sequences (Liang et al., 2007; Hanoulle et al., 2009a) and show here the additional values for the Con1 NS5A domain II (Fig. 15.7). The Con1 domain is characterized by carbon chemical shift values that indicate an

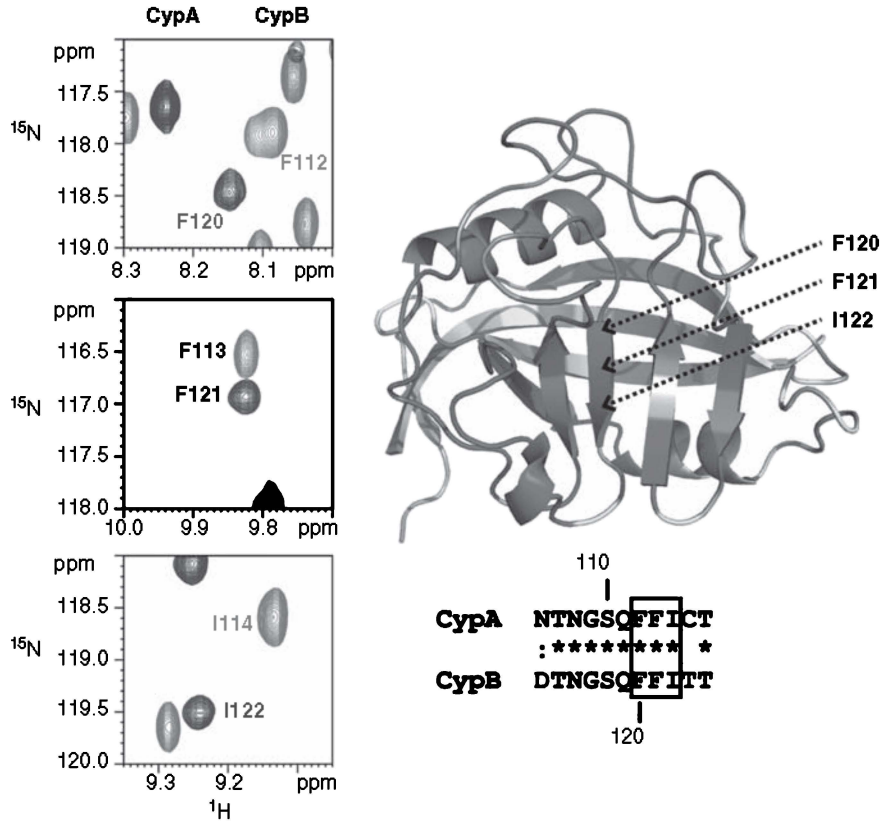


Figure 15.6 Selected regions showing the superposition of the $^{1}H,^{15}$ N-HSQC NMR spectra of CypA (red) and CypB (blue), with the sequence alignments of the peptides considered, and their localization indicated on the structure of CypB, on top of which the sequence conservation with CypA has been reported (from red, fully conserved, to light gray, no conservation). (*See insert for color representation of the figure.*)

α-helical tendency starting after Pro52 and ending with the very positively charged $R_{57}RSRK_{61}$ tetrapeptide. In the PONDR predictions, this stretch was still predicted as disordered, but a dip in the profile (Fig. 15.3a) did hint to some possible ordering. Our previous NMR analysis of the H77 NS5A domain II equally pointed to some structuring in the same zone, although rapid exchange of the amide protons of the four basic residues hampered precise determination of the $^{3}J_{HNH\alpha}$ coupling constants (Liang et al., 2007).

15.6 CONCLUSIONS

The final goal of a comparative analysis of different viral genotypes is to construct a molecular understanding of the clinical differences between the different genotypes,

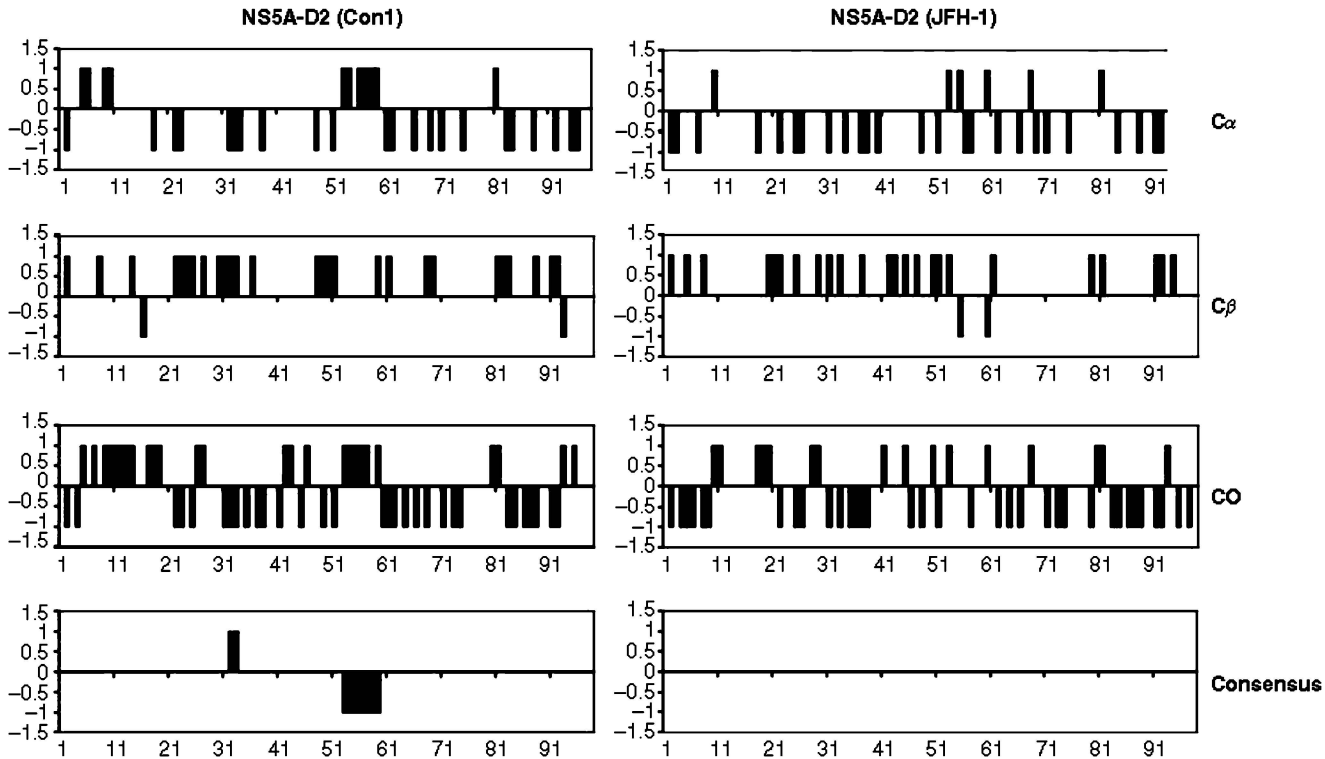


Figure 15.7 CSI analyses for the Con1 (left) and JFH-1 (right) NS5A domains II. $^{13}C\alpha$, $^{13}C\beta$, and ^{13}CO carbon chemical shifts were analyzed using CSI software (Wishart and Sykes, 1994). $^{13}C\alpha$, $^{13}C\beta$, and ^{13}CO CSI were combined to give a consensus value according to (Wishart and Sykes, 1994).

as is clearly the case with respect to the efficiency of the standard practice of care. In this chapter, we have extended this comparative analysis to the structural characterization of disordered proteins over several genotypes. The JFH-1 strain is characterized by a four-residue deletion in its NS5A domain II and equally by a higher degree of disorder compared to the two genotype 1 sequences. Whether these factors contribute to its unique capacity of producing infective particles in a cell system is under further study in our laboratories.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr F. Penin for valuable discussions and for Figure 15.1 This work was supported by CNRS, the University of Lille 1, and by a grant from the French National Agency for Research on AIDS and Viral Hepatitis. The NMR facility was supported by the CNRS, the University of Lille 1, the Pasteur Institute of Lille, the Regional Council Nord-Pas de Calais, and the Fonds Européen de Développement Régional.

ABBREVIATIONS

CD	circular dichroïsm
CSI	chemical shift index
HCV	hepatitis C virus
HSQC	heteronuclear single-quantum coherence spectroscopy
IFN	interferon
IRES	internal ribosome entry site
ISDR	interferon sensitivity-determining region
IUP	intrinsically unstructured protein
JFH-1	Japanese fulminant hepatitis 1
NMR	nuclear magnetic resonance
NS5A	Nonstructural protein 5A
RNA	ribonucleic acid

REFERENCES

Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, Friebe P, Kallis S, Engel U, Bartenschlager R. Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. PLoS Pathog 2008;4(3):e1000035.

Beaulieu PL. Non-nucleoside inhibitors of the HCV NS5B polymerase: progress in the discovery and development of novel agents for the treatment of HCV infections. Curr Opin Investig Drugs 2007;8(8):614–634.

Chatterji U, Bobardt M, Selvarajah S, Yang F, Tang H, Sakamoto N, Vuagniaux G, Parkinson T, Gallay P. The isomerase active site of cyclophilin A is critical for hepatitis C virus replication. J Biol Chem 2009;284(25):16998–17005.

Coelmont L, Kaptein S, Paeshuyse J, Vliegen I, Dumont JM, Vuagniaux G, Neyts J. Debio 025, a cyclophilin binding molecule, is highly efficient in clearing hepatitis C virus (HCV) replicon-containing cells when used alone or in combination with specifically targeted antiviral therapy for HCV (STAT-C) inhibitors. Antimicrob Agents Chemother 2009;53(3):967–976.

Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Le Mercier P, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky JM, Rice CM, Trepo C, Penin F, Deleage G. euHCVdb: the European hepatitis C virus database. Nucleic Acids Res 2007;35(Database issue):D363–D366.

Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J Clin Invest 1995;96(1):224–230.

Flisiak R, Horban A, Gallay P, Bobardt M, Selvarajah S, Wiercinska-Drapalo A, Siwak E, Cielniak I, Higersberger J, Kierkus J, Aeschlimann C, Grosgurin P, Nicolas-Metral V, Dumont JM, Porchet H, Crabbe R, Scalfaro P. The cyclophilin inhibitor Debio-025 shows potent anti-hepatitis C effect in patients coinfected with hepatitis C and human immunodeficiency virus. Hepatology 2008;47(3):817–826.

Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR, Harris M. All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. J Virol 2010;84(18):9267–9277.

Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J Jr. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;347(13):975–982.

Gale MJ, Korth MJ, Tang NM, Tan SL, Hopkins DA, Dever TE, Polyak SJ, Gretch DR, Katze MG Jr. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. Virology 1997;230(2):217–227.

Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell RA, Serrano-Wu MH, Langley DR, Sun JH, O'Boyle DR 2nd, Lemm JA, Wang C, Knipe JO, Chien C, Colonno RJ, Grasela DM, Meanwell NA, Hamann LG. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature 2010; 465(7294):96–100.

Hanoulle X, Badillo A, Verdegem D, Penin F, Lippens G. The domain 2 of the HCV NS5A protein is intrinsically unstructured. Protein Pept Lett 2010;17(8):1012–1018.

Hanoulle X, Badillo A, Wieruszeski JM, Verdegem D, Landrieu I, Bartenschlager R, Penin F, Lippens G. Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B. J Biol Chem 2009a;284(20): 13589–13601.

Hanoulle X, Verdegem D, Badillo A, Wieruszeski JM, Penin F, Lippens G. Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. Biochem Biophys Res Commun 2009b;381(4):634–638.

Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 2008;24(11):1344–1348.

Landrieu I, Hanoulle X, Bonachera F, Hamel A, Sibille N, Yin Y, Wieruszeski JM, Horvath D, Wei Q, Vuagniaux G, Lippens G. Structural basis for the non-immunosuppressive character of the cyclosporin A analogue Debio 025. Biochemistry 2010;49(22):4679–4686.

Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform* 1999;1030–1040.

Liang Y, Kang CB, Yoon HS. Molecular and structural characterization of the domain 2 of hepatitis C virus non-structural protein 5A. Mol Cells 2006;22(1):13–20.

Liang Y, Ye H, Kang CB, Yoon HS. Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. Biochemistry 2007;46(41):11550–11558.

Macdonald A, Crowder K, Street A, McCormick C, Harris M. The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. J Gen Virol 2004;85(Pt 3):721–729.

Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. J Gen Virol 2004;85(Pt 9):2485–2502.

Mikol V, Kallen J, Pflugl G, Walkinshaw MD. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. J Mol Biol 1993;234(4): 1119–1130.

Mikol V, Kallen J, Walkinshaw MD. X-ray structure of a cyclophilin B/cyclosporin complex: comparison with cyclophilin A and delineation of its calcineurin-binding domain. Proc Natl Acad Sci U S A 1994;91(11):5183–5186.

Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nat Rev Microbiol 2007;5(6):453–463.

NIH. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002–June 10–12, 2002. Hepatology 2002;36(5 Suppl 1), S3–20.

Paeshuyse J, Kaul A, De Clercq E, Rosenwirth B, Dumont JM, Scalfaro P, Bartenschlager R, Neyts J. The non-immunosuppressive cyclosporin DEBIO-025 is a potent inhibitor of hepatitis C virus replication in vitro. Hepatology 2006;43(4):761–770.

Pascu M, Martus P, Hohne M, Wiedenmann B, Hopf U, Schreier E, Berg T. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. Gut 2004;53(9):1345–1351.

Romero P, Obradovic Z, Dunker AK. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Inform Ser Workshop Genome Inform* 1997;8110–8124.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38–48.

Shirota Y, Luo H, Qin W, Kaneko S, Yamashita T, Kobayashi K, Murakami S. Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. J Biol Chem 2002;277(13):11149–11155.

Smet C, Leroy A, Sillen A, Wieruszeski JM, Landrieu I, Lippens G. Accepting its random coil nature allows a partial NMR assignment of the neuronal Tau protein. Chembiochem 2004;5(12):1639–1646.

Tellinghuisen TL, Evans MJ, von Hahn T, You S, Rice CM. Studying hepatitis C virus: making the best of a bad virus. J Virol 2007;81(17):8853–8867.

Tellinghuisen TL, Foss KL, Treadaway J. Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. PLoS Pathog 2008;4(3):e1000032.

Tellinghuisen TL, Marcotrigiano J, Rice CM. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. Nature 2005;435(7040): 374–379.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673–4680.

Wakita T, Kato T. Development of an infectious HCV cell culture system. In: Tan S-L, editor. Hepatitis C viruses genomes and molecular biology. Norfolk: Horizon Bioscience; 2006. pp.451–464.

Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994;4(2):171–180.

16

BACTERIOPHAGE λ N PROTEIN DISORDER-ORDER TRANSITIONS UPON INTERACTIONS WITH RNA OR PROTEINS

KRISTIAN SCHWEIMER AND PAUL RÖSCH

16.1 INTRODUCTION

16.1.1 Phage λ and Its Life Cycle

Simplicity and abundance of bacteriophages as compared to other biological systems make them nearly perfect model systems for the study of biological functions. Bacteriophage λ is a virus infecting the bacterium *Escherichia coli*. After the discovery that cell lysis of *E. coli* strain K12 induced by ultraviolet light irradiation was accompanied by the release of phage particles, λ was initially isolated in 1951 (Lederberg, 1951). A phage particle consists of DNA and protein in roughly equal amounts, and phage λ has a characteristic morphology with an icosahedral head of approximately 50-nm diameter and a long, flexible but noncontractile tail of about 150-nm length and 8-nm diameter. The head contains the double-stranded linear DNA of 48,502 base pairs as sequenced by Sanger and coworkers (Sanger et al., 1982). This DNA codes for around 50 proteins.

Gene regulation in phage λ was studied in great detail from the discovery of this phage, and λ indeed became the archetypal model system for this entire field of research (Hendrix et al., 1983; Ptashne, 2004). The special allure of studying this phage is the fact that λ is a member of the class of temperate phages, so that after host cell infection, two alternative pathways for phage development exist. In

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

the lytic pathway, the host cell expresses phage DNA, synthesizes phage proteins, and reassembles and releases new phage particles on bursting (lysis) of the host cell; alternatively; in the lysogenic pathway, the phage DNA is integrated into the host cell DNA where it stays in prophage state as part of the host cell's genome until transition to the lytic pathway is induced by less-favorable environmental conditions.

16.1.1.1 Lysogenic Pathway During lysogeny, the phage genome is part of the complete genome of the host cell, and, accordingly, genes causing phage maturation and cell lysis are not expressed. These mechanisms to stop phage gene expression are particularly useful for survival of the phages in situations in which the host cell is superinfected by non-λ phages. Suppression of λ gene expression is strictly regulated: High accumulation of the protein CII stops expression of the gene for protein Q, which is essential for expression of later genes that would allow phage assembly and host cell lysis. Further, the synthesis of the Int integrase is stimulated to integrate the phage DNA into the host genome. Additional λ-repressor (CI) is expressed, which represses early intermediate promoters by binding to operators oL and oR, switching off expression of other phage genes. The prophage state after integration of the phage genome into the bacterial chromosome is extremely stable through many cycles of host cell division. Damaging of host DNA, for example, by ultraviolet (UV) irradition, activates host protein RecA that proteolytically cleaves λ-repressor, resulting in deblocking of λ-genes and thus initiating the transition to lytic pathway.

16.1.1.2 Lytic Pathway During the lytic cycle, all phage genes are expressed, resulting in assembly of new phages that are released on lysis of the host cell. Virtually, perfect timing and control of gene expression is necessary to avoid premature cell lysis (Greenblatt et al., 1993). The exact timing of the sequential expression of genes involves tight transcriptional control. After initiation of transcription at promoters *pL* and *pR*, early genes are transcribed and the corresponding proteins accumulate. Initially, transcription is attenuated at terminators *tL* and *tR*. These terminators act as check points as the nascent transcript and the template DNA are released from the transcription elongation complex (TEC), preventing the transcription of the full set of phage genes and allowing control of expression of late genes. The essential protein in this early phase is the antitermination λN. After its expression, λN forms a stable complex with the RNA recognition site *nut* (N utilization), the RNA polymerase (RNAP), and a set of host proteins, the Nus factors (Nus, N utilization substances). In this antitermination complex, the RNAP is able to pass the termination signals and thus transcribe genes downstream of the termination signals. This allows the transcription of the *O* and *P* genes, which are required for replicating the phage DNA and including the *Q* gene. *Q* is a second antiterminator and allows transcription to pass terminator tR' and cause expression of the late genes necessary for bacteriophage morphogenesis and host cell lysis.

16.2 THE ANTITERMINATION COMPLEX

The antitermination complex is a stable ribonucleoprotein complex allowing transcription through intrinsic and factor-dependent terminators far downstream of the *nut* site, where the complex assembles (Fig. 16.1a). The key interaction seems to be the binding of λN to RNAP as excess of N is able to transform RNAP *in vitro* to a form that is able to transcribe through proximal terminators (Rees et al., 1996). This interaction, however, is not sufficiently stable to allow processive transcription. In fact, to form a stable antitermination complex, the host cell Nus factors are required (Das, 1993). NusA interacts directly both with the RNAP and the λN, NusB forms a heterodimer with NusE (identical to ribosomal protein S10), and NusG links this heterodimer to RNAP via its amino-terminal domain and to the ribosome via its carboxy-terminal domain, thus physically coupling transcription

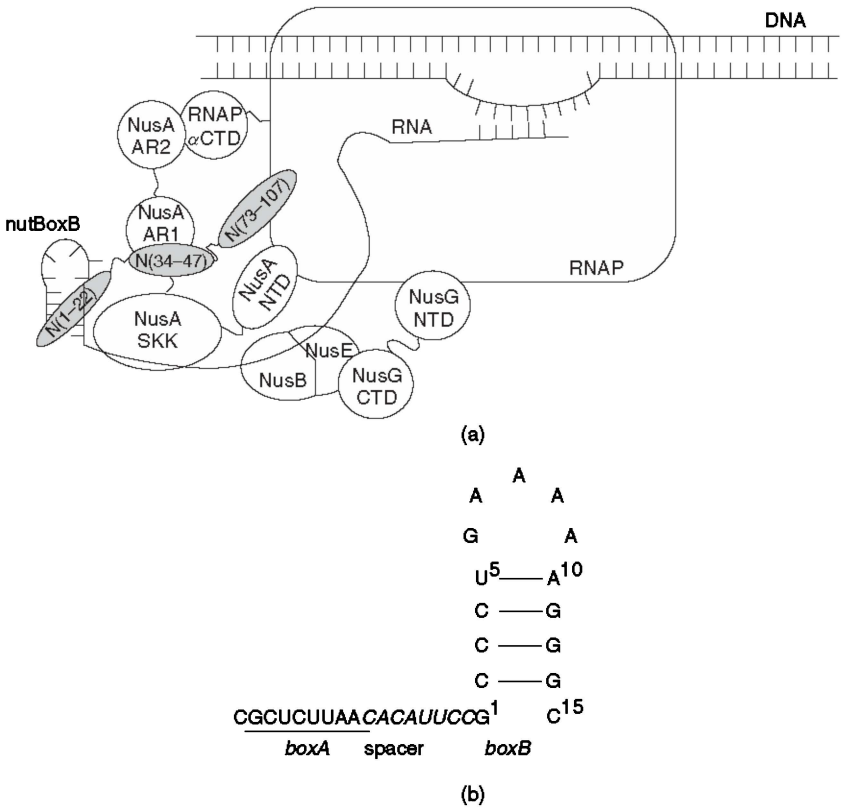


Figure 16.1 (a) Antitermination complex. A schematic summary of protein–protein and protein–nucleic acid interaction within the antitermination complex. λN (gray) is labeled with the different sequence regions interacting with their corresponding binding partners. (b) Sequence of the *nutR* site of bacteriophage λ. *BoxA* interacts with NusB–NusE heterodimer, the spacer binds to NusA, and *boxB* forms the complex with λN.

and translation (Burmann et al., 2010). *nut* RNA (Fig. 16.1b) is suggested to initially assist complex formation by tethering λN to the RNAP. The *nut* site interacts with the λN (Lazinski et al., 1989) via *BoxB*, a stem-loop RNA structure within the *nut* RNA sequence. NusB (Luo et al., 2008) binds to *nut* at the *nutBoxA*, and NusA attaches to the spacer region between *nutBoxA* and *nutBoxB* (Prasch et al., 2009). The interactions within the antitermination complex cannot be pinpointed to a few strong interactions as all protein–protein interactions are only weak, resulting in a highly cooperative assembly of the complex (Greenblatt et al., 1993), and several components are not even connected by a direct interaction, rendering a well-tuned network essential for complex formation as exemplified by the interaction chains RNAP-N-*nutBoxB*, RNAP-N-NusA/RNAP-NusA-N, *nutBoxA*-NusB/NusB-NusE-NusG-RNAP. These numerous interactions explain the overall complex stability in spite of the lack of high affinity binding between single partners. Recent results allow a detailed model of action: NusA binds close to the RNA exit channel of RNAP (Toulokhonov et al., 2001; Ha et al., 2010). During regular termination, NusA facilitates formation of the termination hairpin by interaction with the hairpin's left shoulder (Gusarov and Nudler, 2001). After termination hairpin formation, the TEC dissociates rapidly (Gusarov and Nudler, 1999). Based on cross-linking data, it is suggested that λN suppresses hairpin formation by interaction with the ascending portion of the RNA stem (Gusarov and Nudler, 2001; Nudler and Gottesman, 2002) and that λN promotes formation of additional contacts between NusA and nascent RNA. It is suggested that in this case, NusA interacts with RNA differently than in the formation of the termination hairpin (Gusarov and Nudler, 2001), and different RNA-binding regions of NusA may be used in either situation.

