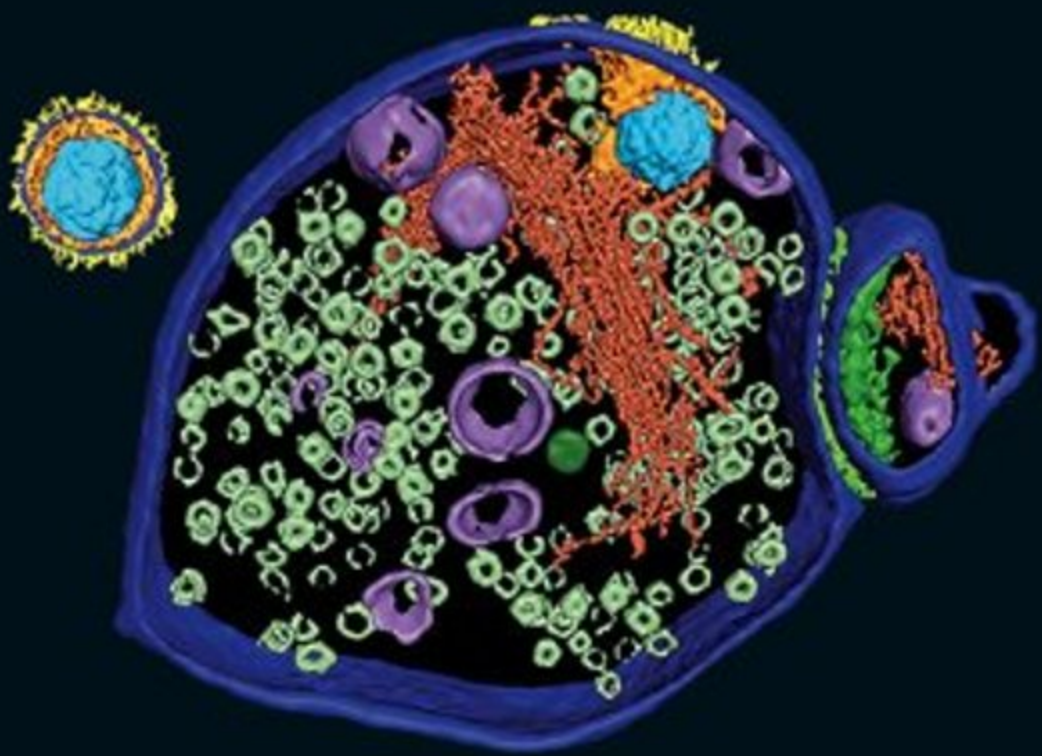
Emerging Topics in Physical Virology



Peter G Stockley • Reidun Twarock
editors

Imperial College Press

Emerging Topics in Physical Virology

This page intentionally left blank

Emerging Topics in Physical Virology

Editors

Peter G Stockley
University of Leeds, UK

Reidun Twarock
University of York, UK

ICP
Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

EMERGING TOPICS IN PHYSICAL VIROLOGY
Copyright © 2010 by Imperial College Press

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-1-84816-464-2
ISBN-10 1-84816-464-5

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore.

Contributors

Peter G. Stockley
Asbury Centre for Structural Molecular Biology,
University of Leeds, Leeds, LS2 9JT, UK
p.g.stockley@leeds.ac.uk

Reidun Twarock
Departments of Mathematics and Biology,
York Centre for Complex Systems Analysis,
University of York,
York, YO10 5DD, UK
rt507@york.ac.uk

Nicola G. A. Abrescia
Structural Biology Unit, CICbioGUNE,
Bizkaia Technology Park,
48160 Derio, Spain
nabrescia@cicbiogune.es

Javier Arsuaga
Department of Mathematics,
San Francisco State University,
San Francisco, CA 94132, USA
jarsuaga@sfsu.edu

Dennis H. Bamford
Institute of Biotechnology and Department of
Biological and Environmental Sciences,
Viikki Biocenter, University of Helsinki,
00014 University of Helsinki, Finland
dennis.bamford@helsinki.fi

Robijn F. Bruinsma
Department of Physics and Astronomy,
University of California, Los Angeles,
Los Angeles, CA 90024, USA
bruinsma@physics.ucla.edu

Martin Castelnovo
Laboratoire de Physique,
Ecole Normale Superieure de Lyon,
69364 Lyon Cedex 07, France
martin.castelnovo@ens-lyon.fr

Alex Evilevitch
Department of Physics,
Carnegie Mellon University,
Pittsburgh, PA 15213, USA
alex.evilevitch@biochemistry.lu.se

Elizabeth E. Fry
Division of Structural Biology
and the Oxford Protein Production Facility,
The Wellcome Trust Centre for Human Genetics,
University of Oxford,
Headington, Oxford, OX3, 7BN, UK
Liz@strubi.ox.ac.uk

William M. Gelbart
Department of Chemistry and Biochemistry,
University of California, Los Angeles,
Los Angeles, CA 90095-1569, USA
gelbart@chem.ucla.edu

Jonathan M. Grimes
Division of Structural Biology
and the Oxford Protein Production Facility,
The Wellcome Trust Centre for Human Genetics,
University of Oxford,
Headington, Oxford, OX3 7BN, UK
jonathan@strubi.ox.ac.uk

Thomas Keef
Department of Mathematics,
University of York, York, YO10 5DD, UK
tk506@york.ac.uk

William S. Klug
Department of Mechanical and Aerospace Engineering,
University of California, Los Angeles,
Los Angeles, CA 90095, USA
klug@seas.ucla.edu

Charles M. Knobler
Department of Chemistry and Biochemistry,
University of California, Los Angeles,
Los Angeles, CA 90095-1569, USA
knobler@chem.ucla.edu

Kristopher J. Koudelka
Department of Cell Biology,
The Scripps Research Institute,
La Jolla, CA 92037, USA
koudelka@scripps.edu

Marianne Manchester
Skaggs School of Pharmacy and Pharmaceutical Sciences,
University of California San Diego,
9500 Gilman Drive, MC 0749,
La Jolla, CA 92093, USA
mmanchester@ucsd.edu

Eric B. Monroe
Department of Microbiology,
University of Alabama at Birmingham,
Birmingham, AL 35294-2170, USA
ebmonroe@uab.edu

Alexander Yu. Morozov
Department of Physics and Astronomy,
University of California, Los Angeles,
Los Angeles, CA 90024, USA
morozov@physics.ucla.edu

J. Zachary Porterfield
University of Oklahoma Health Sciences Center,
Dept of Biochemistry, Oklahoma City,
OK 73104, USA
zach-porterfield@ouhsc.edu

Peter E. Prevelige
Department of Microbiology,
University of Alabama at Birmingham,
Birmingham, AL 35294-2170, USA
prevelige@uab.edu

Neil A. Ranson
Asbury Centre for Structural Molecular Biology,
University of Leeds, Leeds, LS2 9JT, UK
n.a.ranson@leeds.ac.uk

Janne J. Ravantti
Institute of Biotechnology and Department
of Biological and Environmental Sciences,
Viikki Biocenter, University of Helsinki,
00014 University of Helsinki, Finland
janne.ravantti@helsinki.fi

Joaquim Roca
Instituto de Biologia Molecular de Barcelona,
CSIC, Barcelona, Spain
joaquim.roca@ibmb.csic.es

Wouter H. Roos
Fysica van complexe systemen,
Natuur- en Sterrenkunde, Vrije Universiteit,
1081 HV Amsterdam, The Netherlands
wroos@few.vu.nl

Joseph Rudnick
Department of Physics and Astronomy,
University of California, Los Angeles,
Los Angeles, CA 90024, USA
jrudnick@physics.ucla.edu

David I. Stuart
Division of Structural Biology
and the Oxford Protein Production Facility,
The Wellcome Trust Centre for Human Genetics,
University of Oxford,
Headington, Oxford, OX3 7BN, UK
dave@strubi.ox.ac.uk

De Witt Sumners
Department of Mathematics,
Florida State University,
Tallahassee, FL 32306-4510, USA
sumners@math.fsu.edu

Gijs J. L. Wuite
Fysica van complexe systemen,
Natuur- en Sterrenkunde, Vrije Universiteit,
1081 HV Amsterdam, The Netherlands
gwuite@nat.vu.nl

Adam Zlotnick
Department of Biology,
Indiana University,
Bloomington, IN 47405, USA
azlotnic@indiana.edu

This page intentionally left blank

Preface

Emerging Topics in Physical Virology is a state-of-the-art account of recent developments in the analysis and modelling of virus structure, function and dynamics. It pays tribute to the importance of interdisciplinary research by integrating an exposition of experimental techniques such as cryo-electron microscopy, atomic force microscopy and mass spectrometry with mathematical and biophysical modelling techniques. The number of chapters co-authored by experimentalists and theoreticians is testimony to the importance of interdisciplinarity in tackling some of the most challenging and exciting research problems in this area.

The aim of this book is to introduce the reader to recent developments in the field, and to provide a comprehensive review of the results that prompted them. It is therefore not only a primer for researchers working in the analysis and modelling of viruses, but also serves as an introduction for non-experts into this tantalising field of research. The book starts with a description of cryo-electron microscopy by Neil Ranson and Peter Stockley and demonstrates its power in determining the structure and dynamics of viruses. Structural insights gained from X-ray crystallography have revealed an intriguing phenomenon: there is a striking conservation in the topologies of the capsid proteins that form the containers encapsulating viral genomes. This has prompted Dennis Bamford, David Stuart and collaborators to classify viral families into lineages based on the concept of the viral 'self'. An important feature of this conservation of capsid protein folds is that it appears to be non-sequence-specific, i.e. the chemical structures of proteins with homologous folds can often be very different. This suggests that there must be a guiding principle for the formation of the capsid proteins that is independent of their sequences. Thomas Keef and Reidun Twarock suggest that the icosahedral symmetry of many viruses may provide such a guiding principle, and they introduce novel group theoretical techniques to model this effect. Their approach implies that a wide spectrum of viral features can be predicted based on symmetry, and that perhaps the limited number of structural folds is a

consequence of the fact that only a limited number of layouts are possible for viruses with such symmetry.

Atomic force microscopy provides important insights into the mechanical properties of viruses as detailed in the chapter by Wouter Roos and Gijs Wuite. The authors discuss capsid shell structure, presence of encapsidated material, capsid failure, maturation and capsid protein mutations in relation to viral material properties and highlight similarities and differences for different types of viruses. Another important experimental technique in the study of viruses is mass spectrometry. Eric Monroe and Peter Prevelige show how this technique can be used to gain invaluable information on viral proteins. Mass spectrometry can also play a crucial role in the study of virus assembly, i.e. the process of formation of viral particles from their protein building blocks and genomic material. An overview of capsid assembly kinetics is provided by one of the pioneers in this area, Adam Zlotnick, and his collaborator Zachary Porterfield. Their chapter covers both modelling and experimental techniques and provides a comprehensive overview of viral capsid kinetics. An important factor in virus assembly and disassembly is the mechanical stress on different components of the viral capsid. A beautiful account of how stress distributions impact on the assembly and disassembly of viral capsids formed from pentamers and hexamers is provided by Robijn Bruinsma and collaborators.

Viruses may package genomic material in the form of DNA or RNA. An important question concerning the formation of RNA viruses is 'what determines the size of an RNA virus?' It is addressed by Chuck Knobler and Bill Gelbart in their discussion of the correlation between capsid and genome sizes. The impact of genome length versus capsid size on the physics of viral infectivity is also discussed with respect to double-stranded DNA (dsDNA) phages by Alex Evilevitch and Martin Castelnovo. Another intriguing feature of packaged DNA genomic material is its topology. Together with their experimental collaborator Joaquim Roca, Jarvier Arsuaga and De Witt Sumners provide a comprehensive account of the mathematical and experimental analysis of the topology of viral DNA.

A volume on emerging topics in physical virology would not be complete without a discussion of the fascinating applications of viruses and virus-like particles in biomedical nanotechnology that are opened up by

these recent theoretical and experimental approaches. We therefore conclude with a chapter by Kristopher Koudelka and Marianne Manchester that discusses the state of the art in this area.

We would like to express our special thanks to the authors of these fascinating chapters for making this volume possible by sharing their exciting research with us.

Peter Stockley and Reidun Twarock
March 2009

This page intentionally left blank

Contents

Chapter 1

Cryo-Electron Microscopy of Viruses

Neil A. Ranson*,† and Peter G. Stockley*,‡

Cryo-electron microscopy (cryo-EM) is a structural technique that images biological macromolecules in native-like conditions, and has been widely applied to the study of viruses. Virus structures have been determined by cryo-EM at resolutions ranging from molecular (~30–50 Å) to near-atomic (~4 Å). Here we introduce cryo-EM of virus particles for the non-expert reader and review how some of the key cryo-EM studies have advanced our understanding of virus biology. We also describe the latest advances in cryo-EM. These advances are on the one hand driving cryo-EM studies of symmetric viruses towards atomic resolution. On the other, they are developing structural methods that allow the study of individual, pleiomorphic virus particles and the interactions they make with cellular machinery.

1. Introduction

The molecular details of how viruses infect and hijack the cellular processes of their host cells are of critical importance in biology and medicine. It is only through a precise understanding of such events that new anti-viral therapies will be developed. One of the key elements of this growing understanding of the viral life cycle has been an increased understanding of the structure of viruses, and in particular their dynamic properties. Together with X-ray crystallography, cryo-electron microscopy (cryo-EM) techniques have played a central role in studying virus structure. At the same time, viruses have played an equally important role in the development of cryo-EM and single-particle image processing as techniques in modern structural biology. In this chapter we will illustrate these developments for non-expert readers with selected examples.

*Asbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK. E-mails: †n.a.ranson@leeds.ac.uk; ‡p.g.stockley@leeds.ac.uk

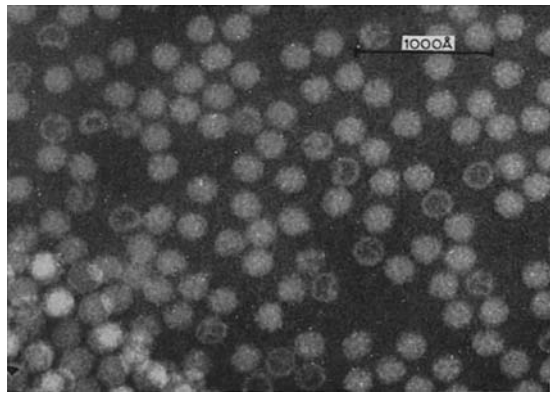


Fig. 1. Electron micrograph of Turnip Yellow Mosaic Virus (TYMV) negatively stained with 1% phosphotungstic acid. From Brenner & Horne (1959). *Biochem. Biophys. Acta*. **34**, p. 103.

Electron microscopy has been used to image viruses for more than 50 years. Initial studies involved the use of negative staining, in which the virus is placed on a continuous carbon film, coated in a heavy metal salt such as uranyl acetate, ammonium molybdate or phosphotungstic acid, and then dried before imaging (Fig. 1). Such treatment embeds the virus in a cast of the heavy metal salt and it is this cast, rather than the biological macromolecule itself, that gives rise to contrast in the images; hence 'negative' staining. Staining methods are rapid, generate high image contrast, and allow relatively large electron doses to be used, producing readily interpretable relatively low-resolution images of virus particles. However, staining also has major disadvantages. Firstly, penetration of stain into the interior of macromolecules is limited by both the structure of those macromolecules and by the size of the stain crystals. In practice this means that negative staining typically reveals only the external envelope of a macromolecule rather than any internal features. Secondly, the drying process can significantly distort the structure being imaged, as specimens are typically flattened onto the carbon support film as they dehydrate (see Fig. 2a). Such structural deformations are exacerbated by the pH and ionic strength of the stain solution, which places the biological macromolecule in a profoundly non-native environment. In summary, negatively stained images of viruses therefore are excellent at identifying the presence of viral particles in both purified and cellular samples, and this remains a major tool in the search for the presence of novel viruses in tissues.

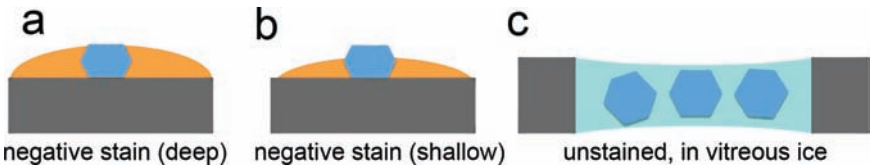


Fig. 2. A schematic representation of specimens for EM studies. (a) A negatively-stained virus (blue hexagon) adsorbed to a carbon support film (grey), and embedded in deep stain (orange). Note the flattening of the hexagonal virus onto the support film. (b) A viruss particle partially embedded in shallow stain. The portion of the virus structure not embedded in stain will not contribute to images of the specimen. (c) A cryo-EM specimen. Virus particles are unstained and trapped in a layer of vitreous ice. The ice layer is formed in a hole in the carbon support film.

A further complication of electron microscopy in general is caused by the nature of image formation in the microscope. EM has a depth of field that is significantly larger than most biological specimens. The consequence of this is that the images formed are projections of the three-dimensional (3-D) object onto the plane of the two-dimensional (2-D) recorded image, which greatly complicates interpretation. Various tricks have been developed to reveal information about the third dimension. One technique often used is metal shadowing at defined angles, which produces a pattern of electron dense material around particles of interest. The lengths of shadowed areas are proportional to the height of the object. More sophisticated image processing techniques have also been developed to reconstruct 3-D objects from a series of 2-D projection images, techniques that are very similar to those used in computed tomography (CT) scanning in medicine. Such methods work particularly well for viruses that have highly symmetric capsids. A major advance has been the use of cryogenic freezing of unstained samples (cryo-EM) and techniques for data collection that allow 3-D reconstructions from samples undamaged by excessive exposure to the electron beam (see below). Such techniques yield structures for viruses at resolutions that can rival those of X-ray crystallography, and of course do not require crystallisation of the samples before structure determination. Fortunately it turns out that in most cases to date structures determined by X-ray and cryo-EM methods are very similar, and X-ray structures can often be fitted into cryo-EM density. This implies that crystal structures reflect viral structures in solution but it has also opened up a wonderful synergy between the two techniques, as there are many conditions in which virologists would like to examine viruses

that are unlikely ever to be accessible by single crystal diffraction studies. Recent developments allow 3-D reconstructions to be determined for single virus particles, further extending the power of cryo-EM to interrogate viral life-cycles. Such techniques allow the determination of tomograms of viruses that lack isometric structures, such as the major pathogens of influenza, HIV and herpes. Viral asymmetry is therefore no longer a barrier to structural studies

2. The Cryo-EM Technique

As a result of the problems described above with interpreting stained images, structural biologists were keen to move to imaging of unstained specimens. The development of cryo-EM built upon the long established concept that low temperature preserves biological specimens. However at the molecular level, the formation of ice crystals during freezing can disrupt biological structures and withdraws water molecules from their hydration shell causing macromolecules to become partially dehydrated. Furthermore, crystalline ice diffracts the electron beam, preventing useable image formation. Dubochet and colleagues at the EMBL in Heidelberg discovered that if a biological specimen was frozen sufficiently rapidly the formation of ice crystals was prevented, preserving the solution conformation and allowing it to be imaged in a thin layer of vitreous ice. The key to this rapid freezing was the use of liquid ethane or propane, cooled to near liquid nitrogen temperatures, to freeze the samples. These coolants have extremely high thermal conductivity, which allow extremely rapid cooling rates to be achieved. Together with the low mass of an aqueous thin film, cooling rates approaching a million degrees per second can be achieved during vitrification. In the vitreous ice layer that results from such rapid cooling, biological macromolecules remain hydrated, and in a structural state essentially identical to that seen in the liquid phase. Typically, the thin film of ice was formed and frozen in a hole in a carbon support film, ameliorating the effects of a support film that can flatten particles and cause preferred orientations to be observed (see Fig. 2). The first practical results of this new technique were images of viruses: unstained, frozen-hydrated Semliki Forest Virus and bacteriophage T4 (see Fig. 3e). Even in these first cryo-EM images, the potential of the technique for studying the structure of viruses was immediately apparent. The images of SFV immediately settled an ongoing debate about whether SFV had a $T = 3$ or a

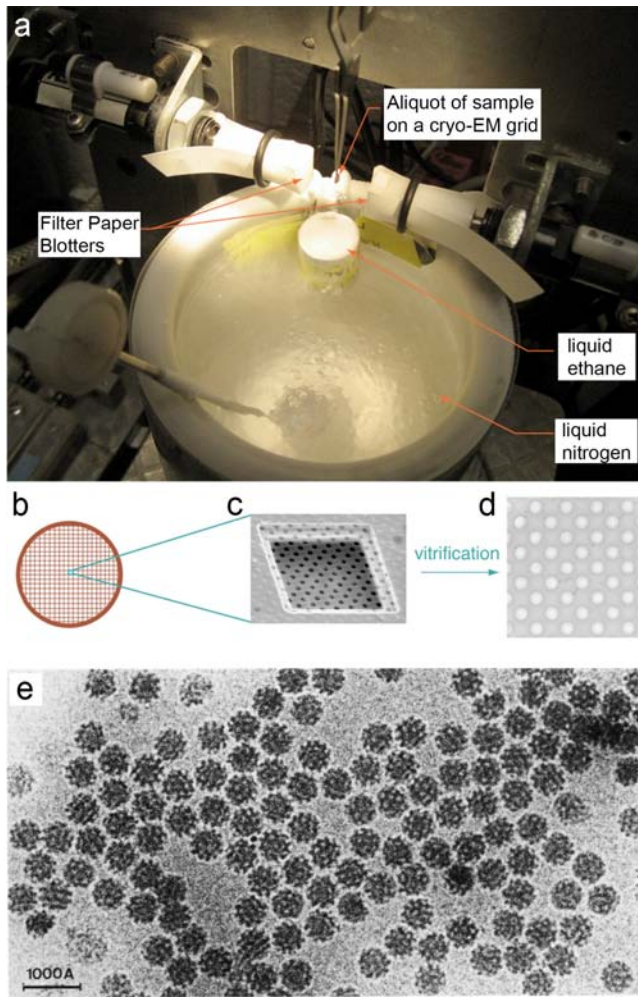


Fig. 3. Cryo-EM sample preparation. (a) A cryo-EM freezing apparatus. An EM grid is held in forceps above a reservoir of liquid nitrogen-cooled liquid ethane. Sample in water of buffer is added, excess liquid is blotted away (here by pneumatically-driven blotters) and then the grid is plunged into the liquid ethane reservoir. (b) A typical EM grid The grids are ~3 mm in diameter and consist of a mesh made from a variety of metals, and in various spacings. (c) Grids covered with a support film lithographically etched to contain a regular array of regularly sized holes are commercially available. The support film shown is a Quantifoil R2/2 grid (Quantifoil Microtools, Gmbh), in which the holes are 2 μm in diameter). (d) A close-up view (~1200x) of such a support film with a thin layer of vitreous ice in the EM is also shown. (e) The first published cryo-EM images of a virus; Semliki Forest virus. From Adrian *et al.*, (1984).

$T = 4$ capsid morphology, unambiguously revealing a $T = 4$ morphology (Adrian *et al.*, 1984).

3. Determining the 3-D Structure of Viruses from EM Data

Like all methods that are in principle suitable for determining the high-resolution structure of biological macromolecules, electron microscopy has both advantages and disadvantages. Perhaps the most notable advantage of an electron microscope might sound trivial but is not; an electron microscope records an image. The process of image formation means that both amplitude and phase information are recorded simultaneously, in contrast to diffraction-based methods where only amplitudes are captured, and phase information has to be derived. However, aberrations in the magnetic lenses used to focus electrons (which allows formation of the image) mean that very small apertures are required in electron microscopes. Small apertures contribute to the large depth of field in EM. Together with the penetrating nature of electrons, this results in the 3-D electron density of the object being projected onto the 2-D plane of the image (c.f. a medical X-ray). The fact that EM generates projection images of biological structures means that it is extremely difficult to interpret individual EM images. However, it also means that true 3-D structural information is accessible. 3-D reconstruction of an object from projections is possible owing to the fact that the Fourier transform of a projection image is a central section of the 3-D Fourier transform of the original object (Fig. 4).

Thus, if all possible views of the object are available, then a full 3-D reconstruction of the object is possible, in principle to the resolution limit of the instrument. For the case of electrons in the 2–300 keV range as used in modern microscopes, this is better than 1 Å resolution. The key problem faced in calculating such a 3-D structure is determining in which orientation the individual projections of the original object were imaged. The orientation of the projected structure defines which central section through the 3-D transform of the object the projection represents. From this it can readily be seen that a full range of all possible orientations leads to the most completely sampled 3-D transform, and hence contributes towards higher resolution. The essence of determining a virus structure

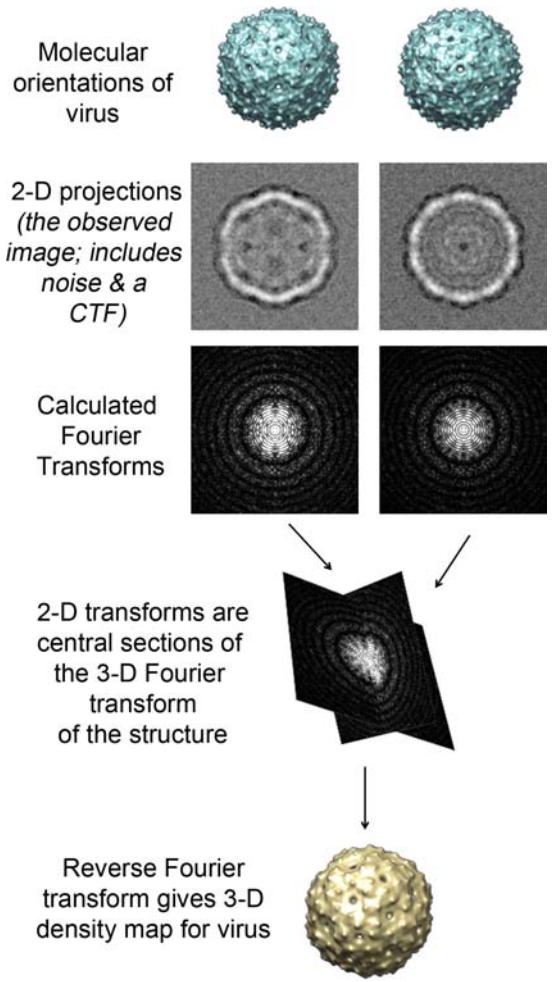


Fig. 4. The principle of how to calculate a 3-D structure from 2-D projections. The virus is randomly oriented in vitreous ice layer of a cryo-EM grid. The recorded images, containing noise and distorted by the effects of the microscope CTF are projections of different views of the virus onto the plane in which the image is recorded. The 2-D Fourier transform of those images are central sections through the 3D Fourier transform of the virus. Knowing the orientation of each particle tells you which central section of the 3-D transform that virus particle represents — i.e. how the slices fit together. The degree to which the 3-D transform is populated is one of the limits on resolution in cryo-EM studies; hence the need for randomly distributed orientations and the advantage of high symmetry. The 3-D object is reconstructed by reverse Fourier transformation to give a 3-D electron density map.

by cryo-EM is therefore simply to determine the orientation of each picture of the virus with as high a degree of accuracy as possible. Details of exactly how the orientation of each imaged virus particle is found are beyond the scope of this article. However, in essence a number of different approaches are possible which share the same basic idea that experimental images of the virus are compared to reprojections of a model structure for the virus. For review, see (Baker *et al.*, 1999; van Heel *et al.*, 2000).

Implicit in the idea that all possible views are required is the fact that structures of highly symmetric particles are more readily calculated, as the symmetry of the particle helps to ensure that all possible views are sampled. Averaging of symmetry-related views also increases the effective size of datasets, meaning that fewer raw images are required for a given resolution than for asymmetric particles. Symmetry averaging does however 'average-away' any asymmetric features of a virus, such as single-copy infectivity or maturation proteins which are commonly essential for viral lifecycles.

The negative-staining method that dominated EM imaging of viruses until the mid-1980s presented major challenges to the application of this approach to 3-D structure determination. Very often biological particles adhere to the carbon support film on the EM grid, deforming the native structure and resulting in preferred particle orientations that make 3-D structure determination difficult because these do not contain enough information. For highly symmetric particles such as isometric viruses, this is not an insurmountable problem, as symmetry ensures an even coverage/sampling of the 3-D transform of the reconstructed density. However, the size of such particles means that obtaining images of viruses fully embedded in stain is extremely difficult, and for larger viruses essentially impossible. Since it is the interaction of electrons with the stain layer that generates contrast, this results in the information from un-embedded portions of the virus not being recorded in the image, i.e. the images represent projections of only part of the structure (see Fig. 2b). Furthermore, owing to the problems associated with staining, only information on the surface envelope of the virus is obtained.

Preservation and imaging of the entire native-like structure is the clear advantage of cryo-EM over staining methods. However, the fact that the specimen is unstained results in a different set of problems for structure determination. Firstly, amplitude contrast in cryo-EM images arises from the scattering power of the atoms found in the specimen, which is related

to their mass. Hence the use of heavy metal salts in negative staining. The masses of atoms typically found in proteins and/or nucleic acids are very similar to those found in the surrounding water and buffer molecules of the vitreous ice layer. This means that although cryo-EM images contain information from the entire macromolecular structure, they typically have extremely low contrast. This problem is made markedly worse by the radiation sensitivity of unstained biological macromolecules, requiring that the electron dose delivered to the specimen be kept low to minimise radiation damage. Together these factors result in the characteristically poor signal to noise ratios (SNR) of cryo-EM images, and necessitate computational averaging of data to improve the SNR to a point where structure determination is possible. In essence, in cryo-EM noisy images of individual molecules have to be explicitly averaged after the data is recorded, whilst in X-ray diffraction averaging is an intrinsic property of the crystal. This averaging creates significant computational challenges that need to be overcome in order to determine cryo-EM structures, especially to high resolution.

To a limited extent, the problem of low amplitude contrast in cryo-EM images can be overcome by tuning the optical properties of the microscope to introduce a second type of contrast into the recorded images: phase contrast. In practice, this is routinely achieved by defocusing the microscope, which introduces low-resolution phase contrast that is often essential to allow relatively small objects, such as virus capsids, to be found in noisy micrographs. It also means that distortion of the observed image by the microscope's contrast transfer function (CTF) becomes significant and problematic (Fig. 5). A CTF is a phenomenon, common to all optical systems, that defines how information is transferred as a function of spatial frequency (i.e. resolution) in Fourier space. The form of this function is of an oscillating sinusoidal variation in information transfer with increasing frequency and decreasing amplitude (Fig. 5f). In real space (i.e. in the observed image) the effects of a CTF are to convolute the image of the object being studied with a point spread function (PSF). A PSF spreads information from each area of the image into surrounding areas and attenuates information at high resolution. The practical consequences of using highly defocused images are therefore that although they have more low-resolution contrast, allowing the particles to be found, the particles themselves have less high resolution information and are profoundly distorted, necessitating computational CTF correction during

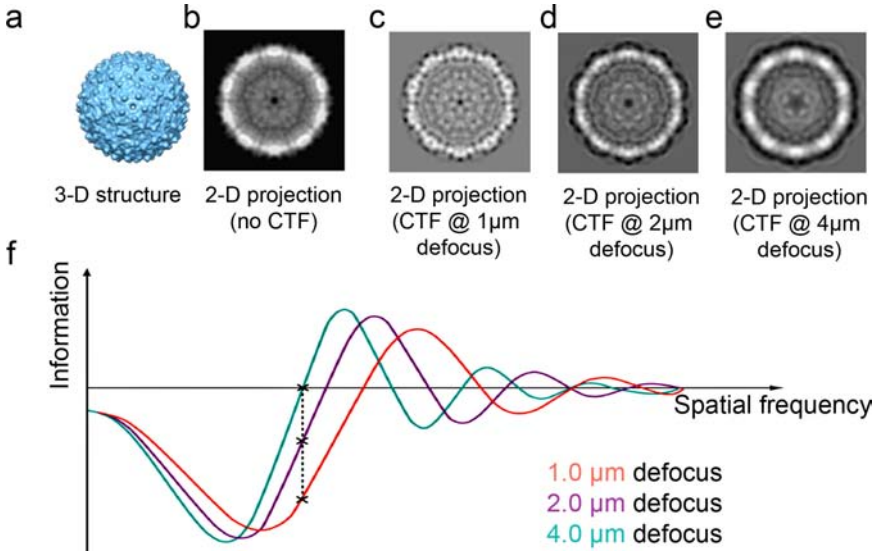


Fig. 5. Contrast transfer functions in cryo-EM. (a) The effects of a CTF on the observed image. The 3-D structure (a) is projected (in a computer) to give a 2-D projection image with no added CTF (b). The effects of a CTF function are then simulated with defocus of 1.0 μm (c), 2.0 μm (d), and 4.0 μm (e). The effects of the Fourier space CTF function on the real image are of a point spread function, which spreads information from the particle into surrounding areas. (f) Plots of typical CTF functions applied in (c–e). A CTF has an oscillating form with increasing frequency and decreasing amplitude — i.e. there are contrast reversals, spatial frequencies where no information is transferred ('zeroes') and attenuation of high-resolution information. Only by combining data at different degrees of defocus can full restoration of information at all resolutions by achieved.

the structure refinement calculations. Such corrections essentially flip the image contrast in regions where the CTF has made it negative, removing the effects of the PSF.