Nus factors, however, convey stability to the antitermination complex. In the standard situation of transcription, the function of Nus factors is modulated as compared to their function in the antitermination complex. Indeed, Nus factor function is sometimes changed to its opposite on formation of the antitermination complex. The C-terminal domain (CTD) of NusG interacts with termination factor rho, but it interacts with NusE in the antitermination complex, suggesting regulation of the transition between elongation and termination by competition between NusE and rho binding to NusG (Burmann et al., 2010). Also, the standard effect of NusA during elongation is to increase the rate of termination at intrinsic terminators (Farnham et al., 1982; Schmidt and Chamberlin, 1987; Gusarov and Nudler, 2001). In N-mediated antitermination, however, this function is just reversed, showing that the Nus factors do not per se comprise the switch between termination and antitermination, but are themselves subject to switching processes such as λN–nut RNA complex formation.

16.3 ANTITERMINATION PROTEIN N

Early studies identified λN as an important factor for transcribing phage genes as the absence of λN reduces transcription of the phage λ genome to about 2%, with the expression limited to early genes (Roberts, 1969). In its key role in the assembly

of the antitermination complex, λN is a classical example of a positive regulator of transcription, stimulating the expression of genes located downstream of termination signals. The translation of λN is affected by an autoregulatory mechanism in which λN represses the expression of its own gene, a process that requires not only binding to *nutL* but also formation of the complete N-modified transcription complex (Wilson et al., 1997; Wilson et al., 2004).

After sequence determination and initial isolation of λN (Franklin and Bennett, 1979; Greenblatt et al., 1980), analogous proteins were found in the λ-related phages p21 and p22, and these proteins are also involved in the lysis-lysogenesis transition. λN is a monomeric protein of 107 amino acids (molecular mass 12.2 kDa; Fig. 16.2). About 22% of the amino acids are arginine and lysine, resulting in a high basicity of the protein (pI 11.5). λN's *in vivo* half-life is only 2 min because of degradation of the protein by the *E. coli* endoprotease Lon (Gottesman et al., 1981), but the protein is heat stable *in vitro* (Greenblatt and Li, 1982). The degradation by Lon protease allows posttranslational regulation and possibly regulates the excision of the λ prophage from the *E. coli* genome during transition from lysogenic to lytic development (Gottesman et al., 1981).

Analysis of the charge–hydrophobicity phase space of several intrinsically disordered proteins and comparison with those of small globular proteins indicated that the former exhibits a large net charge and low overall hydrophobicity (Uversky et al., 2000) and clustered well outside the region of globular proteins in a two-dimensional map that correlates net charge and hydrophobicity. The net charge of λN of +14 and its mean hydrophobicity of 0.3968 positions this protein far outside the region of folded globular proteins, so that a highly flexible, intrinsically disordered state is suggested already on the basis of its sequence. The amino-terminal region shows the highest content of arginines, the percentage of arginines and lysines increases to 32% for the first 22 amino acids, suggesting that especially this region does not adopt a stable fold.

Overall, the protein displays large structural asymmetry as determined by sedimentation and gel filtration experiments (Greenblatt and Li, 1982). Neither full-length λN nor short peptides derived from the λN sequence show characteristic features of secondary structure in circular dichroism (CD) spectra (Van Gilst and von Hippel, 1997; Van Gilst et al., 1997a), and 28% α-helical conformation was detected by CD spectroscopy only at 4°C (Tan and Frankel, 1995).

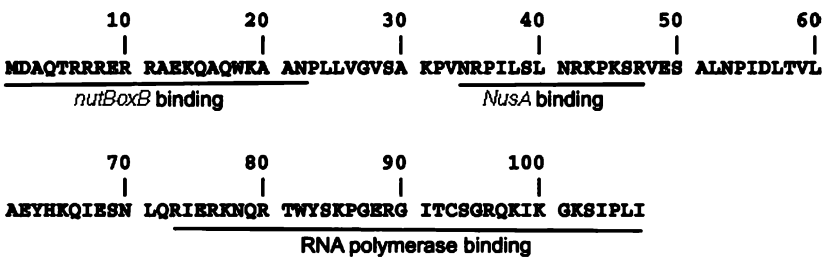


Figure 16.2 Sequence of antitermination λN. The binding sites are labeled and underlined.

For λN, only a narrow chemical shift dispersion was observed in nuclear magnetic resonance (NMR) experiments (Van Gilst et al., 1997a; Mogridge et al., 1998). The dispersion of resonance frequencies of a protein's ^{1}H, ^{15}N, and ^{13}C spins is characteristic of its local fold, with low dispersion indicating the absence of a well-defined local structure and high dispersion indicating the opposite. This easily observed effect is a consequence of differences in the local electronic environment of amino acids caused by tertiary structure formation, for example, by formation of stable hydrogen bonds or defined reorientation of side chains relative to magnetically anisotropic groups such as aromatic rings or carbonyl groups. As a consequence, the shielding of the individual nuclei from the external field differs, and the resulting chemical shifts are well dispersed so that the individual residues can be discerned by their characteristic resonance frequencies. In contrast, in highly flexible peptide chains, the environment for all residues is highly similar and the resulting chemical shift dispersion is poor. Another NMR parameter that may lead to conclusions about a peptide's or a protein's flexibility is ^{15}N spin relaxation, which critically depends on the timescale of rotational reorientation of the $^{15}N-^{1}H^{N}$ bond vectors relative to the static magnetic field. For folded globular proteins, this timescale is governed by the overall rotational reorientation of the folded proteins, characterizing the proteins as stable on this timescale (10^{-9} s). For disordered peptide chains, the reorientation of the $^{15}N-^{1}H^{N}$ bond vectors depends on local dynamics as opposed to overall reorientation, and the former occurs on a considerably faster timescale ($<10^{-9}$ s). Thus, by analyzing ^{15}N relaxation properties, the dynamic features of peptide chains can be derived. NMR relaxation experiments showed that the $N-H^{N}$ bond vector reorientation of backbone amides of λN occurs on a faster timescale than typically found for folded globular proteins of identical size (Prasch et al., 2006). This data is consistent with the notion of λN being a disordered protein, and the full *nutBoxB* RNA-binding capacity of this protein demonstrates that this disordered form is an active and biologically significant state.

Various interaction studies between λN and binding partners identified three binding regions for different partners. The λN arginine- and lysine-rich amino terminus, Met1–Asn22, binds *nutBoxB* RNA; the stretch Asn34–Arg47 is responsible for NusA binding, and residues Arg73–Ile107 form the binding site of RNAP (Van Gilst and von Hippel, 1997; Mogridge et al., 1998). Whereas detailed structural models exist for the first two interactions (Legault et al., 1998; Scharpf et al., 2000; Bonin et al., 2004; Prasch et al., 2006), the binding of λN to RNAP, which is expected to be key in antitermination, is not characterized yet at the atomic level.

16.4 THE λN–RNA INTERACTION

16.4.1 ARM Peptide–RNA Binding

The arginine-rich motif (ARM) as present in the amino terminus of λN is a prominent feature of several viral RNA-binding proteins including and Tat proteins from equine infectious anemia virus (EIAV), human immunodefiency virus (HIV), and bovine immunodeficiency virus (BIV) and HIV1-Rev (Lazinski et al., 1989;

Willbold et al., 1994; Bayer et al., 1995; Tan and Frankel, 1995; Metzger et al., 1997), and the basic arginine residues of the ARM are important for specific interactions. A long flexible aliphatic side chain together with the charged guanidinium group makes arginine special among amino acids (Patel, 1999). The guanidinium allows formation of intermolecular salt bridges and hydrogen bonds that combine with van-der-Waals interactions, permitting arrangements of several arginines in small sequence stretches to form RNA-binding sites with high sequence specificity and high affinity. The function of ARMs does not depend on specific defined secondary or tertiary structure environments in the isolated protein, but defined, rigid structures are exhibited in the context of complex formation with target RNA molecules (Patel, 1999). In general, these are examples of structures not solely defined by the sequence, but induced by the molecular environment provided by an interaction partner.

The structural studies of RNA binding to ARM-containing proteins substantially changed the view of protein–RNA recognition (Patel, 1999). Until then, structural data had indicated that proteins provide surface binding pockets optimized for interaction with minimal elements of nucleic acid architecture. The components undergo only slight structural changes to improve packing and defined interaction at the interface. The complexes of ARM-containing peptides with their cognate RNAs demonstrate that the RNA tertiary structure forms a scaffold that allows wrapping around the peptide. In these cases, the peptides represent the minimal structural binding units of the underlying proteins. These peptides are disordered in solution, but adopt a defined conformation on complex formation. A variety of structures of ARM peptide–RNA complexes that have been determined show that the peptides do not fold into identical structures. Helix formation is found in *nutBoxB*–RNA complexes of λN as well as HK022 Nun (Cai et al., 1998; Legault et al., 1998; Scharpf et al., 2000; Faber et al., 2001) and in the Rev–RRE complex from HIV (Battiste et al., 1996), while BIV Tat adopts a hairpin structure after binding to BIV TAR RNA (Puglisi et al., 1995; Ye et al., 1995). In contrast, binding of peptides derived from HIV Rev and human T-cell leukemia virus type I Rex to designed aptamers results in an extended conformation of the peptide in the bound state (Jiang et al., 1999; Ye et al., 1999).

16.4.2 λN-nut*B* ox*B* Interaction

The role of λN depends on a small site of the transcribed RNA, the *nut* site (Fig. 16.1b). This site consists of two elements, *nutBoxA* and *nutBoxB* (Whalen and Das, 1990; Nodwell and Greenblatt, 1991; Chattopadhyay et al., 1995a; Mogridge et al., 1995). *nutBoxB* forms an RNA hairpin with a stem of five base pairs and a loop of five nucleotides. In solution, *nutBoxB* RNA shows significant flexibility. Analysis of low energy CD spectra of *nutBoxB* RNA in the absence of λN suggests that in spite of this high overall flexibility, a significant fraction of the RNA adopts defined conformations that share characteristic interactions of bases found in the peptide–RNA complex (Johnson et al., 2005). The RNA thus already samples binding competent conformations in the free state. λN binds to *nutBoxB* with an affinity

virtually identical to that of shorter peptides containing only the amino-terminal 22 residues (Dissocation constant 5 nM) (Chattopadhyay et al., 1995a; Tan and Frankel, 1995; Cilley and Williamson, 1997; Van Gilst et al., 1997b). Therefore, the RNA-binding site of λN is completely localized in the amino terminus, and the rest of the protein does not influence binding. No conformational changes were observed outside the ARM region on interaction of λN with *nutBoxB* RNA (Scharpf et al., 2000). CD spectroscopy demonstrates an α-helical conformation of λN in its *nutBoxB* complex (Tan and Frankel, 1995; Su et al., 1997a; Van Gilst et al., 1997b; Scharpf et al., 2000). Helix formation is only induced on specific binding to *nutboxB* RNA, while λN remains unstructured when nonspecifically bound to other RNAs (Van Gilst et al., 1997b). Specificity can be influenced by single-nucleotide exchanges, and the lack of λN structure in nonspecific complexes is accompanied by loss of λN-dependent antitermination activity (Van Gilst et al., 1997b), demonstrating the importance of induced protein structure for biological activity.

Large chemical shift changes in the NMR spectra of λN as well as *nutBoxB* RNA accompany these major structural rearrangements on complex formation (Mogridge et al., 1998). The dramatically increased dispersion of the signals is characteristic for the formation of a well-defined and stable three-dimensional structure. By using peptides corresponding to λN(1–22) and λN(1–36) and by applying multidimensional heteronuclear NMR spectroscopy, the structure of λN–*nutBoxB* could be determined (Legault et al., 1998; Scharpf et al., 2000) (Fig. 16.3). *nutBoxB* RNA adopts a GNRA fold in the pentaloop with a looped-out base (A9) and a regular A-form stem region. GNRA RNA structures are the basic building blocks of RNA hairpin structures interacting with proteins, and they are known for their remarkable stability (Correll and Swinger, 2003). The λN peptides form a bent α-helix, with helical segments λN(4–10) and λN(12–21) at an angle of about 120° There are no significant deviations from the ideal ϕ/ψ angles, and backbone ϕ/ψ angles

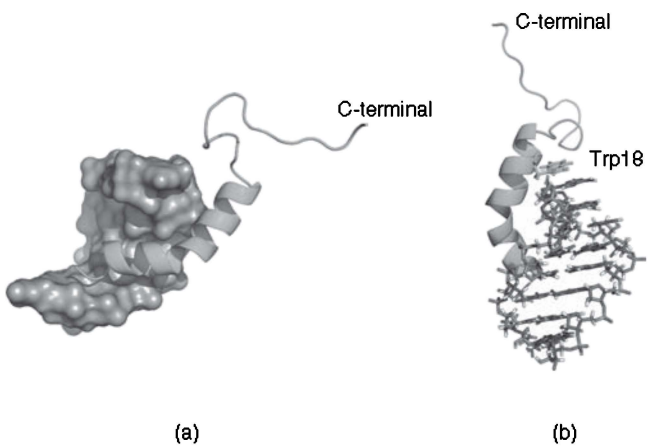


Figure 16.3 The λN(1–36)–*nutBoxB* complex. (a) peptide, cyan; RNA, surface model. (b) RNA, sticks representation, Trp18 of λN, stick representation, highlighting the stacking interaction.

of Arg11 show that the bend results from a single perturbation rather than additive effects along the peptide backbone. At the amino terminus, the α-helix is capped by intramolecular hydrophobic interaction of the methyl group of Met1 with side chains of Thr5 and Arg6. The helix bend that is further stabilized by a hydrogen bond between the side chain of Gln15 and the carbonyl oxygen of Arg8 maximizes λN–*nutBoxB* surface complementarity as exemplified by the essential stacking of Trp18 and A7 of *nutBoxB*. The first helical segment interacts with the major groove of the RNA, but the intermolecular interactions involve only 5′ residues of the RNA, consistent with the observed protection of exclusively 5′-phosphates from ribonuclease degradation (Chattopadhyay et al., 1995a). An important interaction is the packing of the methyl group of Ala3 between nucleotides C2 and C3. This anchors the first helical segment in the major groove. Substitution with larger amino acids (Val, Thr) results in a dramatic decrease of binding affinity (Su et al., 1997a) together with a shortening of the helix at the N-terminus (Scharpf et al., 2000) The arginines and lysines create a positively charged surface on one side of the α-helix to interact with the negatively charged phophodiester backbone of the RNA. Except Lys19, all amino and guanidinium groups are less than 5 Å away from RNA phosphates, allowing electrostatic interactions and possibly hydrogen bonds. The complex structures show fewer intermolecular contacts for Arg6 and Arg10 in comparison to Arg7, Arg8, and Arg11. This explains their importance for binding and activity in antitermination assays as shown in mutation studies (Franklin, 1993; Su et al., 1997a). Several amino acid–RNA interactions are observed in the loop region. Mutations within the *nutBoxB* loop influence the formation of the GNRA tetraloop, for example, hindering the formation of the sheared G–A base pair by mutation of G6 or A10 (Chattopadhyay et al., 1995a; Mogridge et al., 1995; Tan and Frankel, 1995; Cilley and Williamson, 1997; Su et al., 1997b; Van Gilst et al., 1997b), while deletion of the looped-out base reduces the affinity of *nutBoxB* for λN only slightly (Legault et al., 1998). This demonstrates the importance of a well-defined loop conformation for λN recognition.

A characteristic interaction of the λN–RNA complex is the stacking of Trp18 and the base of A7 at the top of the tetraloop. This interaction extends the RNA π-stack by one residue. It was demonstrated that Trp18 is important for RNA binding and antitermination (Franklin, 1993; Su et al., 1997a). Femtosecond optical spectroscopy revealed local dynamics around Trp18 and interconversion between two states on very short (picosecond) timescales (Xia et al., 2003a). Mutations were identified that destabilize this stacking interaction (Xia et al., 2003b). A double mutation Lys14Glu and Gln15Arg still accommodated binding of N peptide to *nutBoxB* RNA, but without Trp18 stacking, as revealed by chemical shift analysis of the tryptophan HE1 resonance by proton NMR spectroscopy as well as by femtosecond fluorescence techniques (Xia et al., 2003a,b, 2005). Analysis of a variety of mutations concerning Trp stacking, affinity to RNA, and antitermination activity demonstrates that biological function directly correlates with this stacking interaction (Xia et al., 2003a,b, 2005). Functionality of a several 100-kDa machinery thus critically depends, among others, on the side chain orientation of a single Trp residue.

```
                10        20
                 |         |
λ     MDAQTRRRERRAEKQAQWKAAN
Nun   LTSRDRRRIARWEKRIAYALKN
P22   GNAKTRRHERRRKLAIERDTI
φ21   GTAKSRYKARRAELIAERRSNE
```

Figure 16.4 Sequence alignment of amino-terminal sequences of the N proteins of bacteriophages λ, P22, and ϕ21, and ARM of the Nun protein of phage HK022.

16.4.3 HK022 Nun-nut*B*oxB Interaction

Typically, the interaction between ARM proteins from λ-like phages (Fig. 16.4) and their target RNAs displays high sequence specificity (Tan and Frankel, 1995). The non-λN proteins recognize λ *nutBoxB* RNA with significantly reduced affinity. This situation is completely different in the case of the termination factor Nun from coliphage HK022. Nun excludes superinfection of HK022 lysogens by λ by promoting transcription termination at or close to λ *nut* sites (Robert et al., 1987) and interacts with *nutBoxB* RNA in a similar manner as λN (Chattopadhyay et al., 1995c). Nun is a 109-amino-acid protein that carries an ARM in the amino-terminal half, sequence region 20–44. Nun and λN bind to λ *nutBoxB* with comparable affinities and are able to compete with each other (Chattopadhyay et al., 1995b; Faber et al., 2001). Nun utilizes the same bacterial host factors as λN, NusA, NusB, NusE, and NusG for assemblage of a TEC, which is, however, modified to act opposite to the λN-containing complex (Das, 1993; Greenblatt et al., 1993). The structure of a complex of a Nun peptide with *nutBoxB* RNA has been solved by NMR spectroscopy (Faber et al., 2001). Like λN, the peptide adopts a bent α-helical conformation, and an aromatic residue (Tyr39) stacks on the base of A9 of the RNA loop. A remarkable difference is the presence of a large solvent-exposed hydrophobic surface. The helix is amphipathic, and the charged residues point to the RNA for specific interactions, while the hydrophobic residues are exposed to the opposite site. It is suggested that this hydrophobic surface forms a protein–protein interaction site, but the binding partner is unknown. The stacking interaction essential for the antitermination function of λN is not important for the termination function of Nun (Burmann et al., 2008). The difference in Nun function compared to λN is attributed to interactions of the carboxy-terminal region with the TEC and arrest of transcription by anchoring RNAP to DNA (Nudler and Gottesman, 2002).

16.4.4 Other N–*B*ox*B* Complexes

Solution structures were determined for the homolog N–*nutBoxB* complexes of lambdoid phages P22 (Cai et al., 1998) and ϕ21 (Cilley and Williamson, 2003). Together with the λN–RNA complex, these structures allow detailed insight into determinants of RNA recognition by different amino acid sequences of N proteins.

In both complexes, the peptides form helical structures with varying degrees of bend and interact with the RNA hairpin loop and the 5′ residues of the RNA stem. The RNA loop of ϕ21 *nutBoxB* contains one additional nucleotide compared to the λ and P22 RNA sequences, and it forms a hexaloop with a distinct U-turn fold. P22 *nutBoxB* differs at the third position of the loop (C instead of A) from the corresponding λ RNA sequence. In the complex with P22-N, the five nucleotides of the loop form a GNRA tetraloop by flipping out one base, but here, the third base is flipped out (3-out) and makes extensive contacts with hydrophobic residues of the P22-N peptide. The specificity for an RNA loop conformation in the complex is dictated by the peptide. In the noncognate λN–P22 *nutBoxB* complex, a 4-out loop conformation is observed (Austin et al., 2003). P22-N peptide causes broadening in NMR signals, consistent with conformational dynamics in the RNA loop. Mutational studies reveal that Gln4 and especially Arg8 are the important residues affecting the RNA loop conformation. Arg8 interacts with the phosphates of A9 and A10. This interaction is only possible in the 4-out loop.

16.5 NusA–λN INTERACTION

16.5.1 NusA

NusA is an essential bacterial transcription factor involved in regulatory processes including transcription pausing and termination (Greenblatt and Li, 1981). *E. coli* NusA is a multidomain protein with a molecular mass of 55 kDa. Six domains are arranged in linearly along the sequence (Mah et al., 1999) (Fig. 16.5a). The amino-terminal domain is responsible for binding to RNAP, and the three subsequent domains represent typical RNA-binding modules (S1, KH1, KH2). The carboxy-terminus forms two domains (acidic repeats AR1 and AR2) that are found in NusA from *E. coli* and other γ-proteobacteria only, but are missing in NusA from all other known bacteria.

In spite of the presence of characteristic RNA-binding domains, wild-type *E. coli* NusA does not interact with RNA, in contrast to AR2 deletion mutants, a behavior suggesting autoinhibition by AR2 masking of the RNA-binding surface. This autoinhibition is released by binding of RNAP αCTD to AR2 (Mah et al., 2000; Schweimer et al., 2011). The binding site for λN could be located in the AR1-AR2 tandem (Mah et al., 1999), and the minimal fragment of λN that binds NusA was found to be the sequence stretch Asn34–Arg47 (Mogridge et al., 1998).

Crystal structures are available for *Thermotoga maritima* and *Mycobacterium tuberculosis* NusA (Gopal et al., 2001; Shin et al., 2003; Worbs et al., 2001) (Fig. 16.5b). A structure of *M. tuberculosis* NusA complexed with single-stranded RNA derived from the *boxC* recognition site required for antitermination in bacterial ribosomal operons was solved by X-ray crystallography (Beuth et al., 2005). This structure reveals RNA binding in an extended manner exclusively to the KH domains. No significant structural rearrangement of the NusA domains was observed, suggesting that the RNA-binding interface of NusA acts as a rigid scaffold for RNA recognition (Beuth et al., 2005).

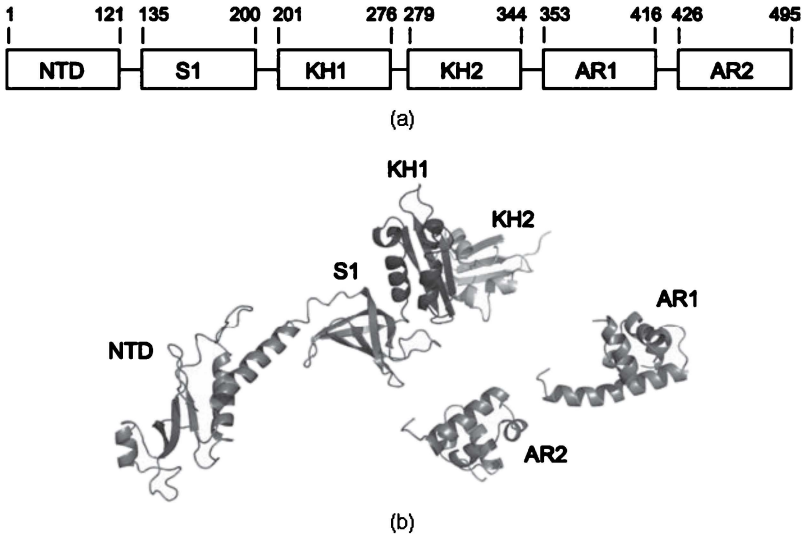


Figure 16.5 (a) Domain organization and structural model of *E. coli* NusA. (b) Structural model of *E. coli* NusA based on the crystal structure of *Thermotoga maritima* (NTD, S1, KH1, and KH2 domains) and solution structures of *E. coli* NusA AR1 and AR2.

E. coli AR1 and AR2 share the same double helix-hairpin-helix $(HhH)_2$ fold (Eisenmann et al., 2005) (Fig. 16.5b). In this structure, two helix-hairpin-helix motifs are packed perpendicularly against each other, forming a compact globular fold. The structure of the NusA acidic repeats deviate only slightly in orientation and length of helices from the *sterile alpha* motif, a well-known protein–protein interaction module. The overall structural similarity of AR1 and AR2 is consistent with their sequence homology of 61%, but the distribution of polar, charged, and hydrophobic residues at the protein surfaces clearly differs. Analysis of rotational reorientation by NMR spin relaxation measurements and different alignment tensors in liquid crystal media as found by analysis of residual dipolar couplings demonstrate that the domains do not interact and that the connecting linker is highly flexible (Eisenmann et al., 2005). NMR titration experiments of isotopic-labeled NusA AR1-AR2 tandem with λN(1–53) peptide did not allow the delineation of the contact interface because of widespread chemical shift changes. Addition of RNAP αCTD to the AR1-AR2 tandem broadened signals from residues of AR2 beyond detection, but left signals from AR1 unaffected, confirming exclusive AR2 binding to αCTD. The NMR titration experiments demonstrate the different binding specificity of AR1 and AR2 despite their high structural similarity.

16.5.2 Structure of a NusA–λN Peptide Complex

Crystallization of the AR1-AR2 tandem in the presence of a λN(34–47) peptide shows two AR1 molecules interacting with the λN peptide, probably a result of

unwanted proteolytic separation of AR1 and AR2 (Bonin et al., 2004). Mutations of individual aminoacids of λN(34–47) to alanine showed that only residues pointing to one of the two AR1 molecules of the crystal structures are essential for binding. Calorimetry reveals the 1:1 binding stoichiometry with a dissociation constant of 3.5 μM (Bonin et al., 2004).

The crystal structure of AR1 in complex with N(34–47) (Bonin et al., 2004) (Fig. 16.6a) superimposes well with the solution structure of AR1 alone (backbone atomic RMSD: 1.2 Å), demonstrating that AR1 acts as a rigid scaffold conferring additional stability to the complex. For λN(34–47), electron density was seen up to residue Leu40 only, but the affinity to AR1 of a peptide spanning residues 31–43 is nearly the same as that of λN(34–47), suggesting that the carboxy-terminal region does not play an essential role in binding. The complex shows a buried interaction surface of 850 $Å^2$. λN(34–47) binds to AR1 as an extended peptide, maximizing the interaction surface with AR1. Single mutations to Ala resulted in loss of binding, demonstrating the importance Arg35, Ile37, Leu38, and Leu40. Arg35 points deep into a crevice of AR1, forming a salt bridge with Asp364, and hydrogen bonds to Arg409. Ile37, Leu38, and Leu40 form important hydrophobic

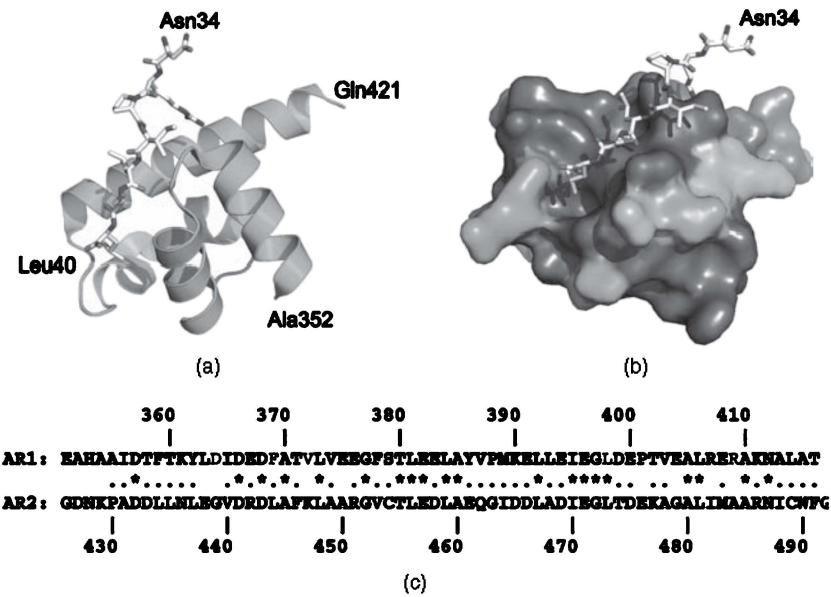


Figure 16.6 The λN–AR1 interaction. (a) Cartoon representation of λN(34–47)–AR1 complex. Residues 41–47 of λN are not shown because of lack of corresponding electron density in the crystal structure. (b) Mapping of λN(34–47)–AR1 interaction on the crystal structure of λN(34–47)(sticks)–AR1(surface) by NMR spectroscopy. Chemical shift changes of AR1 on titration with λN(34–47) are grouped in three categories (weak, light green; medium, green; and strong, dark green) and color coded on AR1 surface. (c) Sequence comparison of AR1 and AR2. Identical residues are labeled by (*) and similar residues are labeled by (·). Residues of AR1 which are important for λN binding are colored gray. (*See insert for color representation of the figure.*)

interactions, for example, Leu40 stacks with the aromatic ring of Phe369 and shows contacts with Val372 and Leu398.

The crystal structure is in line with previous biochemical data. But because of the proteolytically separated AR1 and AR2 domains and the lack of electron density for residues Asn41–Arg47, the role of this sequence stretch of λN and AR2 for complex formation remained undefined. Based on the sequence similarity of AR1 and AR2 as well as the sequence similarity of λN(41–47) to λN(34–40) and the higher binding affinity of Asn34–Arg47 compared to Lys31–Lys43, it was suggested that residues Asn41–Arg47 bind to AR2 in a manner similar to that by which Asn34–Leu40 binds to AR1 (Bonin et al., 2004). An NMR study was performed to answer this question (Prasch et al., 2006). Because of nonideal behavior of full-length λN at the high concentration required for NMR spectroscopy, a peptide containing the ARM and NusA-binding region, λN(1–53) was used. The unbound peptide shows narrow chemical shift dispersion as well as high flexibility of amide bond vectors on the subnanosecond timescale, detected by the heteronuclear steady-state Nuclear Overhauser Effect (NOE) experiment, which are typical features of disordered proteins. On titration of this peptide with the AR1-AR2 tandem, significant chemical shift changes were observed. Methyl groups of Ile37 and Leu40 showed remarkable upfield shifts, which are typically found in proteins with defined tertiary structure. Also, amide NMR resonances ($^{1}H^{N}$ and ^{15}N) exhibit large changes in the region Asn34–Arg47, consistent with all previous data and indicating that the sequence stretch Asn41–Arg47, which was not observed in the crystal structure, may also contribute. The heteronuclear steady-state NOE shows an increase of rigidity on the subnanosecond timescale. These data demonstrate the disorder-order transition of λN on binding to NusA.

To answer the question whether or not AR2 is involved in binding, isotope-labeled AR1-AR2 tandem was titrated with λN peptides so that effects on NMR parameters on complex formation could be unambiguously attributed to the individual domains. To avoid any effects by the positively charged ARM region on the negatively charged acidic repeats, synthetic peptides spanning only the region Asn34–Arg47 of λN were used. On titration of AR1-AR2 tandem with N(34–47), significant chemical shift changes were observed only for residues located in AR1. With the peptide λN(40–47), only small chemical shift changes were observed, and again only for residues of AR1. The NMR titrations confirmed the 1:1 binding stoichiometry, and a dissociation constant of 10 μM was derived from chemical shift changes, reproducing the data determined by isothermal calorimetry (Prasch et al., 2006). From this data, it is clear that AR2 does not play a role in binding of λN.

Mapping the chemical shift changes on the surface of AR1 (Fig. 16.6b) revealed a continuous binding region containing the already known binding region of the crystal structure, but chemical shift changes were also observed for residues located in the carboxy-terminal part of helix 3 and the hairpin between helices 4 and 5. It was not possible to derive any intermolecular distance restraints based on Nuclear Overhauser Effect spectroscopy (NOESY) experiments to determine a detailed structural model for this region. This can be attributed to enhanced flexibility on

fast to intermediate (picoseconds to microseconds) timescales, which reduces NMR signal intensity. This is consistent with the fact that residues Asn34–Leu40 contribute the most relevant amino acids for binding. The structural change of residues Asn41–Arg47 observed in the NMR data together with their low contribution to binding might be interpreted only as a secondary effect of the AR1–λN interaction.

16.5.3 Binding Specificity of NusA Acidic Repeats to λN

The high sequence similarity (Fig. 16.6c) between AR1 and AR2 (31% sequence identity, 60% sequence homology) and their identical fold suggest that only subtle differences between the two domains allow λN recognizing AR1, but not AR2. Asp364 was identified as an important amino acid in AR1 as it forms a salt bridge with Arg35. In AR2, Gly439 makes formation of the equivalent salt bridge impossible. The hydrophobic patch on AR1 necessary for interaction with λN Leu40 is composed of Phe369, Val372, and Leu398. In AR2, the amino acid equivalent to Val372 is Lys447. This charged residue disrupts the hydrophobic patch on AR2, and stable interaction with Leu40 seems unlikely. While sequence similarity of the two acidic repeats is manifested in the same overall structure, binding specificity is determined by differences in amino acid sequence together with the resulting subtle structural changes. Up to now, no function of AR1 outside the antitermination complex is known. AR2 interacts with αCTD of the RNAP; probably, binding of λN to AR2 would render transcriptional processes unproductive. The similarity of AR1 and AR2 suggests gene duplication during evolution, but whereas AR2 acts as autoinhibitory domain and regulates transcriptional processes by interaction with αCTD, AR1 might just be a linker to place AR2 in correct orientation to the other domains of NusA.

16.5.4 Induction of Helical Conformation in λN by AR1 Binding

Interaction of NusA with λN(1–53) containing the NusA binding site and the ARM region of λN causes large chemical shift changes also for most of the residues of the ARM region, although the ARM is not directly involved in NusA recognition. In addition to remarkable shift changes of amide proton and nitrogen resonances, alpha- and carbonyl-carbon resonances show significant downfield shifts compared to their values in the unbound state (Prasch et al., 2006). A chemical shift change of these nuclei toward higher frequency is characteristic for a transition to helical conformation (Wishart et al., 1991). These data demonstrate that on binding to NusA, the λN ARM adopts the helical conformation that it also adopts in the complex of λN and *nutboxB* RNA even in the absence of RNA, suggesting that λN initially binds to NusA and thus preforms the *nutboxB* RNA recognition helix. As λN does not adopt helical conformation when binding to unspecific RNA (Van Gilst et al., 1997b), the induction of helical conformation in λN by NusA might facilitate recognition and hinder unproductive interactions. This could be important when λN is already present before release of *nutBox* RNA from the exit of the RNAP. However, kinetics experiments revealed the λN–*nutBoxB* complex formation rate to be

diffusion limited (Austin et al., 2002) and not limited by required conformational changes within λN. The induction of helical conformation in λN might be a result of electrostatic effects as the highly negatively charged AR1-AR2 tandem could influence the conformation of the arginine-rich sequence stretch. Anchoring λN to AR1 via residues 34–47 may bring relevant regions of λN and NusA sufficiently close to cause this conformational change.

16.6 CONCLUDING REMARKS

The classical structural view of proteins was that of a well-defined conformation coded in the amino acid sequence (Anfinsen, 1973). The formation of stable structures of peptide chains allows the presentation of defined surfaces necessary for recognition of other molecules or for orientation of functional groups, resulting in perfectly tuned catalytic centers of enzymes acting with high processivity. This structure–function paradigm works well for longer polypeptides, in which sufficient intramolecular interactions in order to fold the peptide chain are possible. The antitermination protein λN was one of the first proteins that were observed to be intrinsically disordered as no stable structure on very short time scales was observed, and *in vitro* studies demonstrate the full activity of this disordered state by an disorder-order transition without any time constraints by conformational changes except diffusion of reaction partners. The modular arrangement of different interaction sites in λN allows the proper formation of macromolecular complexes by defined adaptive binding to the target proteins and RNA. This shows that in this case, the intermolecular interactions are essential driving forces for folding as tertiary interactions within λN are simply not sufficient to yield a stable fold. The presence of a stable fold for function might not be necessary in the cases of adaptive recognition. The functional sites of λN resemble short sequential stretches, which are supposed to interact independently in binary complexes. Therefore, any long-range interaction within the protein is not expected. The high flexibility of λN thus allows the stepwise formation of pairwise interactions in order to assemble the fully functional antitermination complex.

REFERENCES

Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181: 223–230.

Austin RJ, Xia T, Ren J, Takahashi TT, Roberts RW. Designed arginine-rich RNA-binding peptides with picomolar affinity. J Am Chem Soc 2002;124:10966–10967.

Austin RJ, Xia T, Ren J, Takahashi TT, Roberts RW. Differential modes of recognition in N peptide-boxB complexes. Biochemistry 2003;42:14957–14967.

Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, et al. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. Science 1996;273:1547–1551.

Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P. Structural studies of HIV-1 tat protein. J Mol Biol 1995;247:529–535.

Beuth B, Pennell S, Arnvig KB, Martin SR, Taylor IA. Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. EMBO J 2005;24:3576–3587.

Bonin I, Muhlberger R, Bourenkov GP, Huber R, Bacher A, Richter G, et al. Structural basis for the interaction of *Escherichia coli* NusA with protein N of phage lambda. Proc Natl Acad Sci U S A 2004;101:13762–13767.

Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, et al. A NusE:NusG complex links transcription and translation. Science 2010;328:501–504.

Burmann BM, Uc-Mass A, Schweimer K, Gottesman ME, Rosch P. The Y39A mutation of HK022 nun disrupts a boxB interaction but preserves termination activity. Biochemistry 2008;47:7335–7341.

Cai Z, Gorin A, Frederick R, Ye X, Hu W, Majumdar A, et al. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. Nat Struct Biol 1998;5:203–212.

Chattopadhyay S, Garcia-Mena J, DeVito J, Wolska K, Das A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. Proc Natl Acad Sci 1995a;92:4061–4065.

Chattopadhyay S, Hung SC, Stuart AC, Palmer AG III, Garcia-Mena J, Das A, Gottesman ME. Interaction between the phage HK022 nun protein and the nut RNA of phage lambda. Proc Natl Acad Sci U S A 1995b;92:12131–12135.

Cilley CD, Williamson JR. Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). RNA 1997;3:57–67.

Cilley CD, Williamson JR. Structural mimicry in the phage phi21N peptide-boxB RNA complex. RNA 2003;9:663–676.

Correll CC, Swinger K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. RNA 2003;9:355–363.

Das A. Control of transcription termination by RNA-binding proteins. Annu Rev Biochem 1993;62:893–930.

Eisenmann A, Schwarz S, Prasch S, Schweimer K, Rosch P. The *E. coli* NusA carboxy-terminal domains are structurally similar and show specific RNAP- and lambdaN interaction. Protein Sci 2005;14:2018–2029.

Faber C, Scharpf M, Becker T, Sticht H, Rosch P. The structure of the coliphage HK022 nun protein-lambda-phage boxB RNA complex. implications for the mechanism of transcription termination. J Biol Chem 2001;276:32064–32070.

Farnham PJ, Greenblatt J, Platt T. Effects of NusA protein on transcription termination in the tryptophan operon of *Escherichia coli*. Cell 1982;29:945–951.

Franklin NC. Clustered arginine residues of bacteriophage λ N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. J Mol Biol 1993;231:343–360.

Franklin NC, Bennett GN. The N protein of bacteriophage lambda, defined by its DNA sequence, is highly basic. Gene 1979;8:107–119.

Gopal B, Haire LF, Gamblin SJ, Dodson EJ, Lane AN, Papavinasasundaram KG, et al. Crystal structure of the transcription elongation/anti-termination factor NusA from mycobacterium tuberculosis at 1.7 A resolution. J Mol Biol 2001;314:1087–1095.

Gottesman S, Gottesman M, Shaw JE, Pearson ML. Protein degradation in *E. coli*: the lon mutation and bacteriophage lambda N and cII protein stability. Cell 1981;24:225–233.

Greenblatt J, Li J. The nusA gene protein of *Escherichia coli*. Its identification and a demonstration that it interacts with the gene N transcription anti-termination protein of bacteriophage lambda. J Mol Biol 1981;147:11–23.

Greenblatt J, Li J. Properties of the N gene transcription antitermination protein of bacteriophage lambda. J Biol Chem 1982;257:362–365.

Greenblatt J, Malnoe P, Li J. Purification of the gene N transcription anti-termination protein of bacteriophage lambda. J Biol Chem 1980;255:1465–1470.

Greenblatt J, Nodwell JR, Mason SW. Transcriptional antitermination. Nature 1993;364: 401–406.

Gusarov I, Nudler E. The mechanism of intrinsic transcription termination. Mol Cell 1999;3:495–504.

Gusarov I, Nudler E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. Cell 2001;107:437–449.

Ha KS, Toulokhonov I, Vassylyev DG, Landick R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. J Mol Biol 2010;401:708–725.

Hendrix, RW, Roberts, JW, Stahl, FW, Robert A Weisberg, RA. Lamda II. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Jiang F, Gorin A, Hu W, Majumdar A, Baskerville S, Xu W, et al. Anchoring an extended HTLV-1 rex peptide within an RNA major groove containing junctional base triples. Structure 1999;7:1461–1472.

Johnson NP, Baase WA, von Hippel PH. Low energy CD of RNA hairpin unveils a loop conformation required for lambdaN antitermination activity. J Biol Chem 2005;280:32177–32183.

Lazinski D, Grzadzielska E, Das A. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell 1989; 59:207–218.

Lederberg EM. Lysogenicity in *E. coli* K.12. Genetics 1951;36:560.

Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. Cell 1998;93:289–299.

Luo X, Hsiao HH, Bubunenko M, Weber G, Court DL, Gottesman ME, et al. Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. Mol Cell 2008;32:791–802.

Mah TF, Kuznedelov K, Mushegian A, Severinov K, Greenblatt J. The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. Genes Dev 2000;14: 2664–2675.

Mah TF, Li J, Davidson AR, Greenblatt J. Functional importance of regions in *Escherichia coli* elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. Mol Microbiol 1999;34:523–537.

Metzger AU, Bayer P, Willbold D, Hoffmann S, Frank RW, Goody RS, et al. The interaction of HIV-1 tat(32–72) with its target RNA: a fluorescence and nuclear magnetic resonance study. Biochem Biophys Res Commun 1997;241:31–36.

Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J. Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. Mol Cell 1998;1:265–275.

Mogridge J, Mah TF, Greenblatt J. A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. Genes Dev 1995;9:2831–2845.

Nodwell JR, Greenblatt J. The nut site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. Genes Dev 1991;5:2141–2151.

Nudler E, Gottesman ME. Transcription termination and anti-termination in *E. coli*. Genes Cells 2002;7:755–768.

Patel DJ. Adaptive recognition in RNA complexes with peptides and protein modules. Curr Opin Struct Biol 1999;9:74–87.

Prasch S, Jurk M, Washburn RS, Gottesman ME, Wöhrl BM, Rösch P. RNA-binding specificity of *E. coli* NusA. Nucl Acids Res 2009;37:4736–4742.

Prasch S, Schwarz S, Eisenmann A, Wöhrl BM, Schweimer K, Rösch P. Interaction of the intrinsically unstructured phage lambda N protein with *E. coli* NusA. Biochemistry 2006;45:4542–4549.

Ptashne, M. A genetic switch, phage lambda revisited. 3rd edn, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2004.

Puglisi JD, Chen L, Blanchard S, Frankel AD. Solution structure of a bovine immunodeficiency virus tat-TAR peptide-RNA complex. Science 1995;270:1200–1203.

Rees WA, Weitzel SE, Yager TD, Das A, von Hippel PH. Bacteriophage lambda N protein alone can induce transcription antitermination *in vitro*. Proc Natl Acad Sci U S A 1996;93:342–346.

Robert J, Sloan SB, Weisberg RA, Gottesman ME, Robledo R, Harbrecht D. The remarkable specificity of a new transcription termination factor suggests that the mechanisms of termination and antitermination are similar. Cell 1987;51:483–492.

Roberts JW. Termination factor for RNA synthesis. Nature 1969;224:1168–1174.

Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage lambda DNA. J Mol Biol 1982;162:729–773.

Scharpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S, Rosch P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. Eur J Biochem 2000;267:2397–2408.

Schmidt MC, Chamberlin MJ. nusA protein of *Escherichia coli* is an efficient transcription termination factor for certain terminator sites. J Mol Biol 1987;195:809–818.

Schweimer K, Prasch S, Sujatha PS, Bubunenko M, Gottesman ME, Rösch P. NusA Interaction with the α Subunit of E. coli RNA Polymerase Is via the UP Element Site and Releases Autoinhibition. Structure 2011;19:945–954.

Shin DH, Nguyen HH, Jancarik J, Yokota H, Kim R, Kim SH. Crystal structure of NusA from thermotoga maritima and functional implication of the N-terminal domain. Biochemistry 2003;42:13429–13437.

Su L, Radek JT, Hallenga K, Hermanto P, Chan G, Labeots LA, et al. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. Biochemistry 1997a;36:12722–12732.

Su L, Radek JT, Hallenga K, Hermanto P, Chan G, Labeots LA, et al. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. Biochemistry 1997b;36:12722–12732.

Tan R, Frankel AD. Structural variety of arginine-rich RNA-binding peptides. Proc Natl Acad Sci U S A 1995;92:5282–5286.

Toulokhonov I, Artsimovitch I, Landick R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science 2001;292:730–733.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.

Van Gilst MR, Rees WA, Das A, von Hippel PH. Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. Biochemistry 1997a;36:1514–1524.

Van Gilst MR, Rees WA, Das A, von Hippel PH. Complexes of N antitermination protein of phage λ with specific and nonspecific RNA target sites on the nascent transcript. Biochemistry 1997b;36:1514–1524.

Van Gilst MR, von Hippel PH. Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. J Mol Biol 1997;274:160–173.

Whalen WA, Das A. Action of an RNA site at a distance: role of the nut genetic signal in transcription antitermination by phage-lambda N gene product. New Biol 1990;2:975–991.

Willbold D, Rosin-Arbesfeld R, Sticht H, Frank R, Rosch P. Structure of the equine infectious anemia virus tat protein. Science 1994;264:1584–1587.

Wilson HR, Kameyama L, Zhou JG, Guarneros G, Court DL. Translational repression by a transcriptional elongation factor. Genes Dev 1997;11:2204–2213.

Wilson HR, Zhou JG, Yu D, Court DL. Translation repression by an RNA polymerase elongation complex. Mol Microbiol 2004;53:821–828.

Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 1991;222:311–333.

Worbs M, Bourenkov GP, Bartunik HD, Huber R, Wahl MC. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol Cell 2001;7:1177–1189.

Xia T, Becker HC, Wan C, Frankel A, Roberts RW, Zewail AH. The RNA-protein complex: direct probing of the interfacial recognition dynamics and its correlation with biological functions. Proc Natl Acad Sci U S A 2003a;100:8119–8123.

Xia T, Frankel A, Takahashi TT, Ren J, Roberts RW. Context and conformation dictate function of a transcription antitermination switch. Nat Struct Biol 2003b;10:812–819.

Xia T, Wan C, Roberts RW, Zewail AH. RNA-protein recognition: single-residue ultrafast dynamical control of structural specificity and function. Proc Natl Acad Sci U S A 2005;102:13013–13018.

Ye X, Gorin A, Frederick R, Hu W, Majumdar A, Xu W, et al. RNA architecture dictates the conformations of a bound peptide. Chem Biol 1999;6:657–669.

Ye X, Kumar RA, Patel DJ. Molecular recognition in the bovine immunodeficiency virus tat peptide-TAR RNA complex. Chem Biol 1995;2:827–840.