4. Complementarity between Cryo-EM versus X-ray Methods

Advances in all aspects of X-ray crystallography, especially in robotic crystallization condition screening, and in the advent of synchrotron radiation sources, have led to ever larger and more complicated macromolecular

complexes being solved by X-ray methods. Such advances are exemplified by recent X-ray structures of large enveloped viruses such as PRD1 and PM2 (Cockburn *et al.*, 2004; Abrescia *et al.*, 2008). Despite such remarkable successes, generally speaking the larger the virus the more difficult it is to grow crystals that diffract to high resolution. Cryo-EM, which, as described above, has no requirement for crystallization, is applicable (and has been applied) to all classes of virus from the smallest (plant satellites) to the largest (mimivirus). Its use has been especially powerful in the study of larger viruses (especially those containing lipid bilayer envelopes), virus-receptor complexes, and transient states in virus maturation pathways. Some notable studies in these areas are described in the following sections.

A recurring theme in these studies is the complementarity between X-ray and cryo-EM techniques. Cryo-EM maps at even quite modest resolutions have been used to phase X-ray data. Fitting of X-ray coordinates into low or intermediate resolution density dramatically enhances the interpretable information content of the cryo-EM structure. X-ray and cryo-EM methods have also been used extensively to map different states in the same maturation pathway.

5. Structure of Large Enveloped Viruses

Several very large viruses have been isolated and their structures determined by cryo-EM. Chilo iridescent virus (CIV) and Paramecium bursaria chlorella virus type 1 (PBCV-1) infect very different hosts (insect and unicellular green algae respectively) but have strikingly similar structural features (Yan *et al.*, 2000). CIV, with a diameter of ~1850 Å and a $T = 147$ morphology, is marginally smaller than PBCV-1 (diameter 1900 Å, with a $T = 169$d morphology). An even bigger virus Phaeocystis pouchetii virus (PpV01) has a capsid with $T = 219$d quasi-symmetry and diameter of ~2200 Å (Yan *et al.*, 2005). However even these giant viruses are dwarfed by another virus discovered at around this time. Mimivirus was first isolated from amoeba in a water-cooling tank, and initially thought to be a bacterium (it is ~3x the size of the smallest known bacterium). The capsid is currently thought to have a $T = \sim 1179$ morphology (Fig. 6; Xiao *et al.*, 2005; Xiao *et al.*, 2009), and packages a dsDNA genome of ~1.2Mb encoding 911 proteins (Raoult & Forterre, 2008). The smallest known

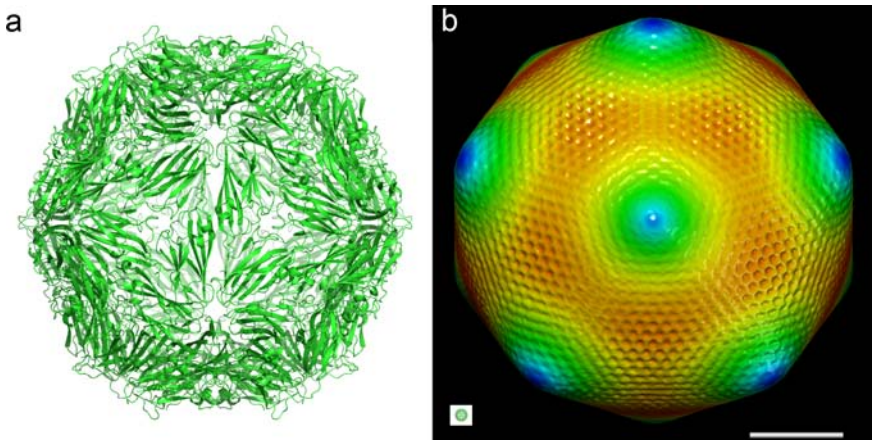


Fig. 6. Extremes of virus size. (a) X-ray structure of Satellite Panicum Mosaic virus (SPMV; Ban & McPherson, 1995), amongst the smallest of known virus particles. SPMV has a $T = 1$ morphology and a diameter of ~160 Å. (b) The largest known virus: Mimivirus. Mimivirus has a $T \sim= 1179$ morphology and a diameter of ~5000 Å. From Xiao *et al.*, 2005; Xiao *et al.*, 2009. Inset in the small box is SPMV shown at the same scale as mimivirus.

free-living organism is a mycoplasma and its genome encodes just 470 proteins. Mimivirus therefore blurs the boundaries between viruses and unicellular organisms. Indeed, mimivirus has since been reported to be itself parasitised by a smaller virus, the first 'virophage' (La Scola *et al.*, 2008).

6. Virus — Receptor Interactions

The interaction of a virus with its cellular receptor is of critical importance in infection, and cryo-EM has played a significant role in elucidating the molecular details of such interactions. One area in which this has been especially apparent is in the binding of rhinovirus, a picornavirus, to intracellular adhesion molecule-1 (ICAM-1). These studies helped to address a fundamental question regarding virus–receptor interactions: viruses are known to mutate rapidly, helping them to evade the host immune response, yet make specific interactions with their receptors. It is, however, difficult to rationalise how such rapid mutation, and the accompanying structural changes that it promotes, can be compatible with the maintenance of a stable, specific interaction with a host molecule. Kolatkar *et al.* (1999), showed how soluble fragments of ICAM-1 bound

in a canyon on the surface of the virus. This surface feature was thought to be too small to allow antibody access, providing spatial separation of the receptor-binding site from the immune response. Surface residues could therefore mutate more rapidly, allowing immune evasion, whilst the receptor-binding site was maintained in its canyon. Although, antibody binding has now been detected to epitopes within the canyon, the basic proposal remains a paradigm for picornavirus — receptor interactions.

Structural studies on poliovirus, another picornavirus, have visualised the virus bound to its cellular receptor incorporated into liposomes (Bubeck *et al.*, 2005). Such a structure represents a virus that is poised to deliver its genomic information across the liposome bilayer, and provides invaluable information about the process of infection via membrane-embedded receptors. The authors first determined the viral site at which receptor binding occurred using a novel post-imaging fiducial marker technique. Briefly they determined the orientation of each image in their dataset, and then added a white 'spot' to each image marking the attachment site. When such 'spotted' images were used in an icosahedral reconstruction a clear density above each five-fold vertex was observed, unambiguously showing that the virus binds its intact membrane-embedded receptor at a five-fold vertex. In turn this knowledge allowed a reconstruction of the virus-receptor-liposome complex to be calculated using C5 (five-fold rotational averaging around the attachment site) symmetry (Fig. 7). The resulting five-fold averaged structure of the complex shows how receptor binding brings one of the twelve five-fold vertices of the capsid into close proximity with the membrane of the liposome, where the lipid bilayer is perturbed. Presumably this perturbation is the first step in allowing the viral RNA to access the interior of the host cell. The same authors have now applied similar methodologies to examine the membrane-attachment complex of Semliki forest virus (SFV), a model for enveloped virus fusion. SFV also approaches the membrane along a five-fold axis that at least suggests this may be a generic feature of receptor recognition and fusion events.

7. Maturation Processes

Another area of virology in which cryo-EM has made a substantial contribution to our expanding knowledge is in viral maturation. Although significant efforts have been directed at HSV-1 nucleocapsid maturation, the most complete description of a viral maturation pathway to date is for

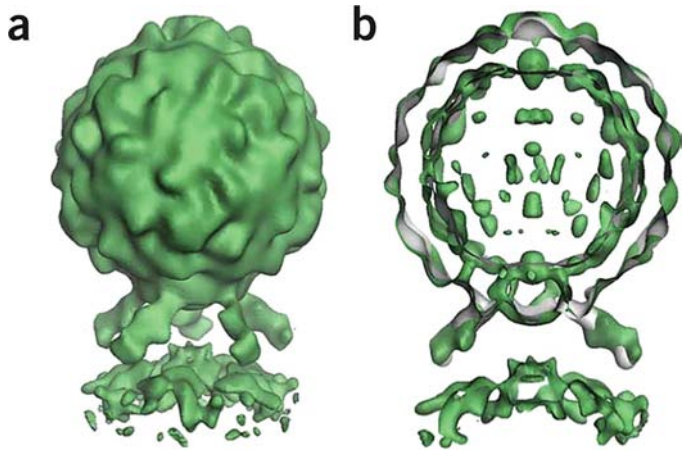


Fig. 7. The five-fold averaged (C5) structure of Poliovirus bound to its receptor incorporated into liposomes. Surface (a) and central slab (b) through the structure of poliovirus bound to its cognate receptor. The grey surface in (b) is the structure of the virus-receptor complex in solution. Extra density is visible in the outer leaflet of the bilayer and may represent perturbation of the membrane upon virus binding. From Bubeck *et al.*, 2005.

Hong Kong 97 (HK97), a tailed bacteriophage with a dsDNA genome. Both HSV-1 and HK97 maturation are reviewed in detail in Steven *et al.*, 2005. The detailed description of HK97 maturation is also interesting because it acts is a paradigm for combining X-ray crystallography and cryo-EM methods to elucidate new biology.

HK97 is extremely well suited to studying maturation processes. The major coat protein, gp5, has been exhaustively studied, and assembles to form icosahedral particles in the absence of its portal protein and tail assembly. Such 'Prohead I' particles (Fig. 8), which have a $T = 7l$ quasi-symmetry, are much better suited to structural studies than tailed particles. Maturation is triggered by an autocatalytic cleavage of gp5 to gp5*, which produces Prohead II. Both Prohead I & II have a relatively small diameter and a thick capsid wall. Intermediate states between the Prohead II and final Head II state have now been identified and structurally characterised using cryo-EM. In vitro, upon acidification to pH 4.15, Prohead II rapidly alters its conformation to a semi-expanded Expansion Intermediate (EI)-I state, then to a structurally similar EI-II state, and finally to an EI-III state that is characterised by its large diameter and thin walls. Upon neutralization, EI-III rapidly assumes the structure of the polyhedral mature head

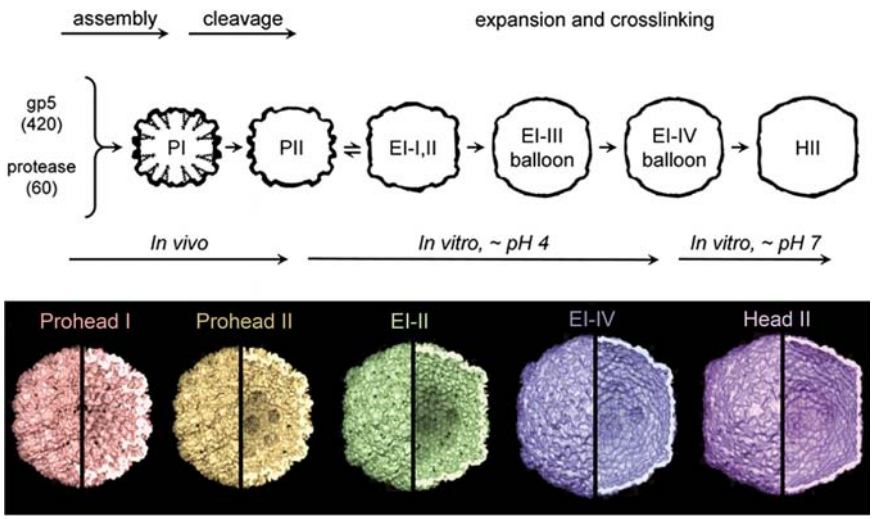


Fig. 8. The maturation of bacteriophage HK97. HK97 assembles to form icosahedral particles (Prohead I) in the absence of its tail. Autocatalytic cleavage of the coat protein forms Prohead II, which then undergoes a sequence of maturation steps (driven by changes in pH) that increase the diameter of the capsid whilst decreasing the thickness of the capsid wall. The left half of each capsid structure shows a surface rendering, whilst the right half shows a cut-away to reveal the thickness of the capsid wall. The schematic portion of the figure was adapted from Lee *et al.*, 2008; the structural portion was adapted from Steven *et al.*, 2005.

(Head-II). The Head-II conformation is characterised by a remarkable cross-linking between coat protein subunits, which arises from the auto-catalytic creation of iso-peptide bonds between each subunit and two of its neighbours (Wikoff *et al.*, 2000). The resulting 'catanenes' of protein subunits are interlinked, giving the capsid a chainmail-like structure. The maturation process also encompasses both major rigid-body rotations and translation of some portions of the coat protein subunit structure, and refolding of the rest.

Another viral system in which cryo-EM has helped to advance our understanding of maturation processes is the Dengue virus, a human pathogen of increasing importance. In common with many enveloped viruses, the acquisition of Dengue virus infectivity is closely associated with changes in pH. Yu *et al.* (2008), determined the structure of Dengue virus in a new conformation promoted by low pH (pH 6.0; Fig. 9). This pH mirrors that found in the Golgi network through which Dengue virus

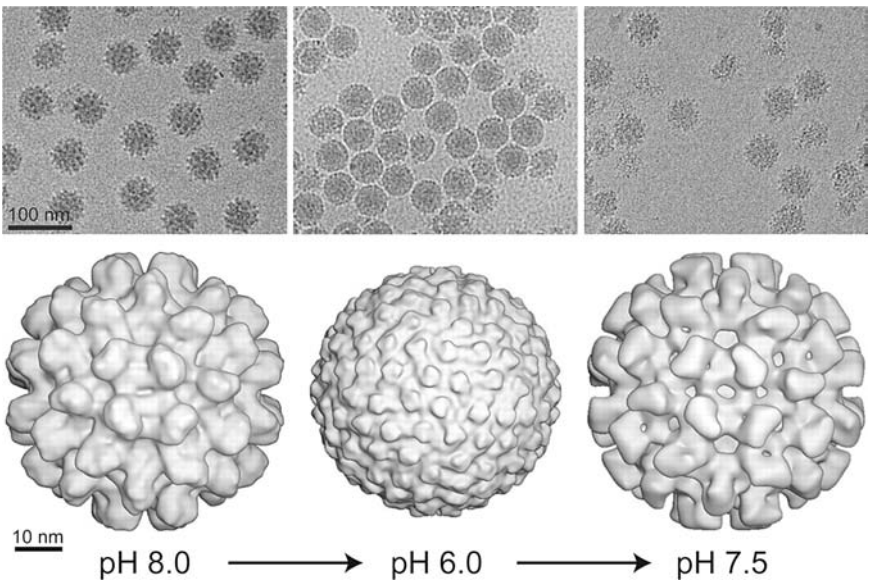


Fig. 9. Cryo-EM characterization of a reversible, pH-induced conformational change in the structure of Dengue virus. The conformational change renders the virus sensitive to proteolytic processing that is essential for the acquisition of virus infectivity. From Yu *et al.*, 2008.

is trafficked, and was shown to render the virus susceptible to proteolytic cleavage by furin, a cellular protease normally localised to the Golgi network. This cleavage liberates a peptide that remains bound to the virus particle at low pH, i.e. whilst the virus is still in the Golgi, but dissociates when the pH is raised, generating the membrane fusion potential of the virus.

8. Structural Information on Packaged Genomes

Both X-ray diffraction and cryo-EM have been spectacularly successful in elucidating details of the protein capsids of a wide range of viruses. However, information on how the genomic material is packaged within those protein capsids remains much scarcer. A wide range of different factors contributes to this lack of information.

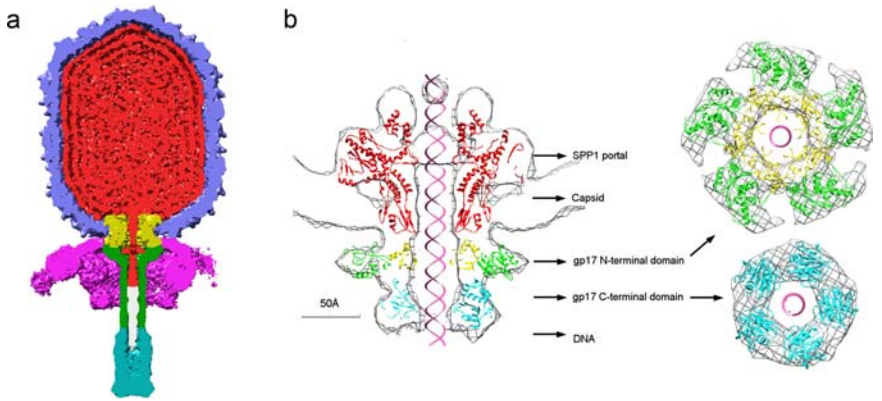


Fig. 10. dsDNA packaging by bacteriophage. (a) The cryo-EM structure of bacteriophage $\phi 29$. The view is of the rear half of the capsid, and packaged dsDNA is visible (in red) as tightly spooled density. From Tang *et al.*, 2008. (b) Details of the dsDNA packaging apparatus of bacteriophage T4. The gp17 packaging motor (cyan, yellow & green) is in the open conformation in the capsid-incorporated pentamer structure (Sun *et al.*, 2008).

For those viruses with a dsDNA genome inserted into preformed capsids by a packaging motor, the enormous pressure generated leads to the structure of the packaged genome being dominated by closely spaced concentric shells of nucleic acid that fully occupy the space encapsidated by the capsid (Fig. 10a). No defined 3D tertiary structure for such a packaged DNA is available, and indeed, it seems plausible that no such defined tertiary structure exists. Much progress has been made on understanding how this enormous packing density is achieved. Sun *et al.* (2008) determined the crystal structure of the packaging motor from bacteriophage T4, and fitted it into a lower resolution cryo-EM reconstruction of the intact virion (Fig. 10b). The packaging motor, gp17, is shown to consist of two domains; an N-terminal domain that provides the energy for dsDNA packaging through the binding and hydrolysis of ATP, and a C-terminal domain that translocates the genome into the capsid. In the X-ray structure of the gp17 monomer, the two domains are in a closed conformation, whilst in the cryo-EM structure of the gp17 pentamer, they are in an open conformation. Based on this and much biochemical work, the authors propose an elegant model for genome packaging driven by ATP-induced opening and closing of the gp17 structure.

The situation for viruses with a single-stranded RNA genome is markedly different. Single-stranded RNA genomes are not inserted into preformed capsids by packaging motors, but rather fold and are packaged as the viral capsid assembles. Single-stranded (ss) RNA has a much greater degree of conformational flexibility that presumably facilitates this coupling of assembly and packaging, and intriguing evidence is beginning to appear that suggests the genome may play a far more active role in the assembly process than has previously been appreciated (Stockley *et al.*, 2007). Indeed it seems plausible that one of the many selective pressures that such viruses experience is the ability of the genomic RNAs to participate in the assembly reaction. In keeping with the lack of a packaging motor to force nucleic acid into a capsid, the density of the packaged genome appears to be far lower. The idea that ssRNA genomes fold as they are packaged is tantalizing, because the defined sequence of the RNA molecule, together with the repeating nature of the protein binding sites for RNA presented by the capsid, raises the possibility that each virus might package its genome in the same or similar way. If this were the case then such a virus should be amenable to structural studies.

However, significant difficulties remain in imaging packaged RNA genomes. The protein capsids of spherical viruses contain a regular array of highly ordered, symmetrically arranged repeats of a single (or small number) of polypeptides. A single-stranded genome however, is an inherently asymmetric structure defined by the linear sequence of its nucleotides. This mismatch in symmetry between capsid and genome almost guarantees that the latter will be less ordered than the former. Furthermore, as all the structural methods used to determine virus structure require this order, it is not surprising that significant amounts of genomic material are rarely resolved in virus structures. Yet more difficulties arise from the mechanics of data collection during X-ray diffraction experiments on viruses. In much early crystallographic work no data at resolutions lower than 20–30 Å were included. This was both to simplify what was at the time a daunting computational task, and because the low-resolution terms are routinely excluded by the beam stop used to allow collection of high-resolution data. However, it is precisely such terms that are most important for describing the structure of imperfectly ordered components of the crystal, such as the packaged genome. The lack of information at appropriate spatial frequencies prevents visualization of density for the imperfectly ordered genome at lower resolution.

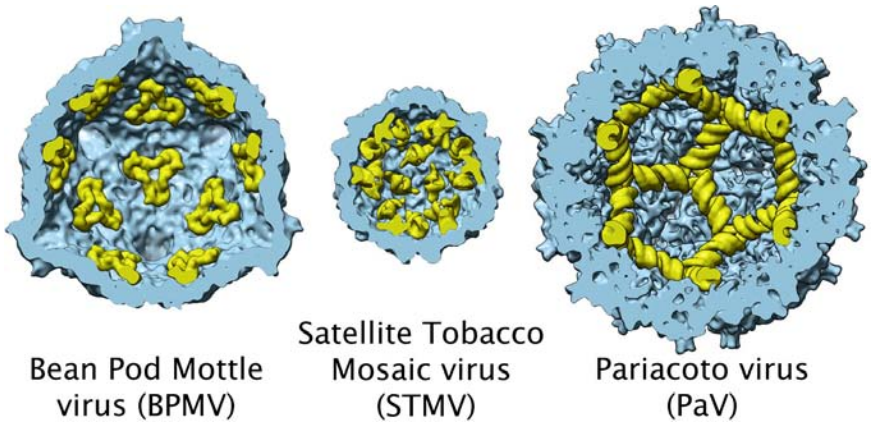


Fig. 11. Structural information on ssRNA genomes within virus particles. The rear half of each virus is shown in a low-resolution surface representation (in blue). Ordered RNA is shown in yellow. Each structure is from X-ray crystallographic data. Adapted from Schneemann *et al.*, 2006.

Despite this array of technical difficulties, there are now a (relatively small) number of structural studies in which significant amounts of density for genomic RNA have been observed (for review see Schneemann *et al.*, 2006). The first virus in which ordered ssRNA was observed was bean pod mottle virus (BPMV; Chen *et al.*, 1989; Fig. 11), a $T = 3$ capsid in which 6 ordered and 5 poorly ordered nucleotides were seen near the icosahedral three-fold axes, accounting for ~20% of the genomic RNA. However, the most complete description of a packaged RNA has been made for the smallest viruses, where the small size of the genomic RNA and its relatively close proximity to the ordering effects of RNA binding sites on the capsid protein shell presumably enhance the degree of order. Satellite tobacco mosaic virus is amongst the smallest of all known viruses. The 1058 nucleotide ssRNA genome encodes for a single gene product, the viral coat protein, which assembles as a $T = 1$ icosahedral particle with a diameter of ~170 Å. These particles are not in themselves infectious, and piggyback their own infectivity to infection by a larger, 'helper' virus (in this case tobacco mosaic virus). The 1.8 Å resolution X-ray structure of STMV reveals ~59% of the genomic RNA occurring as short duplexes beneath the icosahedral two-fold axes of the virion (Larson *et al.*, 1998). Subsequent work has modeled stem loop structures connecting these duplexes,

and now accounts for >80% of the genome (Larson & McPherson, 2001; Fig. 11).

Perhaps the most striking example of a packaged RNA has come from structures of insect viruses in the family Nodaviridae, exemplified by Flock House virus (FHV; Fisher & Johnson 1993) and Pariacoto virus (PaV; Tang *et al.*, 2001). The crystal structure of PaV, which has a $T = 3$ capsid morphology and bipartite, ssRNA genome, is of particular interest. It shows ~25 base pairs of RNA packaged as A-type duplexes beneath the icosahedral two-fold axes. These duplexes extend towards the icosahedral three-fold axes, where icosahedral averaging gives the impression of three-way RNA junctions. The modelled RNA therefore appears as a dodecahedral cage of density, providing an averaged view of ~35% of the PaV genome. The dodecahedral cage is visually seductive, but it is important to recognise that it cannot accurately represent the true situation as not all 20 three-fold contacts can accommodate a three-way junction if the RNA is linear and bipartite. Rather the density weakens near to the three-fold contact, suggesting that at some of the vertices, the RNA dips into the centre of the particle where the remaining ~65% of the genome is presumably located.

In contrast to X-ray crystallography, cryo-EM images simultaneously record information across a very wide range of spatial frequencies. In fact, the signal at low-resolution vastly outweighs that at high resolution, and this is the second major reason for averaging of cryo-EM images: to boost weak information at high spatial frequencies to significant levels. Cryo-EM images are therefore inherently well suited to imaging relatively poorly ordered density at less than atomic resolution, and recent cryo-EM studies have given considerable insight into the packing of genomic ssRNA within capsids.

A cryo-EM study on Pariacoto Virus revealed the same dodecahedral RNA cage as seen in the crystallographic study described above, but also revealed further density at lower radius. This clearly suggests that, the packaged genomic RNA is most ordered close to the lattice of binding sites on the inner surface of the protein capsid. At lower radii, where the steric restraint imposed by binding to the capsid is further removed, order is lower.

A similar theme has now also been observed for a ssRNA bacteriophage, MS2. The crystal structure of the wild-type MS2 virion was amongst the first virus structures to be determined, and contained

almost no information for encapsidated RNA. Subsequent crystallographic studies on recombinant MS2 capsids containing multiple copies of RNA oligonucleotides with defined sequences, each bound to a coat protein dimer, and hence repeated within the asymmetric unit of the virus, revealed the molecular detail of the RNA binding site at atomic resolution (Valegård *et al.*, 1994; Convery *et al.*, 1998). However, it was not until cryo-EM structures became available that the larger genomic RNA began to be visualised (Koning *et al.*, 2003; van den Worm *et al.*, 2006). Initially cryo-EM maps at relatively moderate resolution (17–20 Å) showed a network of RNA bound to the inner surface of the protein capsid. A subsequent study at intermediate resolution (~9 Å) shows the same network of protein-bound RNA, but also a second 'shell' of RNA at lower radius (Fig. 12). The resulting two shells of density, which are connected at the icosahedral five-fold axes, are estimated to encompass the majority of the genomic RNA sequence. The fact that the majority of the genomic RNA is visible in an icosahedrally averaged structure clearly implies that, at least to the resolution of the structure, the RNA is ordered with pseudo-icosahedral symmetry.

The complex fold that the MS2 RNA must adopt to form the observed density has profound implications for the assembly of its virion, and by implication on the assembly of other simple ssRNA viruses. The MS2 genome contains a single copy of a specific packaging signal in the form of a stem loop that enhances the efficiency of assembly and RNA packaging. The majority of the protein — RNA interactions in a wild-type phage particle are therefore 'non-sequence specific', presumably through recognition of other stem loop structures within the genome. The genomic RNA must transit from outer to inner shell and back at the five-fold axes of the virion, where the density has been shown to accommodate a modelled RNA duplex. Formally, there are two possibilities to account for this: after transiting from the outer (protein-bound) to the inner shell, the RNA can return to the outer shell at either the same five-fold vertex, or a different one. The latter seems unlikely, in that for an assembling capsid, other five-fold vertices are not necessarily formed and the RNA can have no information on their locations. The former in contrast could easily be driven by short-range base pairing interactions.

It seems likely that cryo-EM, together with low-resolution crystallography, will reveal increasing details of the structures of other encapsidated ssRNAs in the future.

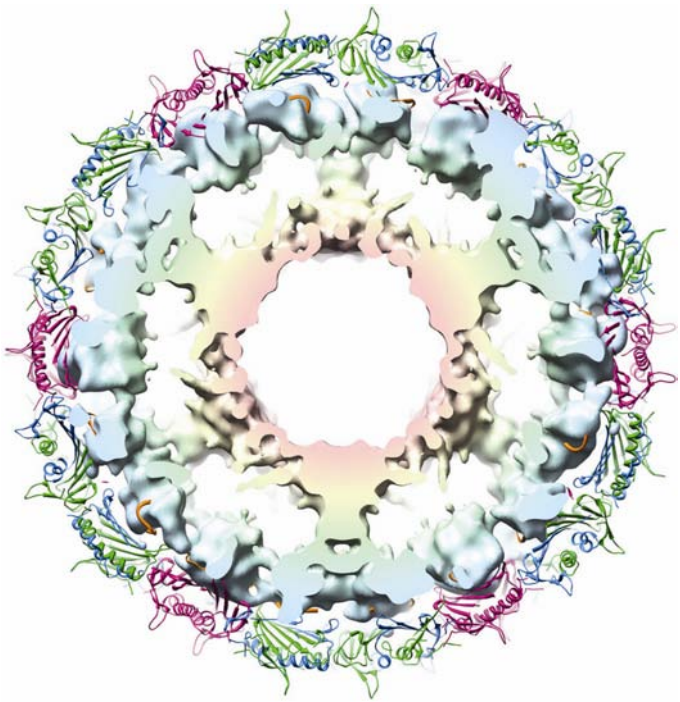


Fig. 12. Packaged, genomic ssRNA in bacteriophage MS2. The crystal structure of the MS2 coat protein shell is show in cartoon representation and coloured (blue/green/red) according to the quasi-equivalent packing in the $T = 3$ capsid. Icosahedrally-averaged cryo-EM density for packaged RNA is shown as a radially-coloured surface. The RNA is packaged in a unique, double-shelled fold, with the two shells being connected by density along each five-fold axis. From Toropova *et al.*, 2008.

9. Prospects for Cryo-EM of Viruses at Atomic Resolution

It is now nearly 30 years since the first structure of a virus at atomic resolution was determined (Harrison *et al.*, 1978). For much of that time, virus structures have also been determined by cryo-EM, but despite these efforts to date no true atomic resolution structure of a virus (higher than $\sim$3 Å) has been published. However, recent advances in instrumentation, data processing algorithms and ready access to computing meant that 2008 saw the publication of three outstanding structural studies on virus particles. For the first time these have raised the realistic prospect of cryo-EM

methods being able routinely to resolve the structures of virus particles at atomic resolution. All three studies are of large (>~700 Å in diameter) icosahedral particles, but each utilized novel technical approaches and revealed new biological insights.

Rotaviruses cause life-threatening diarrhea in humans, especially infants in the third world. They have a dsRNA genome surrounded by three distinct layers of protein. During infection the outer layer is lost, yielding a characteristic double-layered particle (DLP), and it was the structure of this 'rotavirus DLP' that was described in an elegant study from the Grigorieff and Harrison laboratories in Boston (Zhang *et al.*, 2008; Fig. 13). This study was noteworthy for two reasons. Firstly,

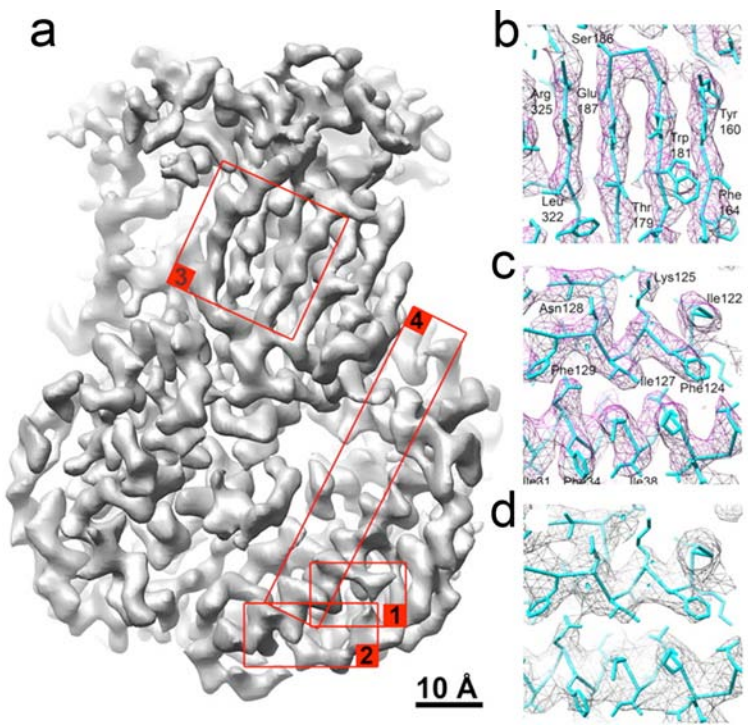


Fig. 13. High-resolution cryo-EM structure of a rotavirus DLP. (a) A trimer of rotavirus VP6 proteins Fourier filtered at ~7 Å resolution. Close up views of the unfiltered structure are shown to highlight the quality of the electron density in β-sheet (b) and α-helical (c) regions of the protein. (d) shows the same region of the protein as in (c), but from a 3.8 Å X-ray map of the same virus. Adapted from Zhang *et al.*, 2008.

they used just 8400 individual virus particles to calculate their structure. However, the authors also made innovative use of the non–icosahedral symmetry in the particle, and averaged across the $T = 13$ asymmetric unit of the DLP. These 8,400 virus particles therefore constitute more than six million individual viral coat proteins, sufficient to produce a 4.1 Å resolution map. Secondly, the authors also reported a previously unpublished structure at 3.8 Å derived from X-ray diffraction data, providing a completely independent validation of their structure and its resolution. Indeed, the comparison between 4.1 Å cryo-EM structure and 3.8 Å structure suggests that the cryo-EM map with a nominally lower resolution may actually be more detailed, indicating that the traditional Fourier Shell Correlation (FSC) method for estimating the resolution of EM structures may be too conservative (Fig. 13). What is certain is that in this resolution range, direct comparisons of the density are more meaningful than somewhat contentious resolution estimates.