17

N-TERMINAL EXTENSION REGION OF HORDEIVIRUS MOVEMENT TGB1 PROTEIN CONSISTS OF TWO DOMAINS WITH DIFFERENT CONTENT OF DISORDERED STRUCTURE

VALENTIN V. MAKAROV, MICHAEL E. TALIANSKY, EUGENY N. DOBROV, AND NATALIA O. KALININA

17.1 INTRODUCTION

To induce disease, plant viruses must spread from the initially infected cells to the rest of the plant. The systemic spread of plant viruses proceeds in several stages. First, the virus (in the form of virion or nucleoprotein complex) moves intracellularly from the sites of replication to plasmodesmata (PD), unique intercellular membranous channels that span cell walls linking the cytoplasm of adjoining cells. Then the virus or nucleoprotein complex moves through PDs to spread from cell to cell (cell-to-cell movement). Virus systemic movement between organs (long-distance movement) occurs through vascular tissue, usually phloem sieve elements. Plant viruses encode special multifunctional protein/proteins defined as *movement proteins* (MPs) that facilitate distinct stages of virus transport in plants. In the genus *Hordeivirus*, viral MPs are represented by three proteins encoded by a "*triple gene block*" (TGB), a conserved module of overlapping genes found in numerous (more

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.

than 8) genera of rod-shaped and filamentous viruses (Morozov and Solovyev, 2003). TGB MPs act in a coordinated manner to assist in cell-to-cell and long-distance virus movement.

It is generally accepted that cell-to-cell movement involves virus-encoded MPs as well as host-encoded components (Atabekov and Taliansky, 1990; Carrington et al., 1996; Roberts and Oparka, 2003; Lucas, 2006; Harries et al., 2010). MPs display a whole spectrum of activities, including interaction with viral RNA and possibly other viral proteins with formation of a virus transport form, interacting with elements of cellular cytoskeleton and endoplasmic reticulum. These proteins also possess PD localization signals; are able to interact with PDs, increasing their size exclusion limit; and take part in ribonucleoprotein (RNP) complex remodeling during its translocation through PD. Much less is known about virus transport through the vascular system, which contains various cell types including bundle sheath cells, vascular parenchyma cells, companion cells, and enucleate sieve elements (Oparka and Turgeon, 1999; Santa Cruz, 1999; Oparka, 2004). Capsid protein (CP) is essential for efficient long-distance transport of plant viruses with only a few exceptions. Evidently, the long-distance plant virus movement also includes numerous interactions between virus proteins and proteins of plant cells participating in transport of macromolecules in uninfected plants. *Hordeivirus* TGB1 protein encoded by the first gene of TGB is of special interest, because it forms nonvirion RNP complex for both stages of systemic spread of virus genome in plants: cell-to-cell and long-distance movement (Lim et al., 2008).

Barley stripe mosaic virus (BSMV) and *Poa* semilatent virus (PSLV), representatives of the genus *Hordeivirus*, have rod-shaped particles (Jackson et al., 2009). The BSMV/PSLV genome consists of three single-stranded positive-sense RNAs, where RNAβ encodes three TGB MPs referred to as TGBp1, TGBp2, and TGBp3 (Solovyev et al., 1996). Analysis of RNA–protein complexes isolated from plants infected with BSMV demonstrated that TGBp1 is the major if not the only virus protein that interacts with viral genomic and subgenomic RNAs in infected cells to form RNP complexes (Brakke et al., 1988; Lim et al., 2008; Jackson et al., 2009). Such TGBp1–RNA RNPs are considered to be a form of viral genome capable of cell-to-cell and long-distance transport in plants (Morozov and Solovyev, 2003; Lim et al., 2008). Small membrane proteins TGBp2 and TGBp3 are necessary for intracellular transport of complexes containing TGBp1 and viral RNA to PDs (Morozov and Solovyev, 2003; Zamyatnin et al., 2004; Haupt et al., 2005; Jackson et al., 2009).

Hordeivirus TGBp1 contains an NTPase/helicase domain (HELD) with seven conserved motifs of superfamily 1 NTPases/helicases (Gorbalenya and Koonin, 1993). This domain exhibits NTPase and RNA helicase activities and binds RNA in a cooperative manner *in vitro* (Bleykasten et al. 1996; Donald et al, 1997; Kalinina et al. 2001, 2002; Leshchiner et al. 2006). Mutations in any of the HELD conserved motifs have been shown to block viral cell-to-cell movement (Donald et al., 1997). Additionally, the *Hordeivirus* TGBp1 has an N-terminal "extension region" variable in sequence and length with multiple RNA-binding sites (Donald et al. 1997; Kalinina et al., 2001; Makarov et al., 2009). As demonstrated for the

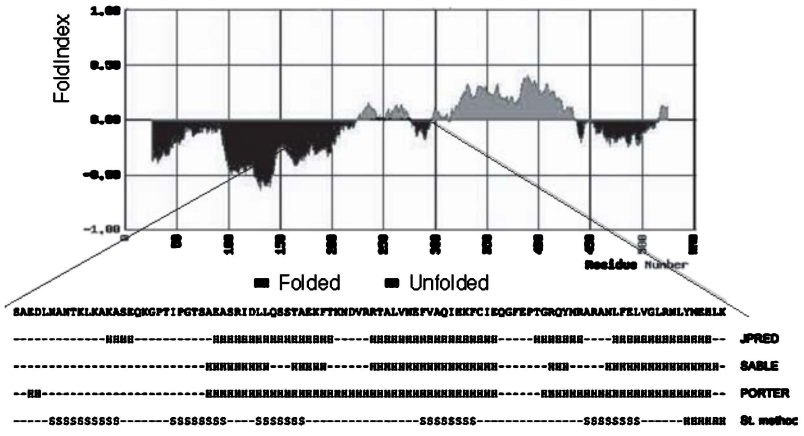
PSLV TGBp1, this protein region binds RNA in a noncooperative manner (Kalinina et al., 2001). The possible functional significance of the RNA binding was revealed by mutagenesis of the PSLV TGBp1 extension domain. Mutations that disrupted any of the two clusters of positively charged amino acids responsible for RNA-binding blocked the long-distance but not cell-to-cell movement of viral infection (Kalinina et al., 2001). We hypothesized that there are two different forms of virus-specific RNP complexes competent for cell-to-cell and long-distance movement and that the requirements of the TGBp1 integrity for long-distance movement are stricter than the ones for cell-to-cell movement (Kalinina et al., 2001).

Despite intensive study of biological properties and biochemical activities of TGBp1s of different viruses with TGB, the MPs and RNP complexes formed by them have never been characterized structurally. Using both computational and experimental approaches, we herein show that the N-terminal region of PSLV TGBp1 possesses large intrinsically disordered regions and consists of two domains: a completely intrinsically disordered extreme N-terminal domain (NTD) and an internal domain (ID), with the structure resembling a partially disordered molten globule. The structural flexibility (disorder) of these domains likely provides additional possibilities for TGBp1 to establish interactions with multiple partners at different stages of formation and movement of transport-competent RNP complexes in the plant.

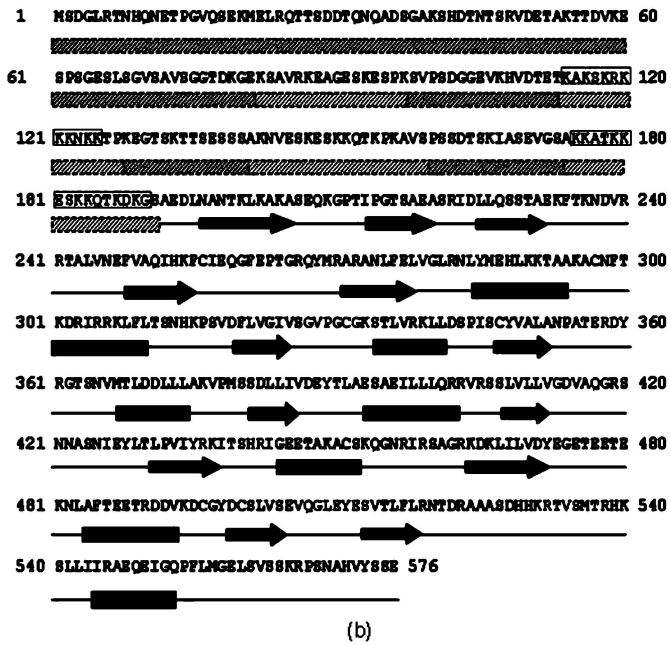
17.2 SECONDARY STRUCTURE PREDICTIONS FOR PSLV TGBP1

The amino acid sequence of PSLV TGBp1 (63 kDa or 63K protein) was analyzed to predict secondary structure elements in the N-terminal extension region (aa residues 1–300) and the C-terminal conserved HELD. Two disorder prediction methods SCRATCH (Cheng et al., 2005) and FoldIndex (Prilusky et al., 2005), confidently predicted that the N-terminal part of the 63K protein (from aa residue 1 to 210) represents an intrinsically disordered protein (IDP) region (Makarov et al., 2009, Fig. 17.1a). The region following the intrinsically disordered one (from aa residue 210 to 300) was predicted as a folded part of the 63K protein containing α-helical segments (Fig. 17.1a). The C-terminal HELD (aa 300–576) was also predicted to be mainly structured. (Fig. 17.1a). In addition to the standard computer programs for prediction of the PSLV TGBp1 secondary structure, a stereochemical method based on periodical sequence patterns of standard structures in proteins (Efimov, 1994) was used. According to this knowledge-based prediction method, the extreme N-terminal region of the 63K protein (from aa 1 to 190/200) was completely intrinsically disordered too (Fig. 17.1b). However, for an ordered sequence with an approximate position between aa 190 and 290, the stereochemical method predicted the existence of β-structure strands in the 195–278-aa segment and two α-helices (aa 284–294 and 301–310) on the border with the HELD (Fig. 17.1b).

Thus, according to the predictions, the N-terminal extension region of the PSLV, TGBp1 might consist of two domains: the extreme N-terminal intrinsically disordered domain (NTD) and the internal ordered domain (ID) preceding the conserved



(a)



(b)

Figure 17.1 Prediction of the structural organization of TGBp1 of *Poa* semilatent virus (a) FoldIndex prediction of ordered and disordered regions in the PSLV TGB1 protein of *Hordeivirus*. Insert: prediction of secondary structure of the central part of PSLV TGBp1 according to JPRED and SABLE web services. (b) PSLV TGBp1 secondary structure prediction using a knowledge-based prediction method (a stereochemical method) (Efimov, 1994). The completely disordered extremely N-terminal part of TGBp1 is shaded. Charcoal-gray and light gray shading marks amino acid clusters with negative and positive charges, respectively. A (aa 114–125) and B (aa 175–189) clusters responsible for RNA binding *in vitro* (Kalinina et al., 2001) are framed. β-strands and α-helices are shown by arrows and rectangles, respectively.

HELD. This suggestion is supported by the data on limited proteolysis of the recombinant full-length PSLV TGBp1 (63K protein) in the course of its expression in *Escherichia coli* (as described in the following section).

17.3 SPONTANEOUS LIMITED PROTEOLYSIS OF THE RECOMBINANT PSLV TGBP1 IN *Escherichia coli*

The expression of the full-length 63K protein in *E. coli* after IPTG induction often resulted in its fragmentation into several cleavage products, the yield of which increased after prolonged induction (Makarov et al., 2009; Fig. 17.2a). Two major proteolytic fragments were copurified with the full-length 63K protein during Ni-NTA agarose chromatography (Fig. 17.2a), showing that these fragments contained the six-histidine tag on the 63K N-terminus. The matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS) analysis of purified six-histidine-tagged proteolytic fragments demonstrated that the larger product migrating in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as a 50-kDa polypeptide corresponded to the N-terminal half of the 576-residue-long 63K protein with its C-terminus located between amino acid positions 279 and 294. The smaller product migrating in SDS-PAGE as 30-kDa polypeptide represented the extreme 63K N-terminal part with its C-terminus located between amino acid positions 188 and 203. Both these fragments of the 63K protein had gel mobilities lower than could be expected based on their molecular masses (34 kDa and 21 kDa, respectively), which might be attributed to the existence of large intrinsically disordered regions in these polypeptides (Fig. 17.2a). It should be noted that the full-length 63K protein also had an abnormal mobility (about 90 kDa) (Fig. 17.2a). In a total extract of *E. coli* expressing the 63K protein, one more proteolytic fragment of approximately 35 kDa was revealed by Western blotting with antibodies against the C-terminal region of the PSLV 63K protein comprising the HELD (Fig. 17.2b). This fragment possibly represented the C-terminal portion of the 63K protein.

Taking into account that proteases preferentially attack interdomain linkers in multidomain polypeptides (Schlotmann and Beyreuther, 1979; Corchero et al., 1996), these data support our suggestion about multidomain organization of PSLV TGBp1. Thus, the 63K protein contains at least three domains: the N-terminal NTD, the central ID, and the C-terminal HELD, which are separated by protease-sensitive linkers at amino acid positions 188–202 and 279–294 (Fig. 17.2c). Experimentally estimated borders of the domains fitted well in the predicted ones. Interestingly, the ID itself could not be detected in our experiments as an individual polypeptide, which presumably reflects its low stability in bacteria.

Deletion mutants comprising the 63K N-terminal extension region (N63K, aa 1–290 of the 63K protein), the NTD (aa 1–190), and the ID (aa 190–290) were constructed, expressed as separate polypeptides in *E. coli*, and purified by affinity chromatography. The recombinant NTD and N63K also had abnormal mobilities in SDS-PAGE (Fig. 17.2d).

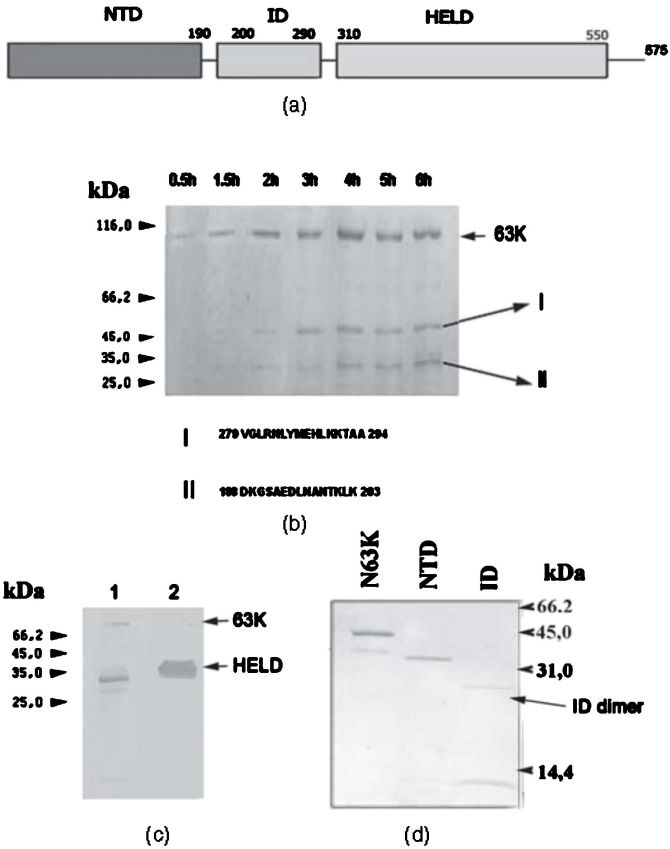


Figure 17.2 Schematic representation of domain organization of PSLV TGBp1 and spontaneous limited proteolysis of the recombinant TGBp1 (63K protein) in *Escherichia coli*. (a) Putative domain structure of the PSLV 63K protein. Figures mark the borders of domains and possible interdomain spacers. N-terminal domain (NTD), internal domain (ID), and NTPase/helicase domain (HELD) are indicated by boxes. (b) Equal volumes of *E. coli* culture carrying a 63K-expressing vector (Kalinina et al., 2001) were taken at different time points (0.5 h–6 h) after induction with IPTG. Protein extracts were purified by Ni-NTA chromatography under denaturing conditions. Proteins were separated by 10% SDS-PAGE and stained with Coomassie Brilliant Blue R-250. The positions of the molecular mass markers (in kDa) are indicated on the left. Major degradation products were cut out from the gel and subjected to enzymatic proteolysis and MS analysis on MALDI-TOF MS system. C-terminal amino acid sequences are shown under the figure. (c) Total protein extract of *E. coli* cells after 2-h induction with IPTG (lane 1) and preparation of recombinant of a C-terminal part of 63K protein (HELD) (lane 2) were separated by 15% SDS-PAGE and then analyzed by Western blot assay with antibodies against HELD. The positions of the molecular mass markers (in kDa) are indicated on the left. The arrows point to the positions of the 63K protein and HELD. HELD was termed C-63K in our previous article (Kalinina et al., 2001). (d) Recombinant proteins corresponding to predicted NTD and ID and the N-terminal region of TGBp1 (N63K) were expressed in *E. coli* as separate polypeptides, purified, and analyzed by 15% SDS-PAGE. The gels were stained with Coomassie Brilliant Blue R-250. Arrow marks the position of the ID dimer. Positions of the molecular mass markers (in kDa) are shown on the right.

17.4 CD AND FTIR SPECTRA OF THE RECOMBINANT PROTEINS CORRESPONDING TO THE N-TERMINAL EXTENSION REGION (N63K) AND PREDICTED DOMAINS NTD AND ID

To experimentally determine the secondary structure of the proposed domains of the PSLV N-terminal extension region, circular dichroism (CD) spectra of corresponding recombinant proteins in the far-ultraviolet (UV) (190–250 nm) region were measured. The CD spectrum of the NTD had a strong negative maximum at 200 nm ($[\theta]_{200} = -14000°$) and a week signal at longer wavelengths (Fig. 17.3a), which is typical for completely disordered proteins (Adler et al., 1973; Uversky, 2002). The ID gave a spectrum characteristic of proteins with significant β-structure content and some fraction of α-helices (Fig. 17.3a). According to current opinion, only α-helix content can be quantitatively determined from the 190–250 nm CD spectra (Sreerama and Woody, 2004). Therefore, we estimated the α-helix content in the ID as 10–15% using the Greenfield–Fasman equation (Greenfield and Fasman, 1969). As β-structure content determinations from CD spectra are highly error prone (Sreerama and Woody, 2004), we employed Fourier transform infrared spectroscopy (FTIR), which has been proved to be more reliable. The infrared absorption spectra were measured in the amide I' band (from 1600 to 1700 cm^{-1}). The FTIR spectral shape within the amide I' band is mainly determined by the secondary structure content, and the contribution of other bands outside amide I' is not significant (Hering et al., 2002). The infrared absorption spectrum of the PSLV ID is shown in Fig. 17.3b. Component analysis of the recorded spectra showed that the β-structure content of the ID is really large (about 37%), which agrees with the ID secondary structure predicted by the knowledge-based prediction method (about 40%) (Fig. 17.1b).

The CD and FTIR spectra of the PSLV TGBp1 deletion mutants confirm the suggestion that the NTD and ID differ in their structure: the NTD is completely disordered, while the ID contains significant amount of β-structure and a fraction of α-helices.

The N63K possessed a CD spectrum similar in shape to that of the NTD, but its molar ellipticity at negative maximum (200 nm) was somewhat lower (about −8000°) possibly because of the presence of the ID (Fig. 17.3a). The observed character of both the NTD and N63K CD spectra testifies to the intrinsically disordered structure of the polypeptides.

17.5 NTD AND N63K AS INTRINSICALLY DISORDERED POLYPEPTIDES

It is well known that proteins possessing fixed tertiary structure are usually denatured cooperatively and irreversibly, while IDPs usually demonstrate a noncooperative and reversible melting (Receveur-Brechot et al., 2006; Uversky, 2009). Figure 17.4a represents the curves for CD heat denaturation of the isolated NTD. It can be seen that on heating from 20°C to 90°C, the molar ellipticity at the negative

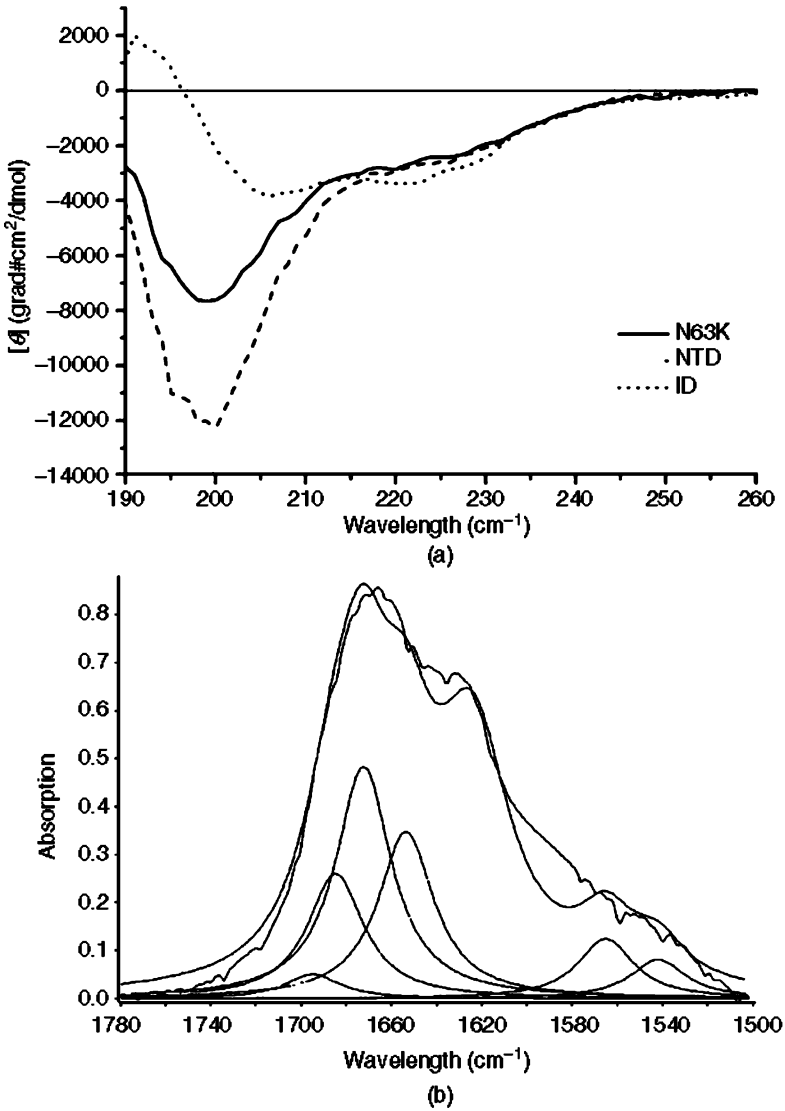


Figure 17.3 Circular dichroism (CD) and FTIR spectra of the recombinant proteins corresponding to predicted domains and the N-terminal region of TGBp1. (a) Far-UV CD spectra of the recombinant proteins (NTD, ID, and N63K) recorded at room temperature. (b) Deconvolved FTIR spectrum of the recombinant ID in the amide I' region. Upper lines indicate the experimental data and the approximation by the component summation; lower lines indicate individual Gaussian components after Fourier transformation.