The second high-resolution study of virus structure by cryo-EM published in the last year was on $\varepsilon 15$, a tailed-bacteriophage with a double-stranded DNA genome the structure of which was determined to ~4.5 Å resolution (Jiang *et al.*, 2008; Fig. 14). This work is noteworthy because in the absence of an atomic resolution crystal structure, the authors were able to trace completely the alpha carbon backbone of the major capsid

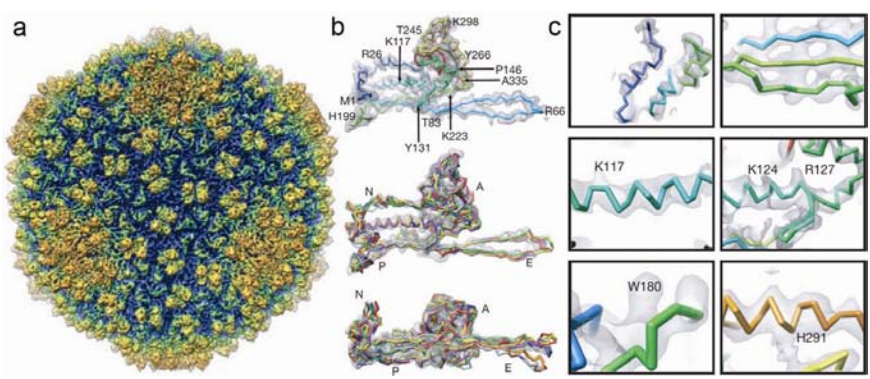


Fig. 14. The 4.5 Å resolution cryo-EM structure of bacteriophage $\varepsilon 15$. (a) A radially-coloured surface representation of the entire virus. (b) Cryo-EM derived models for the coat protein polypeptide backbone. (c) The quality of the map, and the fit of the built model for the polypeptide backbone to experimental electron density. Adapted from Jiang *et al.*, 2008.

protein, gp7 (Fig. 14). Given the extreme difficulty in interpreting density maps in this resolution range, building a model for the entire backbone is a resounding validation of the bioinformatic tools the authors have developed over a number of years. Using these, and homology-based structure prediction, it was also possible to determine the position of a second structural protein gp10. The authors propose that gp10 dimerises, and that this dimer binding plays a critical role in the stabilization of the capsid. This study is also notable for technical reasons, not least the sheer scale of the computational resources assembled for the structure determination. A dataset of approximately 36,000 individual ε15 particles was used. The data were recorded at liquid helium temperatures, and around 3000 film micrographs were used, implying perhaps many more were recorded and rejected. The final structure was calculated from the best ~20,000 molecular views, and required some 1.2M CPU hours of computation.

The final high-resolution virus cryo-EM structure published in 2008 was of cytoplasmic polyhedrosis virus at 3.88 Å resolution (Yu *et al.*, 2008; Fig. 15). This study was notable in that the data were recorded on a CCD camera rather than photographic film. Until now, the consensus within the EM field has broadly been that photographic film, whilst cumbersome and time-consuming to use, was the preferred recording medium for studies in which the highest possible resolution was required. The reasons for this are two-fold. Firstly, the larger field of view of a photographic plate vs. a CCD detector (~8-fold larger for a standard plate scanned with a 7μm step compared to a 16Mpixel CCD camera) means that more particles can be imaged per exposure. Consequently, the large datasets required for high resolution are more easily accumulated. Furthermore, the sampling rate in the digitised image is a fundamental limit on resolution. Together with properties of the detectors that are beyond the scope of this article, the smaller pixel size of scanned film (7 μm *versus* ~15 μm for CCD) leads to a higher theoretical resolution limit in the data.

To counteract this effect, the authors used extremely high magnification (~155,000x) to achieve a sufficiently high sampling rate. As a result, very few virus particles were imaged per micrograph, and assembling and processing the dataset involved represents a prodigiously impressively achievement. With such a detailed density map for CPV, the authors were able to describe a number of novel features of the virus, including both conformational states of the viral capsid shell protein, and details of the viral mRNA capping and release machinery.

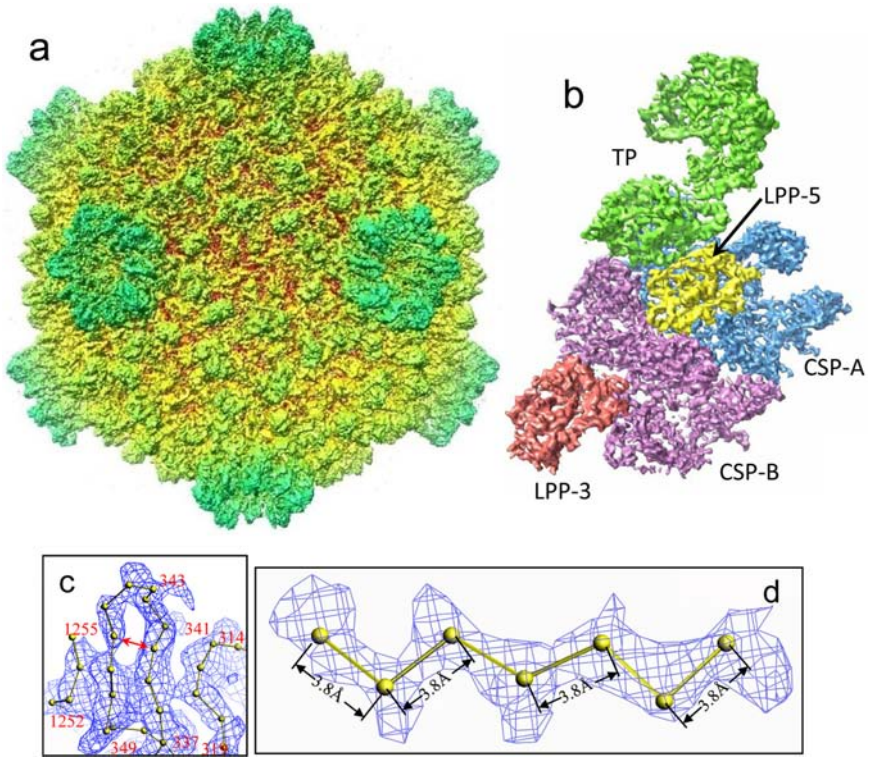


Fig. 15. The cryo-EM structure of cytoplasmic polyhedrosis virus (CPV) at 3.88 Å. (a) A radially-coloured surface representation of CPV. (b) The asymmetric unit of the CPV structure contains one turret protein (in green), two capsid shell proteins (in blue and purple), and two copies of a large protrusion protein (brown & yellow). (c and d) Quality of the EM map and goodness of fit of backbone model. Adapted from Yu *et al.*, 2008.

10. Pleiomorphic Viruses and the Power of Cryo-Electron Tomography

Whilst the highly symmetric viruses described above are representative of many important pathogens, in recent years attention has turned to a second class of viruses that lack such regular structures. Such 'pleiomorphic' viruses include many human pathogens, including many retroviruses such as the human immunodeficiency virus (HIV).

As described above, the basic premise of a single-particle cryo-EM reconstruction is to determine the orientation of each of many copies of a randomly oriented virion. The range of views is essential to recovering 3-D information, and this recovery absolutely relies on the assumption that each image of the virus is an identical copy of the same structure. For pleiomorphic viruses, each copy of the virus is different, and no such assumption can be made. The sometimes remarkable degree of heterogeneity in such viruses prevents averaging of different particles and subsequent 3-D analysis.

However, another electron microscopy technique is emerging as perhaps the only way in which such challenging structural problems can be addressed: cryo-electron tomography (cryo-ET; Medalia *et al.*, 2002). Cryo-ET turns the typical cryo-EM way of data collection on its head: rather than taking images of multiple viruses in different orientations, multiple images of the same viruses are taken in different orientations by physically tilting the specimen in the electron microscope. Cryo-ET is therefore closely analogous to the 3-D imaging of live patients using computed tomography (CT) scanning. In CT scanning however, the detector moves around the patient, which is not possible with an electron microscope as the EM column would sag under the effects of gravity when its axis was not vertical. In cryo-ET a 3-D reconstruction of the object is possible as the multiple images of the same object are related by the known tilt induced during data collection (for review see Subramaniam *et al.*, 2007).

In practice the specimen is tilted and an exposure taken at each tilt angle. Given the discussion above regarding the radiation sensitivity of biological material to high-energy electrons, the accumulated electron dose is obviously a major problem in cryo-ET. To ameliorate the effects of radiation damage, individual images in a tomographic tilt series are taken with extremely low doses limiting the resolution to $\sim$35–50 Å. Indeed such images typically have such poor SNR that they are effectively impossible to align based on the features of the virus particles within them. This problem has been overcome by inclusion of electron-dense fiducial markers, such as colloidal gold particles, to facilitate alignment before 3-D reconstruction.

However, a fundamental limitation remains: to form an image, the electron beam has to pass through the specimen and have minimal inelastic interactions with it. Tomographic specimens are typically quite thick ($>$2000 Å), and become effectively thicker still as the tilt angle is increased. As the specimen becomes thicker, it becomes more difficult to get the beam

through it; secondary (and tertiary) scattering events increase, adding noise to the images. At tilt angles approaching 90 degrees, the specimen is effectively infinitely thick and opaque to the electron beam. In practice therefore, not all tilt angles are achievable, with $\sim \pm 70$ the maximum routinely achieved. The consequence of the highly tilted views being missing is that the resulting 3-D reconstruction contains a missing wedge of information in Fourier space, i.e. a reconstruction that has non-isotropic resolution and is elongated in the direction of the electron beam. Nevertheless, even such low-resolution, noisy and distorted structural information has provided remarkable insight into the biology of a range of pleiomorphic viruses, including human & simian immunodeficiency viruses (HIV & SIV respectively), Herpes Simplex virus (HSV), Rous Sarcoma virus and many others.

Cryo-ET on pleiomorphic viruses was pioneered through a collaboration between Alasdair Steven & Wolfgang Baumeister, who published the first cryo-ET reconstruction of intact HSV particles in ice (Grunewald *et al.*, 2003). HSV is a membrane-bound, pleiomorphic virus, consisting of an icosahedrally ordered nucleocapsid, a membrane envelope that contains many different viral proteins, and a tegument that fills the space between nucleocapsid and membrane (Fig. 16a & b). Using cryo-ET, Grunewald *et al.*, were able to show in individual virions, that the nucleocapsid closely approaches the membrane on one side of the virion, leaving up to 35 nm of tegument separating nucleocapsid and membrane on the other side.

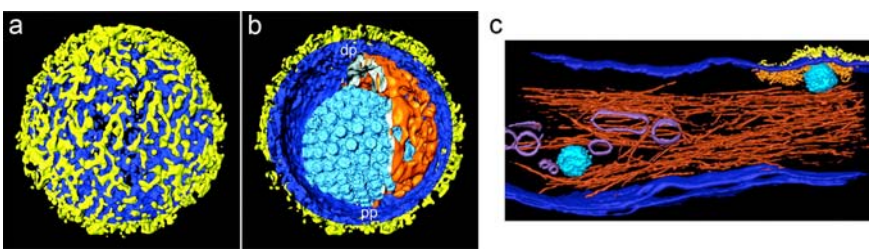


Fig. 16. Cryo-electron tomography of Herpes Simplex virus-1 (HSV1). Surface (a) and section (b) of an asymmetric reconstruction of a single HSV-1 particle. The structure shows the distribution of glycoprotein spikes (yellow), membrane (dark blue), viral tegument (orange) and the icosahedrally symmetric nucleocapsid (light blue). From Grunewald *et al.*, 2003. (c). HSV-1 entry into a mammalian (PtK2) cell. The same colour scheme is used, with the exception that cellular vesicles (purple) and the actin cytoskeleton (dark brown) are also now seen. From Maurer *et al.*, 2008.

They also described short actin-like filaments within the tegument, and the size and heterogeneity (in distribution and structure) of glycoprotein spikes in the membrane envelope.

Cryo-ET continues to play a pivotal role in the latest research on HSV-1, in which remarkable images of the pleiomorphic structure described above entering a target cell have been captured (Maurer *et al.*, 2008; Fig. 16c). They show how incoming capsids fuse with the plasma membrane of their target cells, releasing the nucleocapsid and tegument into the interior. They propose that the side of an HSV particle where the nucleocapsid closely approaches the viral membrane is the site of fusion with the plasma membrane.

For obvious reasons, immunodeficiency viruses have been a major focus of cryo-ET studies in recent years. Immunodeficiency viruses exhibit an extremely large degree of structural heterogeneity. The size of the virion, the number (and distribution) of glycoprotein spikes, and the size and shape of the nucleocapsid core are all variable (Fig. 17).

Much effort has been directed towards the structure of spikes that protrude from the HIV and SIV virions. These spikes, which are formed by trimers of the viral envelope glycoproteins (Env), mediate entry of immunodeficiency viruses into their target cells and are a target for neutralising antibodies. Env is a heterodimer of a trans-membrane glycoprotein (gp41) and a surface glycoprotein (gp120). Recent studies have revealed the trimeric nature of the Env spikes, and Liu *et al.*, (2008) showed a

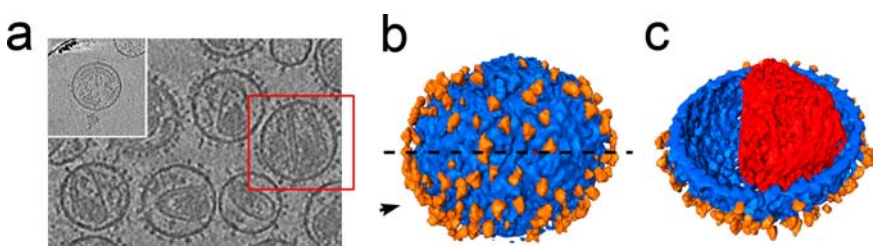


Fig. 17. Cryo-electron tomography of immunodeficiency viruses. (a) A field of mature SIV particles in vitreous ice, showing the extreme structural diversity exhibited by mature immunodeficiency viruses. The red box indicates a single virion, the surface and sectioned view of which is seen in 3-D in (b) and (c) respectively. Glycoprotein spikes are shown in orange, membrane in blue, and core in red. Adapted from Zanetti *et al.*, 2006. Inset in (a) shows a central, 5.6 nm section through a tomographic reconstruction of an immature HIV virion. Adapted from Wright *et al.*, 2007.

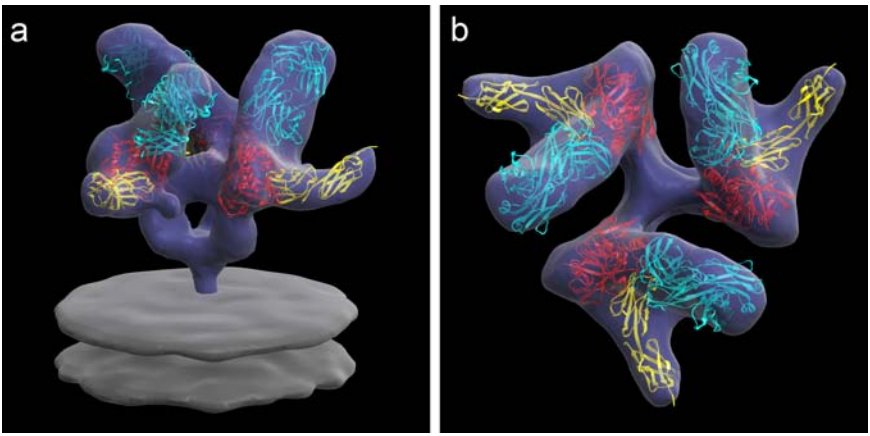


Fig. 18. The structure of the HIV-1 glycoprotein spike in complex with CD4 and Fab fragment. (a) Side and (b) top views of the cryo-EM structure with fitted X-ray structures for gp120 (red), the Fab fragment of a neutralising antibody (cyan) and CD4 (yellow) docked into the EM density. Adapted from Liu *et al.*, 2008.

conformational change in the trimer upon binding of CD4, the initial cellular receptor for HIV binding to its target cells. This conformational change appears to act to draw the cellular and viral membranes together, presumably facilitating membrane fusion and thereby infection.

11. Conclusions

Cryo-EM has become an indispensible tool in structural virology. With its ability to examine the structure of a virus in a range of conditions, and recover structural information without the need to grow crystals, it is ideally suited to looking at the structural dynamics that appear to be essential for infectivity in many viral systems. In many ways cryo-EM is the perfect counterpart to X-ray diffraction methods. We have described a number of studies in which intermediate-resolution information from cryo-EM, and high-resolution information from X-ray crystallography, have been combined to give biological insights impossible from either method alone. Clearly the whole is greater than the sum of the parts, and the resulting synergy has been a key driver in our increased understanding of viral lifecycles.

However, whilst cryo-EM has matured rapidly in the past decade, now is a particularly exciting time to be involved in its use and development. Recent advances in instrumentation, image processing algorithms

and computing mean that there is finally a realistic possibility that cryo-EM will become a tool capable of determining virus structures to atomic resolution. This will be restricted to homogeneous, ordered particles, but nevertheless it may provide a new route to study details of virus structure in systems that have to date been intractable to crystallization. At the other end of the resolution scale, cryo-electron tomography is rapidly developing and is the only tool capable of looking at the structure of individual, pleiomorphic pathogens in their cellular context. Cryo-ET is particularly exciting because it affords the opportunity to bridge a gap in resolution between the high resolution of X-ray, NMR and single-particle cryo-EM methods (2–20 Å) and low-resolution optical methods such as fluorescence microscopy (>∼150 Å). The prospect therefore exists of a continuum of structural information from atoms to cells. Exciting times indeed.

References

1. N. Abrescia, J. Grimes, H. Kivelä, R. Assenberg, G. Sutton, S. Butcher, *et al.* Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Molecular Cell* **31**:749–61 (2008).
2. M. Adrian, J. Dubochet, J. Lepault and A. McDowall. Cryo-electron microscopy of viruses. *Nature* **308**:32–6 (1984).
3. T. Baker, N. Olson and S. Fuller. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiology and Molecular Biology Reviews: MMBR* **63**:862–922 (1999).
4. N. Ban and A. McPherson. The structure of satellite panicum mosaic virus at 1.9 A resolution. *Nat. Struct. Biol.* **2**:882–90 (1995).
5. S. Brenner and R. Horne. A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta.* **34**:103–10 (1959).
6. D. Bubeck, D. Filman and J. Hogle. Cryo-electron microscopy reconstruction of a poliovirus-receptor-membrane complex. *Nat. Struct. Mol. Biol.* **12**:615–8 (2005).
7. Z. Chen, C. Stauffacher, Y. Li, T. Schmidt, W. Bomu, G. Kamer, *et al.* Protein-RNA Interactions in an Icosahedral Virus at 3.0 Å Resolution. *Science* **245**:154–159 (1989).
8. J. Cockburn, N. Abrescia, J. Grimes, G. Sutton, J. Diprose, J. Benevides, *et al.* Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* **432**:122–5 (2004).
9. M.A. Convery, S. Rowsell, N.J. Stonehouse, A.D. Ellington, I. Hirao, J.B. Murray, D.S. Peabody, S.E.V. Phillips and P.G. Stockley. Crystal structure of an RNA aptamer protein complex at 2.8 angstrom resolution. *Nature Structural Biology* **5**:133–139 (1998).
10. A. Fisher and J. Johnson. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361**:176–9 (1993).

11. K. Grünewald, P. Desai, D. Winkler, J. Heymann, D. Belnap, W. Baumeister, *et al.* Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**:1396–1398 (2003).
12. S. Harrison, A. Olson, C. Schutt and F. Winkler. Tomato Bushy Stunt Virus at 2.9A resolution. *Nature* **276**:368–373 (1978).
13. W. Jiang, M. Baker, J. Jakana, P. Weigele, J. King and W. Chiu. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* **451**:1130–1134 (2008).
14. P. Kolatkar, J. Bella, N. Olson, C. Bator, T. Baker and M. Rossmann. Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. *EMBO J.* **18**:6249–6259 (1999).
15. B. La Scola, C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature* **455**:100–104 (2008).
16. S. Larson and A. McPherson. Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.* **11**:59–65 (2001).
17. S. Larson, J. Day, A. Greenwood and A. McPherson. Refined structure of satellite tobacco mosaic virus at 1.8 A resolution. *J. Mol. Biol.* **277**:37–59 (1998).
18. K. Lee, L. Gan, H. Tsuruta, C. Moyer, J. Conway, R. Duda, *et al.* Virus capsid expansion driven by the capture of mobile surface loops. *Structure* **16**:1491–1502 (2008).
19. J. Liu, A. Bartesaghi, M. Borgnia, G. Sapiro and S. Subramaniam. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**:109–113 (2008).
20. U. Maurer, B. Sodeik and K. Grünewald. Native 3D intermediates of membrane fusion in herpes simplex virus 1 entry. *Proc. Natl. Acad. Sci. USA* **105**:10559–10564 (2008).
21. O. Medalia, I. Weber, A. Frangakis, D. Nicastro, G. Gerisch and W. Baumeister. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**:1209–1213 (2002).
22. D. Raoult and P. Forterre. Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* **6**:315–319 (2008).
23. A. Schneemann. The structural and functional role of RNA in icosahedral virus assembly. *Annu. Rev. Microbiol.* **60**:51–67 (2006).
24. A. Steven, J. Heymann, N. Cheng, B. Trus and J. Conway. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr. Opin. Struct. Biol.* **15**:227–236 (2005).
25. P.G. Stockley, O. Rolfsson, G.S. Thompson, G. Basnak, S. Francese, N.J. Stonehouse, S.W. Homans and A.E. Ashcroft. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T = 3 viral capsid. *Journal of Molecular Biology* **369**:541–552 (2007).
26. S. Subramaniam, A. Bartesaghi, J. Liu, A. Bennett and R. Sougrat. Electron tomography of viruses. *Curr. Opin. Struct. Biol.* **17**:596–602 (2007).
27. S. Sun, K. Kondabagil, B. Draper, T. Alam, V. Bowman, Z. Zhang, *et al.* The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135**:1251–1262 (2008).
28. J. Tang, N. Olson, P. Jardine, S. Grimes, D. Anderson and T. Baker. DNA poised for release in bacteriophage phi29. *Structure* **16**:935–943 (2008).

29. L. Tang, K. Johnson, L. Ball, T. Lin, M. Yeager and J. Johnson. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat. Struct. Biol.* **8**:77–83 (2001).
30. K. Toropova, G. Basnak, R. Twarock, P. Stockley and N. Ranson. The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**:824–836 (2008).
31. K. Valegard, J.B. Murray, P.G. Stockley, N.J. Stonehouse and L. Liljas. Crystal-structure of a bacteriophage-RNA coat protein-operator complex. *Nature* **371**:623–626 (1994).
32. S. van den Worm, R. Koning, H. Warmenhoven, H. Koerten and J. van Duin. Cryo electron microscopy reconstructions of the Leviviridae unveil the densest icosahedral RNA packing possible. *J. Mol. Biol.* **363**:858–865 (2006).
33. M. van Heel, B. Gowen, R. Matadeen, E. Orlova, R. Finn, T. Pape, *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Quart. Rev. Biophys.* **33**:307–369 (2000).
34. W. Wikoff, L. Liljas, R. Duda, H. Tsuruta, R. Hendrix and J. Johnson. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**:2129–2133 (2000).
35. E. Wright, J. Schooler, H. Ding, C. Kieffer, C. Fillmore, W. Sundquist, *et al.* Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* **26**:2218–2226 (2007).
36. C. Xiao, P. Chipman, A. Battisti, V. Bowman, P. Renesto, D. Raoult, *et al.* Cryo-electron microscopy of the giant Mimivirus. *J. Mol. Biol.* **353**:493–496 (2005).
37. C. Xiao, Y.G. Kuznetsov, S. Sun, S.L. Hafenstein, V.A. Kostyuchenko, P.R. Chipman, *et al.* Structural studies of the giant mimivirus. *PLoS Biol.* **7**:e92 (2009).
38. X. Yan, P. Chipman, T. Castberg, G. Bratbak and T. Baker. The marine algal virus PpV01 has an icosahedral capsid with T = 219 quasisymmetry. *J. Virol.* **79**:9236–9243 (2005).
39. X. Yan, N. Olson, J. Van Etten, M. Bergoin, M. Rossmann and T. Baker. Structure and assembly of large lipid-containing dsDNA viruses. *Nat. Struct. Biol.* **7**:101–103 (2000).
40. I.-M. Yu, W. Zhang, H. Holdaway, L. Li, V. Kostyuchenko, P. Chipman, *et al.* Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* **319**:1834–1837 (2008).
41. X. Yu, L. Jin and Z. Zhou. 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **453**:415–419 (2008).
42. G. Zanetti, J. Briggs, K. Grünewald, Q. Sattentau and S. Fuller. Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS pathogens* **2**(8):e83 (2006).
43. X. Zhang, E. Settembre, C. Xu, P. Dormitzer, R. Bellamy, S. Harrison, *et al.* Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* **105**:1867–1872 (2008).

Chapter 2

What Does it Take to Make a Virus: The Concept of the Viral 'Self'

Nicola G. A. Abrescia[*,§], Jonathan M. Grimes[†], Elizabeth E. Fry[†], Janne J. Ravantti[‡], Dennis H. Bamford[‡] and David. I. Stuart[†,¶]

Structural analyses now provide a framework to understand some aspects of virus evolution, where the three dimensional structures of virus capsids essentially substitutes for a fossil record. It has emerged that whilst many aspects of viruses are subject to horizontal gene transfer, for many viruses the core structures responsible for capsid assembly and genome packaging are highly conserved. We define this set of indispensable properties as the virus 'self'. A 'viral lineage' then comprises a set of viruses sharing a recognizably common 'self'. In this review we summarize the history and current status of the endeavour to develop a structure-based taxonomy based around this concept, highlighting not only some of the insights that this has brought, but also the problems that have emerged and the prospects for a systematic structure-based classification of the virosphere.

1. Introduction

Recent results, primarily structural, and new ideas have rekindled the debate on the origins and evolution of viruses. Viruses are the most

[*]Structural Biology Unit, CICbioGUNE, Bizkaia Technology Park, Bld.800, 48160 Derio, Spain.
[†]Division of Structural Biology and the Oxford Protein Production Facility, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford, OX3 7BN, UK.
[‡]Institute of Biotechnology and Department of Biological and Environmental Sciences, Viikki Biocenter, University of Helsinki P.O. Box 56, Viikinkaari 5, 00014 University of Helsinki, Finland.
[§]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain.
E-mail: [¶]dave@strubi.ox.ac.uk

abundant forms of life, infecting practically all cellular organisms and permeating the environment, constituting a pervasive virosphere. Large numbers also reside integrated into the genomes of their host organisms. The human suffering and losses in agricultural production due to viral infections underscore their impact on the human condition and furthermore the emerging picture is that there is a considerable contribution of viruses to carbon cycling. It is obvious that to understand biology we must have a deep understanding of viruses — where they came from and how they operate and evolve.

Sequence information has significantly advanced our understanding of the relatedness of virus isolates. The gradual accumulation of such information, in the context of the Baltimore classification scheme, which groups viruses according to their strategy of replication (i.e., genome nucleic acid type[1]), and host range, has allowed most known viruses to be grouped into families. At present there are 73 virus families recognized[2] and it is likely that these encompass most extant viruses.

In contrast to the good progress in classifying within families it has proved much more difficult to find a secure basis for grouping families. Some tentative steps have been taken with the definition of a limited number of virus orders which encompass a bigger portion of the virosphere. The problem that underlies the struggle to group virus families is that, in general, there is almost no detectable sequence similarity at either the nucleic acid or protein level between virus families. Where there are clear exceptions to this (for example tailed phages are grouped into three different families) it can be argued that this reflects a recrudescence of the ancient taxonomic battle between 'splitters' and 'lumpers', where the splitters have gained the upper hand. Several factors contribute to the much more rapid dissolution of such similarities during virus evolution than cellular evolution (the rate of RNA virus evolution often being reckoned to be up to a million-fold greater than for higher eukaryotic hosts), so that the absence of such evidence cannot be taken to imply that no evolutionary relationships exist. An alternative approach to classification has long been that of morphology (the Linnean approach, defining structure-based homology). The shortage of detailed information available has traditionally limited the usefulness of such an approach to virus classification. Now, however, the accumulation of atomic resolution structures of virion component proteins and, even more usefully, entire virions, allows extremely detailed morphological comparisons to be made

for a significant number of viruses (the number of high resolution capsid structures is ~80). These comparisons have led to some surprising new observations.

An immediate result from these structure-based analyses is that it appears that the common practice of using host taxonomy as a classification criteria for viruses (i.e., animal, plant, bacterial, archaeal viruses, *etc.*) is of little value since capsid structures and assembly principles are shared within structure-based viral lineages whose members often infect hosts from all three domains of life (bacteria, archaea and eukarya). These observations lead to the suggestion that viruses are ancient, predating the separation of the three cellular domains of life and that LUCA (the last universal common ancestor of all cellular life) was already infected by a number of structurally distinct viral species for which no universal common viral ancestor can be meaningfully proposed. We have discussed these issues previously.[3–6] Following on from the idea of the very ancient origins of viral life, the role of viruses in shaping the development of cellular life has also been actively discussed; see for example the proposal on the origin of the cellular nucleus as a consequence of primordial virus intervention.[7] Although the methods for protein structure-based taxonomy are still at a primitive stage of development, it is evident that such new information will eventually and necessarily change the way we classify viruses and may provide a more objective criteria for lumping or splitting at the family and higher levels.

Here we summarize recent experimental results and present some quantitative analyses of these findings using a structure-based approach to virus phylogeny, with the aim of further clarifying the idea of viral lineages, as well as extending the concept of the viral 'self',[4,5] i.e., when does a replicon become a virus and what is the relationship between the viral 'self' elements and the other elements that viruses carry in their genomes.

2. Towards the Concept of Viral Lineage?

Some of the earliest structural analyses of viruses demonstrated that the capsid protein fold and the arrangement of protein subunits could be conserved between viruses that belong to different families and infect very different hosts. Thus the first plant and animal viruses (i.e., tomato bushy stunt virus,[8] southern bean mosaic virus,[9] human rhinovirus[10] and

poliovirus[11]) possessed striking similarities that were recognized as probably signifying descent from a common ancestor,[12] despite the absence of detectable sequence similarity between these plant and animal viruses, which were assigned to different families. However, since these were some of the first virus structures determined it was not easy to disentangle similarities that were the residue of ancestral similarity from those that arose because there was simply no other way of forming a suitable viral capsid, and might therefore reflect convergence from separate ancestors. Years passed before a key observation was made, that the enteric membrane containing bacteriophage PRD1 and human adenovirus share a common coat protein fold (the so-called double-barrel[13]) arranged in similar trimers to form hexagonal capsomer structures suitable for building the icosahedral plates from which both of these virions are constructed. By this time sufficient diversity in virus molecular architecture had been observed that there could be no doubting that these structures shared a common ancestry.

Such observations raised the question of whether the astronomical number of viruses in the virosphere are described by a relatively small number of structural principles governing how to build the viral capsid and perhaps how to encapsidate the viral genome. Although we only have structural information about a tiny fraction of all viruses, the number of protein folds capable of assembling to viral capsids is likely to be very limited, including only a small subset of all protein folds. Even though it is evident that novel virus coat protein folds will be discovered in the future it is intriguing to carry out an exercise where all currently known virus coat protein folds (and virion architectures when available) are compared and categorized. Figure 1 presents a simplified summary of such an analysis and reveals four separate viral 'lineages' for icosahedral viruses. We will consider each of these in turn, then assess how incomplete this analysis is and try to judge what will be the likely limits for such analyses.

3. The Double-Barrel Paradigm Lineage

The unexpected similarity between the capsid proteins of PRD1 and adenovirus prompted a re-analysis of the low resolution structure of one of the largest viruses known, Chlorella virus PBCV-1, a phycodna virus, and the proposition that it belonged to the same lineage,[3] which was subsequently

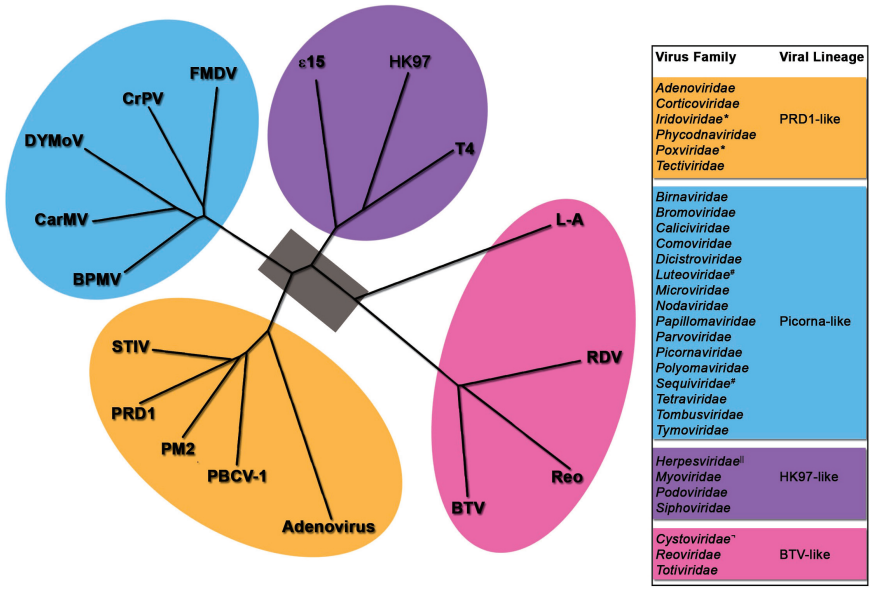


Fig. 1. Structure-based phylogenetic tree showing the four different lineages so far detected by structural comparison of viral capsid proteins. The matrix of evolutionary distances has been calculated using SHP[5,64,65] and the tree drawn using PHYLIP.[66] Only some of the family members are represented in the tree whilst the full list for each clade is given in the table (right). Those families included on the basis of sequence similarity or structural information obtained by electron microscopy have been marked. The grey rectangle at the centre of the tree indicates that the lineages should not be considered as sharing a common ancestor, i.e., viruses are polyphyletic in origin.

The PDB codes used for each virus representative are: Foot-and-Mouth disease virus (FMDV; 1zba), Bean pod mottle virus (BPMV; 1pgl), Carnation mottle virus (CarMV; 1opo), Desmodium yellow mottle virus (DYMoV; 1ddl), Cricket paralysis virus (CrPV; 1b35), bacteriophage HK97 (1ogh), bacteriophage T4 (1yue), bacteriophage ε15 (3c5b), Sulfolobus icosahedral turret virus (STIV; 2bbd), bacteriophage PRD1 (1cjd), Paramecium bursaria Chlorella virus 1 (PBCV-1; 1j5q), bacteriophage PM2 (2vvf), Adenovirus (1p2z), L-A virus (1m1c), Rice dwarf virus (RDV; 1uf2), Blue tongue virus (BTV; 2btv), and Reovirus (Reo; 1ej6).

verified.[14] This, together with an understanding of the likely assembly pathway of PRD1, which offered an eminently scaleable mechanism for particle size determination by using a molecular tape measure,[15] suggested that there might be a spectrum of viruses belonging to this lineage, including some of the largest viruses known. Further structural

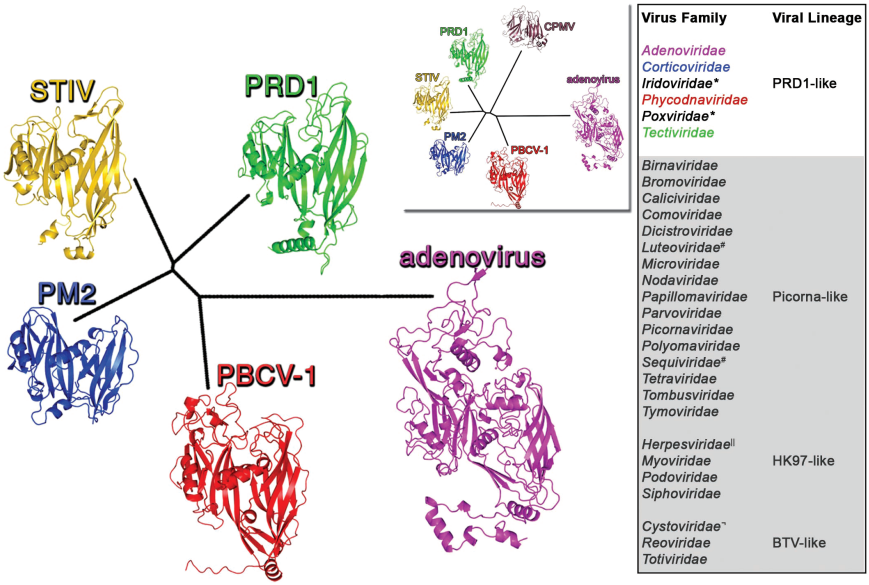


Fig. 2. Structure-based phylogenetic tree (calculated as for Fig. 1) for the PRD1-like viral lineage calculated using the double β-barrel major capsid proteins of the family members for which the structure is known at atomic resolution (represented as cartoon models). The lack of any loop elaboration at the top or bottom of the two jelly-rolls is apparent in the PM2 capsid protein (blue) when compared to the other members. The table on the right highlights all the virus families included in the PRD1-like lineage. Members of these families infect organisms across the three domains of life. For the *Iridoviridae* and *Poxviridae* there is some structural evidence[22,67] to suggest their membership to this clade. The PDB codes for each virus representative are as Fig. 1. The inset phylogenetic tree includes Cowpea mosaic virus double jelly-roll (PDB code: 1ny7; large subunit S) to show that, for its size, it is an outlier from the other double jelly-rolls.

studies and bioinformatics analyses of viral genomes demonstrated that there is a major lineage of viruses with similar coat protein topology and virion architecture (Fig. 2). These viruses infect hosts residing in all three domains of life. This lineage is now postulated to also include, in addition to PRD1, adenovirus, Chlorella virus PBCV-1,[14] viruses such as the iridoviruses, mimivirus,[6] and archaeal viruses STIV,[16] TKV4 and MVV.[17,18] Such viruses had not previously been considered to be related. The lineage has been extended to include phage Bam35[19] infecting gram positive

bacterial hosts (PRD1 infects gram negative bacterial hosts) and, more surprisingly, the marine bacteriophage PM2 belonging to its own family *Corticoviridae* has very recently been shown to have the PRD1-type double barrel coat protein topology[20] albeit a minimal one.[15] PM2 retains some ancestral properties, possessing a double jelly-roll with the individual jelly-rolls almost identical to each other (suggestive of a gene duplication event) and an assembly pathway where the different morphological stages are linked together.[20] PM2-type viruses are commonly found integrated in the genomes of aquatic bacteria.[21] Perhaps most surprising is the suggestion, albeit based on very low resolution electron microscopy data,[22] which is in line with the earlier bioinformatics analysis,[6] that the apparently pleomorphic vaccinia viruses share the double-barrelled, pseudo-hexagonal structural building block characteristic of this lineage but it is used only transiently during the virion assembly.

The current extent of and, where calculable, the structure-based relationships within this lineage are shown in Fig. 2. This structural similarity between proteins seen in viruses infecting organisms in all three domains of life suggests that this is a very ancient lineage, whose origins predate the separation of the domains of life.

4. Picorna-Like Group — Single Barrel, Single Lineage?

The hypothesis that the double jelly-roll emerged as duplication event of a single jelly-roll has been recently reinforced by observations on PM2.[20] The double jelly-roll may therefore be seen as a specialised version of the single β-barrel which is found mainly in RNA viruses. However, the jelly-roll is a sufficiently simple and common folding motif that we cannot infer from it alone that the two types of structure share a common ancestor. In fact there are characteristic differences between the two types of structures. Perhaps the most obvious is that for most of the single jelly roll capsids the axis of the β-barrels tends to be in the plane of the capsid shell, whereas the double-barrels are usually roughly perpendicular. However this is not a hard and fast rule (e.g., picornavirus VP1 jelly rolls can be almost perpendicular) and detailed structural signatures are a more compelling route to defining similarities and differences. A detailed discussion

of this type is beyond the scope of this review; however, we note that there is a series of signature features that designate these proteins as distinct. For instance, whilst the BIDG strands superpose well, there is a clear difference in the relative orientation of the CHEF strands (nomenclature as defined in Ref. 13). On this basis we classify the concatenated β-barrel structure of cowpea mottle virus as a relatively recent duplication event within the picorna-like lineage, rather than a progenitor of the double β-barrel lineage, which we believe is much older (Figs. 2 and 3 provide a quantitative justification for this assignment).

At present we have tentatively assigned 16 virus families to the single jelly roll picornavirus-like lineage. *Picornaviruses* are positive-sense, single-stranded RNA (ssRNA) non-enveloped viruses infecting vertebrates; examples of the family are poliovirus,[11] Foot-and-Mouth disease virus,[23] and human rhinovirus[10] on which a wealth of structural information has been accumulated. Traditional classification methods have recognised a larger group of positive-sense ssRNA viruses infecting animals, plants and insects as picorna-like, defining a superfamily recently expanded to include viruses of the marine community.[24] Most of the phylogenetic analyses on this group of viruses have been carried out using the RNA-dependent RNA polymerase (RdRp) sequence as a molecular marker. A more recent study has included three additional genes encoding for a chymotrypsin-like protease (3CPro), a superfamily 3 helicase (S3H) and a genome-linked protein (VPg) as possible signature markers for membership in the picorna-like superfamily.[25] However, in their analyses structurally unrelated viruses were also grouped together. Our approach, based on a very different set of comparisons between virus 'selves', groups the recognised members of the picorna-like superfamily in the same lineage (Fig. 3). However, the set of families that we identify as being possible members of this lineage extends beyond those proposed by others. At present we have not assembled sufficient detailed structural data to be confident about these assignments. Nevertheless, the structural superimpositions provide a set of core data with which to address the question and we think an inclusive analysis is a useful starting point. However, this raises various issues, for instance the capsid based comparison suggests that the lineage includes some DNA viruses (e.g., mammalian parvoviruses and bacterial microviruses) and whilst most members of the lineage are a similar size (capsid shell of the order of 30 nm across) polyoma and papilloma

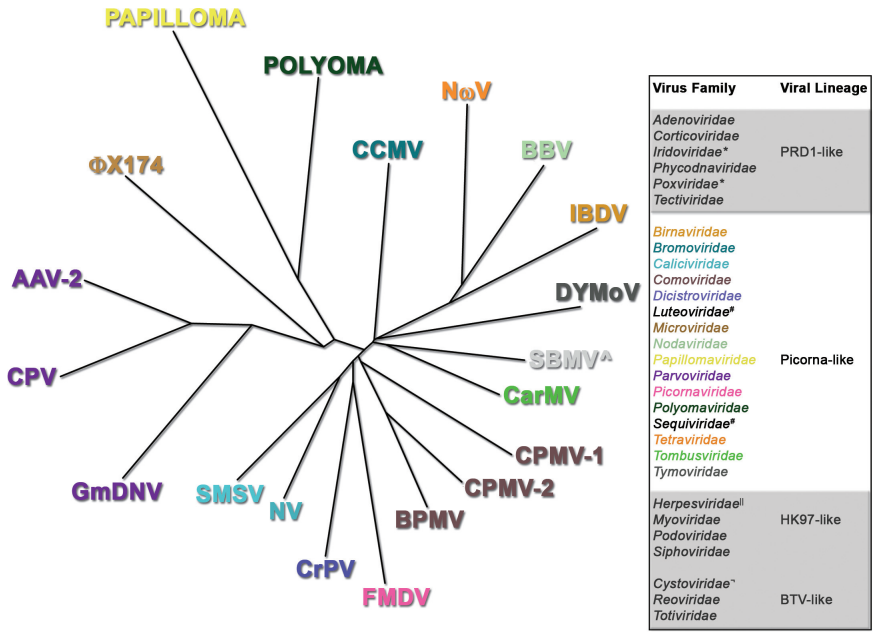


Fig. 3. Picorna-like tree (calculated as for Fig. 1). The *Parvoviridae* along with *Papillomaviridae* and *Polyoviridae* are at the boundaries of the lineage and they have been included in the tree after removal of a large loop insertion (cf. in AAV-2 residues 289–505 and in Papilloma residues 162–245). For the *Tetraviridae*, *Birnaviridae* and *Nodaviridae* insertions have been deleted before superimposition (*cf.* in NωV residues 163–217 and 281–412). Southern Beam Mosaic Virus (SBMV) has also been included in the tree but marked because it doesn't represent a family (genus *Sobemovirus*). Same colours indicate viruses belonging to the same family (*cf.* Parvoviruses). In the table-key for this lineage the *Luteoviridae* and *Sequiviridae* families have been added on the basis of sequence similarity with *Sobemoviruses* but no structural information is available. In the tree the individual jelly-rolls of Cowpea mosaic virus (CPMV-1 and CPMV-2) have also been included to highlight their membership of this viral lineage rather than to the PRD1-like viruses.

The PDB codes used for each virus representative are the same as in Fig. 1 and Southern bean mosaic virus (SBMV; 4sbv), Infectious bursal disease virus (IBDV; 1wcd), Black beetle virus (BBV; 2bbv), Nudaurelia capensis ω virus (NωV; 1ohf), Cowpea chlorotic mottle virus (CCMV; 1za7), Papilloma virus (1dzl), Polyomavirus (1cn3), bacteriophage $\Phi \times 174$ (2bpa), Adeno-associated virus 2 (AAV-2; 1lp3), Canine Parvo virus (CPV; 4dpv), Galleria mellonella densovirus (GmDNV; 1dnv), San Miguel sea lion virus (SMSV; 2gh8), Norwalk virus (NV; 1ihm).

viruses are clearly substantially bigger (the capsid proteins possess very large inserted loops).

No matter where the boundaries of this lineage are drawn the data shown in Fig. 3 provide a useful objective and quantitative benchmark for the splitting into genera and families. Again it is outside the scope of this review to fully address this but we note that the diversity seen within the parvovirus family (in particular the insect virus GmDNV is quite distinct from the mammalian viruses human AAV-2 and canine PV) is significantly greater than that between the family *Tombusviridae* and the sobemoviruses (which are classified as a genera distinct from currently assigned families). Thus our list of members of the Picorna-like family includes *Luteoviruses* and *Sequiviruses* on the basis of 33% sequence similarity with the Rice yellow mottle virus (genus *Sobemovirus*), as well as sharing morphology and genome organisation.[2] We expect that structure determinations may not support the current splitting of these plant viruses into separate families.

5. More Phages, a Different Lineage

Another distinct virion structure-based lineage of icosahedral viruses embraces all tailed double stranded DNA (dsDNA) bacteriophages. The canonical coat protein topology for this lineage is the one originally determined for phage HK97,[26] and all tailed phage coat protein structures (determined using either X-ray diffraction or electron microscopy) resemble HK97 (phages T4,[27] P22,[28] phi29,[29] T5,[30] Epsilon 15[31]). Overall we now have structural data covering members of all three of the families of tailed bacteriophages (*Podo- Sipho-* and *Myo- viridae*) and all are structurally very similar (Fig. 4). Furthermore, members of two of these families infect both prokaryotic and archaeal hosts. Whilst there is no structural information for a validated archaeal-infecting virus the structure of a particle isolated from the archaea *Pyrococcus furiosus* belongs to this lineage, although it is not established as a true virus.[32] The first link to viruses infecting eukaryotic hosts was, perhaps surprisingly, to herpes viruses as the shell-forming domain of the protein making the icosahedral capsid appears to possess the canonical HK97 fold.[33] There are further similarities, for instance both herpes viruses and tailed phages are dsDNA viruses which appear to package their DNA into an empty premade capsid using somewhat similar molecular machines.

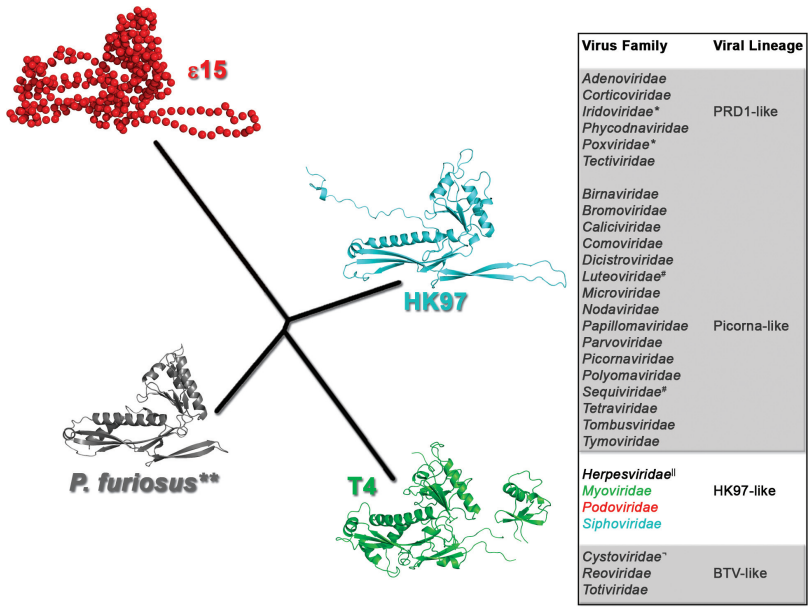


Fig. 4. Phylogenetic tree for the HK97-lineage (calculated as for Fig. 1). It includes the capsid protein of bacteriophage Epsilon15 obtained by electron microscopy at 4.5 Å (since only Cα positions are available these are represented as small spheres)[31] and the archaeal protein from *Pyrococcus furiosus* that self-assembles in virus-like particles obtained by crystallography at 3.6 Å resolution.[32] Strikingly HK97 major capsid protein gp5 is closer to the archaeal protein than to its counterparts from the other families of tailed phages. In the Table-key the *Herpesviridae* family is also included due to some structural evidences of possessing the capsid protein of similar fold.[33] The PDB codes of the structures used in the tree are: HK97 (1ohg), T4 (1yue), ε15 (3c5b), *P. furiosus* (2e0z).

6. dsRNA Genome — A Constraint on Architecture?

It has also been noticed that double stranded RNA (dsRNA) viruses of the *Reo, Toti* and *Cystoviridae* families (members infecting plants, animals, fungi and bacteria) share a common life style as well as a related capsid architecture[34] (Fig. 5A). The problem these viruses share is that exposed genomic dsRNA will illicit strong apoptotic responses in a eukaryotic host. This means that many of these viruses maintain at least an inner capsid in the infected cell which acts as a transcription factory. This innermost

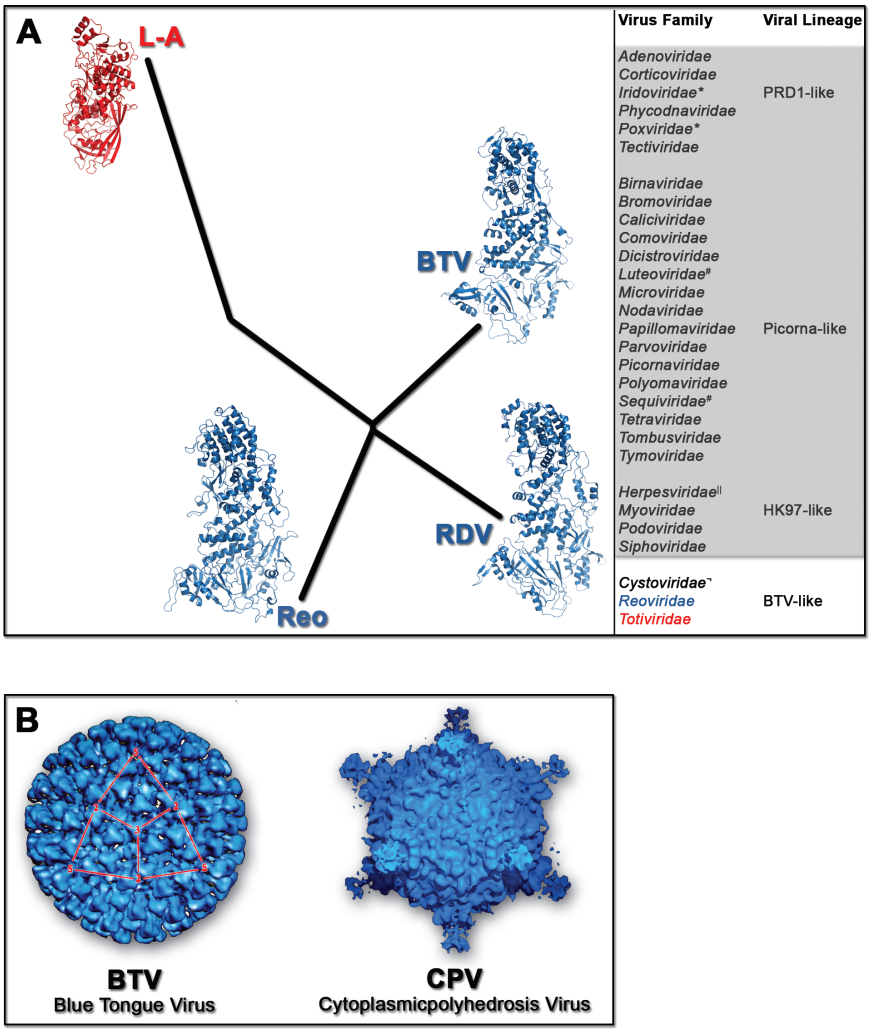


Fig. 5. BTV-like lineage (calculated as for Fig. 1). A) Phylogenetic tree comprising the *Totiviriadae* and the *Reoviridae*, the only two families so far implicated in this lineage by atomic structures. In the table-key on the right the *Cystoviridae* have also been ascribed to this lineage.[5] B) Side-by-side comparison of electron densities corresponding to two *Reoviridae* family members, Blue-tongue-virus (BTV)[68] and Cypovirus (CPV)[69] showing the strikingly morphological differences. The red lines with numbers mark respectively the icosahedral asymmetric unit within the virus facet and the icosahedral symmetry axes.

portion of the architecture defines the virus self, and it can be seen in the members of *Reoviridae* that this is retained very strongly despite massive alterations in the outer capsid layers (Fig. 5B). What remains a somewhat open question is how far does this lineage reach across the dsRNA viruses? The yeast L-A virus[35] has a series of features in common — a 120 copy inner layer and a broadly similar three-dimensional structure, however, the issue is that the degree of structural similarity is sufficiently limited that current algorithms find it hard to meaningfully align the structures. At present we favour the single lineage interpretation, but further evidence (e.g., a serial homologue) is required. Further, a bacteriophage (phi6) has been tentatively placed in this lineage, and in this case the immediate problem is the lack of high resolution structural information. Finally there are some families of dsRNA viruses that are not assigned to this lineage, in some cases because of a lack of structural knowledge. However in the case of the birnaviruses,[36] it has been proposed that there may be structural similarities with the BTV-like lineage. Nonetheless, in our opinion the similarity seems weak and a stronger link appears to be with the picorna-like lineage (see Fig. 3). Consequently a definitive structure-based assignment is difficult.

7. Enveloped Viruses — Infectious Vesicles?

A major outstanding problem for structure-based classification is how to deal with enveloped viruses. These viruses have an external lipid-membrane (usually containing glycoproteins and surrounding a nucleocapsid) and have so far proved refractory to crystallographic analysis. Indeed even the isolated structural proteins of these viruses tend to be difficult to solve crystallographically. Nevertheless quasi-atomic information has been obtained using hybrid methods by combining cryo-electron microscopy of the entire particles with the atomic structures of individual viral proteins.[37] An example of this combination of techniques is the study of Dengue virus which has elucidated key aspects of the flavivirus lifecycles.[38,39] So, given that structural information is being accumulated, what should we compare? By analogy with the other lineages we might expect the nucleocapsid to be at the heart of the viral self. We have investigated this for the negative sense ssRNA viruses (Fig. 6). We find that between viruses recognised as being related, e.g., Rabies virus[40] and

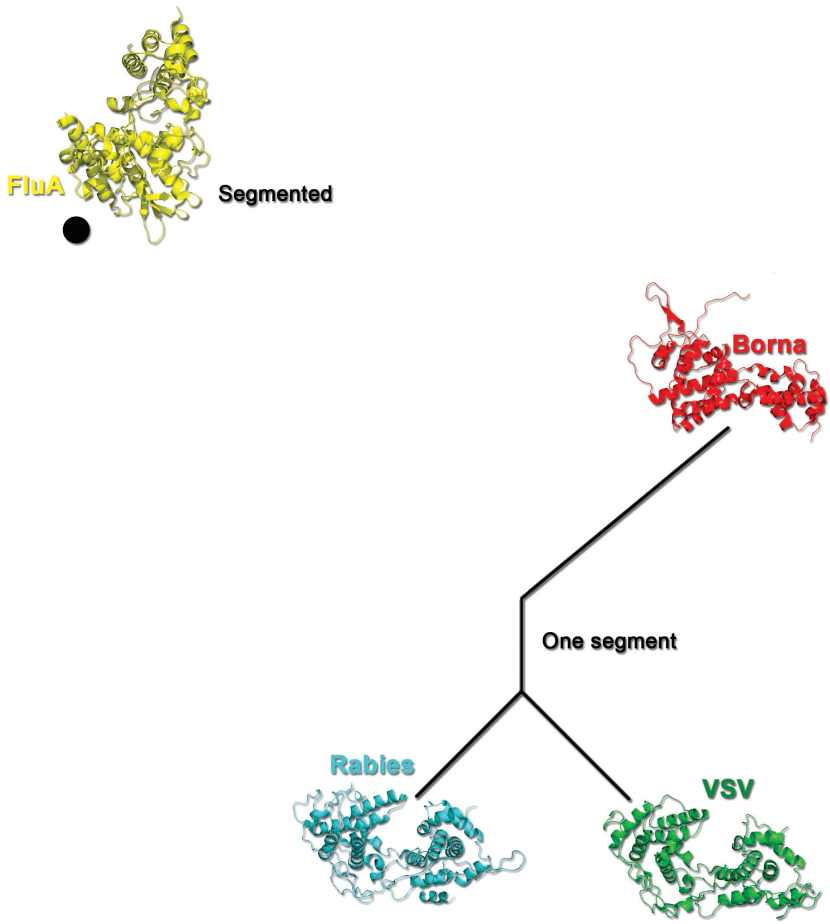


Fig. 6. Structure-based phylogenetic tree using the nucleoprotein structures of Rabies virus (PDB code: 2gtt),[40] Vesicular Stomatitis virus (VSV) (PDB code: 2gic)[41] and Borna Disease virus.[42] At a distance not linked to the tree, the structure of influenza A virus nucleoprotein (PDB code: 2iqh)[70] is shown; the dot indicates where the branch would end if connected; since the similarity is not compelling this structure is not included in the tree.

vesicular stomatitis virus (VSV),[41] there is strong similarity. If we consider bornavirus, whilst this structure has not been reported as similar to others,[42] we find a compelling similarity with the equivalent protein of rabies and VSV (Fig. 6). These three viruses all belong to the order *Mononegavirales* and the structural similarity confirms that grouping. If

we include a virus with a segmented genome, influenza virus, we find that there is no detectable similarity (this may be judged from Fig. 6 where the structures are all orientated so as to achieve the best mutual structural alignment). In this case the nucleocapsid comparison provides interesting confirmation of a relationship implicit in the present taxonomy, but does not support a broader grouping. In contrast nucleocapsid classification appears inappropriate for some other enveloped viruses, for instance the two groups of viruses which until very recently were classified together as the *Togaviridae* family: alphaviruses (now *Togaviridae* family) and flaviviruses (now *Flaviviridae* family).[43] The alphavirus internal nucleocapsid is ordered,[44] whilst that of the flaviviruses appears disordered.[45] Furthermore structural analysis of the individual capsid subunits reveals that they are unrelated. The capsid protein of the alphavirus Semliki Forest Virus (SFV) possesses the characteristic structure (two β-barrel sub-domains) and functions of a serine proteinase[46] whereas the flavivirus West Nile virus capsid contains four α-helices.[47] However if we shift focus to the envelope glycoprotein, which forms an ordered outer shell and is responsible for membrane fusion (and host-cell binding) there is astonishing structural similarity between the corresponding proteins (Fig. 7), suggesting that this protein may be a more appropriate marker of self.

A reasonable smattering of viral fusion proteins have now been analysed by crystallography (including representatives of the *Togaviridae*,[48] *Rhabdoviridae*,[49,50] *Herpesviridae*,[51] *Filoviridae*,[42,52] *Flaviviridae*,[38,53,54] *Baculoviridae*,[55] *Orthomyxoviridae*,[56] *Paramyxoviridae*[57,58]), and there are clear mechanistic commonalities between these proteins.[59] The structures of these viral fusion machines show massive structural differences, which have led to them being classified into three types (types I–III with the alpha/flavi-viruses belonging to type II). Despite the huge structural differences there has recently been a suggestion that they may share common elements and perhaps a common ancestry. This proposal was prompted by the recent crystal structure of baculovirus envelope protein gp64,[55] a type III fusion protein. Figure 8 compares the type III fusion protein structures and it is worth underlining that this type of fusion protein is used by virus families that span RNA (VSV) and DNA genomes (baculo- and herpes-viruses). Furthermore sequence comparisons indicate that the baculovirus protein is similar to that of Thogotovirus (also called influenza virus D).[55] Since in almost all other aspects these viruses are radically different it seems sensible to assume that the presence of a

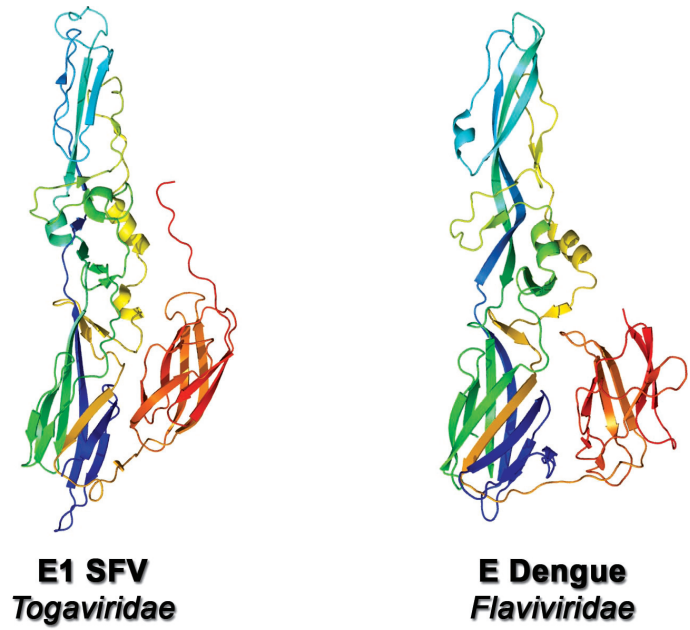


Fig. 7. Side-by-side comparison of the E1 glycoprotein of Semliki Forest Virus (SFV) (PDB code: 1rer) of the *Togaviridae* family and the glycoprotein E of Dengue type 2 (PDB code: 1ok8) of the *Flaviviridae* family, represented as cartoon and rainbow coloured from the N-terminus (blue) to the C-terminus (red).

closely related fusion protein reflects the ability of both viruses to infect insects and has arisen by gene transfer.