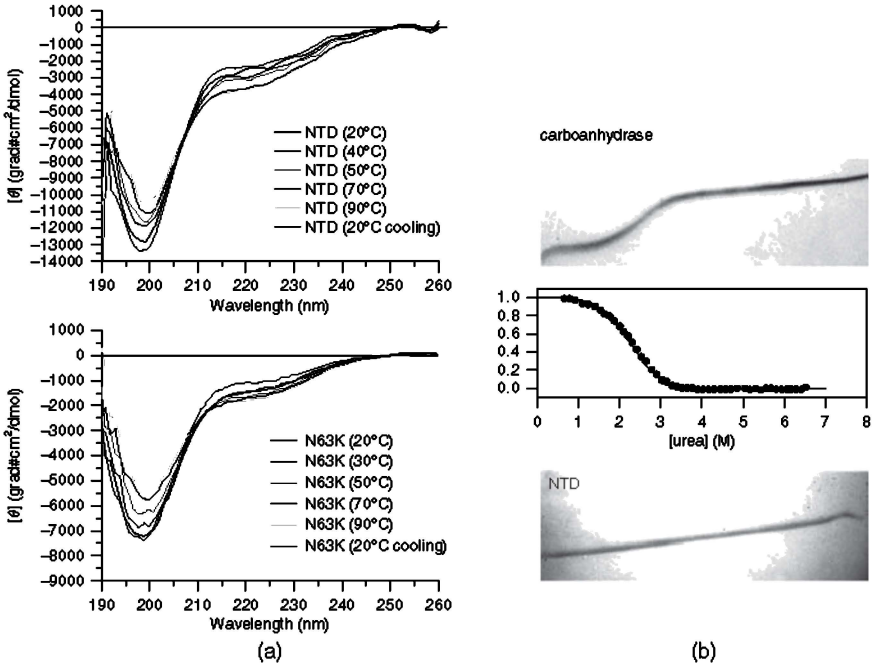


Figure 17.4 Disordered structure of the NTD and the N-terminal region of TGBp1: heat denaturation of polypeptides (NTD and N63K) monitored by CD spectroscopy and Creighton electrophoresis assay. (a) Reversible heat denaturation of the NTD and the N63K monitored by CD spectroscopy in the far-UV region (190–260 nm). Spectra were measured at +20°C, +40°, +50°, +70°, +90°C, and after heat denaturation followed by renaturation at +20°. (b) Creighton electrophoresis assay. Figure represents the results of electrophoresis in polyacrylamide gels containing a gradient of urea concentration (8–0 M) perpendicular to the direction of electrophoresis and the graph of dependence of the CA folding on urea concentration.

maximum at 200 nm decreased (in absolute values) from −12700° to −10500°, while $[\theta]_{220}$ increased from −2500° to −3800° On rapid (20 min) cooling from 90°C to 20°C, the NTD CD spectrum completely returned to initial values. Similar results were obtained on melting of the N63K (Fig. 17.4b). For this protein, $[\theta]_{200}$ dropped on heating to 90° from −7200° to −5200° and the shoulder at 220 nm increased from −1500° to −1850° The N63K heat denaturation was also completely reversible.

To further study structural characteristics of the NTD and the N63K, we performed Creighton electrophoresis in polyacrylamide gels containing a gradient of urea concentration (8–0 M) perpendicular to the direction of electrophoresis. Proteins applied across the top of such gels migrate in the presence of continuously varying urea concentration, and a profile of the unfolding transition is generated directly (Goldenberg and Creighton, 1984). Carboanhydrase (CA) was used as a

positive control: it behaved as a tightly packed globular protein under conditions of Creighton electrophoresis (pH, ionic strength, etc). Figure 17.4c shows the results of experiments for the CA and the NTD. CA is a rather stable protein, and its transition from denatured to native state is highly cooperative, with free energy changes typical for globular proteins (Santoro et al., 1988). On the contrary, the NTD demonstrated only gradual transition, which was not completed even in 0 M urea. Such a noncooperative transition is characteristic for nonglobular partially unstructured proteins (Baryshnikova et al., 2005). Calculation of free energy change could not be performed in this case. The N63K displayed in Creighton electrophoresis characteristics similar to those of the NTD, confirming a disordered nature of this polypeptide too (data not shown).

Thus, the NTD is completely intrinsically disordered as shown by predictions with different web-based services, CD spectra measurements, reversible heat denaturation monitored by CD spectroscopy, behavior in Creighton electrophoresis, and abnormal mobility in SDS-PAGE. Computational tools for the detection of disordered regions in proteins revealed that: (i) disordered regions should have a biased composition and (ii) that they usually contain either small number of or no hydrophobic clusters (Coeytaux and Poupon, 2005). In line with these observations, the NTD has low content of hydrophobic amino acids and demonstrates complete absence of aromatic amino acids. Moreover, it is characterized by a rather high content of lysine (18.5%), serine (16.9%), and threonine (11.1%). The NTD contains a region at the extreme N-terminus (aa residues 1–81) with a calculated p*I* value of 4.62 and serine and threonine contents of 14.8% and 13.6%, respectively, followed by interspersed amino acid clusters (14–16 aa in length) with negative and positive charges. The theoretically calculated p*I* values of these clusters are 9.53 (aa 82–97), 4.31 (aa 98–113), 10.87 (aa 114–129), 6.28 (aa 130–145), 10.18 (aa 146–161), 4.37 (aa 162–175), and 10.22 (aa 176–189). These clusters are shown in Fig. 17.1b. Two of these positively charged clusters marked in Fig. 17.1b as A and B are responsible for RNA binding *in vitro* (Kalinina et al., 2001). Theoretically calculated p*I* value of the whole NTD is 9.54, and its molecular mass is 21.0 kDa. Analysis of NTD far-UV CD spectra in terms of double wavelength plot ($[\theta]_{222}$ versus $[\theta]_{200}$) allows the classification of IDPs into coil-like and premolten globulelike subclasses (Uversky, 2002). According to this plot, the NTD contains no ordered elements and belongs to the coil-like subclass.

17.6 ID IS MOSTLY DISORDERED WITH ORDER-PRONE SEGMENTS

Computational predictions and optical methods demonstrate that the ID represents a protein region with a pronounced secondary structure containing 10–15% of α-helices and about 40% of β-structure elements. On heating from 20° to 90°C, ID displayed significant decrease in intensity of CD negative peak at 206 nm: $[\theta]_{206}$ dropped from −4000° to −2800°, and only partial restoration of initial spectrum took place on cooling (data not shown). This partial restoration may be considered

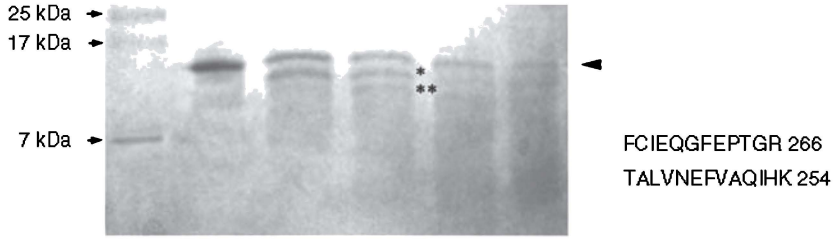


Figure 17.5 Limited trypsin digestion of the ID. About 3 mkg of the ID samples were incubated with 0.05 mkg of trypsin for different times at 37°C, and the samples were analyzed by 15% SDS-PAGE. The asterisks mark positions of major tryptic fragments. C-terminal amino acid sequences of the tryptic peptide fragments are showed on the right. The arrowheads mark the position of the ID monomer and dimer.

as an argument for the presence of globular elements in the ID. Preliminary results of one-dimensional nuclear magnetic resonance (NMR) assay also support such suggestion: the ID spectrum was typical for proteins with a high percentage of intrinsic disorder, but some segments of the polypeptide possess elements of tertiary structure (Kutushenko V., personal communication).

Limited trypsin digestion is usually used to map the folding properties of proteins. It is known that trypsin first attacks disordered or flexible regions of proteins (Fontana et al., 2004). After 10-min incubation of the ID with trypsin, two products migrating faster in SDS-PAGE than the ID monomer appeared in the gel (Fig. 17.5). MALDI-TOF MS analysis of these products demonstrated that the N-termini of these polypeptides coincided with that of the recombinant ID (aa residue 190 of the full-length 63K protein), while their C-termini corresponded to amino acid residues 266 and 254 for products marked by single asterisk and double asterisk, respectively. Combining these results with the data of computational predictions, we propose that the ordered portion is located between residues 215 and 266; for example, the central part of ID polypeptide chain could have a loose tertiary structure.

Analysis of the ID CD spectrum using web-service K2D2 (www.olic.ca/projects/k2d2) shows a high percentage of spatial disordered structure (about 48%). Sedimentation velocity experiments were performed to evaluate the shape of the ID monomer under nondenaturing conditions. The results of the ID monomer analysis gave the $S_{20,w}$ value of 1.0 ± 0.1S (Fig. 17.6a). This value is rather small compared with that calculated for globular compact molecule of a protein with the same molecular mass (11.2 kDa), which should be about 1.5 S. Thus the results of analytical ultracentrifugation imply an elongated structure for the ID monomer. The structure of the ID might be significantly unstable. The state of the ID was characterized by "double wavelength" plot ($[\theta]_{222}$ versus $[\theta]_{200}$) as premolten or molten globule (Uversky, 2002).

We also studied an influence of 2,2,2-trifluoroethanol (TFE) on the ID secondary structure. It is well known that TFE may stabilize existing α-helices and induce formation of additional α-helical regions in proteins from disordered segments (Shiraki et al., 1995). On addition of 33% TFE, the position of negative maximum in the ID CD spectrum shifted from 206 nm to 208 nm and the ellipticity at this peak increased from −4200° to −7600° (Fig. 17.6b). Besides, on TFE addition,

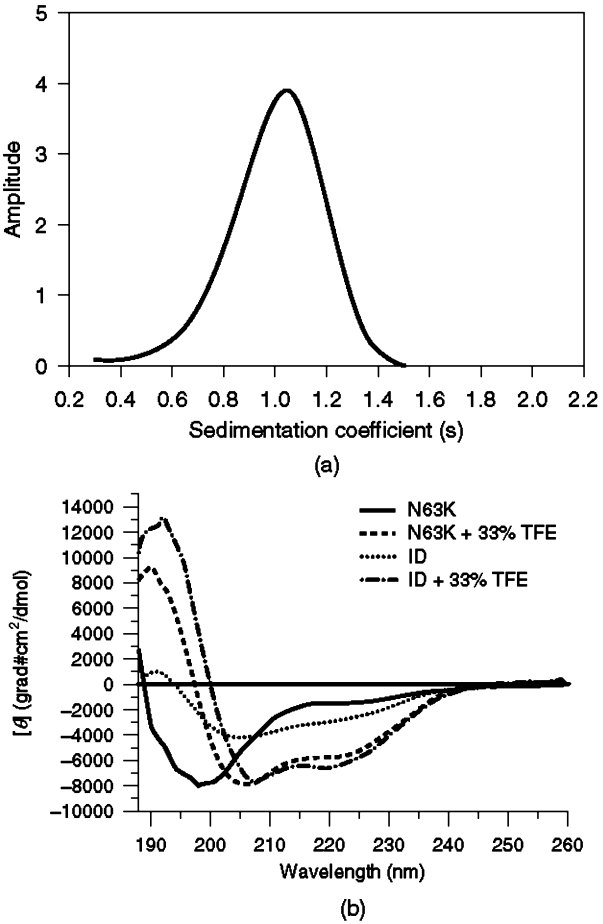


Figure 17.6 ID unstable structure: analytical ultracentrifugation analysis and CD spectroscopy of the ID and the N63K in the presence of TFE. (a) Analytical ultracentrifugation analysis of the ID monomers. (b) CD spectroscopy of ID and the N63K in the presence of 33% TFE. The far-UV CD spectra of the ID and the N63K were recorded at room temperature without (dotted and solid lines, respectively) and in the presence of 33% TFE (dash-dotted and dashed lines, respectively). (c) Comparison of the experimental N63K CD spectrum (solid line) and the N63K CD spectrum calculated as a sum of the isolated the NTD and ID spectra (dashed line).

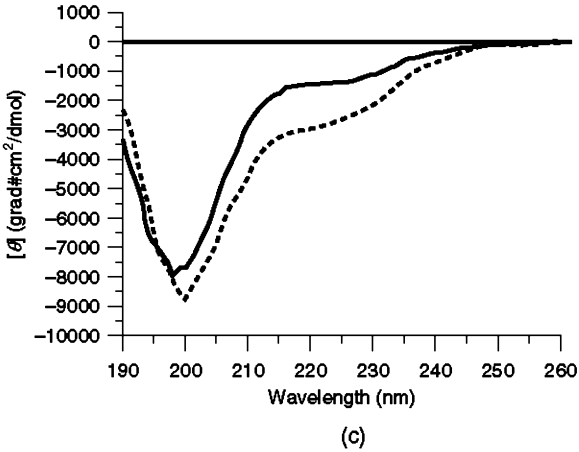


Figure 17.6 (*Continued*)

the intensity of positive peak at about 190 nm increased from +1000° to +13000° (Fig. 17.6b). Both these alterations are characteristic for transition of a significant portion of a protein molecule from disordered to α-helical structure. Similar changes in a CD spectrum in the presence of TFE were found for the N63K, which also showed increased α-helical content in response to addition of 33% TFE (Fig. 17.6b). Therefore, it appears that both the ID and the N63K contain disordered regions, which are able to form α-helices in favorable conditions. The N63K's capacity to form α-helices may be induced by the presence of the ID.

Our results demonstrate that the N-terminal region of PSLV TGBp1 has all the properties of IDP in spite of the presence of the ID moiety (Figs. 17.3a and 17.4b). We compared the experimental N63K CD spectrum with the one calculated as a sum of the isolated NTD and ID spectra (taking into account their respective lengths in the N63K). The calculated spectrum in 220-nm region gave almost two times higher (absolute) molar ellipticities than the experimental N63K spectrum (Fig. 17.6c). These data suggest that the ID as a component of the N63K loses a significant part of structured elements and becomes practically totally disordered.

17.7 PROPERTIES OF NTD, ID AND N63K STUDIED BY DYNAMIC LASER LIGHT SCATTERING (DLS) AND ATOMIC FORCE MICROSCOPY (AFM)

The method of dynamic laser light scattering (DLS) is used for the estimation of the size distribution of individual protein molecules and protein complexes in solution (Schmitz, 1990; Barilla et al., 2005; Tonges et al., 2006; Nemykh et al., 2008).

Results obtained using DLS are represented as the amount of a certain size particles, where the particle size is its hydrodynamic diameter approximated by globular particle parameters, or as the volume occupied by particles of this size. Results of our DLS experiments showed that all domains of PSLV TGBp1 including the NTD, the ID, and the HELD are able to form homo-oligomeric complexes of different sizes. These data are consistent with our previous observations when the oligomerization/multimerization of the domains was demonstrated by ultracentrifugation of recombinant proteins in sucrose gradients (Makarov et al., 2009).

The average diameter of the NTD particles at 25°C was 15 ± 3 nm (Fig. 17.7a, solid line). The addition of SDS to 0.01% (not inducing formation of detergent micellae), resulted in conversion of complexes into particles of 6.5 ± 1.5 nm (Fig. 17.7a, dashed line). The size of these particles is similar to the diameter of potato virus X (PVX) CP molecules (molecular mass 25 kDa) measured by DLS (about 7 nm), which presumably have extended structure (Nemykh et al., 2008). Our results could also be explained either by the extended structure of the NTD monomers or by formation of the protein dimers.

The average size of the complexes formed by the isolated ID at 25°C was 50 ± 15 nm (Makarov et al., 2009, 2010; Fig. 17.7b, solid line). Large protein complexes with diameter of 300 ± 150 nm have also been revealed. Although their number is low (less than 0.1%), they occupy a considerable volume in the analyzed preparation as shown in Fig. 17.7b. The size of complexes and their ratio in preparations depended on incubation temperature. When the temperature was decreased to 4°C, complexes of 300 ± 150 nm became the main component of the preparations (up to 99%). In contrast, on increasing the temperature to 37°C, small particles of about 15 nm became prevalent (Makarov et al., 2010). The addition of 0.01% SDS resulted in the conversion of the complexes into particles of 5.0 ± 2.0 nm (Fig. 7b, dashed line). These particles might be the ID dimers, because the stable dimers formed by this domain were observed on electrophoresis in denaturing SDS-PAGE (Fig. 17.2d).

Properties of the recombinant protein corresponding to the N-terminal extension region of the PSLV TGB1 protein (N63K) were generally similar to the properties of the ID. The DLS analysis revealed in N63K preparations two types of particles of 60 ± 20 nm (over 99% by number) and $(250–300) \pm 100$ nm (less than 0.1%) at 25°C (Makarov et al., 2010; Fig. 17.7c, solid line). Addition of 0.01% SDS converted the N63K complexes into particles of 4.0 ± 1.5–nm diameter (Fig. 17.7c, dashed line), while RNase treatment had practically no effect on the complex size (data not shown). Decreasing the temperature to 4°C was accompanied by protein aggregation, and in this case, complexes of 350 ± 150 nm became prevalent. In contrast to the ID, at 37°C, the particles of 60 ± 20 nm were preserved, while large complexes disappeared completely (Makarov et al. 2010). Evidently, particles formed by the extended N-terminal portion of the TGBp1 were more stable than structures formed by the ID.

Atomic force microscopy (AFM) was used to visualize structures formed by different domains of PSLV TGBp1. AFM image of the NTD preparation is shown in Fig. 17.8a. Only small globules 1.5 ± 0.2 nm in height and apparent diameter,

measured at half height of about 15.0 ± 2.0 nm were found in the NTD preparation. The ID preparations represented a set of particles with heterogeneous type and size. The preparation contained a significant amount of globules of different sizes: globules of 1.5 ± 0.2 nm as well as smaller (from 0.4 nm) and larger (up to 3.8 nm) globular particles were found. Along with globules, the preparation contained filamentous structures with a height of 1.4 ± 0.4 nm, diameter of about

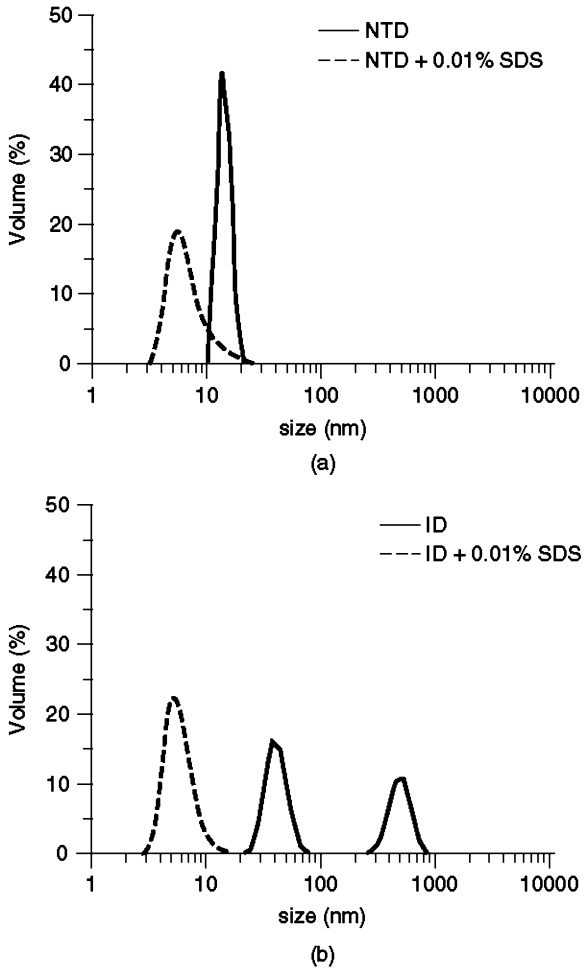


Figure 17.7 Measurement of hydrodynamic diameter (size) of particles in the NTD, ID, and N63K preparations without and in the presence of 0.01% SDS by dynamic laser light scattering (DLS) method. Recombinant proteins were purified by Ni-NTA chromatography under denaturing conditions, renatured, and analyzed by DLS method. Size (nm) is the hydrodynamic diameter. Size distribution by volume of particles is present for the NTD (a), ID (b), and N63K (c) preparations at 25°C without (solid lines) and in the presence of 0.01% SDS (dashed lines).

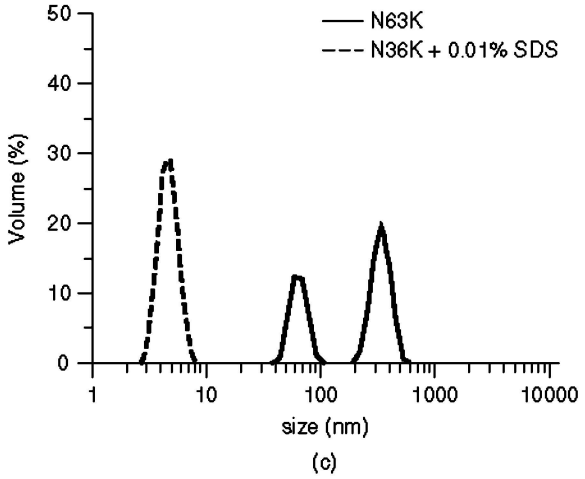


Figure 17.7 (*Continued*)

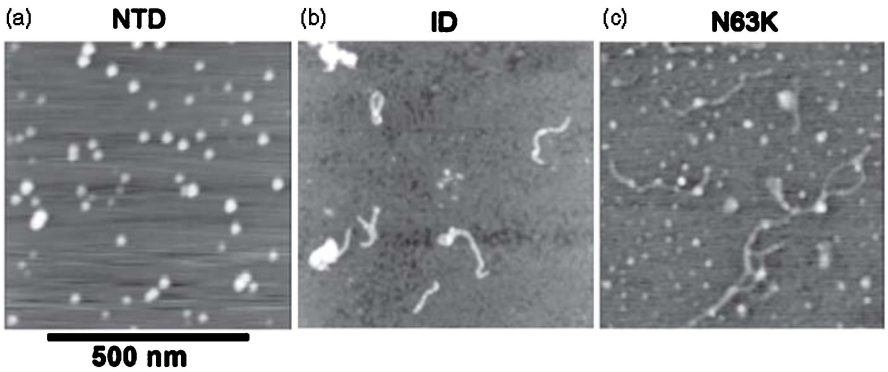


Figure 17.8 Topographic images of the particles present in the NTD, ID, and N63K preparations using a Nanoscope III atomic force microscope operating in tapping mode. The image size is $0.8 \times 0.8\ \mu m^2$. NTD (a), ID (b), and N63K (c).

25.0 ± 7.0 nm, and length of up to 300 nm (Fig. 17.8b). These filamentous structures often formed snarls and a certain number of larger complexes defined as snarl aggregates (Makarov et al., 2010).

It is quite likely that large complexes, observed by DLS technique represent snarl aggregates observed by AFM. Similar aggregates and aggregates consisting of low-molecular-weight oligomers are formed in protein preparations at lower temperatures. The complexes, detected by DLS as structures with mean size of about 50 nm, are most likely visualized as filamentous structures. This suggestion may be supported by the fact that 550-nm long filamentous virions of PVX are measured by DLS as particles with hydrodynamic diameter 40–50 nm (Nemykh et al., 2008).

The N63K preparations contained globules of 1.2 ± 0.2 to 3.0 ± 0.6 nm in height and filamentous structures up to 500 nm in length, 1.2 ± 0.5 nm in height, and apparent diameter of about 35.0 ± 10.0 nm, often branched (Fig. 17.8c). Thus, the N-terminal region of PSLV TGB1 protein is also able to form filamentous structures, and this property is probably determined by the ID.

The morphology of filamentous complexes is similar to that of RNP complexes obtained *in vitro* in the presence of plant viral MPs and RNA (Kiselyova et al., 2001; Kim et al., 2004). The so-called nucleoproteins forming internal nucleocapsid in a number of animal viruses with single-stranded negative-sense RNA genomes, such as influenza and rabies viruses, represent a well-studied class of viral proteins able to form extended filamentous structures (Ruigrok and Baudin, 1995; Iseni et al., 1998). *In vitro* formation of extended structures, morphologically identical to viral RNP complexes (nucleocapsids), was observed both in protein preparations isolated from nucleocapsid and in preparations of recombinant nucleoproteins (Iseni et al., 1998). The authors believe that formation of RNP-like complexes is determined by the ability of these viral nucleoproteins to self-assemble. Properties of the ID and the N-terminal region of PSLV TGBp1 are similar to those of the viral nucleoproteins: the nucleoproteins also aggregate on decrease in temperature, whereas dissociation of complexes is observed when temperature is increased. RNase treatment does not have effect on complex size (Makarov et al., 2010).