The systematic comparison of gp64 with the available crystal structures of type I, II and III fusion proteins[55] led to the observation of pervasive structural homology at the domain-level inferring that these viral machines might have a common origin. Although the mechanism of membrane fusion seems therefore to use mostly a common mechanism, involving a trimeric final state of the fusion protein this may not be a universal mechanism of protein-membrane fusion. Vaccinia virus (*Poxviridae* family), for example, seems to have adopted a different solution for fusion with the cell during entry requiring a set of 8 viral proteins forming the so called entry-fusion-complex (EFC).[60]

It appears that for the enveloped viruses it may be impossible to construct a generally useful phylogeny on the basis of any one viral protein and it is possible that the powerful mechanism for animal/insect cell

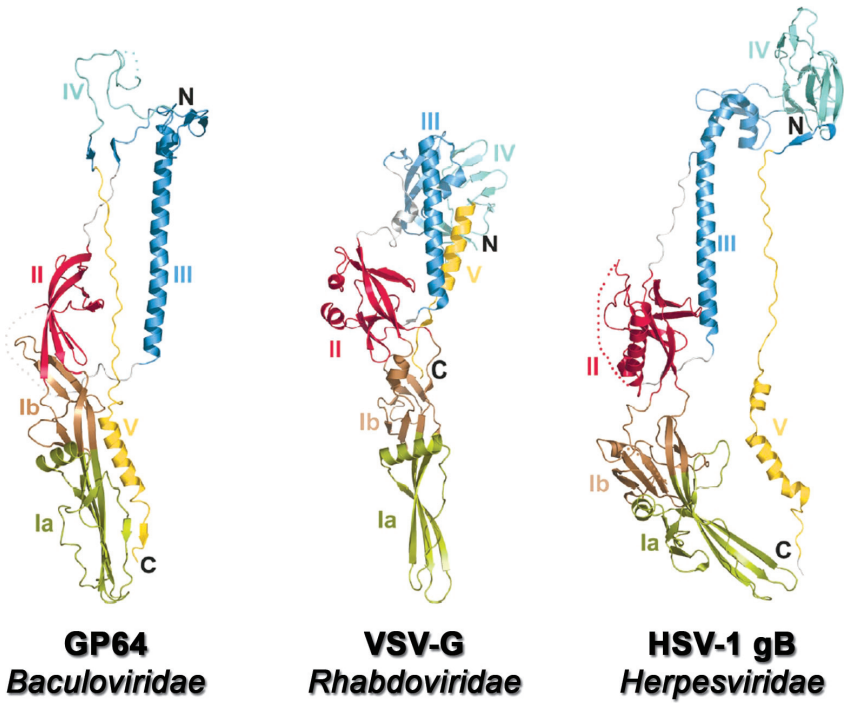


Fig. 8. Comparison of the structures of Baculovirus gp64, Vesicular stomatitis virus (VSV) G and Herpesvirus (HSV-1) gB protomers adapted from Ref. 55. The domain colouring corresponds to gp64 domain structure: pale-green (Ia: residues 68–166), light-brown (Ib: residues 60–67 and 167–218), red (II: residues 45–60 and residues 219–271), light-blue (III: residues 21–39, 299–374 and 409–413), cyan (IV: residues 375–408), in light-grey residues 40–44 and 272–298. VSV-G domain nomenclature was molded to the one of gp64 and HSV-1 gB.

entry offered by the fusion protein, with, in general, little requirement for large scale structural organisation may have facilitated the proliferation of these viruses. This may have allowed the virus self to be stripped down to a structurally malleable element that can be relatively easily proliferated by horizontal gene transfer. The subsequent differentiation of these viruses has then, for several virus families, been achieved by the use of an additional protein to recognise the cell receptor (e.g., the G proteins of paramyxoviruses[61]), or in the case of baculovirus perhaps by dispensing with receptor binding.[55]

8. The Viral Self and Emerging Viruses

It appears that the majority of known virus types fall into a very limited number of lineages. The number of essential proteins responsible for the virion assembly (including the factors involved in genome packaging) in each lineage is also reasonably small. If we consider that a virus is a virus only when it has the capacity to make a virion (quite a sensible assumption for a virus) we will also simultaneously define the determinants necessary and sufficient to turn a replicon into a virus. We have previously designated such determinants, which must be coupled in evolutionary space to ensure virion formation, the viral self.[3,4] It is obvious that there are many other obligatory viral functions such as replication and host cell recognition which are quite distinct from those defining the viral self. The discovery of PM2-type prophages in bacterial genomes has revealed the presence of several replication and lysis systems in this group of viruses.[21] Tailed phages phi29 and relatives use protein primed replication as opposed to the rolling circle replication systems used in other tailed phages.[62] It is obvious that in a given viral lineage (as defined here) different replication, entry, release etc. mechanisms are employed revealing that there is no necessary coupling of such properties with the self (and consequently these are designated as non-self). Thus a replicon (plasmid, transposon, etc.) can turn into a virus by obtaining the self determinants. However, in terms of the virion only a limited number of functional (and ancient) architectural types will be available and these were probably thoroughly explored at the time life was beginning on earth, so that it is unlikely that many new virus lineages (as defined here) will emerge subsequently. However, a myriad of non-self properties can be combined with the self determinants. An additional example of this is the spike protein (needed to bind to the host cell receptor) of adenovirus which is structurally similar to that of reovirus (a dsRNA virus, see above) and most probably achieved by horizontal gene flow.[63]

It should also be noted that a way for recognition of the genome for packaging and successful delivery to the host cell for infection has to be established. As a consequence the properties of the viruses in the same structural lineage (as defined here) may be extremely different. It also follows that malfunction of any of the self determinants would immediately remove the replicon from the domain of viruses. In every viral life cycle this is tested and only functional viruses continue vertically in the chain of viruses established early in the history of life.

9. Conclusions

Since the virion coat fold is vertically inherited and strongly conserved because of tight structural constraints, it follows that a capsid centred virus clustering will uncover the deep history of viruses. Due to the limited number of high resolution viral structures available we have only the first glimpses of the structural order of viruses. Although certain virus lineages are reasonably clearly defined, others such as the enveloped viruses still pose a real challenge. Whether the viral self hypothesis turns out to be universally applicable remains an open question, nevertheless the power of the hypothesis is demonstrated by the fact that, excluding enveloped viruses, we have already assigned nearly half of the 73 currently accepted virus families to lineages.

Whilst the virion architecture is the defining feature of a virus lineage, we would expect that within a lineage the members may have a variety of entry, replication and exit mechanisms. Such properties, along with more detailed virion analysis, would then be available for subdividing the lineages. Eventually when sequence information allows comparisons, fine tuning of virus classification can be made. We propose that this strategy may support a quantitative and comprehensive taxonomy of viruses. In addition this structural approach may have implications even for understanding the origin of life itself.

Acknowledgements

This work was supported by Academy of Finland Center of Excellence Program in Virus Research (2006–2011) grants 1129684, 1210253 (D.H.B.), by the UK Medical Research Council and European Commission through the SPINE2–COMPLEXES project LSHG–CT–2006–031220 (D.I.S., E.E.F. & J.M.G).

References

1. D. Baltimore. Expression of animal virus genomes. *Bacteriol. Rev.* **35**:235 (1971).
2. C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball (Eds.). *Virus Taxonomy: Classification and Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses*, Elsevier Academic Press Inc., San Diego, California (2005).

3. D.H. Bamford, R. M. Burnett and D.I. Stuart. Evolution of viral structure. *Theor. Popul. Biol.* **61**:461 (2002).
4. D.H. Bamford. Do viruses form lineages across different domains of life? *Res. Microbiol.* **154**:231 (2003).
5. D.H. Bamford, J.M. Grimes and D.I. Stuart. What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**:655 (2005).
6. S.D. Benson, J.K. Bamford, D.H. Bamford and R.M. Burnett. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* **16**:673 (2004).
7. P. Forterre. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**:5 (2006).
8. F.K. Winkler, C.E. Schutt, S.C. Harrison and G. Bricogne. Tomato bushy stunt virus at 5.5 Angstroms resolution. *Nature* **265**:509 (1977).
9. C. Abad-Zapatero, S.S. Abdel-Meguid, J.E. Johnson, A.G.W. Leslie, I. Rayment, M.G. Rossmann, D. Suck and T. Tsukihara. Structure of southern bean mosaic virus at 2.8 [angst] resolution. *Nature* **286**:33 (1980).
10. M.G. Rossmann, E. Arnold, J.W. Erickson, E.A. Frankenberger, J.P. Griffith, H.J. Hecht, J.E. Johnson, G. Kamer, M. Luo, A.G. Mosser, *et al.* Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**:145 (1985).
11. J.M. Hogle, M. Chow and D.J. Filman. Three-dimensional structure of poliovirus at 2.9 A resolution. *Science* **229**:1358 (1985).
12. M.G. Rossmann and J.E. Johnson. Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58**:533 (1989).
13. S.D. Benson, J.K. Bamford, D.H. Bamford and R.M. Burnett. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**:825 (1999).
14. N. Nandhagopal, A.A. Simpson, J.R. Gurnon, X. Yan, T.S. Baker, M.V. Graves, J.L. Van Etten and M.G. Rossmann. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl. Acad. Sci. USA* **99**:14758 (2002).
15. N.G. Abrescia, J.J. Cockburn, J.M. Grimes, G.C. Sutton, J.M. Diprose, S.J. Butcher, S.D. Fuller, C. San Martin, R.M. Burnett, D.I. Stuart, D.H. Bamford and J.K. Bamford. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**:68 (2004).
16. R. Khayat, L. Tang, E.T. Larson, C.M. Lawrence, M. Young and J.E. Johnson. Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl. Acad. Sci. USA* **102**:18944 (2005).
17. M. Krupovic and D.H. Bamford. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* **375**:292 (2008).
18. M. Krupovic and D.H. Bamford, Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat. Rev. Microbiol.* **6**:941 (2008).
19. P.A. Laurinmaki, J.T. Huiskonen, D.H. Bamford and S.J. Butcher. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure* **13**:1819 (2005).

20. N.G. Abrescia, J.M. Grimes, H.M. Kivela, R. Assenberg, G.C. Sutton, S.J. Butcher, J.K. Bamford, D.H. Bamford and D.I. Stuart. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol. Cell.* **31**:749 (2008).
21. M. Krupovic and D.H. Bamford. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics* **8**:236 (2007).
22. J.K. Hyun, F. Coulibaly, A.P. Turner, E.N. Baker, A.A. Mercer and A.K. Mitra. The structure of a putative scaffolding protein of immature poxvirus particles as determined by electron microscopy suggests similarity with capsid proteins of large icosahedral DNA viruses. *J. Virol.* **81**:11075 (2007).
23. R. Acharya, E. Fry, D. Stuart, G. Fox, D. Rowlands and F. Brown. The three-dimensional structure of foot-and-mouth disease virus at 2.9 A resolution. *Nature* **337**:709 (1989).
24. A.I. Culley, A.S. Lang and C.A. Suttle. High diversity of unknown picorna-like viruses in the sea. *Nature* **424**:1054 (2003).
25. E.V. Koonin, Y.I. Wolf, K. Nagasaki and V.V. Dolja. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat. Rev. Microbiol* **6**:925 (2008).
26. W.R. Wikoff, L. Liljas, R.L. Duda, H. Tsuruta, R.W. Hendrix and J.E. Johnson. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**:2129 (2000).
27. A. Fokine, P.G. Leiman, M.M. Shneider, B. Ahvazi, K.M. Boeshans, A.C. Steven, L.W. Black, V.V. Mesyanzhinov and M.G. Rossmann. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl. Acad. Sci. USA* **102**:7163 (2005).
28. W. Jiang, Z. Li, Z. Zhang, M.L. Baker, P.E. Prevelige, Jr. and W. Chiu. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat. Struct. Biol.* **10**:131 (2003).
29. M.C. Morais, K.H. Choi, J.S. Koti, P.R. Chipman, D.L. Anderson and M.G. Rossmann. Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol. Cell* **18**:149 (2005).
30. G. Effantin, P. Boulanger, E. Neumann, L. Letellier and J.F. Conway. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J. Mol. Biol.* **361**:993 (2006).
31. W. Jiang, M.L. Baker, J. Jakana, P.R. Weigele, J. King and W. Chiu. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* **451**:1130 (2008).
32. F. Akita, K.T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara and A. Nakagawa. The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. *J. Mol. Biol.* **368**:1469 (2007).
33. M.L. Baker, W. Jiang, F.J. Rixon and W. Chiu. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**:14967 (2005).
34. P. Mertens. The dsRNA viruses. *Virus Res.* **101**:3 (2004).

35. H. Naitow, J. Tang, M. Canady, R.B. Wickner and J.E. Johnson. L-A virus at 3.4 A resolution reveals particle architecture and mRNA decapping mechanism. *Nat. Struct. Biol.* **9**:725 (2002).
36. F. Coulibaly, C. Chevalier, I. Gutsche, J. Pous, J. Navaza, S. Bressanelli, B. Delmas and F.A. Rey. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell* **120**:761 (2005).
37. R.J. Gilbert, J.M. Grimes and D.I. Stuart. Hybrid vigor: hybrid methods in viral structure determination. *Adv. Protein Chem.* **64**:37 (2003).
38. L. Li, S.M. Lok, I.M. Yu, Y. Zhang, R.J. Kuhn, J. Chen and M.G. Rossmann. The flavivirus precursor membrane-envelope protein complex: structure and maturation. *Science* **319**:1830 (2008).
39. I.M. Yu, W. Zhang, H.A. Holdaway, L. Li, V.A. Kostyuchenko, P.R. Chipman, R.J. Kuhn, M.G. Rossmann and J. Chen. Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* **319**: 1834 (2008).
40. A.A. Albertini, A.K. Wernimont, T. Muziol, R.B. Ravelli, C.R. Clapier, G. Schoehn, W. Weissenhorn and R.W. Ruigrok. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**:360 (2006).
41. T.J. Green, X. Zhang, G.W. Wertz and M. Luo. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**:357 (2006).
42. M.G. Rudolph, I. Kraus, A. Dickmanns, M. Eickmann, W. Garten and R. Ficner. Crystal structure of the borna disease virus nucleoprotein. *Structure* **11**:1219 (2003).
43. R.J. Kuhn, W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, T.S. Baker and J.H. Strauss. Structure of dengue virus: implications for flavivirus organization, maturation and fusion. *Cell* **108**:717 (2002).
44. E.J. Mancini, M. Clarke, B.E. Gowen, T. Rutten and S.D. Fuller. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol. Cell* **5**:255 (2000).
45. W. Zhang, P.R. Chipman, J. Corver, P.R. Johnson, Y. Zhang, S. Mukhopadhyay, T.S. Baker, J.H. Strauss, M.G. Rossmann and R.J. Kuhn. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nat. Struct. Biol.* **10**:907 (2003).
46. H.K. Choi, G. Lu, S. Lee, G. Wengler and M.G. Rossmann. Structure of Semliki Forest virus core protein. *Proteins* **27**:345 (1997).
47. T. Dokland, M. Walsh, J.M. Mackenzie, A.A. Khromykh, K.H. Ee and S. Wang. West Nile virus core protein; tetramer structure and ribbon formation. *Structure* **12**:1157 (2004).
48. D.L. Gibbons, M.C. Vaney, A. Roussel, A. Vigouroux, B. Reilly, J. Lepault, M. Kielian and F.A. Rey. Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* **427**:320 (2004).
49. S. Roche, S. Bressanelli, F.A. Rey and Y. Gaudin. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**:187 (2006).
50. S. Roche, F.A. Rey, Y. Gaudin and S. Bressanelli. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* **315**: 843 (2007).

51. E.E. Heldwein, H. Lou, F.C. Bender, G.H. Cohen, R.J. Eisenberg and S.C. Harrison. Crystal structure of glycoprotein B from herpes simplex virus 1. *Science* **313**:217 (2006).
52. J.E. Lee, M.L. Fusco, A.J. Hessell, W.B. Oswald, D.R. Burton and E.O. Saphire. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**:177 (2008).
53. F.A. Rey, F.X. Heinz, C. Mandl, C. Kunz and S.C. Harrison. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature* **375**:291 (1995).
54. Y. Modis, S. Ogata, D. Clements and S.C. Harrison. Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**:313 (2004).
55. J. Kadlec, S. Loureiro, N.G. Abrescia, D.I. Stuart and I.M. Jones. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. *Nat. Struct. Mol. Biol.* **15**:1024 (2008).
56. P.A. Bullough, F.M. Hughson, J.J. Skehel and D.C. Wiley. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**:37 (1994).
57. H.S. Yin, R.G. Paterson, X. Wen, R.A. Lamb and T.S. Jardetzky. Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc. Natl. Acad. Sci. USA* **102**:9288 (2005).
58. H.S. Yin, X. Wen, R.G. Paterson, R.A. Lamb and T.S. Jardetzky. Structure of the parainfluenza virus 5 F protein in its metastable. prefusion conformation. *Nature* **439**:38 (2006).
59. S.C. Harrison. Viral membrane fusion. *Nat. Struct. Mol. Biol.* **15**:690 (2008).
60. T.R. Wagenaar and B. Moss. Association of vaccinia virus fusion regulatory proteins with the multicomponent entry/fusion complex. *J. Virol.* **81**:6286 (2007).
61. T.A. Bowden, A.R. Aricescu, R.J. Gilbert, J.M. Grimes, E.Y. Jones and D.I. Stuart. Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. *Nat. Struct. Mol. Biol.* **15**:567 (2008).
62. H.W. Ackermann. Bacteriophage observations and evolution. *Res. Microbiol.* **154**:245 (2003).
63. J.D. Chappell, A.E. Prota, T.S. Dermody and T. Stehle. Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO J.* **21**:1 (2002).
64. D.I. Stuart, M. Levine, H. Muirhead and D.K. Stammers. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J. Mol. Biol.* **134**:109 (1979).
65. N. Riffel, K. Harlos, O. Iourin, Z. Rao, A. Kingsman, D. Stuart, and E. Fry. Atomic resolution structure of Moloney murine leukemia virus matrix protein and its relationship to other retroviral matrix proteins. *Structure* **10**:1627 (2002).
66. J. Felsenstein. PHYLIP — Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164 (1989).
67. X. Yan, Z. Yu, P. Zhang, A.J. Battisti, H.A. Holdaway, P.R. Chipman, C. Bajaj, M. Bergoin, M.G. Rossmann and T.S. Baker. The capsid proteins of a large, icosahedral dsDNA virus. *J. Mol. Biol.* **385**:1287 (2009).
68. J.M. Grimes, J. Jakana, M. Ghosh, A.K. Basak, P. Roy, W. Chiu, D.I. Stuart and B.V. Prasad. An atomic model of the outer layer of the bluetongue virus

core derived from X-ray crystallography and electron cryomicroscopy. *Structure* **5**:885 (1997).
69. C.L. Hill, T.F. Booth, B.V. Prasad, J.M. Grimes, P.P. Mertens, G.C. Sutton and D.I. Stuart. The structure of a cypovirus and the functional organization of dsRNA viruses. *Nat. Struct. Biol.* **6**:565 (1999).
70. Q. Ye, R.M. Krug and Y.J. Tao. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**:1078 (2006).

Chapter 3

Beyond Quasi-Equivalence: New Insights Into Viral Architecture via Affine Extended Symmetry Groups

Thomas Keef[*,‡] and Reidun Twarock[*,†,§]

Since the seminal work of Caspar and Klug on the structure of the protein containers that encapsulate and hence provide protection for the viral genome, it has been recognised that icosahedral symmetry is crucial for the prediction of viral architecture. In particular, their theory of quasi-equivalence invokes icosahedral symmetry to pinpoint the surface structures of viral capsids in terms of triangulations that schematically encode the locations of the protein subunits in the capsids. Whilst this approach is capable of predicting the relative locations of the proteins in the capsids, information on their tertiary structures and the organisation of the viral genome within the capsid are inaccessible. We present here a mathematical framework based on affine extensions of the icosahedral group that has been developed to predict a wide spectrum of features of the three-dimensional structure of simple viruses. This approach implies that the predictions of Caspar and Klug's quasi-equivalence theory are the consequences of a deeper level of structural organisation in viruses that orchestrates the full three-dimensional structure of simple viruses.

1. Introduction — Symmetry in Virus Architecture

Models for the structure and life cycle of viruses are invaluable tools in the analysis of the plethora of structural and functional data on viruses and provide profound insights in key mechanisms in virology. As observed by

[*]Department of Mathematics, University of York.
[†]Department of Biology, University of York, York YO10 5DD, U.K.
E-mails: [‡]tk506@york.ac.uk; [§]rt507@york. ac.uk

Crick and Watson in 1952,[1] the principle of genetic economy requires protein containers to be built from a small number of identical protein building blocks organised according to symmetry. Since the icosahedral group is the symmetry group with the largest number (60) of elements in three dimensions, layouts based on this symmetry optimise the ratio between capsid size and number of distinct building blocks. It is therefore not surprising that icosahedral symmetry occurs predominantly in viruses (for a review see Ref. 2). An energetic argument for the origin of icosahedral symmetry in viruses was moreover provided by Zandi *et al.*[3]

Caspar and Klug's seminal quasi-equivalence theory[4] is the first theory that makes use of icosahedral symmetry to predict virus structure. They argue that viral capsids formed from more than 60 identical copies of capsid protein should be organised according to triangulations, with proteins located in the corners of the triangular facets of the tessellation. In this way, they predict that viruses are formed from 12 clusters of five proteins (pentamers) and $10(T-1)$ clusters of six (hexamers), where T is called the triangulation number (in short T-number) and indicates the number of triangular facets per face of an inscribed icosahedron. T-numbers are established as a fundamental concept in virology, and are universally used for the analysis and classification of structural data on viruses.

However, experiments concerning the cancer–causing *Polyoma-* and *Papillomaviridae* have shown that there are viruses that do not follow the organisation predicted by this theory.[5,6] In particular, these viruses are formed from pentamers throughout (for example, 72 pentamers in the case of Human Papilloma Virus) and their surface structures can hence not be described *via* triangulations. It has been shown[7–9] that Caspar and Klug's quasi-equivalence theory can be generalised to incorporate these cases by allowing for more general types of tessellations. In particular, inspired by quasilattices as known from the study of quasicrystals,[10] i.e., alloys with atomic arrangements exhibiting long-range order but no periodicity,[11] as well as Penrose tilings,[12] it has been possible to derive surface lattices that model the structural organisation of *Polyoma-* and *Papillomaviridae*. This sparked a new approach, called Viral Tiling theory (VTT), in which viral capsids are modeled in terms of tilings[7,8] respecting icosahedral symmetry. VTT triggered various applications, including the construction of models for the assembly of protein containers from the capsid proteins,[13,14] descriptions of malformations during assembly,[15,16] models for special

bonding configurations (crosslinking structures)[17] and the analysis and prediction of vibrational motions of viral capsids.[18]

A common feature of Caspar and Klug's quasi-equivalence theory and its generalisations in terms of tilings is the fact that they all use icosahedral symmetry as the maximal symmetry content in the theory. Therefore they describe viral capsids schematically in terms of surface lattices, rather than as objects extended in space. As a consequence, the predictive power of these theories is limited to specifying the locations and types of the protein clusters in the capsids and their relative orientations. Information on the tertiary structures of the capsid proteins, the thickness of the capsid, the organisation of the genomic material or on the relative sizes of different particles in a family are by construction inaccessible in the framework of these theories.

A further generalisation was hence needed in order to predict finer details of viral architecture. Inspiration for this was provided by Janner's work which shows that lattice-like structures obtainable from a higher-dimensional lattice with icosahedral symmetry *via* projection can be used to model the layouts of proteins[19–22] or protein assemblies such as viral capsids.[23,24] Janner's theory is important because it shows that such lattice points indeed are the appropriate concept to model these biological structures. However, it is only descriptive as there is no mechanism in his theory to predict which subsets of lattice points of the higher dimensional lattice are relevant for virus architecture.

In order to derive such predictive information, the mechanisms of this organisation have to be better understood. Indeed, as we show here the answer is contained in the underlying symmetry group. As known from the theory of quasicrystals,[25] and nested carbon cage structures called carbon onions,[26] the symmetry of an extended structure exhibiting a certain symmetry (such as icosahedral symmetry) at different radial levels is encoded by an affine extension (or, affinisation) of that symmetry group. We therefore explored the use of affine extensions of the icosahedral group in the context of virology. An application of the standard procedure for such an affinisation showed that the tilings describing the surface structures of *Papilloma-* and *Polyomaviridae* in VTT were encoded by an affine extension of the icosahedral group[9] and that this fact could be used to predict the relative radii of the different protein containers that are observed during the *in vitro* self-assembly of the capsid proteins of viruses in these families.

It soon became clear that, even though powerful for the modeling of some viruses, the standard approach of affine extension based on root systems of Coxeter groups was too restrictive to describe all viruses. Keef and Twarock therefore identified the mathematical property of these affine extensions important for applications in virology, and then classified all possible affine extensions of the icosahedral group that fulfill it.[27,28] This resulted in a finite library of 3-D point-arrays that constitute geometric boundary conditions on real viral architectures and encode all possible layouts of simple viruses.[29] This approach provides, for the first time, predictive information on the structures of individual capsid proteins and the packaging arrangements of their genomes. A recent study of a number of test cases demonstrates that it predicts a wide spectrum of viral architectural features in unprecedented detail.[29]

In this review we start by outlining the various stages of generalising Caspar and Klug's quasi-equivalence theory that have led to these latest developments. We start by discussing Viral Tiling theory and highlight the enhanced predictive power gained by considering surface tessellations other than triangulations. We then introduce affine extensions of symmetry groups and explain how they can be used to obtain a library of layouts for simple viruses. Applications to a number of test viruses are discussed to highlight the versatility of this approach in predicting a wide spectrum of viral structural and dynamic features. We conclude with a discussion of why symmetry, and especially affine extended symmetry, plays such a fundamental role in virology.

2. The Surface Structures of Viral Capsids: Viral Tiling Theory

The theory of quasi-equivalence introduced by Caspar and Klug[4] provides a framework for the classification and prediction of viral capsid structures. In its strict interpretation, quasi-equivalence requires that the local environment around each protein is qualitatively the same: for example, zooming in on proteins organised according to a triangulation, one sees each protein located in a corner of a triangular facet, and hence in an environment of "similar type". A triangulation is hence one way of realising the concept of quasi-equivalence. A classification of triangulations compatible with icosahedral symmetry has led to a classification of virus structures in

terms of T-numbers, with viral capsids predicted to be composed of 12 pentamers and $10(T-1)$ hexamers.

Viral Tiling theory recognises that there exist other tessellations than triangulations with the property of respecting icosahedral symmetry. For example, all tessellations in terms of one type of building block are quasi-equivalent in the above sense. As we discuss below, these tessellations usually predict the same numbers and types of capsomeres as the triangulations in Caspar and Klug's approach; however, the predictions on the relative orientations of the capsomeres are different. This has important implications for the building blocks of assembly as we discuss below. Moreover, Viral Tiling theory also permits tessellations in terms of more than one building block, which are no longer quasi-equivalent in its strict sense, as long as they respect icosahedral symmetry. The capsomere types predicted by these approaches do not exhibit the characteristic distribution of 12 pentamers and $10(T-1)$ hexamers known from Caspar-Klug theory. Such tessellations are in particular relevant for the modeling of the surface structures of the cancer-causing *Papillomaviridae* as we discuss in the second subsection below.

2.1. *Quasi-Equivalent Tessellations Beyond Triangulations*

For viruses with capsids exhibiting the correct number of pentamers and hexamers according to Caspar–Klug theory, VTT distinguishes between different scenarios for the relative orientations of these capsomeres. This is due to the fact that quasi-equivalent structures can be modeled in terms of different types of tessellations: besides the triangulations used in Caspar–Klug theory, tessellations in terms of other building blocks such as kite and rhombic shapes, are possible. This is illustrated in Fig. 1 for the case of a $T = 3$ capsid, where in each case, two of the 20 triangular faces are shown that are needed to build an icosahedron. Besides the triangulation known from Caspar-Klug theory, which is shown in Fig. 1(a) superimposed in red on two large black triangular faces, tessellations in terms of rhombs (Fig. 1(b)) or kites (Fig. 1(c)) are possible. Superimposing the implied locations of the proteins schematically on the tiles as discs, we see that all three tessellations encode a hexagonal cluster (here shown in red/green) in the middle of the triangular faces.

Note, however, that the relative orientations of these hexagonal clusters with respect to the pentagonal clusters at the 5-fold axes (blue circles

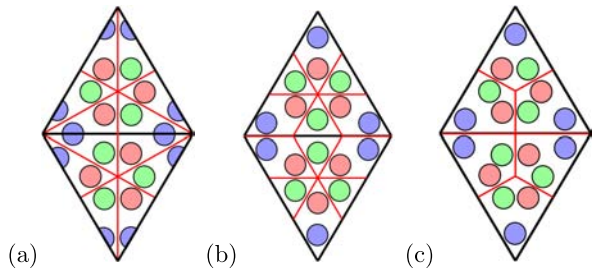


Fig. 1. Different types of quasi-equivalent tessellations for a $T = 3$ viral capsid: (a) a triangulation, (b) a rhombic tiling (c) a kite tiling. Note that the predictions on the relative orientations of the capsomeres are different in each case.

represent proteins forming part of these pentamers) differ in these three cases: In the triangulation two proteins in a hexamer face two proteins in a pentamer, whilst in the rhombic tessellation one protein in a hexamer faces one protein in a pentamer, and in the kite tessellation two proteins in a hexamer face one protein in a pentamer. The three tessellations hence encode different information on the relative orientations of the capsomeres. Examples of $T = 3$ viruses following these different layouts in the organisation of their capsid proteins are discussed in Subsection 4.2: Pariacoto virus is organised according to the triangulation in (a), Bacteriophage MS2 according to the rhombic tiling in (b) and Seneca Valley Virus according to the kite tiling in (c).

This information on the relative orientations of the capsmeres is important as it provides clues on the dominant bonds between proteins or the types of clusters from which the particle is formed: A rhombic tiling is most likely formed from protein dimers, whilst viral capsid following one of the other two types are more likely to be formed from either trimers, or, pentamers and hexamers as their units of assembly. The spectrum of possibilities of capsomere orientations, and hence clues on the dominant bonds and building blocks of assembly, follow from a classification of the different types of quasi-equivalent tessellations in VTT.

2.2. *All-Pentamer Capsids in Viral Tiling Theory*

As discussed earlier, it is moreover possible to relax quasi-equivalence in its strict sense and allow also for tessellations in terms of more than one different type of building block, as long as these tessellations respect

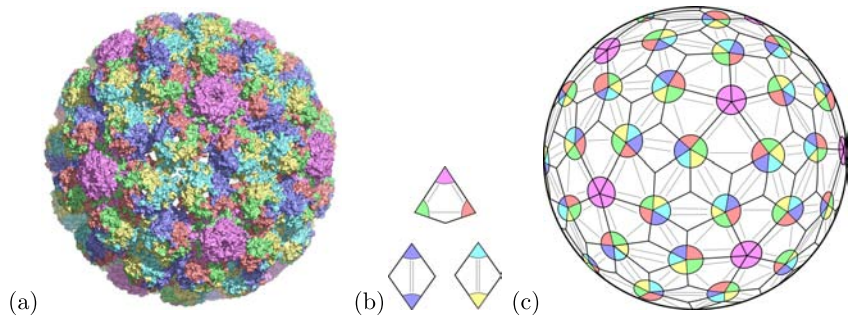


Fig. 2. (a) The crystal structure of SV40, a virus with only pentameric capsomeres. (b) The kite tile and the two rhomb tiles used in VTT to describe the surface structure of SV40. The coloured corners schematically indicate the locations of the protein subunits (colour coding matching that in (a)). The grey lines schematically indicate the locations of the C-terminal arm extensions, and hence the bonds between the protein subunits. (c) The tiling shown as a spherical tiling, with coloured discs at the tile vertices indicating the locations and orientations of the pentameric capsomeres.

icosahedral symmetry. This is important for viruses that do not follow the characteristic pattern of 12 pentamers and $10(T-1)$ hexamers of Caspar-Klug theory, such as for example viruses in the families of *Polyoma-* and *Papillomaviridae*,[5,6] which have a structure entirely composed of pentamers. Note that the locations of the capsomeres are the same as predicted by a $T = 7d$ (triangular) Caspar–Klug lattice, but a triangulation is not appropriate to describe the types of the capsomeres (see Fig. 2(a)).

The inspiration for the construction of tessellations with more than one building block came from Penrose tilings[10,12] which embody quasilattices, that have local five-fold symmetries and long-range order but lack periodicity. Figure 2(a) shows the crystal structure of SV40 (PDB-ID 1sva) and the tiling associated with it.[7,8] The tiling is composed of three tiles, these being a kite shaped tile and two rhombic shaped tiles, which are shown in Fig. 2(b) with locations of proteins indicated schematically in the same colour scheme as in Fig. 2(a). Besides the locations of the proteins (shown schematically in the corners of the tiles), tiles provide clues on the interactions between the proteins represented by them. In particular, the kite tile indicates that there is a trimer interaction between the three proteins represented by this tile, whilst the rhombic tiles indicate a dimer interactions between its two proteins. In the biological structure they manifest

themselves *via* C-terminal arm extensions, and their locations are indicated schematically by grey lines on the tiles.

3. Generalisation of the Symmetry Group via Affine Extension

The approaches discussed above model viral capsids in terms of surface lattices that provide information on the locations of the capsid proteins. Since surface lattices are two-dimensional objects, details about the three-dimensional structure of the proteins such as their tertiary structures or of the organisation of the viral genome within the particle can not be predicted with these approaches. This is owing to the use of symmetry: icosahedral symmetry encodes how objects (such as capsid proteins) at a fixed distance from the center of the virus are mapped to other objects at the same distance *via* rotations about five-, three- and two-fold symmetry axes. However, it does not contain any information on how these objects are organised at different radial levels, or how they relate to other structures within the particle. Such information can be derived *via* an extension of the symmetry group by an operation which no longer preserves distance from the center of the object, such as a translation. This addition of a further (non-compact) generator to a symmetry group is known as affine extension, or affinisation, of the symmetry group.