Thus, the results of DLS and AFM experiments support our hypothesis that the N-terminal extension region of PSLV TGBp1 plays a structural role in the formation of viral RNP complex, on the one hand exhibiting the ability for multimerization to filamentous RNP-like structures and on the other hand demonstrating RNA-binding properties. We have shown that both domains are able to nonspecifically bind single- and double-stranded RNAs. The NTD interacts with RNA noncooperatively similar to the N63K and the ID binds RNA in a cooperative manner (Makarov et al., 2009).

Interestingly the N-terminal extension region of the *Hordeivirus* movement protein TGB1 partially has a counterpart in the YB-1 protein—a main structural protein of eukaryotic untranslatable RNP complexes, "informosomes". The YB-1 protein is composed of two different domains with RNA-binding activities—a cold-shock domain rich in β-structure and an intrinsically disordered domain with interspersed amino acid clusters with negative and positive charges (Skabkin et al., 2004). Additionally, YB-1 protein has been found to form homomultimeric complexes with molecular mass up to 800 kDa (Evdokimova et al., 1995). The PSLV TGBp1 N-terminal extension region is organized in a similar way, with RNA binding and protein multimerization functions associated with the structured ID and another RNA-binding activity associated with the intrinsically disordered NTD. The YB-1 protein and related proteins are proposed to act as RNA chaperones to keep a pool of specific mRNAs in conformation inaccessible for ribosomes. The available data allow us to propose that the PSLV TGBp1 N-terminal extension region may share some properties with RNA chaperones. Further research is required to verify this hypothesis and to determine the step(s) of viral transport

in plants where this activity is required. Informosomes (mRNP complexes) move to specific compartments of the cell and may be activated at a certain stage of the cell ontogenesis (Skabkin et al., 2004). It has been suggested that untranslatable *Hordeivirus* TGBp1–RNA complex must be transported to cell wall and translocated through PD into an uninfected cell in a form activated for translation (Morozov and Solovyev, 2003). Another common feature of these proteins is a close structural–functional connection with NTPase/helicases: this enzyme is an integral part of mRNP complexes formed by YB-1 (Weston and Sommerville, 2006) and a third domain of the *Hordeivirus* TGBp1.

17.8 STRUCTURAL ORGANIZATION OF THE TGB1 N-TERMINAL EXTENTION REGIONS OF OTHER VIRUSES WITH HORDEI-LIKE TGB

Amino acid sequence analyses of TGB-encoded MPs from different virus genera as well as functional studies, revealed two distinct TGB types termed 'hordei-like' and 'potex-like' TGB (Morozov and Solovyev, 2003). The TGBp1 of viruses with potex-like TGB contain no N-terminal extension region and the NTPase/helicase domain comprises the entire TGBp1 sequence. On the contrary, the TGBp1s of viruses with hordei-like TGB contain N-terminal extension regions and are found not only in the genus *Hordeivirus* but also in the genera *Pomovirus, Pecluvirus*, and *Benyvirus* (Morozov and Solovyev, 2003). Fig. 17.9 represents the prediction of secondary structural elements in TGBp1s encoded by viruses with hordei-like TGB, namely PSLV and BSMV (genus *Hordeivirus*), potato mop-top virus (PMTV, genus *Pomovirus*), peanut clump virus (PCV, genus *Pecluvirus*) and beet necrotic yellow vein virus (BNYVV, genus *Benyvirus*). TGBp1s of all viruses with hordei-like TGB likely have two domains in the N-terminal extension region: a full-size ID (90/100 aa) and an NTD of different lengths (from 30 aa in the case of BNYVV TGBp1 to about 300 aa in the case of PSLV TGBp1). It should be pointed out that the NTD domains of these viruses have no sequence similarity except for the presence of arginine/lysine-rich clusters (Morozov and Solovyev, 2003), but intrinsically disordered structure is a characteristic feature of all these NTDs. At the same time, multiple alignment of sequences comprising IDs shows that ordered regions preceding HELD in TGB1 proteins of viruses with hordei-like TGB have an obvious sequence similarity in a half of ID and a subsequent linker region between ID and HELD (Morozov and Solovyev, 2003).

17.9 RELATIONSHIP BETWEEN STRUCTURE OF TGB1 MOVEMENT PROTEINS AND THEIR FUNCTION IN VIRUS TRANSPORT IN PLANTS

Comparison of transport systems of different TGB-containing viruses revealed important characteristics that could be connected with the presence or absence of N-terminal extension region in TGBp1s and with the length of this region.

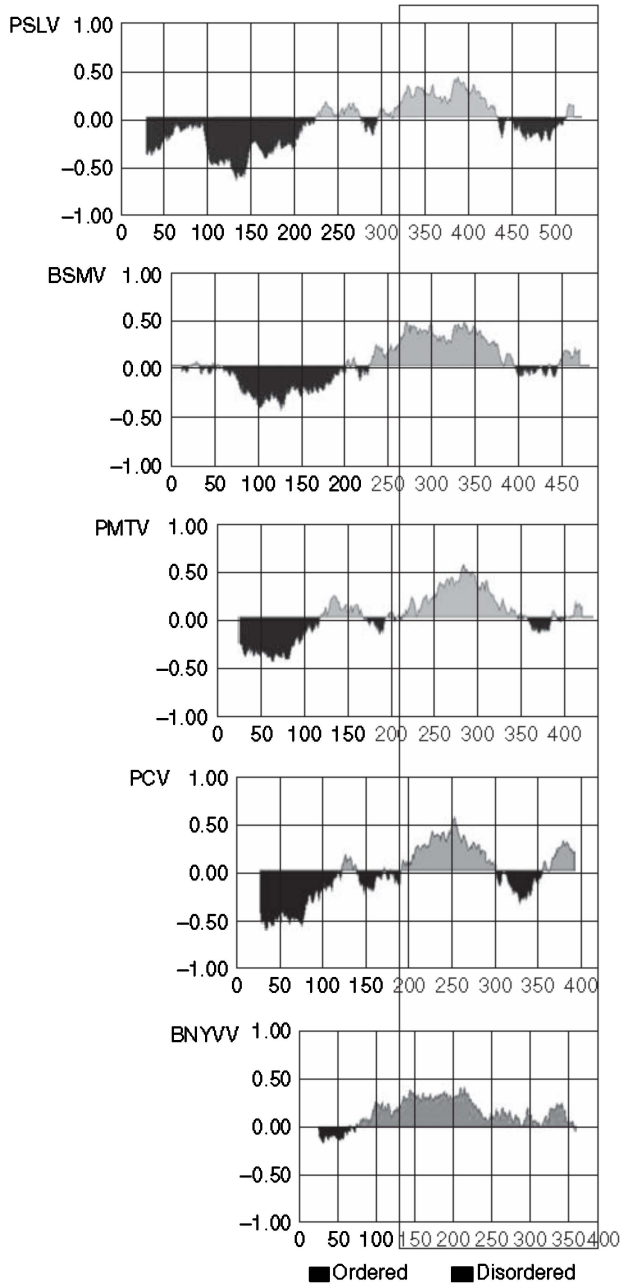


Figure 17.9 FoldIndex prediction of multidomain organization of TGB1 proteins of viruses with hordei-like TGB. Foldindex prediction of ordered and disordered regions in representatives of hordei-like TGBp1s encoded by PSLV and barley stripe mosaic virus (BMSV, genus *Hordeivirus*), potato mop-top virus (PMTV, genus *Pomovirus*), peanut clump virus (PCV, genus *Pecluvirus*) and beet necrotic yellow vein virus (BNYVV, genus *Benyvirus*). The frame includes the NTPase/helicase domains (HELDs) of TGB1 proteins.

Viruses with potex-like TGB containing no N-terminal extension region require coat protein for viral cell-to-cell and long-distance movement (Chapman et al., 1992). It is supposed that the transport form of these virus genomes is virions (Santa Cruz et al., 1998; Morozov and Solovyev, 2003) or virionlike RNP complexes that are called single-tailed particles (viral RNA partly encapsidated by CP beginning from 5′-terminus) (Karpova et al., 2006). As it has been shown for PVX potexvirus, TGBp1 binds to the end of virion or virion-like particle containing a 5'-RNA terminus, and this interaction induces fast cooperative destabilization (remodeling) of the whole long particle resulting in its translational activation (Atabekov et al., 2001; Kiselyova et al., 2003; Rodionova et al., 2003; Karpova et al., 2006).

Hordeiviruses are characterized by the longest known TGBp1 N-terminal extension region. In this case, CP is dispensable for both cell-to-cell and systemic viral movement that is believed to occur in the form of RNP complexes formed by TGBp1 and genomic and subgenomic RNAs (Brakke et al., 1988; Petty and Jackson, 1990; Lim et al., 2008; Jackson et al., 2009). TGBp1s of BNYVV (genus *Benyvirus*) and PCV (genus *Pecluvirus*) are characterized by shorter N-terminal extension regions comprising ID and a truncated NTD. In BNYVV and PCV, CP is dispensable for viral cell-to-cell movement (as in hordeiviruses) but is required for vascular transport (Schmitt et al., 1992; Tamada et al., 1996; Herzog et al., 1998). Apparently, transport forms of BNYVV/PCV genomes involved in local and systemic movement are structurally different. It must be RNP complex for cell-to-cell movement and presumably modified virions for long-distance movement. PMTV (genus *Pomovirus*) provides the example of an intermediate TGB transport system. The PMTV TGBp1 has an NTD similar in size to that of PCV. Nevertheless, the PMTV TGBp1 is able to function similar to *Hordeivirus* TGBp1: two PMTV genomic RNAs coding for replicase and TGB proteins move locally and systemically in the absence of viral CP, presumably in the form of TGBp1-formed RNP complexes (Savenkov et al., 2003). However, the PMTV TGBp1 has an additional function in viral systemic spread: in wild-type infections, the genomic RNA component encoding CP and CP-readthrough (CP-RT) moves long distance in the form of a virus particle, one end of which is believed to be associated with CP-RT and TGBp1 (Torrance et al., 2009). Thus, PMTV TGBp1 can function in systemic transport in two dissimilar ways, either forming RNP complexes like *Hordeivirus* TGBp1 or interacting with one end of the virion similar to potexvirus TGBp1. The described results are summarized in Table 17.1.

These data support our hypothesis that partially disordered ID domain with molten globulelike structure can directly participate in the formation of RNP complex, which is essential for both virus cell-to-cell (all viruses with hordei-like TGB) and long-distance movement (representatives of genera *Hordeivirus* and *Pomovirus*). The presence of unstructured segments in this domain may be critical for the formation of transport complexes because of their necessity for interactions with viral RNA and the protein self-assembly into large filamentous particles. Intrinsically disordered domains are often found in proteins capable of self-assembly into large multimeric complexes such as viral capsids and nucleocapsids (Namba, 2001; Longhi et al., 2003). The PSLV ID shares a number of properties with the

TABLE 17.1 Correlation Between the Length of the N-Terminal Extension Region of TGBP1s of Viruses with TGB and Types of Virus Transport Forms Competent for Cell-To-Cell and Long-Distance Movement

Virus	NTD Length (aa Residues)	Type of Virus Transport-Competent Form			
		RNP Complex		Virion	
		Cell-To-Cell Movement	Long-Distance Movement	Cell-To-Cell Movement	Long-Distance Movement
PSLV	190	+	+	−	−
BSMV	170	+	+	−	−
PMTV	135	+	+	−	+
PCV	101	+	−	−	+
BNYVV	30	+	−	−	+
PVX	0	−	−	+	+

so-called nucleoproteins, viral proteins forming internal nucleocapsid of viruses with negative-sense RNA genomes, as it has been noted above (Section 17.7). The other important property of the PSLV ID may be its ability to interact with other structural domains of the 63K protein. Protein–protein interactions between the ID and the NTD and between the ID and the HELD in solution were demonstrated (Makarov et al., 2009). One can speculate that these interactions (RNA–protein and protein–protein) might be the basis for remodeling of the TGBp1-formed RNP complexes during different phases of cell-to-cell and long-distance movement.

According to all available data, the NTD represents the completely intrinsically disordered protein domain, often being a characteristic feature of multifunctional polypeptides. A large degree of flexibility allows such proteins or protein domains to interact with numerous structurally distinct ligands such as proteins or nucleic acids, overcoming steric restrictions. Interaction with a target often induces a transition from a disordered structure to an ordered one that allows the amendment of structural and functional states of the same polypeptide (Uversky, 2002; Fink, 2005). *Hordeivirus* PSLV NTD possesses several activities *in vitro*: it is able to bind single- and double-stranded RNA nonspecifically and noncooperatively, to form homo-oligomeric complexes and to interact with the ID (Makarov et al., 2009). To date, we know almost nothing about functional role of NTD during viral life cycle. However, the results described above point out that NTD in the case of hordeiviruses and pomoviruses could play an important role in virus phloem transport. Deletion of most part of *Pomovirus* PMTV NTD (84 amino acid residues from the N-terminus) results in the block of virus long-distance but not cell-to-cell movement (Wright et al., 2010). Similar effect on long-distance movement was observed in the case of mutations in any of the two RNA-binding basic clusters of the *Hordeivirus* PSLV NTD (Kalinina et al., 2001). In addition, it has been shown that the NTDs of both these viruses provide TGBp1 localization to the nucleolus of plant cell (Wright et al., 2010; Semashko et al., unpublished results). Further experiments are required to understand the relationship between localization of

TGBp1 in the nucleolus and phloem transport of virus. At the moment, there is only one system in which it was demonstrated that localization of viral MPs in the nucleolus is absolutely necessary for the virus long-distance transport.

We have recently shown that *Umbravirus* MP (open reading frame (ORF)-3) moves to the nucleolus, hijacks a major nucleolar protein, fibrillarin, and relocalizes it from the nucleolus to the cytoplasm, where both proteins take part in the formation of viral RNA-ORF-3-fibrillarin complexes that are capable of long-distance movement (Kim et al., 2007a, 2007b; Canetta et al., 2008). We have also demonstrated that PSLV TBGp1 is able to interact both *in vivo* and *in vitro* with fibrillarin and that the sites of this interaction are mapped within the NTD (Semashko et al., unpublished data). Although a role of interaction between fibrillarin and PSLV TGBp1 is not clear and should be elucidated in future, one could speculate that during *Hordeivirus* infection, fibrillarin together with TGBp1 encapsidates viral RNA forming specific RNP complex facilitating long-distance transport, which is structurally different from the RNP complex for cell-to-cell movement. Finally, another property that has been shown for the PMTV NTD is the capacity to associate with plant cell microtubules (Wright et al., 2010).

Further investigations of the structure and functions of *Hordeivirus* RNP complexes at different stages of virus replication and transport in infected plants and discovery of viral and cellular partners of Hordeivirus-like TGBp1s are challenges for the future that will give novel insights into understanding the functional role of long disordered regions in these MPs.

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research, grant 10-04-10-04-00522a (N. K. and V. M.), the Scottish Government (M. T.), and the Russian Ministry of Education and Science, contract 1 02.740.11.5145 (N. K. V. M., and M. T.). We thank Dr. V. Kutyshenko and N.Molochkov for performing analytical ultracentrifugation and Creighton electrophoresis assay and Dr. A. Kharlanov for performing FTIR assay. We also thank Dr. S. Morozov and A.Solovyev for help and fruitful discussion.

ABBREVIATIONS

TGB	triple gene block
TGBp1 TGBp2 TGBp3	proteins encoded by three overlapping genes of TGB
RNP	ribonucleoprotein
PSLV	*Poa* semilatent virus
N63K	N-terminal region of the 63K protein including NTD and ID
NTD	N-terminal domain
ID	internal domain

PD	plasmodesmata
MP	movement protein
CP	capsid protein
BSMV	Barley stripe mosaic virus
HELD	NTPase/helicase domain
IDP	intrinsically disordered protein
IPTG	isopropyl-β-D-thiogalactoside
MS	mass spectrometry
Ni-NTA	nickel-nitrilotriacetic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
CD	circular dichroism
FTIR	Fourier transform infrared spectrometry
NMR	nuclear magnetic resonance
TFE	2,2,2 trifluoroethanol
DLS	dynamic laser light scattering
AFM	atomic force microscopy
PMTV	Potato mop-top virus
BNYVV	Beet necrotic yellow vein virus

REFERENCES

Adler AJ, Greenfield NJ, Fasman GD. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods Enzymol 1973;27:675–735.

Atabekov JG, Taliansky ME. Expression of a plant virus-coded transport function by different viral genomes. Adv Virus Res 1990;38:201–248.

Atabekov JG, Rodionova NP, Karpova OV, Kozlovsky SV, Novikov VK, Arkhipenko MV. Translational activation of encapsidated potato virus X by coat protein phosphorylation. Virology 2001;286:466–474.

Barilla D, Rosenberg MF, Nobbmann U, Hayes F. Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. EMBO J 2005;24:1453–1464.

Baryshnikova EN, Melnik BS, Finkelstein AV, Semisotnov GV, Bychkova VE. Three-state protein folding: experimental determination of free-energy profile. Protein Sci 2005;14:2658–2667.

Bleykasten C, Gilmer D, Guilley H, Richards KE, Jonard G. Beet necrotic yellow vein virus 42kDa triple gene block protein binds nucleic acid in vitro. J Gen Virol 1996;77:889–897.

Brakke MK, Ball EM, Langenberg WG. A non-capsid protein associated with unencapsidated virus RNA in barley infected with barley stripe mosaic virus. J Gen Virol 1988;69:481–491.

Canetta E, Kim SH, Kalinina NO, Shaw J, Adya AK, Gillespie T, Brown JWS, Taliansky M. A plant virus movement protein forms ringlike complexes with the major nucleolar protein, fibrillarin, in vitro. J Mol Biol 2008;376:932–937.

Carrington JC, Kasschau KD, Mahajan SK, Schaad MC. Cell-to-Cell and Long-Distance Transport of Viruses in Plants. Plant Cell 1996;8:1669–1681.

Chapman S, Hills G, Watts J, Baulcombe D. Mutational analysis of the coat protein gene of potato virus X, effects on virion morphology and viral pathogenicity. Virology 1992;191:223–230.

Cheng J, Randall A, Sweredoski M, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 2005;33:72–76.

Coeytaux K, Poupon A. Prediction of unfolded segments in a protein sequence based on amino acid composition. Bioinformatics 2005;21:1891–1900.

Corchero JL, Viaplana E, Benito A, Villaverde A. The position of the heterologous domain can influence the solubility and proteolysis of b-galactosidase fusion proteins in *E. coli*. J Biotechnol 1996;48:191–200.

Donald RG, Lawrence DM, Jackson AO. The barley stripe mosaic virus 58-kilodalton beta(b) protein is a multifunctional RNA binding protein. J Virol 1997;71: 1538–1546.

Efimov AV. Favoured structural motifs in globular proteins. Structure 1994;2(11): 999–1002.

Evdokimova VM, Wei CL, Sitikov AS, Simonenko PN, Lazarev OA, Vasilenko KS, Ustinov VA, Hershey JW, Ovchinnikov LP. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Ybox binding transcription factor family. J Biol Chem 1995;270:3186–3192.

Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15(1):35–41.

Fontana A, de Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol 2004;51:299–321.

Goldenberg DP, Creighton TE. Gel electrophoresis in studies of protein conformation and folding. Anal Biochem 1984;138:1–18.

Gorbalenya AE, Koonin EV. Helicases, amino acid sequence comparisons and structure-function relationships. Curr Opin Cell Biol 1993;3:419–429.

Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 1969;8:4108–4116.

Harries PA, Schoelz JE, Nelson RS. Intracellular transport of viruses and their components: utilizing the cytoskeleton and membrane highways. Mol Plant Microbe Interact 2010;23:1381–1393.

Haupt S, Cowan GH, Ziegler A, Roberts AG, Oparka KJ, Torrance L. Two plant-viral movement proteins traffic in the endocytic recycling pathway. Plant Cell 2005; 17:164–181.

Hering JA, Innocent PR, Haris PI. Automatic amide I frequency selection for rapid quantification of protein secondary structure from Fourier transform infrared spectra of proteins. Proteomics 2002;2:839–849.

Herzog E, Hemmer O, Hauser S, Meyer G, Bouzoubaa S, Fritsch C. Identification of genes involved in replication and movement of peanut clump virus. Virology 1998;248: 312–322.

Iseni F, Barge A, Baudin F, Blonder D, Ruigrok RW. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. J Gen Virol 1998;79: 2909–2919.

Jackson D, Lim H-S, Bragg J, Ganesan U, Lee M-Y. Hordeivirus replication, movement, and pathogenesis. Annu Rev Phytopathol 2009;47:385–422.

Kalinina NO, Rakitina DV, Yelina NE, Zamyatnin AA Jr., Stroganova TA, Klinov DV, Prokhorov VV, Ustinova SV, Chernov BK, Schiemann J, Solovyev AG, Morozov S Yu. RNA binding properties of the 63-kDa protein encoded by the triple gene block of poa semilatent hordeivirus. J Gen Virol 2001;82:2569–2578.

Kalinina NO, Rakitina DV, Solovyev AG, Schiemann J, Morozov SY. RNA helicase activity of the plant virus movement proteins encoded by the first gene of the triple gene block. Virology 2002;296:321–329.

Karpova OV, Zayakina OV, Arkhipenko MV, Sheval EV, Kiselyova OI, Poljakov VYu, Yaminsky IV, Rodionova NP, Atabekov JG. Potato virus X RNA-mediated assembly of single-tailed ternary 'coat protein-RNA-movement protein' complexes. J Gen Virol 2006;87:2731–2740.

Kim S-H, Kalinina NO, Andreev I, Ryabov EV, Fitzgerald AG, Taliansky ME, Palukaitis P. The C-terminal 33 amino acids of the cucumber mosaic virus 3a protein affect virus movement, RNA binding and inhibition of infection and translation. J Gen Virol 2004;85:221–230.

Kim SH, MacFarlane S, Kalinina NO, Rakitina DV, Ryabov EV, Gillespie T, Haupt S, Brown JWS, Taliansky M. Interaction of a plant virus-encoded protein with the major nucleolar protein fibrillarin is required for systemic virus infection. Proc Natl Acad Sci USA 2007b;104:11115–11120.

Kim SH, Ryabov EV, Kalinina NO, Rakitina DV, Gillespie T, MacFarlane S, Haupt S, Brown JWS, Taliansky M. Cajal bodies and the nucleolus are required for a plant virus systemic infection. EMBO J 2007a;26:2169–2179.

Kiselyova OI, Yaminsky IV, Karpova OV, Rodionova NP, Kozlovsky SV, Arkhipenko MV, Atabekov JG. AFM study of potato virus X disassembly induced by movement protein. J Mol Biol 2003;332:321–325.

Kiselyova OI, Yaminsky IV, Karger EM, Frolova OY, Dorokhov YL, Atabekov JG. Visualization by atomic force microscopy of tobacco mosaic virus movement protein-RNA complexes formed in vitro. J Gen Virol 2001;6:1503–1508.

Leshchiner AD, Solovyev AG, Morozov SY, Kalinina NO. A minimal region in the NTPase/helicase domain of the TGBp1 plant virus movement protein is responsible for ATPase activity and cooperative RNA binding. J Gen Virol 2006;87:3087–3095.

Lim H-S, Bragg JN, Ganesan U, Lawrence DM, Yu J, Isogai M, Hammond J, Jackson AO. Triple gene block protein interactions involved in movement of barley stripe mosaic virus. J Virol 2008;82:4991–5006.

Longhi S, Receveur-Bre'chot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 2003;278:18638–18648.

Lucas WJ. Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. Virology 2006;344:169–184.

Makarov VV, Obraztsova EA, Solovyev AG, Morozov SY, Taliansky ME, Yaminsky IV, Kalinina NO. The internal domain of hordeivirus movement protein TGB1 forms in vitro filamentous structures. Biochemistry (Mosc) 2010;75(6):752–758.

Makarov VV, Rybakova EN, Efimov AV, Dobrov EN, Serebryakova MV, Solovyev AG, Yaminsky IV, Taliansky ME, Morozov SYu, Kalinina NO. Domain organization of the N-terminal portion of hordeivirus movement protein TGBp1. J Gen Virol 2009;90: 3022–3032.

Morozov SY, Solovyev AG. Triple gene block: modular design of a multifunctional machine for plant virus movement. J Gen Virol 2003;84:1351–1366.

Namba K. Roles of partly unfolded conformations in macromolecular self-assembly. Genes Cells 2001;6:1–12.

Nemykh MA, Efimov AV, Novikov VK, Orlov VN, Arutyunyan AM, Drachev AV, Lukashina EV, Baratova LA, Dobrov EN. One more probable structural transition in potato virus X virions and a revised model of the virus coat protein structure. Virology 2008;373:61–71.

Oparka KJ. Getting the message across: how do plant cells exchange macromolecular complexes? Trends Plant Sci 2004;9:33–41.

Oparka KJ, Turgeon R. Sieve elements and companion cells-traffic control centers of the phloem. Plant Cell 1999;11:739–750.

Petty IT, Jackson AO. Mutational analysis of barley stripe mosaic virus RNA beta. Virology 1990;179:712–718.

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 2005;21:3435–3438.

Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assesing protein disorder and induced folding. Proteins 2006;62:24–45.

Roberts AG, Oparka KJ. Plasmodesmata and the control of symplastic transport. Plant Cell Environ 2003;26:103–124.

Rodionova NP, Karpova OV, Kozlovsky SV, Zayakina OV, Arkhipenko MV, Atabekov JG. Linear remodeling of helical virus by movement protein binding. J Mol Biol 2003;333:565–572.

Ruigrok RWH, Baudin F. Structure of influenza virus ribonucleoprotein particles. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. J Gen Virol 1995;76:1009–1014.

Santa Cruz S, Roberts AG, Prior DA, Chapman S, Oparka KJ. Cell-to-cell and phloem-mediated transport of potato virus X. The role of virions Plant Cell 1998;10:495–510.

Santa Cruz S. Perspective: phloem transport of viruses and macromolecules - what goes in must come out. Trends Microbiol 1999;7:237–241.

Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl a-chymotrypsin using different denaturants. Biochemistry 1988;27:8063–8068.

Savenkov EI, Germundsson A, Zamyatnin AA Jr., Sandgren M, Valkonen JP. Potato mop-top virus: the coat protein-encoding RNA and the gene for cysteine-rich protein are dispensable for systemic virus movement in *Nicotiana benthamiana*. J Gen Virol 2003;84:1001–1005.

Schlotmann M, Beyreuther K. Degradation of the DNA binding domain of wild-type and i2d lac repressors in *Escherichia coli*. Eur J Biochem 1979;95:39–49.

Schmitt C, Balmori E, Jonard G, Richards KE, Guilley H. In vitro mutagenesis of biologically active transcripts of beet necrotic yellow vein virus RNA 2: evidence that a domain of the 75-kDa readthrough protein is important for efficient virus assembly. Proc Natl Acad Sci USA 1992;89:5715–5719.

Schmitz SK. An introduction to dynamic light scattering by macromolecules, New York: Academic Press; 1990.

Shiraki K, Nishikawa K, Goto Y. Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding. J Mol Biol 1995;245:180–194.

Skabkin MA, Kiselyova OI, Chernov KG, Sorokin AV, Dubrovin EV, Yaminsky IV, Vasiliev VD, Ovchinnikov LP. Structural organization of mRNA complexes with major core mRNP protein YB-1. Nucleic Acids Res 2004;32:5621–5635.

Solovyev AG, Savenkov EL, Agranovsky AA, Morozov SYu. Comparisons of the genomic cis-elements and coding regions in RNAb components of the hordeiviruses barley stripe mosaic virus lychnis ringspot virus, and poa semilatent virus. Virology 1996;219:9–18.

Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. Methods Enzymol 2004;383:318–351.

Tamada T, Schmitt C, Saito M, Guilley H, Richards K, Jonard G. High resolution analysis of the readthrough domain of beet necrotic yellow vein virus readthrough protein: a KTER motif is important for efficient transmission of the virus by Polymyxa betae. J Gen Virol 1996;77:1359–1367.

Tonges L, Lingor P, Egle R, Dietz GP, Fahr A, Bahr M. Stearylated octaarginine and artificial virus-like particles for transfection of siRNA into primary rat neurons. RNA 2006;12:1431–1438.

Torrance L, Lukhovitskaya NI, Schepetilnikov MV, Cowan GH, Ziegler A, Savenkov EI. Unusual long-distance movement strategies of Potato mop-top virus RNAs in Nicotiana benthamiana. Mol Plant Microbe Interact 2009;22:381–390.

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.

Uversky VN. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. Protein J 2009;28:305–325. Review.

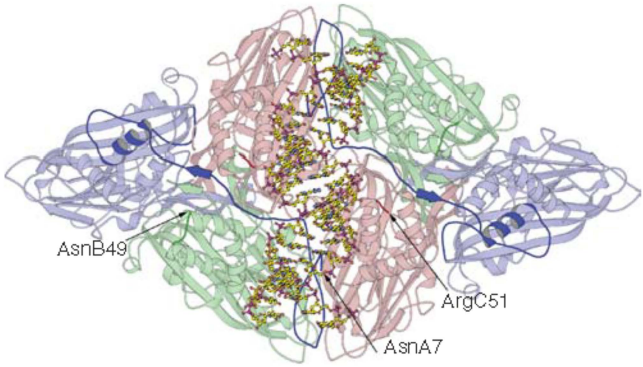
Weston A, Sommerville J. Xp54 and related (DDX6-like) RNA helicases: roles in messenger RNP assembly, translation regulation and RNA degradation. Nucleic Acids Res 2006;34:3082–3094.

Wright K, Cowan G, Lukhovitskaya N, Tilsner J, Roberts A, Savenkov E, Torrance L. The N-terminal domain of PMTV TGB1 movement protein is required for nucleolar localization, microtubule association and long-distance movement. Mol Plant-Microbe Int 2010;23:1486–1497.

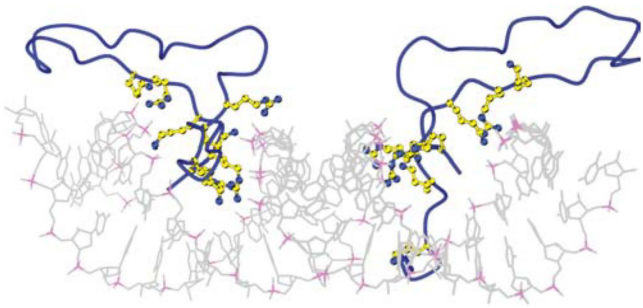
Zamyatnin AA Jr., Solovyev AG, Savenkov EI, Germundsson A, Sandgren M, Valkonen JP, Morozov SY. Transient coexpression of individual genes encoded by the triple gene block of potato mop-top virus reveals requirements for TGBp1 trafficking. Mol Plant Microbe Interact 2004;17:921–930.

INDEX

Flexible Viruses: Structural Disorder in Viral Proteins, First Edition.
Edited by Vladimir N. Uversky and Sonia Longhi.



(a)



(b)

Figure 2.3 See full caption on page 38.

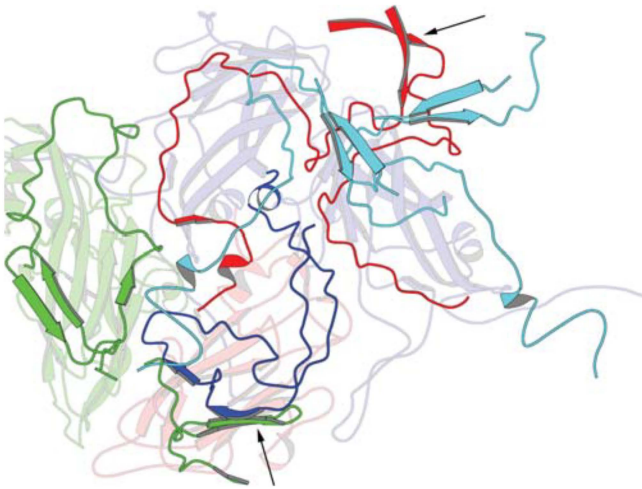
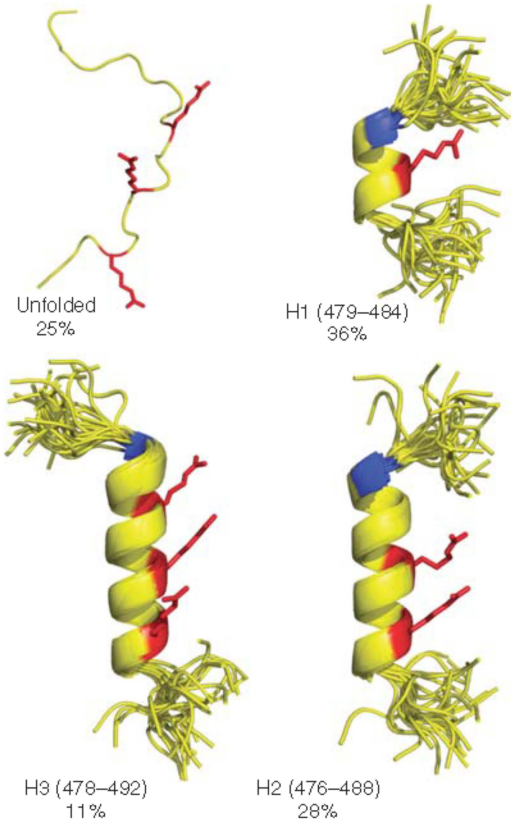


Figure 2.6 See full caption on page 43.



^{471}TNDEDVSDIERRIAMRLAERRQEDSA^{496}
TNDEDVSDIERRIAMRLAERRQEDSA H3 – 11%
TNDEDVSDIERRIAMRLAERRQEDSA H2 – 28%
TNDEDVSDIERRIAMRLAERRQEDSA H1 – 36%

Figure 4.11 See full caption on page 109.

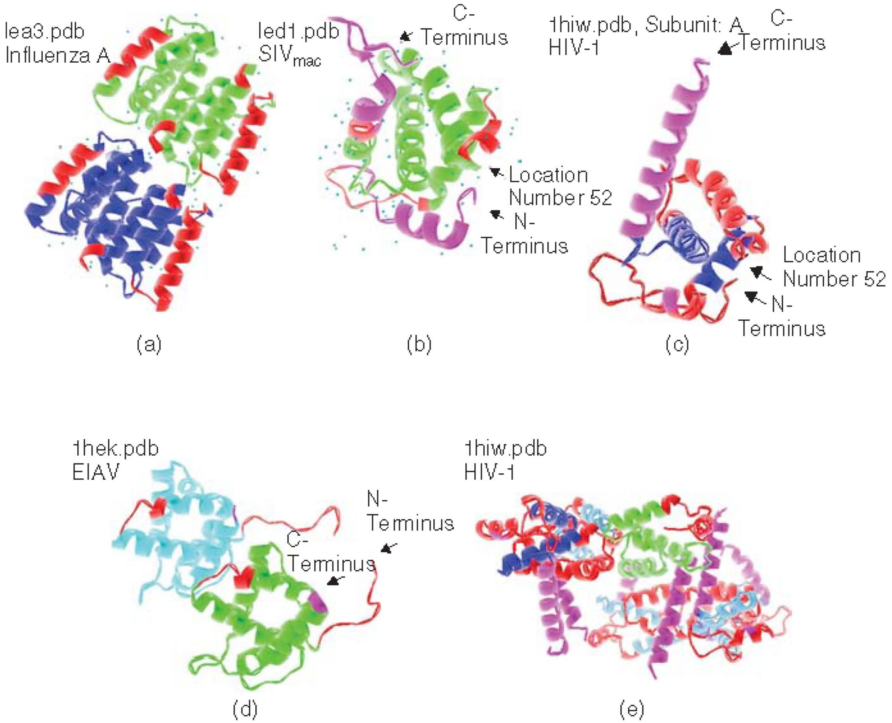


Figure 6.6 See full caption on page 156.

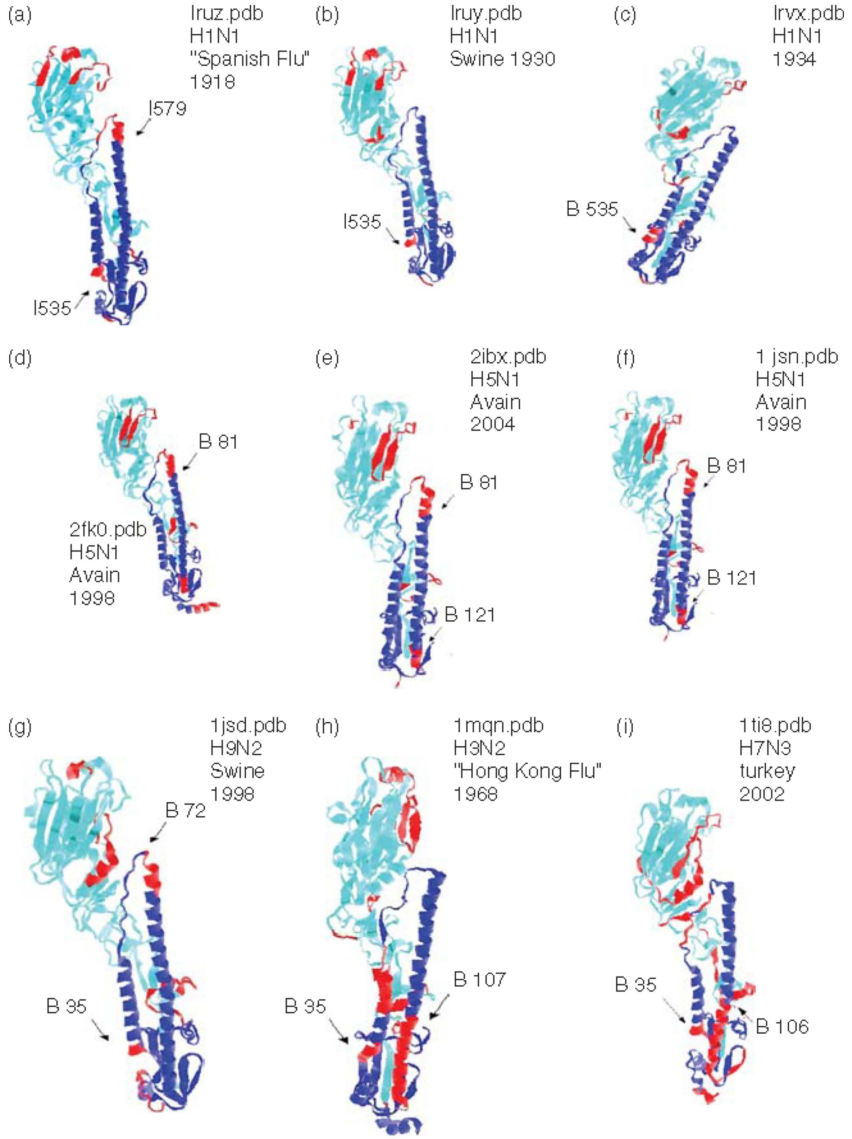


Figure 7.4 See full caption on page 186.

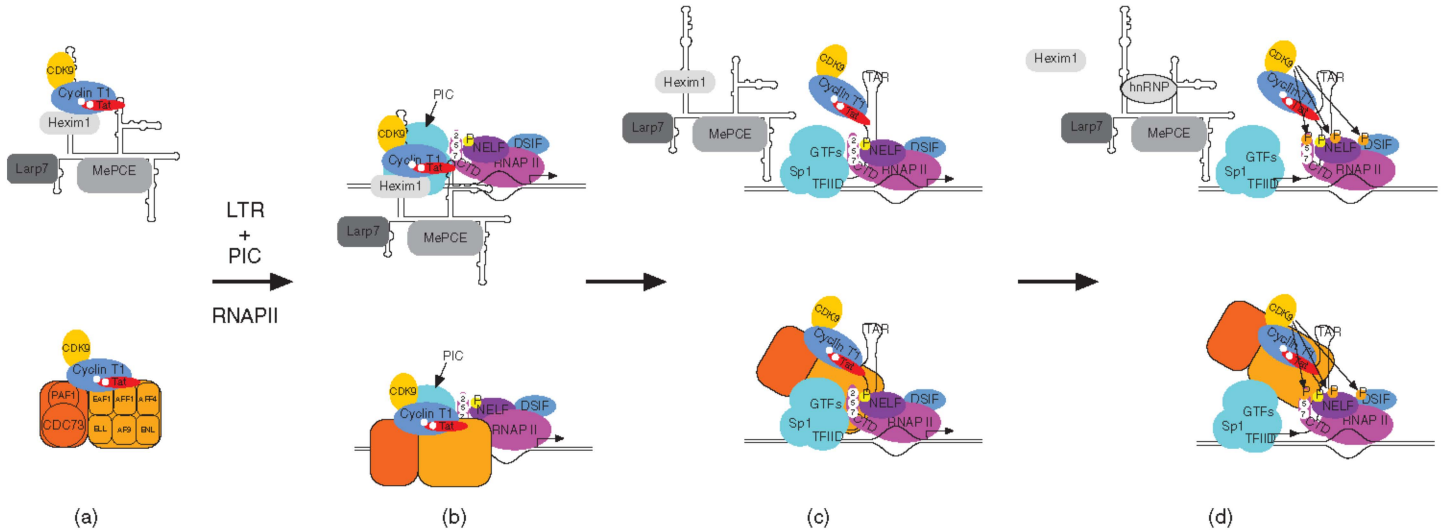


Figure 9.6 See full caption on page 233.

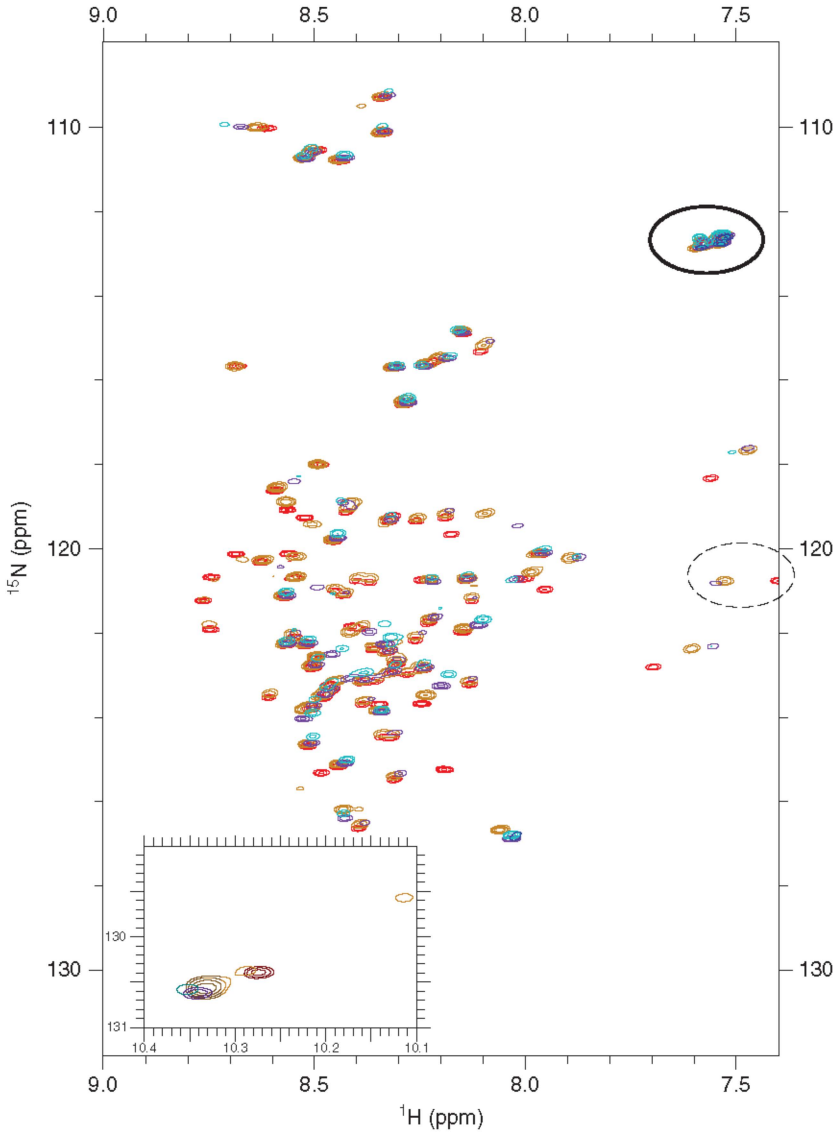


Figure 9.11 See full caption on page 241.

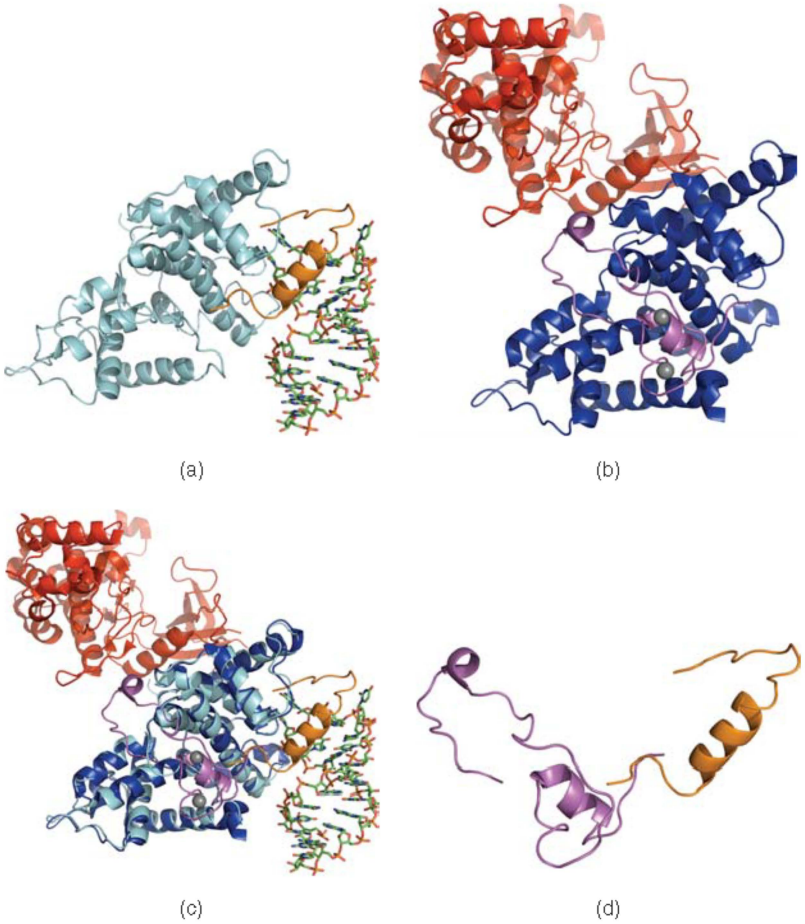
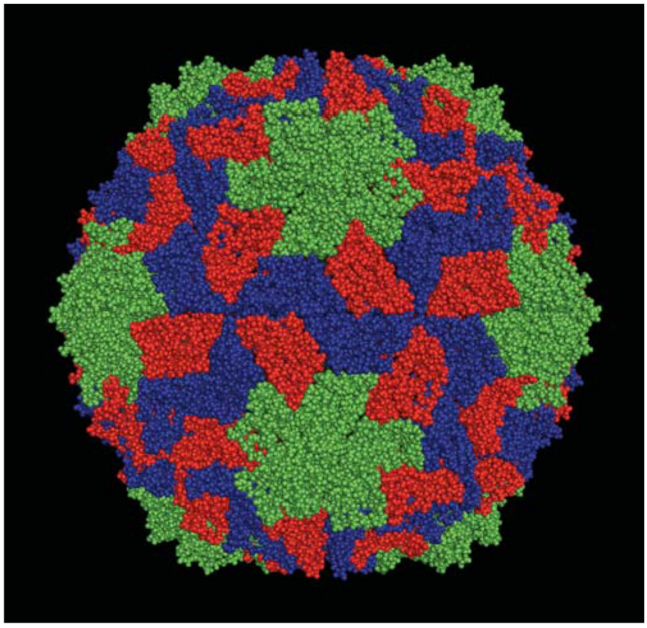
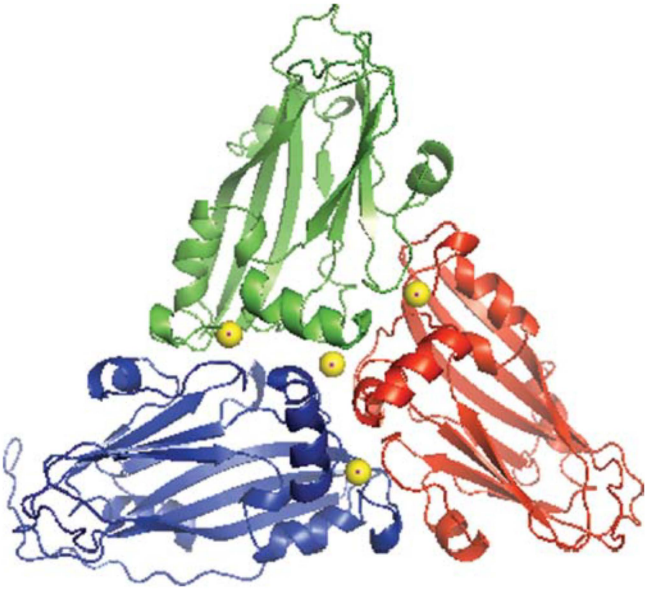


Figure 9.12 See full caption on page 243.