For some symmetry groups, called crystallographic symmetry groups, it is possible to find affine extensions which generate a lattice with the same symmetry as the original object. We demonstrate this for the example of hexagonal symmetry in Fig. 3(a). For objects which do not have this crystallographic property it is not possible to obtain a lattice, regardless of what translation is chosen. However knowledge of what makes translations suitable in the crystallographic case has led to a new technique which produces affine extensions with lattice-like properties for noncrystallographic symmetry groups.[27] An example of such a noncrystallographic setting is shown in Fig. 3(b) based on a pentagon embodying five-fold symmetry. As for the case of the hexagon in Fig. 3(a) the translation length T has been chosen to be the same as the radial distance to the vertices of the original object. In both cases the original object is left unchanged by a rotation, R_H and R_P for the hexagon and pentagon respectively. The vertices have been numbered H_i and P_i, $i \in \mathbb{Z}$. If the vector from the center

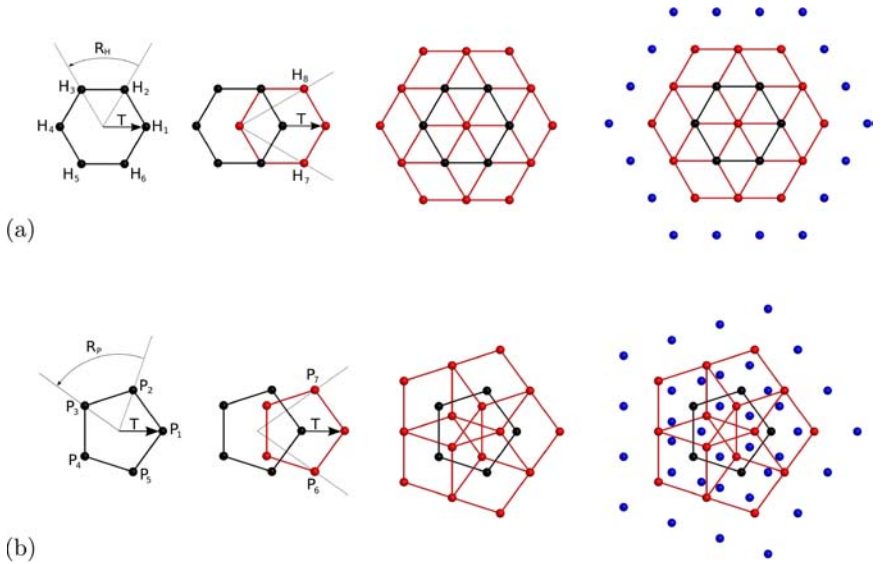


Fig. 3. The construction of lattice and quasilattice structures *via* the repeated action of generators of affine extended symmetry groups. (a) The case of hexagonal symmetry as an example of a crystallographic symmetry. (b) The case of pentagonal symmetry as an example of a noncrystallographic symmetry. In both cases, the points obtained after the first translation are shown in red, and the new points created after a second translation are coloured blue.

of the object to any of its vertices is denoted $\vec{H}_i$ (or $\vec{P}_i$) the rotation R_H (or R_P) will map each vertex onto another, i.e., $\vec{H}_3 = R_H \vec{H}_2$. In each case, after translation, it is possible to rotate one vertex in the translated object to another using the rotation associated with the original object, for example $\vec{H}_8 = R_H \vec{H}_7$. Therefore, there are points that are produced under different combinations of generators of the affine extended group. This is the property required to ensure the lattice-like property of the affine extension. The new vertices in Fig. 3 have been coloured according to the lowest number of translations required to obtain a vertex in that position *via* the affine extended symmetry group: red indicates the points created by the first translation and blue the points created by the second translation.

There is usually more than one translation for the affine extension of a symmetry group, which produce different lattice-like configurations. Two such translations are shown in Fig. 4 for the case of pentagonal symmetry.

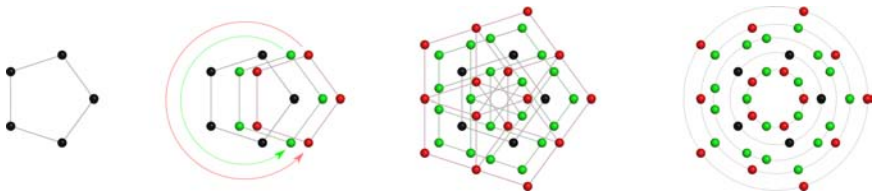


Fig. 4. Two translations resulting in different lattice-like point sets in the case of pentagonal symmetry. The green points originate from a translation of $\frac{1}{\tau}$ times the length of a vector pointing from the center of the pentagon to one of its vertices. The red points stem from a translation of the length of a vector pointing to a pentagonal vertex.

Green points are generated by a translation of length $\frac{1}{\tau}T$ where

$$\tau = 0.5 * (1 + \sqrt{5}) \approx 1.618$$

(the golden ratio) and T is the length of the vector from the center of the pentagon to any of the pentagonal vertices, whilst red points are created by a translation of length T.

In analogy to this, Ref. 27 provides a classification of the affine translations leading to lattice-like structures with icosahedral symmetry. Instead of the pentagon in the above example, the icosahedron, the dodecahedron and the icosidodecahedron displayed in Fig. 5 are used in this construction. This is because vectors pointing to the vertices of these polyhedra correspond to projections of the bases of the three cubic lattices in six-dimensions, which is the smallest dimension containing lattices with icosahedral symmetry and which are hence the most generic objects

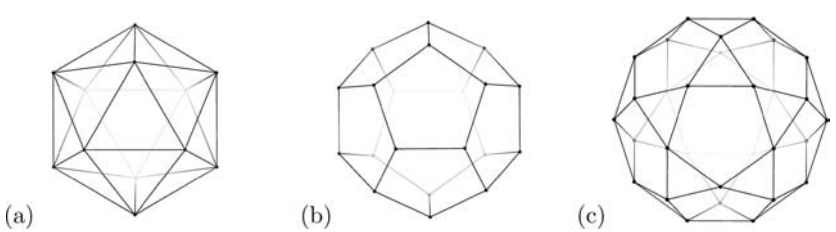


Fig. 5. The icosahedron (a), dodecahedron (b) and icosidodecahedron (c) that have been used in the construction of affine extensions of the icosahedral group.

for producing lattice-like structures with icosahedral symmetry in three dimensions.

As before, for each translation operator one obtains, *via* an iterative application of the original symmetry group and this translation to any vertex of the start configuration, a unique point array. A classification of translation operators leading to affine extensions of the icosahedral group hence results in a collection of point arrays that encode all possible ways of organising structures with icosahedral symmetry in three dimensions. We call the collection of these point arrays the library of point sets. Three examples of point arrays in this library are shown in Fig. 6 to demonstrate their differences in both structure and cardinality (number of points).

4. Applications to Viruses

In this section we explore how the additional information encoded by the affine extensions of the icosahedral group can be used to make predictions beyond those of quasi-equivalence. We will start by discussing how the different particle sizes observed in the polymorphic assembly of *Papilloma-* and *Polyomaviridae* can be inferred *via* this approach. We then show based on a number of test cases that the genome organisation in simple viral particles, as well as features of the tertiary structures of the proteins, can be predicted in striking detail. We then discuss implications for viral dynamics, and conclude by showing that this approach is a generic tool that can be applied also to general protein assemblies with symmetry outside virology.

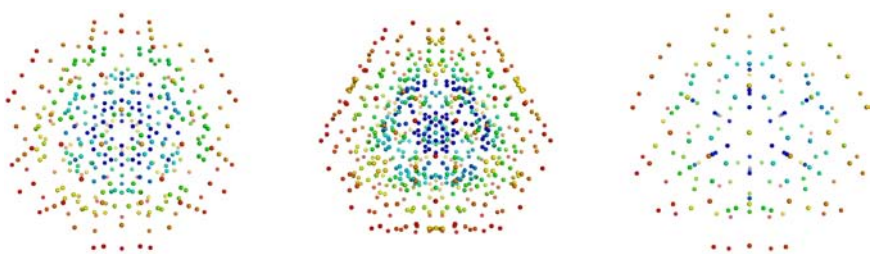


Fig. 6. Three point arrays in the library of point sets obtained by different affine extensions of icosahedral symmetry. Points have been coloured radially from blue (at the center) to red.

4.1. *Prediction of Particle Sizes in Assembly Polymorphism*

The major capsid proteins of Simian Virus 40 and Polyomavirus are able to assemble polymorphic particles *in vitro*,[30,31] which are either spherical or tubular. For the spherical particles, three sizes of particles are known to occur: a small and a large particle, each with icosahedral symmetry, and a medium-sized particle with octahedral symmetry. We have shown in Sec. 2 that the surface lattice of each individual particle can be modeled in the framework of Viral Tiling theory, and the corresponding tilings are displayed in Fig. 7(a–c).

In order to predict the relative sizes of these particles, an affine extension of the icosahedral group has been used to generate a point array that contains all these tilings simultaneously.[9] Since the vertices of all three tilings are subsets of this same point array, the sizes of the three different types of particles are fixed in relation to each other. Figure 7(d) shows the tilings in the radial distribution implied by this approach. There is hence only one free parameter that relates the overall mathematical structure, i.e., relative radii of the particles collectively, to the biological setting. This factor has been computed and has been shown to be in excellent agreement with the experimental data.[9,30,31]

4.2. *Prediction of Genome Organisation*

Affine extensions of icosahedral symmetry have moreover been used to predict features of the internal organisation of virus particles based on the

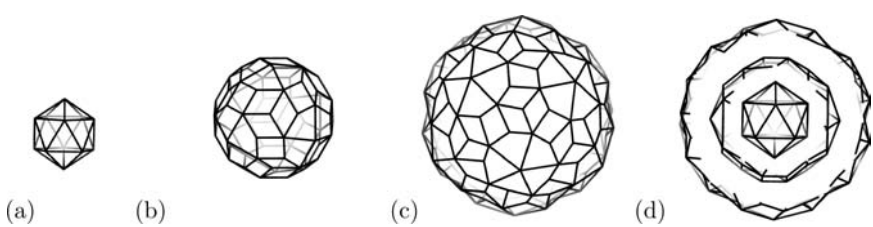


Fig. 7. (a)–(c) The tilings representing the three spherical particles observed in polymorphic assembly of Simian Virus 40 and Polyomavirus. (d) All three tilings shown superimposed in the radial distribution implied by an affine extension of the icosahedral group.

structure of their capsid proteins. In Ref. 29 an automated algorithm has been developed that searches through each element of the library of point sets implied by the affine extended symmetry group, and determines the point array which best represents the viral capsid proteins. In particular, the algorithm first locates the most prominent external features of the virus, and determines the point array which, when scaled correctly, has external points as close to these as possible. A root mean square deviation (RMSD) analysis between the atoms of the capsid proteins and the point arrays is then carried out and the point set with the best combined score of distance to prominent external features and RMSD produces a ranking of best fit for the point arrays. Since only information about points in the array overlapping with the capsid area is used to determine the best-fit point array, internal points are hence a prediction of the theory. In the following, we demonstrate the predictive power of this approach based on a number of test cases.

We start by considering bacteriophage MS2,[32,33] a $T = 3$ virus with 180 capsid proteins assembled as 90 non-covalent dimers around a single-stranded RNA (ssRNA) genome. The automated algorithm determined the best-fit point array in the library for this particle based on its crystal structure (PDB-ID 1zdh, see Fig. 8(a)). The best-fit point array (combined score 7.14; next best score 7.46 — note that we are not comparing like with like as in crystallography applications and that values are therefore higher than RMSDs in those situations) is shown superimposed on the crystal structure in Fig. 8(b). Note that its outermost points are closely located to the (flexible) N-terminal β-hairpins of the capsid proteins.

Toropova *et al.* have published a $\sim$9 Å Cryo-electron microscopy (cryo-EM) map which shows that most of the genome is organised in two concentric shells of density that are connected at the five-fold axes by columns of density consistent with an A-type RNA duplex structure.[33] The points of the best-fit array have been plotted against these data in Fig. 8(c). Observe that the points provide a layout for the organisation of the protein container as well as the RNA organisation within the particle. The implication is that affine-extended icosahedral symmetry restricts the size, shape and placement of all viral components in this phage.

Note that both X-ray and cryo-EM densities were icosahedrally averaged during structure determination, so it is not surprising that the viral molecular components exhibit this symmetry. However, such

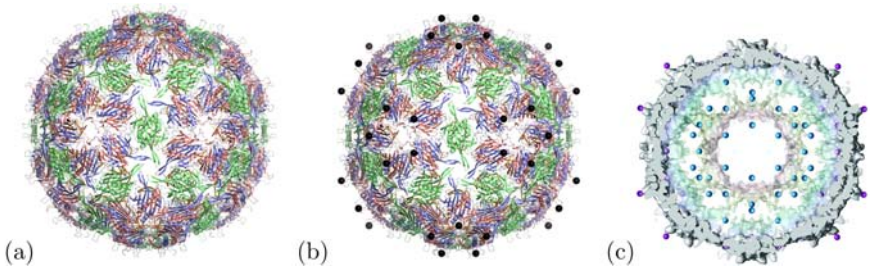


Fig. 8. (a) The crystal structure of bacteriophage MS2 (PDB-ID 1zdh). (b) The locations of the outermost points of the best-fit point array are located at the β-hairpins of the capsid proteins, which are the radially most distant features of the MS2 particle. (c) A cross-section view along a two-fold symmetry axis demonstrates that the point array embodying the affine extended symmetry provides a layout for the organisation of material boundaries in bacteriophage MS2.

averaging places no constraints whatsoever on the radial position of molecular components and does not imply in which form icosahedral symmetry is realised at different radial levels. The symmetry approach, which accurately determines both the radial distribution and the structural features of the packaged genome based only on the crystal structure of the protein capsid, indeed predicts details of genome organisation that could not be inferred *via* previous approaches.

The second test case is Pariacotovirus, a $T = 3$ insect virus with capsid proteins organised as trimers around an ssRNA genome, see Fig. 9(a) which is based on PDB-ID 1f8v. We have chosen this virus as an example because $\approx$35% of its genome is organised as a dodecahedral duplex RNA cage for which high resolution data is available,[34] hence allowing us to probe if our approach can predict finer details of this organisation. Once again our automated algorithm determined one best-fit point array (combined score 6.16; next best 6.19), which is shown in Fig. 9(b) superimposed on the biological data. The point array has been scaled to the top of the local three-fold axes at the centers of the protein trimers. There are further points in the array located on the boundaries between capsid proteins, as well as points delimiting the shape, size and orientation of the dodecahedral RNA cage (Fig. 9(c)). Note that since only the points overlapping with the protein container have been used to determine the best-fit array, all internal points are predictions of our formalism. For example, the points located at the three-way junctions of the

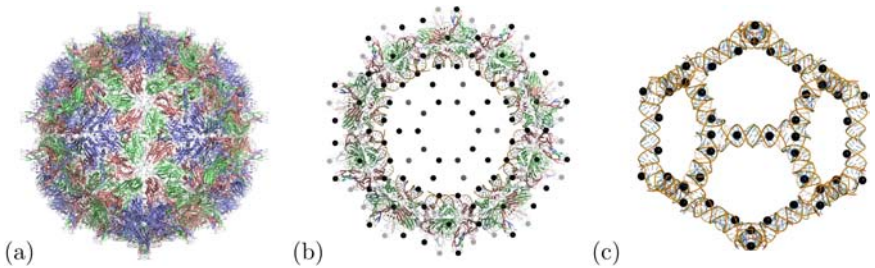


Fig. 9. (a) The crystal structure of Pariacoto virus (PDB-ID 1f8v). (b) The point array superimposed on the crystal structure in a cross-section viewed along a five-fold symmetry axis. The outermost points of the array are located on top of the protruding towers of protein at the local three-fold symmetry axes. The points predicted by the affine extended symmetry are located on material boundaries and encode the layout of the particle. Only the RNA contained in the dodecahedral cage in (c) is shown, and internal points predict the organisation of this RNA. (c) A close-up view of the RNA cage. The points predicted by affine extended symmetry are located in the minor grooves of the RNA duplex.

dodecahedral RNA container and the points in the minor grooves of the A-duplex structures in the modeled density (see Fig. 9(c)) are a predictions of our approach. Since the conformation and dimensions of an A-type RNA duplex are fixed in nature, this striking match between point array and data implies that there is an inherent molecular scaling in viruses that is rooted in affine extended symmetry. This has not been appreciated previously.

Given this remarkable result for Pariacoto virus, it is surprising that in most viruses no clear internal features have been observed *via* cryo-EM. In order to provide an explanation for why this might be the case, we next discuss Seneca Valley Virus (SVV). SVV is a member of the *Picornaviridae* family of major human and animal pathogens, that also include Poliovirus and Foot and Mouth Disease Virus. A characteristic feature of this particle is that part of its RNA is organised in a cage close to the protein container has recently been resolved *via* X-ray crystallography,[35] whilst the structure of the remaining RNA further inside the particle could not be determined to the same resolution. For this virus, our best-fit algorithm returned nine best-fit point arrays (combined score 3.19; next best 3.3).[29] An inspection of these arrays shows that they have the same RMSD value because they have identical points in the area of the protein container

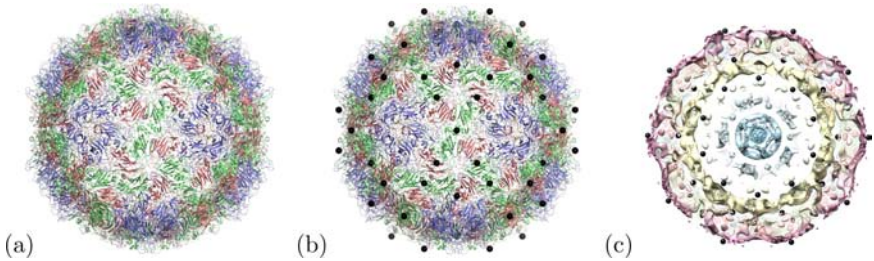


Fig. 10. (a) The crystal structure of Seneca Valley Virus (PDB-ID 3cji). (b) The locations of the outermost points of the best-fit point array. (c) The points common to all best-fit point arrays superimposed on cryo-EM data provided by Vijay Reddy in a cross sectional view along a five-fold axis of symmetry.

based on which the RMSD analysis is performed. From a symmetry point of view, this implies that there are nine equally probable configurations for the genomic material given the structure of the protein container of this particle. A closer inspection of these nine point arrays shows that also the first layer of points internal to the points describing the protein container is common to all nine best-fit point arrays. This layer of points corresponds to the RNA shell resolved in Ref. 35. At all lower radial levels the point sets differ and it is therefore perhaps not surprising that internal features of individual particles could not be resolved *via* experimental techniques relying on icosahedral averaging. In order to see whether this phenomenon also occurs in other viruses in this family, we have moreover analysed Polio virus (data not shown). Also for this virus we found a number of best-fit point arrays that differ only in the area occupied by the genomic material, suggesting perhaps that also in other viruses the lack of information on genome organisation may be due to the heterogeneity of the particles.

4.3. *Predictions of Protein Structure*

In the previous subsection we have discussed how affine extended symmetries can be used to predict the organisation of genomic material in viral particles. Since the symmetry principle constrains the full three-dimensional structure of the virus, it also imposes constraints on the tertiary structures of the capsid proteins themselves. In order to

demonstrate this, we use Tomato Bushy Stunt virus (TBSV) as a test case. TBSV, a member of the *Tombusviridae* affecting tomato and pepper plants, was the first icosahedral virus resolved by X-ray crystallography.[36] We probe here whether the symmetry recognisees the three domain architecture of its capsid proteins: two domains in the capsid area resolved by crystallography, and one in the interior of the capsid interacting with the viral RNA.

The X-ray crystal data (PDB-ID 1tbv) has been analyzed *via* our automated algorithm. The combined scores of best-fit array (6.91, shown in cyan) and second-best-fit array (7.01, magenta) are similar, because both sets share a significant number of points. Both point sets are shown superimposed on the X-ray crystal data in Fig. 11(b), and Fig. 11(b–f) highlight differences between them. Note that the point array shown in magenta has points located at the hinge between the P and S domains of the capsid proteins over the global two-fold vertices, and the point array shown in cyan has points located on the hinges of the remaining proteins in the viral capsid.

The two point sets predict different internal radii for the viral RNA, one with an internal radius of 33 Å (magenta) and another with radius 54 Å (cyan). We compare these predictions with values given in experimental papers for two members of the genus Tombusvirus, TBSV[37] and Cucumber Necrosis Virus (CNV),[38] a virus with a very similar protein structure to TBSV. These papers indicate for TBSV, density (suggested to be protein) at a radial distribution between ≈50–95 Å and some internal density with an internal radius of about 30 Å. Moreover, for CNV there are two shells of density internal to the capsid protein, one between 50 and 72 Å and one on the underside of the capsid proteins. These observations agree with the radii of the internal points in the two point sets, especially since the points at radial level 54 Å are contained in both sets. This suggests that the predictions of affine extended symmetry are once again in good agreement with experimental results.

4.4. *Implications for Viral Dynamics*

The ultimate goal of our symmetry analysis is to use the information encoded by the affine extended symmetries to understand virus dynamics. In particular, we are interested in modeling the maturation or swelling of viral particles. As an example, we discuss here Cowpea

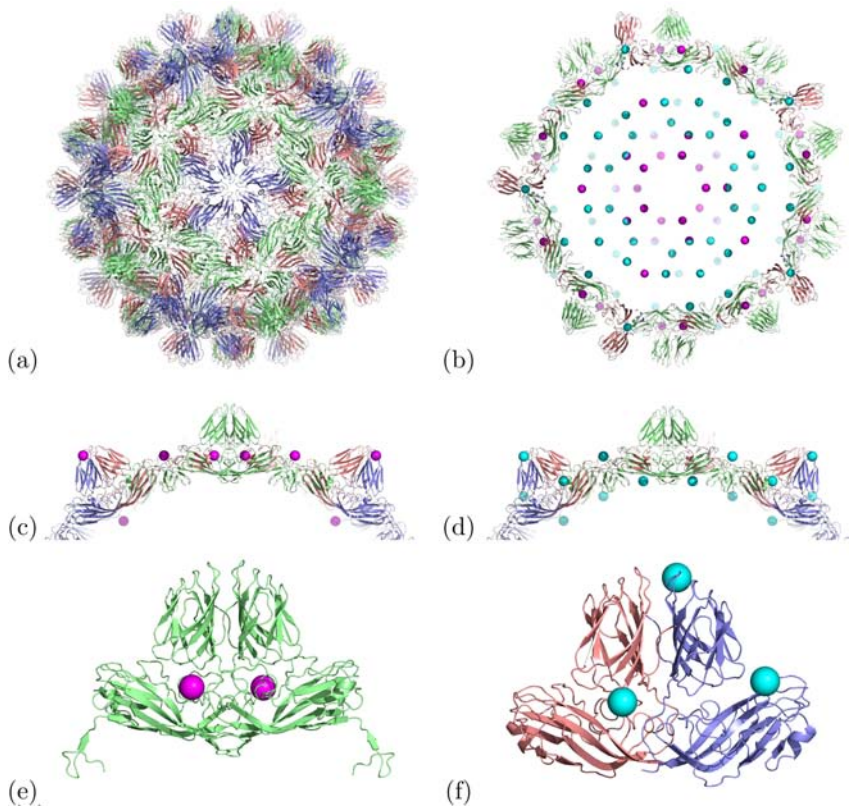


Fig. 11. (a) The crystal structure of Tomato Bushy Stunt virus (PDB-ID 1tbv). (b) The two point sets with the best combined score are shown in cyan and magenta in a five-fold view. (c), (d): Two portions of the protein container shown in magnification: The magenta points are located at the hinge between the P and S domains of the capsid proteins over the global two-fold axes and the cyan points highlight the domain boundaries in the remaining capsid proteins. (e), (f): Zooming in even further on individual capsid proteins to demonstrate the locations of the points with respect to domain boundaries.

Chlorotic Mottle virus (CCMV), a plant virus in the *Bromoviridae* family that can undergo a reversible pH and metal ion dependent structural transition.[39,40] During its swelling transition, CCMV expands by about 10% predominantly around the three-fold axis. The X-ray crystal structure for the native form (PDB-ID 1cwp) and also a model of the swollen form (VIPER-ID ccmv-swln-1) have been deposited on the VIPERdb

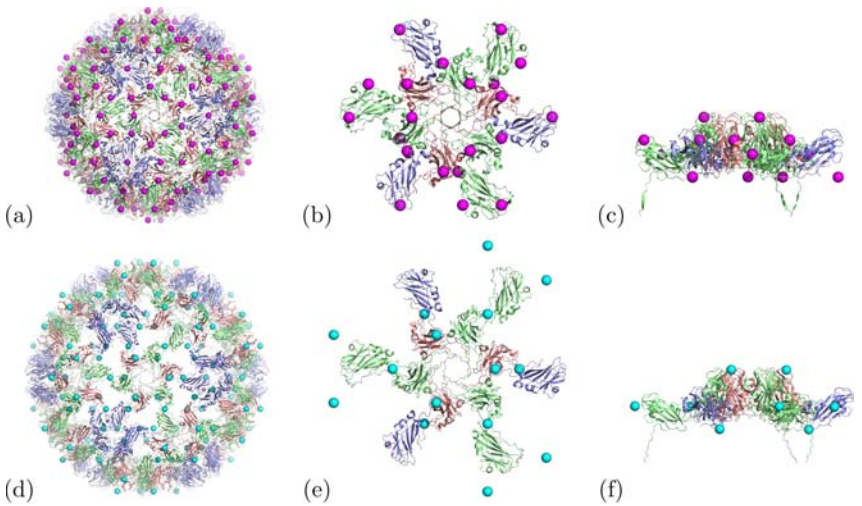


Fig. 12. Cowpea Chlorotic Mottle virus is a $T = 3$ virus that undergoes a swelling transformation. (a–c) shows the best-fit point array for the particle in native form, with (b) and (c) zooming in on a hexamer in a top and side view respectively. (d–f) shows the best-fit point array for the particle in native form, with (b) and (c) close-up views of the same hexamer in its expanded form.

website.[41] Figures 12(a),(d) show the point arrays obtained *via* our best-fit algorithm superimposed on the native (combined score 7.37; next best 7.48) and swollen (combined score 5.11; next best 5.44) forms of CCMV. The point sets each match to the top and bottom of the protein shell in either the native or the swollen form. By zooming in on a hexamer in Fig. 12(b–c), (e–f), it becomes apparent how the different affine extended symmetries are realised by the proteins in each case. The fact that the affine extended symmetry describes both start and end state of the structural transition well implies that both correspond to local energy minima in a free energy landscape. Such information can be used in combination with further energetic arguments (work in progress) to understand better the transition between the native and the expanded form of CCMV.

5. Beyond Virology: Applications to Protein Assemblies with Symmetry

Affine extensions of noncrystallographic symmetry groups are generic tools that can also be applied in the context of other protein assemblies.

As an example, we discuss here our results for the GroEL–GroES chaperonin complex. The chaperonin protein GroEL forms a dual ringed tetradecameric complex, with both the *cis* and *trans* ring consisting of seven subunits.[42] The cochaperonin protein complex GroES is a heptameric protein assembly that acts as a lid enabling GroEL to perform correctly, mediating protein folding within a bacterial cell. Here we work with PDB-ID 2c7c for the GroEL–ATP_7–GroES complex. Since the GroEL–GroES complex has axial seven-fold radial symmetry, which like five-fold symmetry, is noncrystallographic, it has been possible to investigate the organisation of proteins along this symmetry axis using a similar technique to the one used for viral capsids. Since the GroEL–GroES complex only has one seven-fold symmetry axis, affine extensions of this symmetry only provide information about individual planar slices perpendicular to this symmetry axis. In analogy to the case of five-fold symmetry discussed earlier, a library of planar point sets has been derived and the theory predicts that each slice along the length of the molecule and perpendicular to its symmetry axis follows one of the layouts encoded by this library. To test this, an automated algorithm has been created that matches different slices (b–k in Fig. 13(a)) to the library of point sets. The best-fit point arrays are shown superimposed on the biological data in Figs. 13(b–k). Each of these point sets provides a template for the structural organisation of the individual proteins on that slice, implying that the protein complex is organised according to the layouts encoded by affine extensions of seven-fold symmetry. This suggests that affine extensions of noncrystallgraphic symmetries are of interest beyond virology and provide valuable insights into symmetric protein assemblies in a general context.

6. Concluding Remarks — Why is Symmetry Fundamental in Virology?

In previous sections we have provided evidence for the fact that viruses follow the layouts implied by affine extensions of the icosahedral group. This begs the question: "Why does symmetry play this fundamental role in virology?"

It has long been recognised that icosahedral symmetry is important in virus architecture because of the concept of genetic economy: Structures

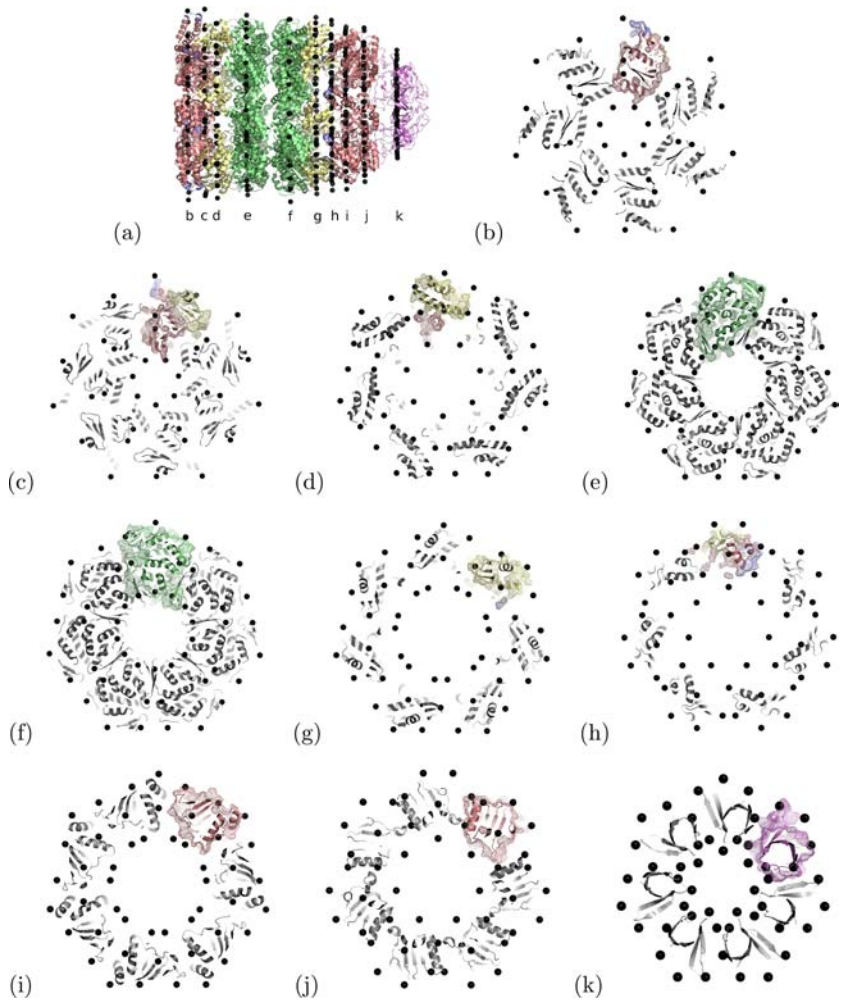


Fig. 13. The GroEL–GroES chaperonin protein complex is composed of 21 proteins and has axial seven-fold symmetry. Affine extensions of seven-fold symmetry encode the layouts of the proteins in planar slices perpendicular to the symmetry axis. (a) shows the locations of the slices, labeled b–k, and (b–k) display the best-fit point array for each of them.

organised according to symmetry can be built from a relatively small number of different building blocks. This makes the coding for building blocks, as well as the assembly of the particles from these building blocks, very efficient. Here we carry this argument one step further and argue that,

similarly, the templates provided by affine extensions of icosahedral symmetry provide a guiding principle for assembly of the entire viral particle and hence enable more efficient assembly, especially in the cases where genomic material and proteins co-assemble.

Moreover, energetic arguments related to the surface lattices encoding capsid organisation, have shown that capsid structures with icosahedral symmetry correspond to energy minima in a free energy landscape.[3] Similarly, we argue that the structures of viruses obeying the layouts implied by affine extensions of this symmetry correspond to local minima in a free energy landscape. Since such minima correspond to stable configurations, this implies that viruses maximise stability by maximising symmetry. This is plausible given that stability is important for the virus when transporting its genetic material between hosts.

Of course, viruses may deviate from their maximal symmetry configuration at an energy cost. This occurs, for example, when viruses undergo structural transitions from one symmetric configuration to another, such as in the example of swelling in CCMV discussed earlier. Such transitions are important for virus function, for example during infection. However, the fact that the start and the end state of such transitions correspond to energy minima implies that the results of these transitions (i.e., the end states) are ensured, as the virus is certain to settle into the energetically minimal configuration at the end of the transition. The use of symmetry in the full three-dimensional structure of the virus, i.e., the use of an affine extended symmetry, is hence a smart way of ensuring the desired functional outcome of a dynamic event.

This has major implications for viral evolution. Viral protein folds are striking in their topological conservation despite very low levels of primary sequence identity. This has been seen as a consequence of divergent evolution.[43,44] However, the inherent molecular scaling implied by the 3-D symmetry principle suggests it may actually be a consequence of the need to maximise symmetry throughout a viral particle as affine extended symmetry would severely restrict the folds that could be adopted by the capsid proteins. The affine extended symmetry may hence act as selective pressure during virus evolution. This has consequences also for our understanding of the variation of immuno-dominant viral epitopes and immune neutralisation, and may hence open up novel ways of anti-viral drug design.

Acknowledgements

We would like to thank Neil Ranson and Jess Wardman for help with the figures. Moreover, funding provided by a Leverhulme Trust Research Leadership Award to RT is gratefully acknowledged.

References

1. F.H.C. Crick and J.D. Watson. The structure of small viruses. *Nature* **177**:473–475 (1956).
2. S. Casjens. *Virus Structure and Assembly*. Jones & Bartlett, Boston, Mass. (1985).
3. R. Zandi, D. Reguera, R.F. Bruinsma, W.M. Gelbart and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA* **101**(44):15556–560 (2004).
4. D.L.D. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**:1–24 (1962).
5. I. Rayment, T. Baker, D.L.D. Caspar and W. T. Murakami. Polyoma virus capsid structure at 22.5 Å resolution. *Nature* **295**:110–115 (1982).
6. R.C. Liddington, Y. Yan, J. Moulai, R. Sahli, T.L. Benjamin and S.C. Harrison. Structure of Simian Virus 40 at 3.8-Å resolution. *Nature* **354**:278–284 (1991).
7. R. Twarock. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **226**(4):477–482 (2004).
8. R. Twarock. The architecture of viral capsids based on tiling theory. *J. Theor. Med.* **6**:87–90 (2005).
9. T. Keef, R. Twarock and K.M. ElSawy. A new series of polyhedra as blueprints for viral capsids in the family of *Papovaviridae*. *J. Theor. Biol.* **253**(4):808–816 (2008).
10. M. Senechal. *Quasicrystals and Geometry*. Cam. Univ. Press (1996).
11. D. Shechtman, I. Blech, D. Gratias and J. Cahn. Metallic phase with long-range order and no translational symmetry. *Phys. Rev. Lett.* **53**:1951–1953 (1984).
12. P. Penrose. The role of aesthetics in pure and applied mathematical research. *Bull. Inst. Math. Appl.* **10**:266–271 (1974).
13. T. Keef, A. Taormina and R. Twarock. Assembly models for papovaviridae based on tiling theory. *J. Phys. Biol.* **2**:175–188 (2005).
14. T. Keef, C. Micheletti and R. Twarock. Master equation approach to viral capsid assembly. *J. Theor. Biol.* **242**(3): 713–721 (2006).
15. R. Twarock. Mathematical models for tubular structures in the family of *Papovaviridae Bull. Math. Biol.* **68**(5): 973–987 (2005).
16. T. Keef, A. Taormina and R. Twarock. Classification of capped tubular viral particles in the family of *Papovaviridae*. *J. Phys. Condens. Matter* **18**:S375–S378 (2006).
17. R. Twarock and R.W. Hendrix. Crosslinking in viral capsids *via* tiling theory. *J. Theor. Biol.* **240**:419–424 (2006).

18. K. ElSawy, A. Taormina, R. Twarock and L. Vaughan. Dynamical implications of viral tiling theory. *J. Theor. Biol.* **252**(2):357–369 (2008).
19. A. Janner. Strongly correlated structure of axial-symmetric proteins. I. orthorhombic, tetragonal, trigonal and hexagonal symmetries. *Acta Crystallographica D61* **62**(5):247–255 (2005).
20. A. Janner. Strongly correlated structure of axial-symmetric proteins. II. pentagonal, heptagonal, octagonal, nonagonal and ondecagonal symmetries. *Acta Crystallographica D61* **62**(5):256–268 (2005).
21. A. Janner. Strongly correlated structure of axial-symmetric proteins. III. complexes with DNA/RNA. *Acta Crystallographica D61* **62**(5):269–277 (2005).
22. A. Janner. Polygrammal symmetries in biomacromolecules: Heptagonal poly d(As^4T).poly d(As^4T) and heptameric α-Hemolysin. *Struct. Chem.* **13**:277–287 (2002).
23. A. Janner. Crystallographic structural organisation of human Rhinovirus serotype 16, 14, 3, 2 and 1a. *Acta Crystallographica A* **62**:270–286 (2006).
24. A. Janner. Towards a classification of icosahedral viruses in terms of indexed polyhedra. *Acta Crystallographica A* **62**(5):319–330 (2006).
25. J. Patera and R. Twarock. Affine extensions of noncrystallographic Coxeter groups and quasicrystals. *J. Phys. A* **35**:1551–1574 (2002).
26. R. Twarock. New group structures for carbon onions and carbon nanotubes via affine extensions of noncrystallographic Coxeter groups. *Phys. Lett. A* **300**:437–444 (2002).
27. T. Keef and R. Twarock. Affine extensions of the icosahedral group with applications to the three-dimensional organisation of simple viruses. *J. Math. Biol.* **59**:287–313 (2009).
28. T. Keef and R. Twarock. New insights into viral architecture via affine extended symmetry groups. *Comp. Math. Methods in Med.* **9**:221–229 (2008).
29. T. Keef, J. Wardman, N. Ranson, P. Stockley and R. Twarock. Beyond quasi-equivalence: 3-D geometry provides novel insights into virus structure. In Preparation (2009).
30. S.N. Kanesashi, K. Ishizu, M.A. Kawano, S.I. Han, S. Tomita, H. Watanabe, K. Kataoka and H. Handa. Simian virus 40 VP1 capsid protein forms polymorphic assemblies *in vitro*. *J. Gen. Virol.* **84**:1899–1905 (2003).
31. D.M. Salunke, D.L.D. Caspar and R.L. Garcea. Polymorphism in the assembly of Polyomavirus capsid protein VP1. *Biophys. J.* **56**:887–900 (1989).
32. K. Valegard, J. Murray, P. Stockley, N. Stonehouse and L. Liljas. Crystal structure of an RNA bacteriophage coat protein operator complex, *Nature* **371**:623–626 (2002).
33. K. Toropova, G. Basnak, R. Twarock, P. Stockley and N. Ranson. The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly. *J. Mol. Biol.* **375**(3):824–836 (2008).
34. L. Tang, K. Johnson, L. Ball, T. Lin, M. Yeager and J. Johnson. The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA., *Nat. Struct. Biol.* **1**(8):77–83 (2001).

35. S. Venkataraman, S.P. Reddy, J. Loo, N. Idamakanti, P.L. Hallenbeck and V.S. Reddy. Structure of Seneca Valley Virus-001: An oncolytic Picornavirus representing a new genus. *Structure* **16**:1555–1561 (2008).
36. S.C. Harrison, A.J. Olson, C.E. Schutt, F.K. Winkler and G. Bricogne. Tomato Bushy Stunt Virus at 2.9 Å resolution. *Nature* **276**:368–373 (1978).
37. P.A. Timmins, D. Wild and J. Witz. The three-dimensional distribution of RNA and protein in the interior of Tomato Bushy Stunt virus: a neutron low-resolution single-crystal diffraction study. *Structure* **2**:1191–1201 (1994).
38. U. Katpally, K. Kakani, R. Reade, K. Dryden, D. Rochon and T.J. Smith. Structures of $T = 1$ and $T = 3$ particles of Cucumber Necrosis Virus: Evidence of internal scaffolding. *J. Mol. Biol.* **365**:502–512 (2007).
39. J. Speir, S. Munshi, G. Wang, T. Baker and J. Johnson. Structures of the native and swollen forms of Cowpea Chlorotic Mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**(1):63–78 (1995).
40. F. Tama and C. Brooks III. The mechanism and pathway of pH induced swelling in Cowpea Chlorotic Mottle virus. *J. Mol. Biol.* **318**:733–747 (2002).
41. V.S. Reddy, P. Natarajan, B. Okerberg, K. Li, K. Damodaran, R. Morton, C. 3rd Brooks and J. Johnson. Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.* **75**(24):11943–11947 (2001).
42. N. Ranson, G. Farr, A. Roseman, B. Gowen, W. Fenton, A. Horwich and H. Saibil. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* **107**(7):869–79 (2001).
43. D. Bamford, R. Burnett and D. Stuart. Evolution of viral structure. *Theor. Pop. Biol.* **61**:461 (2002).
44. D. Bamford, J. Grimes and D.I. Stuart. What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**:655 (2005).

Chapter 4

Mechanical Properties of Viruses

Wouter H. Roos*,† and Gijs J. L. Wuite*,‡

If a virus releases its genomic content prematurely, it loses its infective capability. Yet, the viral shell does need to open at a specific place and time to ensure genome delivery into a new host. Hence, the chemical and mechanical properties of capsids are carefully tuned to fulfill these constraints. Knowledge of these properties will help to elucidate the viral infectious pathway, to develop virus based therapies and to facilitate the use of viruses in nanotechnology. Here we focus on the material properties of viruses mainly based on data obtained by mechanical manipulation of single viral particles. The main tool for such experiments is the atomic force microscope (AFM) and the experimental basis of these measurements will be explained. Next, aspects of the capsid shell structure, presence of encapsidated material, capsid failure, maturation and capsid protein mutations will be discussed in relation to the viral material properties. By comparing the data of various viruses, similarities and differences in the mechanical properties will be highlighted.

1. Introduction

There is a growing research interest in medicine and nanotechnology to *use* viruses as functional nanocontainers, instead of *fighting* them as pathogens.[1–4] Yet, using viruses for constructive purposes is not new and has at least been around since early after the discovery of bacteriophages in the beginning of the 20th century.[5] Various antibacterial therapies based on bacteriophages were then tested and developed. However, because of the discovery and subsequent success and widespread availability of antibiotics, the research into phage therapy dwindled in most of Western

*Fysica van complexe systemen, Natuur- en Sterrenkunde, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands.
E-mails: †wroos@few.vu.nl; ‡gwuite@nat.vu.nl

Europe and the United States. In the rest of the world it was still actively pursued, but that has not led to the development of worldwide available and administered phage therapies. Notably, a renewed research focus on phage therapy as an alternative for antibiotics has developed over the last years[5–7] and it has, for instance, recently been shown how *campylobacter*[8] and *salmonella*[9] colonization in broiler chicken can be reduced by phage therapy. Next to using phages for antibacterial therapies, eukaryotic viruses have been used for oncolytic therapies. Preclinical and clinical anticancer trials have been conducted with various oncolytic viruses and this approach seems to have the potential to replace some of the current standard cancer therapies.[1,10]

Besides biomedical applications, viruses are increasingly being used as building blocks and tools in materials science, chemistry, electronics and physics.[2,3] The capsid interior, exterior and the interface between the capsomers can independently be functionalised to create multi-functional biotemplates.[4] Genetically engineered M13 bacteriophages have, for instance, been used to create self-ordering systems of ZnS quantum dots at the nano- and micro- meter scale[11] and mutant capsid forms of the plant virus cowpea mosaic virus (CPMV) have been used as a scaffold to adhere gold nanoparticles at defined spacings.[12] Furthermore, CPMV capsids have been biotinylated to form self-assembled multi-layer arrays of viral nanoparticles connected by streptavidin.[13] The mechanical properties of these thin-film assemblies can be tuned *via* the grafting density and the linker length of the biotin moieties.[14] Cowpea chlorotic mottle virus (CCMV) has been used as a nanocontainer for the encapsulation of guest material with potential uses in drug delivery and catalysis.[15,16] For instance, CCMV and other viruses have shown the potential to serve as self-assembling nanoreactors for synthesis of nanomaterials in a constrained environment.[17,18] These are just a few examples of how viruses are used for various nanotechnological applications.

In order to fully exploit the potential use of viruses in medicine and nanotechnology, it is essential to have a fundamental biological, chemical and physical understanding of viral structure and function. To elucidate the biomechanics of viruses, various bulk and single particle studies have been performed. Heat shock treatment of phage P22 procapsids[19] and chemical treatment with 2 M GuHCl of herpes simplex virus type 1 capsids[20] have shown to result in the loss of the capsid pentamers. Phage stability and rupture has been studied by osmotic shock experiments[21,22] and the

mechanical properties of crystals of satellite tobacco mosaic virus and films of wiseana iridovirus were analysed by Brillouin light scattering.[23,77] The drawback of these multi-particle techniques is that they always average over many particles and information on the behavior of a single particle is lost. Moreover, these techniques don't reveal how viral capsids (dynamically) deform and fail under directed force. Such data would elucidate many material properties of viruses but it can only be obtained by techniques that permit measurements on single viral particles. Nanoindentation of individual virus particles using atomic force microscopy can accomplish this.[22,24,25] In this chapter we will discuss this method and the information that can be extracted with it.

2. Nanoindentation by AFM

Nanoindentation experiments have been reported for the characterisation of various kinds of (macro)molecular assemblies including microtubules,[26,27] peptide nanotubes,[28] carbon nanotubes,[29] micro- and nano- capsules[30–33] and viruses.[22,25,34,35] Figure 1 shows a schematic of an AFM set-up that performs a nanoindentation measurement.

2.1. *Sample Preparation*

Before starting the nanoindentation measurements in liquid, the viral particles need to be immobilised on a surface. This can be accomplished by adhering the particles onto a hydrophobic surface or onto a poly-lysine coated surface. The protocols are as follows: i) Hydrophobic surface[25,34,35,37] — glass cover slips are cleaned by overnight immersion in a solution consisting of 90% ethanol and 10% saturated KOH. After thorough rinsing with deionised water the cover slips are vacuum-dried, blow-dried or left to dry under the hood. Alternatively the cover slips can be cleaned by boiling in an HCl solution, followed by drying. The dried samples are placed in a sealed container under a saturated hexamethyldisilazane (HMDS) vapour pressure and left to incubate overnight. The container is then opened under the hood to let the alkylsilane evaporate and the hydrophobic cover slips are ready to use; ii) Poly-lysine coated surface[38] — A 0.01% poly-L-lysine solution is incubated on top of freshly cleaved mica for 15 min. A brief rinse with deionised water and subsequent air drying of the mica makes the sample ready for use.

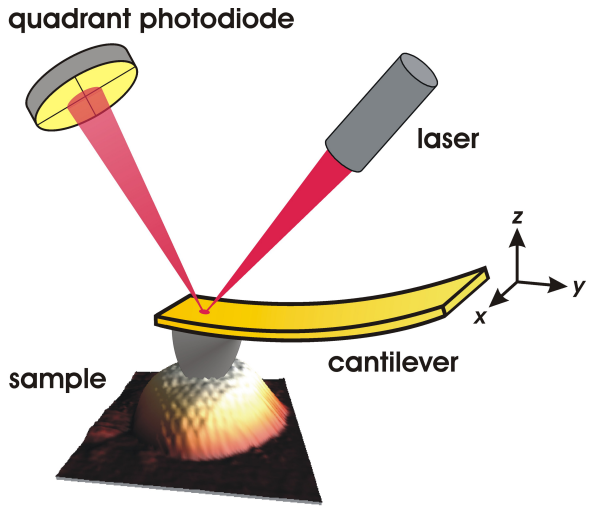


Fig. 1. A schematic illustration of the AFM experiments. The tip (depicted here in grey) is located at the freely moving end of the cantilever. This tip, which has typically a radius of curvature of 10–20 nm at its extremity, is the part of the AFM that directly contacts the sample. During imaging, the cantilever scans over the sample in the x-y plane and adjusts its height (z-direction) according to the surface features. As the surface topography varies, the force on the cantilever changes and therefore its bending. This is recorded by the quadrant photodiode as a change in signal because the angle under which the laser light is reflected changes when the cantilever bends. By a feed-back loop the height is adjusted to reduce damage to the tip and sample, while keeping the sensitivity high. The x-, y-, z- movements are typically performed with a piezoelectric crystal (not shown). Depending on the manufacturing scheme either the cantilever or the sample is scanned. The cantilever and sample can be immersed in liquid to perform measurements under physiological relevant conditions. After imaging a virus, the tip can be directed to the middle of the particle and a nanoindentation experiment can be started as is shown in the schematic. While increasing the load on the particle, the cantilever bends increasingly, which is registered as a change in signal on the quadrant photodiode. After calibration of the system the force in Newton is obtained. Adapted from reference 36.

Freshly cleaved mica has been used to deposit tobacco mosaic virus (TMV),[39] but the adsorption of murine leukemia virus (MLV) onto freshly cleaved mica pretreated with 20 mM $MnCl_2$ resulted in deformation of the particles.[35] So care should be taken in using mica and, as with glass cover slips, the height of the adhered particles should be carefully compared to values known from literature.

After sample preparation the virus solution can be applied to the surface to incubate for typically 20 to 60 min. Next, the surface can be rinsed gently with buffer to remove unabsorbed particles, or imaging can be started directly. The tip can be prewetted with buffer before starting the measurements.[34]

2.2. *Imaging*

Before starting the nanoindentation measurements the particles are imaged to check whether they have the correct shape and size. This imaging is normally performed by tapping[40] or jumping mode AFM,[41] but TMV imaging has even been performed in contact mode, without apparent distortions of the particles. However, contact mode AFM is generally most likely to damage the viruses during imaging and therefore tapping or jumping mode AFM is preferred. Note, however, that when using tapping mode, one needs to choose the proper cantilever and tune the scanning parameters carefully because the peak forces during oscillation of the cantilever tip can easily damage the viruses.[42] Figure 2 shows AFM images of surface adhered capsids, obtained by jumping mode imaging.

2.3. *Indenting*

After imaging a capsid at high resolution, the tip can be placed in the centre of the particle and force-indentation (FZ) curves can be obtained. A typical FZ curve is shown in Fig. 3. The majority of the particles studied so far show an initial linear indentation behaviour (a discussion on nonlinearities can be found in reference 43). From the spring constant of this cantilever-virus system, k_{tot}, and the cantilever spring constant, k_{cant}, one can calculate the virus spring constant, k_{vir}, using Hooke's law: $k_{vir} = k_{cant}{}^{*}k_{tot}/(k_{cant} - k_{tot})$. k_{vir} can then be used to determine the Young's (or elastic) modulus, E, of the viral capsid material. Using continuum elastic theory of thin shells E is calculated as

$$k_{vir} = \alpha E h^2/R, \tag{1}$$

where α is a proportionality factor, h is the shell thickness, and R is the capsid radius.[44] Finite element modelling has shown that the proportionality factor α is around unity for various capsids.[43,45] Even though most capsids have a shell thicker (i.e., $h/R > 0.1$) than assumed limiting for the thin shell formula, Eq. (1) can still be used as a first approximation.

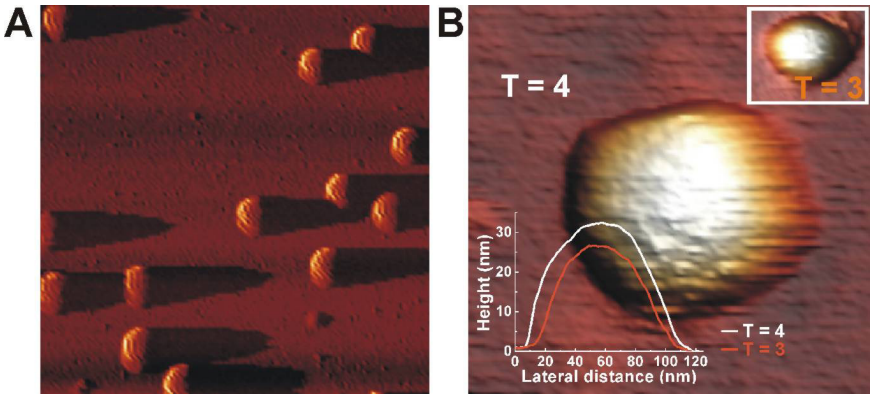


Fig. 2. AFM jumping mode images of viral capsids. A) Normal force overview image of several viruses adhered to a hydrophobic glass cover slip. The appearance of a black 'shadow' next to the particles can be attributed to the AFM imaging and feed-back parameters. These are optimised to image the particles with a minimal interaction force to reduce capsid damage before the start of the nanoindentation measurements. Image size $2 \times 2\,\mu$m. B) Zoomed in height images of hepatitis B virus capsids (3D rendered images with shadowing). A T = 4 and a T = 3 (inset) capsid are shown together with the corresponding height profiles, demonstrating the difference in diameter between the two morphologies. Due to the convolution of the sample and the tip in AFM imaging the lateral distances are overestimated, but the height measurements are reliable to determine the particle diameter. Adapted from Ref. 37.

Modeling,[46] in particular finite element modelling, of the mechanics of viral capsids has proven to be useful for extracting Young's moduli. Typically, modelling has been done on idealised particles with a constant shell thickness, but recent models have incorporated the heterogeneous structure of viral shells as well.[47] Alternatively the Hertz model has been used to extract E of the thick wall of TMV capsids.[39] Other parameters that can be extracted from the nanoindentation experiments are the capsid breaking force and the maximum indentation at breakage.

3. Comparing Viral Material Properties

Viral particles generally have a sphere-like or tube-like appearance though other geometries, for instance cones, are also found. Sphere-like particles can be described by their Casper and Klug triangulation number. This T-number structure contains information on the number of pentamers

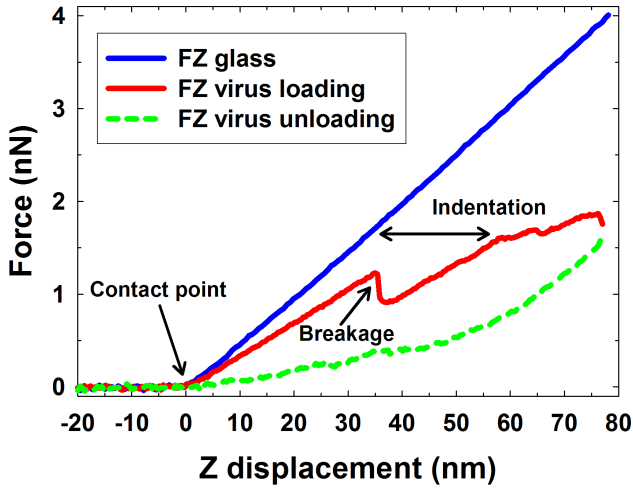


Fig. 3. Force-indentation (FZ) curve of a nanoindentation measurement. The reference curve FZ glass and the loading and unloading curve on the capsid are shown. Various parameters can be obtained from these curves. The linear part of the indentation curve gives, together with the reference curve and the cantilever spring constant, the capsid stiffness. Furthermore, one can read of the breaking force, which is 1.2 nN in this example and the indentation of the particle. The hysteresis between the loading and unloading curve shows the irreversibility of indentation of this capsid.

and hexamers in icosahedral shells. The higher the T-number, the more hexamers a particle contains. Nguyen *et al.*[48] show an illustration of how to calculate a particle's T-number.

Quite a number of theoretical/modelling studies of the mechanical properties of viruses have appeared recently (see for instance Refs. 46, 47, 49–51. However, here we will focus on the reported experimental studies. In this section, we will compare the measured material and physical properties of various viruses. In particular we discuss: i) the assembly process, ii) presence of encapsidated material, iii) failure of viral shells, iv) maturation and v) protein engineering to change the viral mechanical properties.

3.1. *Assembly Around the Genome versus Use of a Packaging Motor*

Some viruses self-assemble around the viral RNA or DNA, while other viruses self-assemble without the genome. In the latter case the genomic material is packaged after assembly with the help of a packaging

motor.[22,52–54] It appears that viruses that package their genome with the help of a packaging motor keep the genome under high pressure.[55–57] Do these viruses have strong capsids in order to contain their DNA? Is there a difference between the mechanical properties of the different virus types?

Examples of sphere-like viruses that assemble around their RNA are the plant virus CCMV and the human virus hepatitis B virus (HBV). The capsids of both these viruses have a Young's modulus between 0.1 and 0.4 GPa.[37,43,45] The bacteriophages Φ29 and λ are examples of sphere-like viruses (phage λ is icosahedral and phage Φ29 has a slightly elongated shape) that use a packaging motor to internalize their genome. The Young's modulus of their capsids lies in between 1.0 and 5 GPa.[25,43,58] There are some indications that the T = 1 minute virus of mice (MVM) also uses a packaging motor to internalize its genome and interestingly its Young's modulus has a value similar to bacteriophages (1.3 GPa).[34] To analyse if there is a correlation between the packing density and the Young's modulus of a capsid, we calculated the packing density of icosahedral viruses, in analogy to Purohit *et al.*[59] Capsids are modelled as perfect spheres and the inner radius of the capsids is used for an accurate estimate for the capsid volume V_{cap}. The viral genome is modelled as a cylinder with radii r_{gen} of 1 nm and 0.75 nm for respectively double-stranded (ds) and single-stranded (ss) genomes. The distance between the bases/base pairs is taken as 0.34 nm. The relative packing density ρ_{pack} can now be calculated by[59]

$$\rho_{pack} = \frac{0.34 N_b \pi r_{gen}^2}{V_{cap}} \tag{2}$$

with N_b either the number of base pairs for a ds genome or the number of bases for a ss genome. Figure 4 indeed shows a strong correlation between packing density and E. Of course, this graph is based on measurements of only a very small proportion of known viruses and with the advancement of mechanical studies on various viruses one will be able to show more clearly common features and differences.

Interestingly, the Young's modulus of TMV nanotubes turns out to have a value of 0.9–1.0 GPa.[39] How does this value, of a capsid that self-assembles around its genome, compare to the values mentioned above? The particles discussed until now all have a sphere-like structure, but TMV has a rod-like structure. It self-assembles in a tube with a diameter of

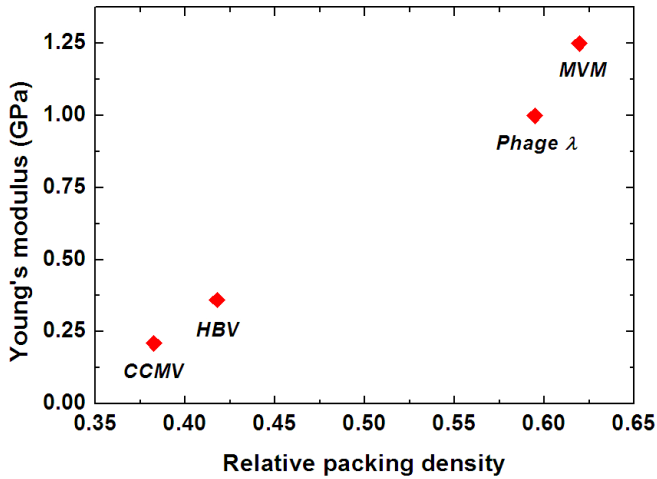


Fig. 4. Young's modulus of icosahedral capsids as a function of genome packing density. The relative packing density was calculated by Eq. (2), using inner capsid radii of 10.4 nm, 12.5 nm, 27.5 nm and 10.5 nm respectively for CCMV,[43] HBV T = 4,[37,60] phage λ[61] and MVM.[34] The ssRNA of CCMV is approximated as having a length of 3 kb[62] and a length of 5 kb is used for the ssDNA of MVM.[63] The dsDNA of phage λ has a length of 48.5 kbp. HBV capsids form in a T = 3 and a T = 4 morphology, but the latter is much more abundant in virions.[64] Therefore we plot here the results for the T = 4 morphology. HBV self-assembles around the viral RNA, which is then retrotranscribed inside the capsid to partially double stranded DNA.[60] As the short strand can approach the length of the long strand,[65] the genome is modelled as completely double stranded with a length of 3.2 kbp. The Young's modulus of CCMV capsids is taken as the average value from references 43 and 45 and for the other capsids the values from the following references are used: HBV,[37] phage λ[58] and MVM.[34]

~18 nm and a length of ~300 nm. It seems that TMV does not need to be strong to keep its genome inside, but requires capsid strength to be able to withstand axial deformations of its long cylindrical structure. This might be reflected in its Young's modulus.

3.2. *Influence of Encapsidated Material on Viral Mechanical Properties*

Experiments on CCMV have shown that the presence of a genome inside the capsids influences the viral mechanical properties. The CCMV genome

consists of ssRNA and, depending on the assembly conditions, CCMV can form empty or full particles which are morphologically indistinguishable.[62] Full capsids have a higher spring constant (0.20 vs. 0.15 N/m) and a higher breaking force (0.81 vs. 0.60 nN) than empty capsids.[45] No difference of the mechanical properties was found between pushing along the 2-fold, 3-fold or 5-fold axis.

Whereas empty and full CCMV capsids exhibit isotropic mechanical properties, this is not the case for MVM. The stiffness of empty MVM capsids is isotropic, but capsids with DNA inside experienced an anisotropic reinforcement of their mechanical characteristics.[34] The spring constant along the 5-fold, 3-fold and 2-fold axis increased by respectively 3%, 40% and 140% as a result of genome encapsidation. By a protein engineering approach (see also Sec. 3.5) it was shown that DNA linkage to the capsid interior is responsible for this anisotropic reinforcement of the viral mechanical properties.[66] Whereas these two examples illustrate the effect of either the presence or the absence of genome inside capsids, experiments on phage λ reveal a DNA length dependent influence on viral mechanical properties.

This influence of the phage λ dsDNA is tested using shorter-genome mutants. The mechanics of phage λ mutants carrying 78% or 94% of the wild-type DNA length was compared to that of empty and wild-type phage λ capsids. The wild-type capsids possess a spring constant of 0.23 N/m and a breaking force of 1.6 nN.[58] These values are roughly double the values of those for the empty particles. Surprisingly it turned out that the spring constant and the breaking force of the 78% and 94% DNA length mutants were essentially the same as those of the empty particles. It seems that the presence of the DNA is not observed while indenting the mutant particles. Apparently the DNA can move around freely in the shorter genome mutants and only for the wild-type, containing 100% of the genomic DNA, is the effect of the genome detected. Ivanovska *et al.*[58] explain this by using an analytical model, including DNA-hydrating water molecules describing the osmotic pressure inside the capsids. As this pressure increases exponentially with the packaged DNA density, there is a considerable difference between the 94% mutant and the wild-type particles. This could also explain why packaging of phage λ with a DNA bigger than 105% of the wild-type length is difficult.[67] The osmotic pressure is then probably too high inside the capsid and it breaks as a result of internal pressure build-up.

3.3. *Capsid Failure*

Some capsids can be pushed several times with a low force without inflicting permanent damage.[25,37] Large indentations can induce damage i.e., when one pushes hard enough, many capsids break as can be seen in Fig. 3. Empty bacteriophage λ capsids and Φ29 procapsids break at a force of 0.8 nN and 1.5 nN respectively, for loading rates of ~1.5 nN/s.[58] Increasing the loading rate (pushing faster) also increases the breaking force as is, for instance, observed for Φ29 procapsids and CCMV capsids.[25,36,45] This behaviour, described by Bell,[68] of the dependence of loading rate on bond/interaction breaking is also observed in dynamic force spectroscopy where bonds are pulled apart.[69–71] On the other hand, some particles such as CCMV capsids at pH 6 do not break at all and deform linearly until the particles are flattened completely.[72] This deformation is completely elastic as follows from its reversible indentation behaviour. Surprisingly CCMV capsids at pH 5 do break while pushing with forces above 0.6 nN.[45]

To understand these differences the Föppl-von Kármán (FvK) number of the viruses can be considered. The FvK number is the dimensionless ratio of stretching and bending energies of a shell and has been employed for theoretical descriptions of the morphology and deformation of icosahedral viral particles.[48,50,51,73] Particles with a FvK value above a critical value of 154 have a more faceted shape, whereas below this value they appear more spherical.[73] By performing finite element simulations on thin-shelled particles it was shown that CCMV at pH 5 behaves like a particle with a FvK value of around 900.[72] In both the simulations and experiments, a discontinuity in the force-indentation curves is observed. At pH 6 these capsids show an indentation behaviour more comparable to particles with an FvK number of 100.

3.4. *Maturation*

A prerequisite for infectivity of retroviruses is a maturation transition which they undergo after budding from the cell. The mechanics of this transition has been studied for murine leukemia virus (MLV) and human immunodeficiency virus (HIV).[35,74] Maturation of retroviruses goes along with dramatic morphological changes of the particles in parallel with mechanical changes, as shown for MLV and HIV. In these cases it results in a decrease of the particle spring constant. It was shown that the immature HIV particles were relatively stiff as a result of interactions of the

cytoplasmic tail of the envelope glycoprotein with the viral Gag protein. The results from these studies imply that rigidity might be related to cell entry activity. Flexible particles enter cells more efficiently than rigid particles. This illustrates the necessity for these retroviruses to reduce their spring constant during maturation to optimise infectivity.