(a)



(b)

Figure 10.1 See full caption on page 259.

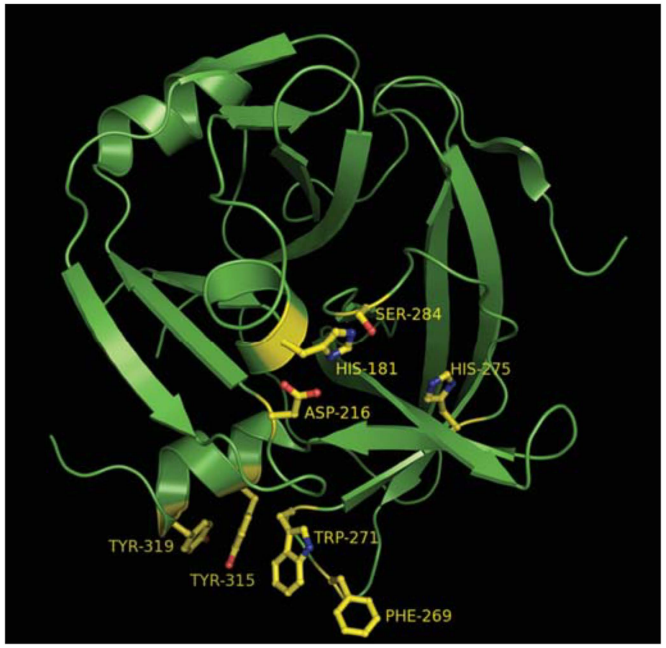


Figure 10.7 See full caption on page 266.

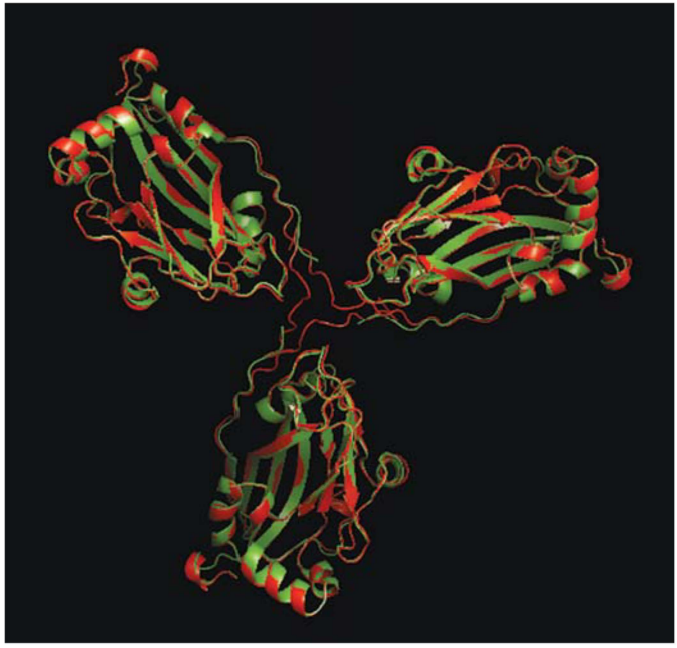
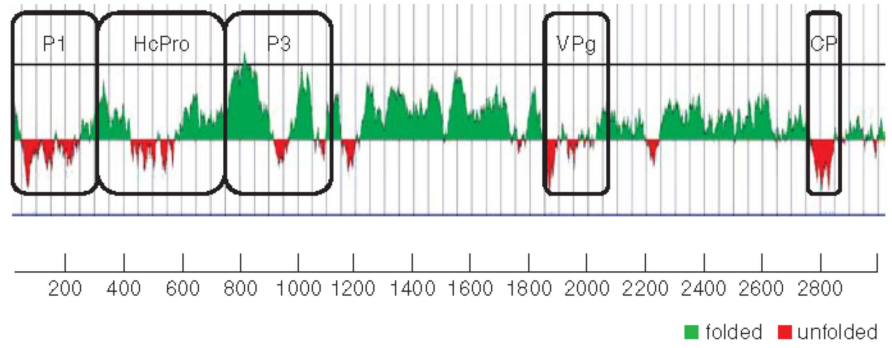
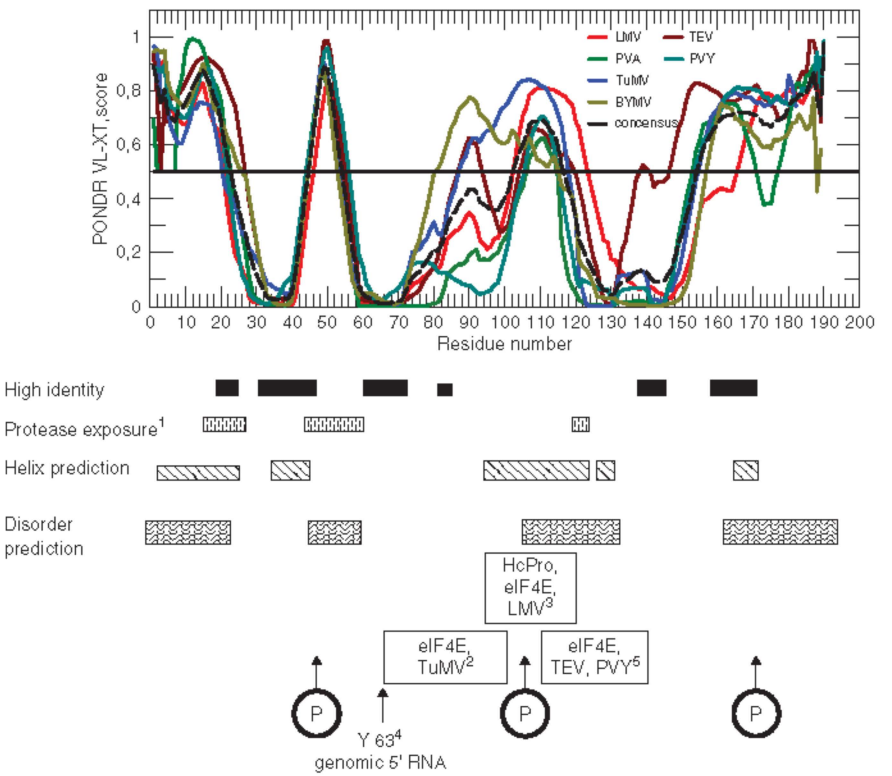


Figure 10.13 See full caption on page 272.



(a)



(b)

Figure 11.12 See full caption on page 301.

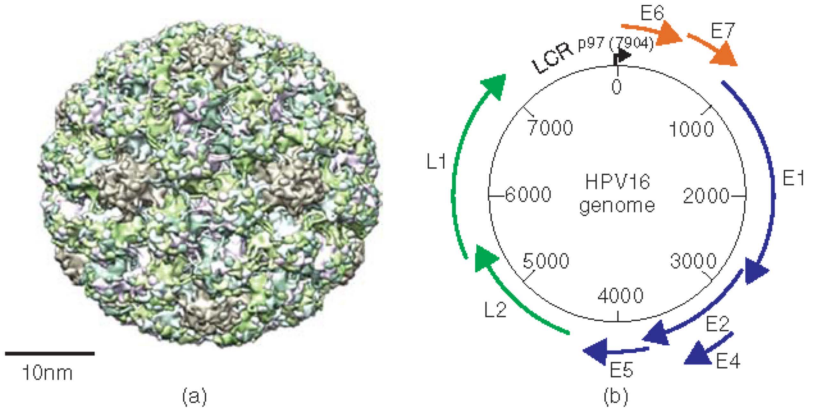


Figure 12.1 See full caption on page 316.

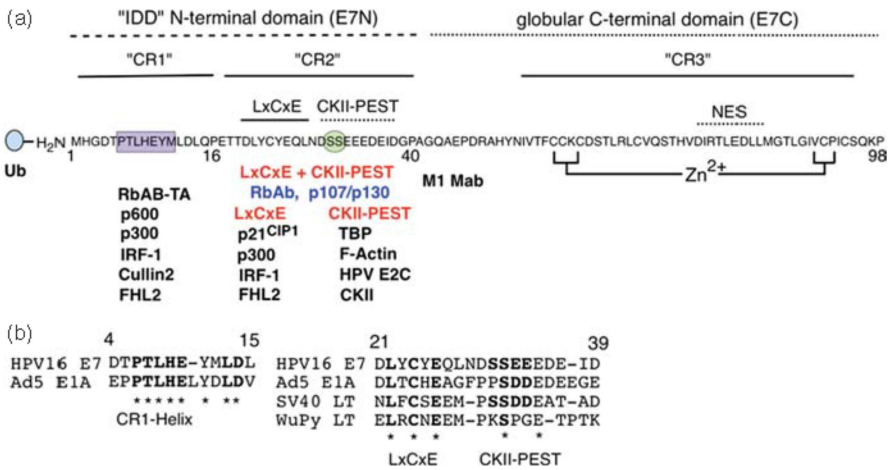


Figure 12.2 See full caption on page 319.

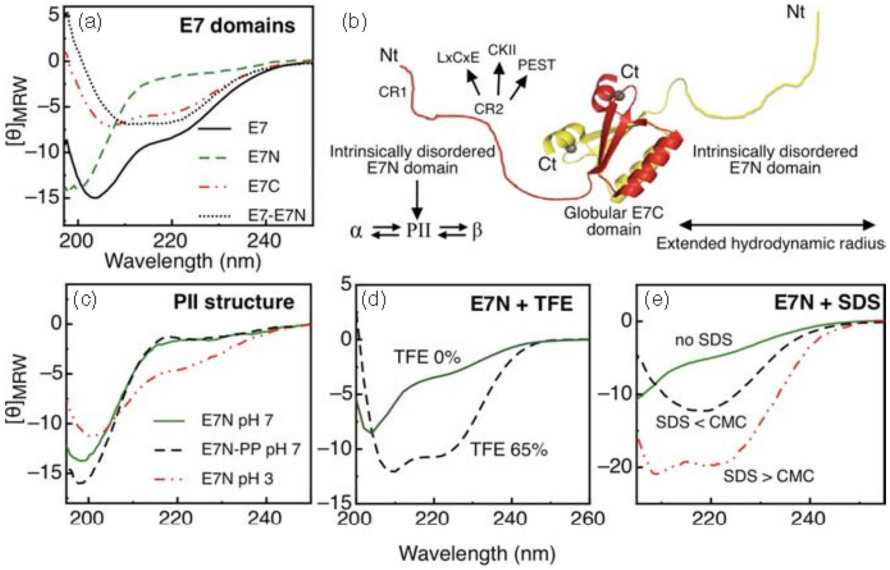


Figure 12.5 See full caption on page 327.

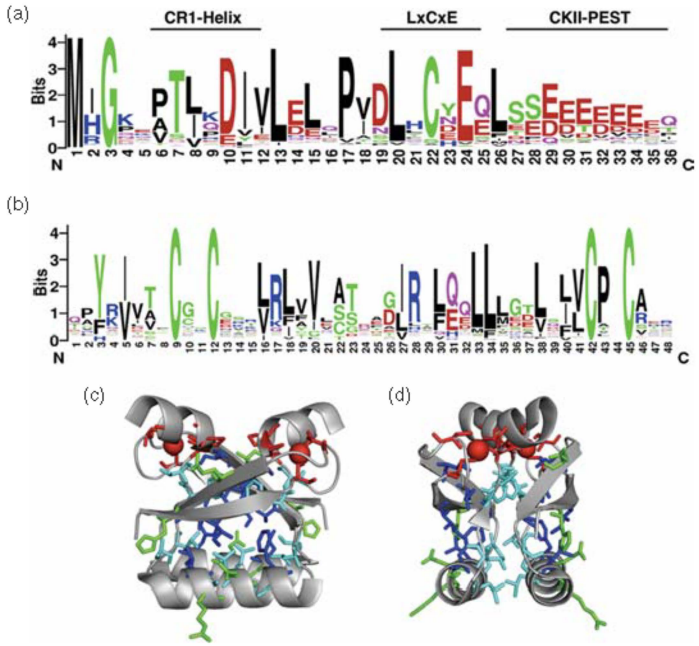


Figure 12.7 See full caption on page 335.

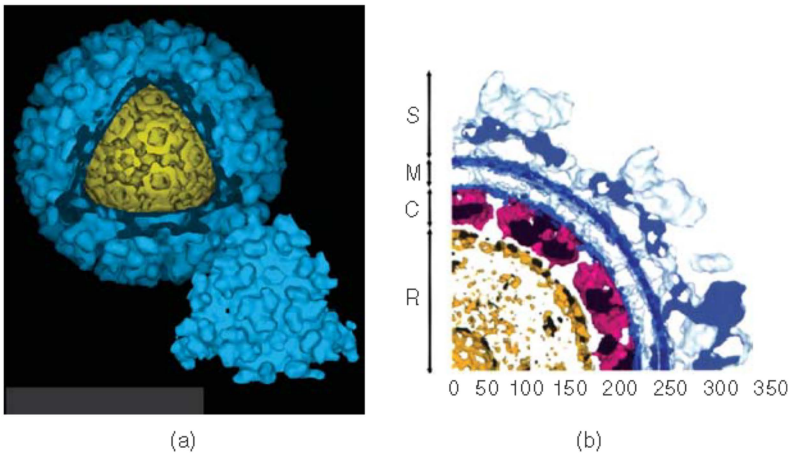
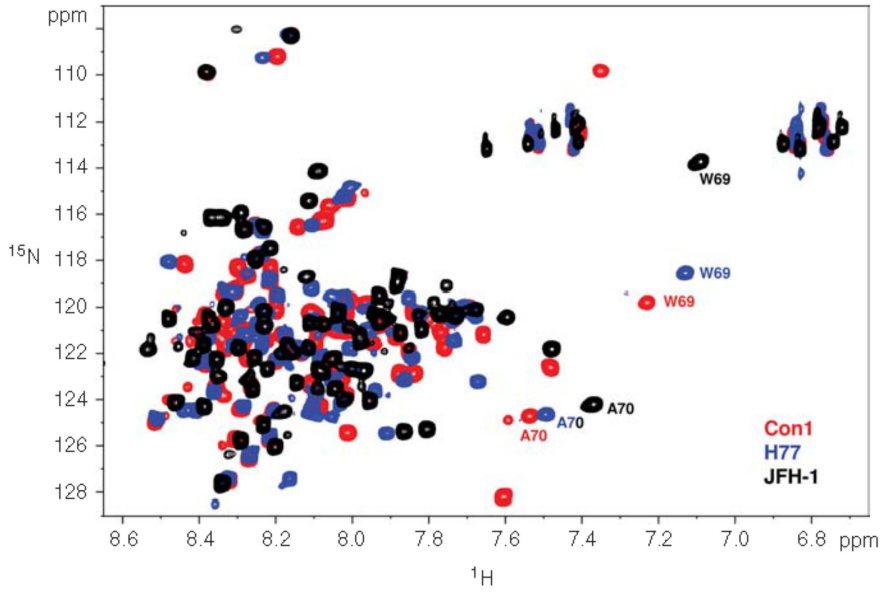
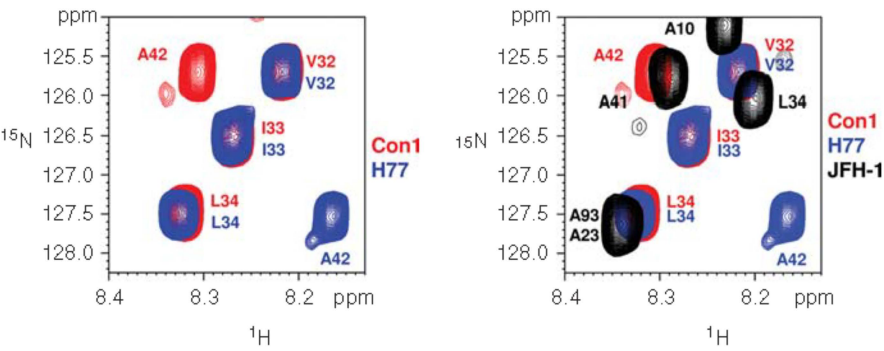


Figure 13.4 See full caption on page 352.



(a)



(b)

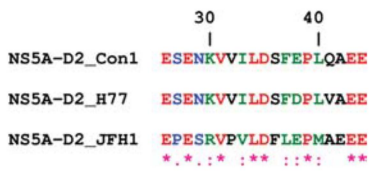


Figure 15.5 See full caption on page 419.

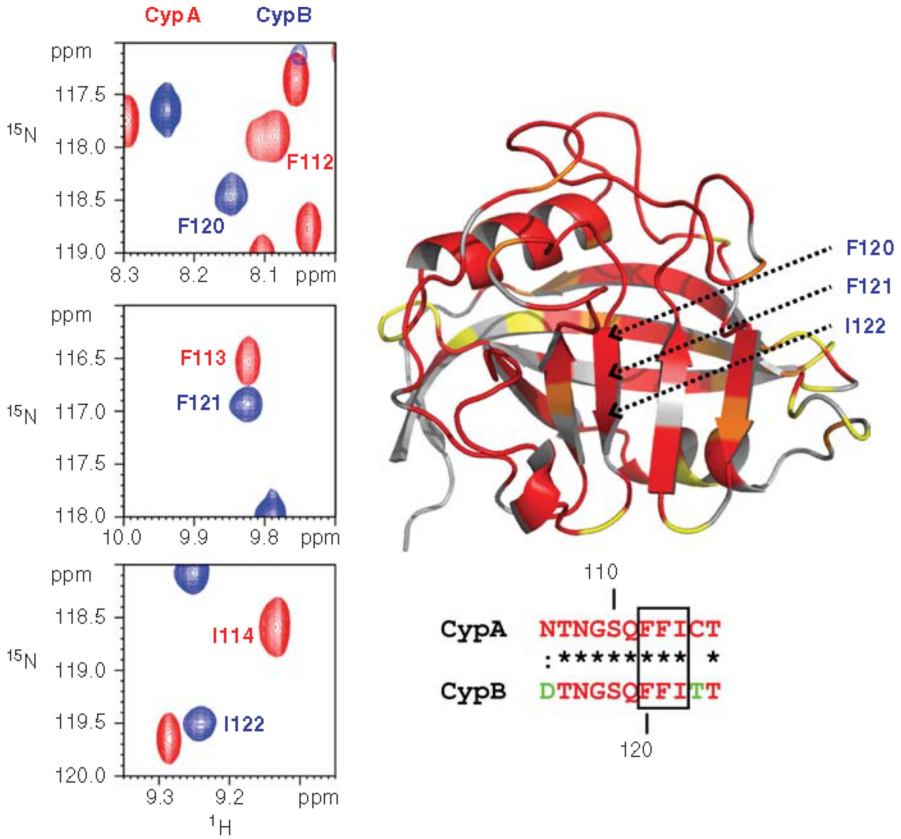


Figure 15.6 See full caption on page 421.

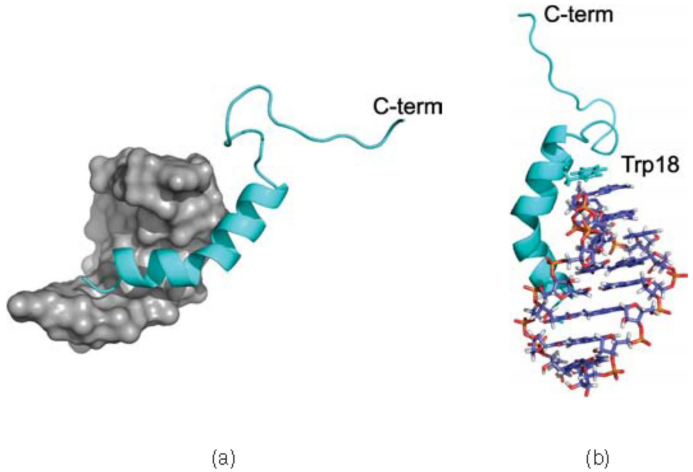


Figure 16.3 See full caption on page 434.

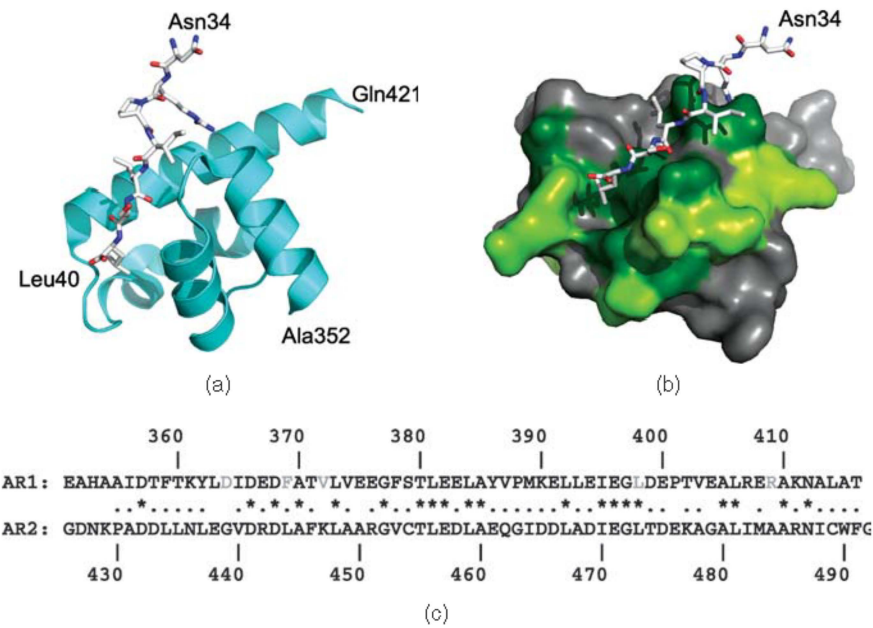


Figure 16.6 See full caption on page 439.