3.5. *Protein Engineering of Capsids*

Protein engineering has been used to increase or decrease the strength of viral particles. For instance, a single point mutation in the capsid protein increases the spring constant and breaking force of CCMV capsids.[45] Experiments on mutant HIV particles revealed a mechanical effect of the interaction between the viral Gag protein and the envelope glycoproteins and its importance during maturation (see also Sec. 3.4).[74] Deletion of the cytoplasmic tail of the envelope glycoprotein resulted in a more than 8-fold decrease in the Young's modulus of immature HIV particles. Strikingly, this mutation left the material properties of mature particles unchanged, probably because of the breakage of the interactions between Gag and envelope glycoproteins during maturation, resulting from the proteolytic processing of Gag. In MVM an anisotropic increase of the capsid stiffness occurs upon DNA encapsidation.[34] This anisotropic increase was explained by the presence of short, crystallographically visible DNA patches near the capsid inner wall, which were proposed to interact with each other. By specifically truncating selected amino acid side chains, the major interactions between the DNA and the inner capsid wall were inhibited.[66] No effects were observed on the material properties of empty capsids. However, the difference in spring constant between empty and full capsids was greatly reduced.

These examples illustrate how a combination of protein engineering and nanoindentation experiments can be used to elucidate the origin of specific mechanical properties of viruses. Furthermore, this approach can be used to optimise the material properties of viral particles for their use in nanotechnology and medicine.

4. Conclusions and Outlook

We have shown how nanoindentation techniques can be used to elucidate the mechanical properties of viruses. This approach advanced our

knowledge of fundamental properties of viral particles such as maturation, the effects of encapsidated material on virus strength and the assembly process. In a wider context, these measurements also broaden our understanding of viral infectivity, capsid failure and the effects of mutations in capsid shells.

The FvK approach discussed in Sec. 3.3 has proven to be a valuable tool to describe the deformation of capsids and it will be interesting to find out how broadly this method can be applied to other viruses. Modelling of viral particles becomes increasingly more refined by the application of non-linear finite element analyses[43] and the use of heterogeneous capsid models.[47] In addition molecular dynamics simulations of viruses are becoming feasible due to increasing computing power. Such simulations can be performed on small viruses in an all-atom approach[75,78] and in a coarse-grained approach[76] for bigger viruses.

Taken together, we are seeing a strong increase in experimental, theoretical and modelling activity of the mechanical properties of viruses. Nanoindentation with an atomic force microscope is the preferred tool to investigate these properties, as this technique allows experiments under physiological conditions (i.e., in liquid) with high force and position resolution. The insights obtained from these experiments enhance our knowledge of the viral infectious pathway and enable the advancement of development and application of viruses and virus-like particles in nanotechnology and medicine.[1–4]

Acknowledgements

This work was supported by the Nederlandse Organisatie voor Wetenschappelijk onderzoek (NWO) through a CW-ECHO and a Rubicon grant.

References

1. K.A. Parato, D. Senger, P.A.J. Forsyth and J.C. Bell. Recent progress in the battle between oncolytic viruses and tumours. *Nat. Rev. Cancer* **5**:965–976 (2005).
2. P. Singh, M.J. Gonzalez and M. Manchester. Viruses and their uses in nanotechnology. *Drug Dev. Res.* **67**:23–41 (2006).
3. M. Fischlechner and E. Donath. Viruses as building blocks for materials and devices. *Angew. Chem. Int. Edit.* **46**:3184–3193 (2007).

4. M. Young, D. Willits, M. Uchida and T. Douglas. Plant viruses as biotemplates for materials and their use in nanotechnology. *Annu. Rev. Phytopathol.* **46**:361–384 (2008).
5. W.C. Summers. Bacteriophage therapy. *Annu. Rev. Microbiol.* **55**:437–451 (2001).
6. B.R. Levin and J.J. Bull. Population and evolutionary dynamics of phage therapy. *Nat. Rev. Microbiol.* **2**:166–173 (2004).
7. M. Skurnik and E. Strauch. Phage therapy: Facts and fiction. *Int. J. Med. Microbiol.* **296**:5–14 (2006).
8. J.A. Wagenaar, M.A.P. Van Bergen, M.A. Mueller, T.M. Wassenaar and R.A. Carlton. Phage therapy reduces Campylobacter jejuni colonization in broilers. *Vet. Microbiol.* **109**:275–283 (2005).
9. R.J. Atterbury, M.A.P. Van Bergen, F. Ortiz, M.A. Lovell, J.A. Harris, A. De Boer, J.A. Wagenaar, V.M. Allen and P.A. Barrow. Bacteriophage therapy to reduce Salmonella colonization of broiler chickens. *Appl. Environ. Microbiol.* **73**:4543–4549 (2007).
10. T.C. Liu and D. Kirn. Gene therapy progress and prospects cancer: oncolytic viruses. *Gene Ther.* **15**:877–884 (2008).
11. S.W. Lee, C.B. Mao, C.E. Flynn and A.M. Belcher. Ordering of quantum dots using genetically engineered viruses. *Science* **296**:892–895 (2002).
12. A.S. Blum, C.M. Soto, C.D. Wilson, J.D. Cole, M. Kim, B. Gnade, A. Chatterji, W.F. Ochoa, T.W. Lin, J.E. Johnson and B.R. Ratna. Cowpea mosaic virus as a scaffold for 3-D patterning of gold nanoparticles. *Nano Lett.* **4**:867–870 (2004).
13. N.F. Steinmetz, G. Calder, G.P. Lomonossoff and D.J. Evans. Plant viral capsids as nanobuilding blocks: Construction of arrays on solid supports. *Langmuir* **22**:10032–10037 (2006).
14. N.F. Steinmetz, E. Bock, R.P. Richter, J.P. Spatz, G.P. Lomonossoff and D.J. Evans. Assembly of multilayer arrays of viral nanoparticles via biospecific recognition: A quartz crystal microbalance with dissipation monitoring study. *Biomacromolecules* **9**:456–462 (2008).
15. T. Douglas and M. Young. Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **393**:152–155 (1998).
16. S.E. Aniagyei, C. DuFort, C.C. Kao and B. Dragnea. Self-assembly approaches to nanomaterial encapsulation in viral protein cages. *J. Mater. Chem.* **18**:3763–3774 (2008).
17. T. Douglas, E. Strable, D. Willits, A. Aitouchen, M. Libera and M. Young. Protein engineering of a viral cage for constrained nanomaterials synthesis. *Adv. Mater.* **14**:415–418 (2002).
18. D.M. Vriezema, M.C. Aragones, J. Elemans, J. Cornelissen, A.E. Rowan and R.J.M. Nolte. Self-assembled nanoreactors. *Chem. Rev.* **105**:1445–1489 (2005).
19. C.M. Teschke, A. McGough and P.A. Thuman-Commike. Penton release from p22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophys. J.* **84**:2585–2592 (2003).
20. W.W. Newcomb and J.C. Brown. Structure of the herpes-simplex virus capsid — effects of extraction with guanidine-hydrochloride and partial reconstitution of extracted capsids. *J. Virol.* **65**:613–620 (1991).

21. T.F. Anderson, C. Rappaport and N.A. Muscatine. On the structure and osmotic properties of phage particles. *Annales de l'Institut Pasteur* **84**:5–15 (1953).
22. W.H. Roos, I.L. Ivanovska, A. Evilevitch and G.J.L. Wuite. Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. *Cell. Mol. Life Sci.* **64**:1484–1497 (2007).
23. R.D. Hartschuh, S.P. Wargacki, H. Xiong, J. Neiswinger, A. Kisliuk, S. Sihn, V. Ward, R.A. Vaia and A.P. Sokolov. How rigid are viruses. *Phys. Rev. E* **78**:021907 (2008).
24. G. Binnig, C.F. Quate and C. Gerber. Atomic force microscope. *Phys. Rev. Lett.* **56**:930–933 (1986).
25. I.L. Ivanovska, P.J. de Pablo, B. Ibarra, G. Sgalari, F.C. MacKintosh, J.L. Carrascosa, C.F. Schmidt and G.J.L. Wuite. Bacteriophage capsids: Tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci. USA* **101**:7600–7605 (2004).
26. A. Vinckier, C. Dumortier, Y. Engelborghs and L. Hellemans. Dynamical and mechanical study of immobilized microtubules with atomic force microscopy. *J. Vac. Sci. Technol. B* **14**:1427–1431 (1996).
27. P.J. de Pablo, I.A.T. Schaap, F.C. MacKintosh and C.F. Schmidt. Deformation and collapse of microtubules on the nanometer scale. *Phys. Rev. Lett.* **91**, 098101 (2003).
28. N. Kol, L. Adler-Abramovich, D. Barlam, R Z. Shneck, E. Gazit and I. Rousso. Self-assembled peptide nanotubes are uniquely rigid bioinspired supramolecular structures. *Nano Lett.* **5**:1343–1346 (2005).
29. I. Palaci, S. Fedrigo, H. Brune, C. Klinke, M. Chen and E. Riedo. Radial elasticity of multiwalled carbon nanotubes. *Phys. Rev. Lett.* **94**:175502 (2005).
30. F. Dubreuil, N. Elsner and A. Fery. Elastic properties of polyelectrolyte capsules studied by atomic-force microscopy and RICM. *Eur. Phys. J. E.* **12**:215–221 (2003).
31. A. Fery and R. Weinkamer. Mechanical properties of micro- and nanocapsules: Single-capsule measurements. *Polymer* **48**:7221–7235 (2007).
32. C. Picart, B. Senger, K. Sengupta, F. Dubreuil and A. Fery. Measuring mechanical properties of polyelectrolyte multilayer thin films: Novel methods based on AFM and optical techniques. *Colloid. Surf. A: Physicochem. Eng. Asp.* **303**:30–36 (2007).
33. C.I. Zoldesi, I.L. Ivanovska, C. Quilliet, G.J.L. Wuite and A. Imhof. Elastic properties of hollow colloidal particles. *Phys. Rev. E* **78**:051401 (2008).
34. C. Carrasco, A. Carreira, I.A.T. Schaap, P.A. Serena, J. Gomez-Herrero, M.G. Mateu and P.J. Pablo. DNA-mediated anisotropic mechanical reinforcement of a virus. *Proc. Natl. Acad. Sci. USA* **103**:13706–13711 (2006).
35. N. Kol, M. Gladnikoff, D. Barlam, R Z. Shneck, A. Rein and I. Rousso. Mechanical properties of murine leukemia virus particles: Effect of maturation. *Biophys. J.* **91**:767–774 (2006).
36. W.H. Roos and G.J.L. Wuite. Nanoindentation studies reveal material properties of viruses. *Adv. Mater.* **21**:1187–1192 (2009).
37. C. Uetrecht, C. Versluis, N.R. Watts, W.H. Roos, G.J. Wuite, P.T. Wingfield, A.C. Steven and A.J. Heck. High-resolution mass spectrometry of viral assemblies:

Molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. USA* **105**:9216–9220 (2008).
38. I. Liashkovich, W. Hafezi, J.E. Kuhn, H. Oberleithner, A. Kramer and V. Shahin. Exceptional mechanical and structural stability of HSV-1 unveiled with fluid atomic force microscopy. *J. Cell Sci.* **121**:2287–2292 (2008).
39. Y. Zhao, Z.B. Ge and J.Y. Fang. Elastic modulus of viral nanotubes. *Phys. Rev. E* **78**, 031914 (2008).
40. C.A.J. Putman, K.O. Vanderwerf, B.G. Degrooth, N.F. Vanhulst and J. Greve. Tapping mode atomic-force microscopy in liquid. *Appl. Phys. Lett.* **64**:2454–2456 (1994).
41. P.J. de Pablo, J. Colchero, J. Gomez-Herrero and A.M. Baro. Jumping mode scanning force microscopy. *Appl. Phys. Lett.* **73**:3300–3302 (1998).
42. X. Xu, C. Carrasco, P.J. de Pablo, J. Gomez-Herrero and A. Raman. Unmasking imaging forces on soft biological samples in liquids when using dynamic atomic force microscopy: A case study on viral capsids. *Biophys. J.* **95**, 2520–2528 (2008).
43. M.M. Gibbons and W.S. Klug. Nonlinear finite-element analysis of nanoindentation of viral capsids. *Phys. Rev. E.* **75**:031901 (2007).
44. L.D. Landau and E.M. Lifshitz. *Theory of Elasticity*, Elsevier, Oxford (1986).
45. J.P. Michel, I.L. Ivanovska, M.M. Gibbons, W.S. Klug, C.M. Knobler, G.J.L. Wuite and C.F. Schmidt. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc. Natl. Acad. Sci. USA* **103**:6184–6189 (2006).
46. M.M. Gibbons and W.S. Klug. Mechanical modeling of viral capsids. *J. Mater. Sci.* **42**:8995–9004 (2007).
47. M.M. Gibbons and W.S. Klug. Influence of nonuniform geometry on nanoindentation of viral capsids. *Biophys. J.* **95**:3640–3649 (2008).
48. T.T. Nguyen, R.F. Bruinsma and W.M. Gelbart. Elasticity theory and shape transitions of viral shells. *Phys. Rev. E* **72**:051923 (2005).
49. R. Zandi and D. Reguera. Mechanical properties of viral capsids. *Phys. Rev. E.* **72**:021917 (2005).
50. G.A. Vliegenthart and G. Gompper. Mechanical deformation of spherical viruses with icosahedral symmetry. *Biophys. J.* **91**:834–841 (2006).
51. M. Buenemann and P. Lenz. Mechanical limits of viral capsids. *Proc. Natl. Acad. Sci. USA* **104**:9925–9930 (2007).
52. W.C. Earnshaw and S.R. Casjens. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**:319–331 (1980).
53. L.W. Black. DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43**:267–292 (1989).
54. S.D. Moore and P.E. Prevelige, Jr. DNA packaging: a new class of molecular motors. *Curr. Biol.* **12**:R96–98 (2002).
55. D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson and C. Bustamante. The bacteriophage phi 29 portal motor can package DNA against a large internal force. *Nature* **413**:748–752 (2001).
56. D.N. Fuller, D.M. Raymer, J.P. Rickgauer, R.M. Robertson, C.E. Catalano, D.L. Anderson, S. Grimes and D.E. Smith. Measurements of single DNA

molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* **373**:1113–1122 (2007).
57. A. Evilevitch, L. Lavelle, C.M. Knobler, E. Raspaud and W.M. Gelbart. Osmotic pressure inhibition of DNA ejection from phage. *Proc. Natl. Acad. Sci. USA* **100**:9292–9295 (2003).
58. I. Ivanovska, G. Wuite, B. Jonsson and A. Evilevitch. Internal DNA pressure modifies stability of WT phage. *Proc. Natl. Acad. Sci. USA* **104**:9603–9608 (2007).
59. P.K. Purohit, M.M. Inamdar, P.D. Grayson, T.M. Squires, J. Kondev and R. Phillips. Forces during bacteriophage DNA packaging and ejection. *Biophys. J.* **88**:851–866 (2005).
60. A. Zlotnick, N. Cheng, J.F. Conway, F.P. Booy, A.C. Steven, S.J. Stahl and P.T. Wingfield. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**:7412–7421 (1996).
61. T. Dokland and H. Murialdo. Structural transitions during maturation of bacteriophage lambda capsids. *J. Mol. Biol.* **233**:682–694 (1993).
62. J.M. Fox, G. Wang, J.A. Speir, N.H. Olson, J.E. Johnson, T.S. Baker and M.J. Young. Comparison of the native CCMV virion with *in vitro* assembled CCMV virions by cryoelectron microscopy and image reconstruction. *Virology* **244**:212–218 (1998).
63. S.F. Cotmore and P. Tattersall. Encapsidation of minute virus of mice DNA: Aspects of the translocation mechanism revealed by the structure of partially packaged genomes. *Virology* **336**:100–112 (2005).
64. A.M. Roseman, J.A. Berriman, S.A. Wynne, P.J.G. Butler and R.A. Crowther. A structural model for maturation of the hepatitis B virus core. *Proc. Natl. Acad. Sci. USA* **102**:15821–15826 (2005).
65. P. Tiollais, C. Pourcel and A. Dejean. The hepatitis-B virus. *Nature* **317**:489–495 (1985).
66. C. Carrasco, M. Castellanos, P.J. de Pablo and M.G. Mateu. Manipulation of the mechanical properties of a virus by protein engineering. *Proc. Natl. Acad. Sci. USA* **105**:4150–4155 (2008).
67. M. Feiss, R.A. Fisher, M.A. Crayton and C. Egner. Packaging of bacteriophage-lambda chromosome — effect of chromosome length. *Virology* 77:281–293 (1977).
68. G.I. Bell. Models for specific adhesion of cells to cells. *Science* **200**:618–627 (1978).
69. R. Merkel, P. Nassoy, A. Leung, K. Ritchie and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **397**:50–53 (1999).
70. E. Evans. Probing the relation between force — Lifetime — and chemistry in single molecular bonds. *Annual Review of Biophysics and Biomolecular Structure* **30**:105–128 (2001).
71. P. Hinterdorfer and Y.F. Dufrene. Detection and localization of single molecular recognition events using atomic force microscopy. *Nat. Methods* **3**:347–355 (2006).

72. W.S. Klug, R.F. Bruinsma, J.P. Michel, C.M. Knobler, I.L. Ivanovska, C.F. Schmidt and G.J.L. Wuite. Failure of viral shells. *Phys. Rev. Lett.* **97**:228101 (2006).
73. J. Lidmar, L. Mirny and D.R. Nelson. Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E* **68**:051910 (2003).
74. N. Kol, Y. Shi, M. Tsvitov, D. Barlam, R Z. Shneck, M.S. Kay and I. Rousso. A stiffness switch in human immunodeficiency virus. *Biophys. J.* **92**:1777–1783 (2007).
75. P.L. Freddolino, A.S. Arkhipov, S.B. Larson, A. McPherson and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **14**:437–449 (2006).
76. A. Arkhipov, P.L. Freddolino and K. Schulten. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **14**, 1767–1777 (2006).
77. B. Stephanidis, S. Adichtchev, P. Gouet, A. McPherson and A. Mermet. Elastic properties of viruses. *Biophys. J.* **93**:1354–1359 (2007).
78. M. Zink and H. Grubmuller. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. *Biophys. J.* **96**:1350–1363 (2009).

Chapter 5

Investigating Viral Structure, Function and Dynamics with Mass Spectrometry

Eric B. Monroe* and Peter E. Prevelige*,†

The development of soft ionization techniques has made it possible to examine biological macromolecules using mass spectrometry. Proteomic approaches have been developed which are capable of identifying and chemically characterizing a wide variety of biological molecules in an efficient manner. A variety of mass spectroscopic experimental approaches have been developed which allow the investigator to obtain information about protein structure, dynamics and protein/protein interactions. This review focuses on the application of these techniques to the study of viral replication.

1. Introduction

The study of viruses has provided invaluable insights into the intrinsic relationship between molecular structure and function in regard to understanding pathological disease and has provided model systems for the molecular study of an array of cellular processes. Despite the wealth of information developed from previous research, the well characterized viral proteins and their interactions with viral and host factors represent only a small fraction of the viral proteins from a small percentage of the identified viruses.

The application of mass spectrometry (MS) to the wide scale study of viral proteins provides a toolbox with which to study aspects of the viral life cycle at the molecular level. These MS studies include proteomic studies

*Department of Microbiology, University of Alabama at Birmingham.
E-mail: †prevelige@uab.edu

of viral or host proteins present in viral particles and host cells, quantitative studies of changes in protein expression levels following infection, structural studies of viral proteins and their dynamics and interactions with both host and viral proteins. This array of MS experiments will be discussed in greater detail throughout this chapter largely in regard to their utility as a means of understanding the viral life cycle including aspects of cellular entry, virion disassembly, replication, assembly, cellular release, viral maturation and immune avoidance. In general, these studies involve experiments designed to identify what proteins are present, how they are structured both as isolated proteins and in macromolecular complexes, what proteins they interact with and how these proteins act in dynamic processes such as viral entry, replication, maturation and immunological avoidance.

The application of MS to the study of viruses, at the most basic level, involves the study of an array of molecules ranging from lipids from enveloped viruses and peptides from digested proteins to intact proteins, macromolecular complexes, and even intact viruses. This assortment of molecular species precludes the use of a single methodology and, in practice, requires the selective application of multiple preparation and MS strategies. Although it is difficult to provide a detailed and complete study of the array of available studies, we aim to present a general overview of the state-of-the-art application of MS to the study of viruses. A brief introduction and discussion of instrumentation and tools will be included in addition to discussion of previously published research regarding the application of virus MS. In addition, we hope to show the various means by which mass spectrometry can compliment a wide range of molecular and structural biological approaches including mutational analysis, protein and genetic analysis, electron microscopy, nuclear magnetic resonance (NMR) and crystallography to enhance the understanding of virus dynamics.

2. Mass Spectrometry Overview

The vast array of experiments performed with MS requires a number of different instruments and preparative strategies. A general overview of mass spectrometry may assist in understanding the results and experiments detailed throughout this chapter. For more detailed explanations of MS, several excellent sources are available.[1–3]

Mass spectrometry directly measures the mass of molecular or atomic ions in a multiplexed manner allowing the detection of tens to hundreds of compounds within a single experiment. Generally speaking, MS first requires that the analyte be transferred from the condensed phase to the gas phase by an ionization method. Owing to the focus of virus MS on molecular and macromolecular compounds this is almost exclusively performed by electrospray ionization (ESI) or matrix assisted laser desorption-ionization (MALDI) methods. The development of these soft ionization techniques made possible the analysis of intact peptides, proteins, and even protein complexes. This achievement resulted in the recognition of Fenn and Tanaka by the Nobel committee with the 2002 Nobel Prize in Chemistry. Following ionization, the resulting ions are then separated in a vacuum on the basis of their mass-to-charge ratio (m/z). As the name suggests, the m/z is based on the mass of the atomic or molecular ion as determined by the chemical composition of the analyte, and the charge on the ionized species, which depends on the method of ionization and the properties of the analyte and the surrounding matrix. Following separation, ions are detected in a manner often tailored to the separation technique.

The ionization strategy of ESI is the most common and versatile ionization method used for biological mass spectrometry. Developed by Fenn in the 1980s, ESI uses an electric field to create an aerosol of highly charged droplets composed of volatile solvents and analytes.[4] These droplets subsequently undergo size reduction, using a combination of solvent evaporation and coulombic explosions, until analyte molecules are ultimately introduced into the gas phase as shown in Fig. 1A. In practice, a high electric potential (1–4 kV) is applied between an emitter such as a metal, fused silica or a metal–coated pulled glass capillary filled with analyte, and the entrance to a mass analyzer. ESI produces primarily multiply charged ions. This complicates spectral interpretation as a single analyte may be represented by multiple charge states but the analysis of higher charge states permits the detection of higher mass compounds and eases fragmentation studies, since multiply charged molecules tend to fragment more completely. ESI is a very 'soft' ionization method that is characterized by little or no analyte fragmentation during ionization and is capable of ionizing intact macromolecular complexes up to and including intact viral particles.

Developed by Tanaka,[5] Karas and Hillenkamp[6] in the second half of the 1980s, MALDI MS utilizes a laser and a small molecule organic matrix to ionize intact biomolecules. Briefly, samples are prepared by mixing the

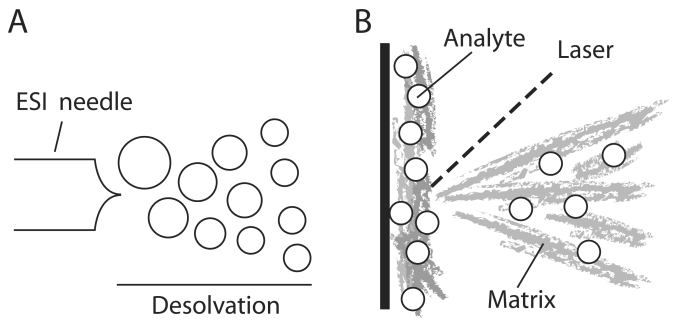


Fig. 1. Schematic of major ionization strategies of (A) ESI and (B) MALDI. ESI relies on a high potential to ionize analytes in a liquid phase followed by desolvation prior to entering the mass spectrometer. MALDI uses a laser to desorb and ionize analytes embedded in a small molecule matrix, which is also desorbed during the process.

analytes with a low molecular weight organic matrix. A wide variety of crystalline and liquid matrices, which generally include derivatives of benzoic and cinnamic acids or glycerol, are available to analyze a broad spectrum of materials. Not surprisingly, some matrices are more effective in the ionization of peptides and proteins, while others are better suited for analysis of oligonucleotides or polymers. Following sample preparation with the appropriate matrix, nanosecond ultraviolet or infrared laser pulses are projected at the sample. The radiation is absorbed by the matrix, causing rapid, localized heating and subsequent ejection of both neutral and charged analyte molecules, matrix molecules, and matrix–analyte clusters (Fig. 1B). As a result of these particle interactions, analyte ionization occurs both on the target in the matrix crystals and within the desorption plume following the laser pulse. This process results in the formation of predominantly singly charged ions. As a result of the rapid expansion and cooling of the of the ejection plume, there is little fragmentation of analyte molecules during the MALDI desorption/ionization process.

As mentioned previously, following ionization, analytes must be separated on the basis of the mass-to-charge ratio. In regard to the experiments discussed within this chapter, several types of mass analyzers are used: time-of-flight (ToF), ion traps (IT), hybrid IT-ToF, and Fourier transform ion cyclotron resonance (FT ICR).

With ToF mass measurements, a packet of ions is accelerated to a constant kinetic energy and then moves into a field free region. The ions separate while traveling along the high vacuum flight tube. Separation

occurs as smaller ions have a higher velocity than larger molecules with the same kinetic energy. During formation, ions have a small spread of kinetic energies causing a spread of flight times. By using a reflectron, this effect may be reduced and a higher resolution spectrum results by accounting for such variations. The detector sits at the end of the flight tube and registers impacts to record the flight time of individual ions. ToF instruments have no theoretical upper mass limit and are commonly used in MALDI mass spectrometers and non-covalent ESI MS experiments.

The ion trap (IT) mass analyzer operates by confining ions in an electrostatic trap by applying a radio frequency (RF) potential to a series of electrodes. To detect the trapped ions, this RF potential is modulated or a supplemental voltage is applied to endcap electrodes to sequentially eject the ions from the trap to a detector. Ion traps are highly beneficial for the study of small and highly charged molecules. Ion traps have a reduced m/z region available for analysis compared to ToF but, unlike ToF instruments, enable multiple levels of analyte fragmentation (MS^n) to increase the amount of information available when attempting to identify an unknown. Many newer ToF instruments are also capable of fragmentation but often require the addition of specialized ion optics to isolate and induce fragmentation of select analytes.

Hybrid ion trap (multiple traps in a single instrument) or quadrupole ToF mass spectrometers utilize a quadrupole mass filter or ion trap between the ion source and the flight tube. The ion trap is used to filter, store and often fragment ions before passing the ions into the ToF portion of the analyzer for more precise mass separation and analysis. These instruments tend to have higher mass accuracy than ToF only instruments and have the added benefit of MS^n capabilities, as multiple fragmentation cycles may occur in the ion trap with the resulting fragments being analyzed by ToF.

The use of FT ICR mass analyzers presents the highest performance mass analyzer with the highest mass resolution and mass accuracy of current instrumentation. Unlike the previous mass analyzers which depend on ion-detector impacts, FT ICR mass spectrometers separate and detect ions based on their motion within an ion cyclotron in the presence of a large magnetic field. The motion of ions induces a cyclical signal in detectors present in the detection cell. As the motion of ions in the cyclotron is a function of their m/z, the frequency of the induced signals may be deconvoluted by applying a Fourier transform to the transient signals to produce a mass spectrum. The largest advantage of this approach is the

high mass accuracy and precision obtainable with FT ICR MS. Often an ion trap is placed prior to the ICR cell for similar purposes as in the hybrid instruments. Additionally, molecules may be fragmented and ejected in a similar manner as an ion trap within the ICR cell such that intact proteins may be studied and the amino acid structure characterized completely in a single 'top-down' experiment.[7]

The use of fragmentation is valuable in the study and identification of the amino acid sequence of proteins and the identification of smaller molecules such as glycans, lipids or small molecules. In the case of IT and FT ICR, molecules may be isolated and fragmented within the trap or cell and then analyzed. In the case of ToF, specialized ion optics and a collision cell are placed within the flight path of the ions. Selected ions may then be isolated by a short period of time-of-flight and then reaccelerated into the remaining portion of the flight tube. Another common methodology is to place a similar collision cell between ion traps or the ion trap/quadrupole in hybrid instruments. A large number of fragmentation methodologies have been developed, as reviewed elsewhere.[8,9] In the case of protein analysis, this fragmentation produces a predictable series of ions along the peptide backbone. By observing the difference between fragment ions, the specific sequence of amino acids may be observed for identifying unknown compounds and verifying molecular identity.

The use of specific ion detectors is much less important in regard to experimental design and largely linked to the mass analyzer. Microchannel plates are often fitted to the end of the field free flight tube of ToF mass spectrometers while electron multipliers such as Faraday cups are commonly used with ion traps. As mentioned previously, detection and mass separation are performed simultaneously in FT ICR MS. However, all detectors enable the production of mass spectra which plot the relative intensity of the produced signal for the continuum of m/z values collected. Although quantitative experiments are possible with MS through additional careful sample preparation, ionization and detection efficiencies often vary across molecules such that MS is often used primarily in a qualitative manner.

3. The -omics of Viruses

The majority of applications of MS to the study of viruses focus on the detection and characterization of a wide range of biological molecules.

Often this involves identifying viral proteins present in viral particles or how they interact with host proteins through proteomic means. The incorporation of host proteins and lipids into viral particles is also being studied. The detection of post and co-translational modifications such as glycosylation are also rapidly growing fields, as is the study of host responses to viral infection.[10] The complexity of many of these experiments can be daunting but several MS approaches have become relatively standard for such studies. Beyond understanding just what molecules are present it is possible to determine how various virus or host related components change during, or as a result of, viral infection with these -omics approaches.

The seemingly wide berth of these experiments presents significant challenges for the authors when attempting to distill them into a single experimental protocol, especially in regard to the preparation of the samples. As such, we will stick to explaining the general analysis strategy from the point of having a virus or infected material in a vial as a result of virus particle or protein purification. Treatment of the samples with a reducing agent to disrupt disulfide bonds or endoproteases such as trypsin to aid in the identification of proteins (by increasing the quality of collected fragmentation spectra) is often highly beneficial in proteomic experiments prior to MS analysis. Samples are often separated with liquid chromatography via reverse phase columns (i.e., C18) in order to separate analytes and address the inherent sample complexity. In order to aid ionization, the organic/aqueous solvents often contain a small fraction (0.1%) of formic acid as an ion pairing reagent. As analytes exit the column, they may be examined with an ESI MS inline or fractionated for offline MALDI or ESI MS. During analysis, fragmentation protocols are very valuable to help identify the compounds giving rise to the individual signals detected by the mass spectrometer. As could be expected, the analysis of the data sets produced by these experiments can be quite complex. A wide variety of freely available and commercial bioinformatic tools have been produced to aid in the identification of proteins and include algorithms capable of the semi-automated annotation of MS/MS spectra.[11,12] Although many of these tools are capable of *de novo* annotation of fragmentation spectra, they are most useful when genetic information is available such that they may query a database of known or predicted gene products from the virus, similar viruses (for homologous protein identification) and host proteins.

A large portion of MS –omics type experiments have focused on identifying viral proteins. Although the genetic sequences of many viruses are

known, MS analysis enables the study of translated viral proteins and particularly those that are incorporated into released virions. Such studies have been performed on a wide range of viruses and viral families including bacteriophages,[13–15] white spot syndrome virus,[16,17] coronoviruses,[18] iridoviruses,[19] herpesviruses,[20,21] adenoviruses,[22] and poxviruses.[23–25]

A study of protein processing and incorporation showed that P2 and P4 bacteriophage virions, as well as procapsid shells isolated from viral infections, contain an N-terminal fragment of the autoproteolytic scaffolding protein gpO.[13] This cleavage was observed to occur between residues 141 and 142 in gpO and, following further mutational and co-expression experiments, allowed for the development of a model explaining the formation of procapsid heads of the appropriate size in P2 and P4. Several studies have examined the protein composition of the causative agent of severe acute respiratory syndrome (SARS).[18,26,27] These studies independently identified four structural proteins with MS: the spike (S), nucleocapsid (N), membrane (M) and envelope (E).[26,27] Additional examination uncovered four glycosylation sites on S,[27] 12 glycosylation sites on N,[18] and a phosphorylation site on M.[26] Numerous isoforms of N were also observed and identified as caspase-3 and -6 cleavage products using recombinant N protein *in vitro*.[27]

Orthopoxviruses are among the largest and most complex of the animal viruses and have garnered a large amount of interest in regard to emerging viruses and bioterrorism. Senkevich and coworkers examined the proteins that form a stable complex which initiates cell entry and membrane fusion in vaccinia.[23] In addition to four proteins (A21, A28, H2 and L5) that had previously been implicated in these processes, four previously uncharacterized putative membrane proteins (A16, G3, G9, J5) were observed to form a stable heterocomplex as an entry-fusion complex. These proteins are conserved across the poxviruses suggesting nonredundant functions and that the cell entry mechanism for poxviruses evolved prior to the evolutionary split between vertebrate and invertebrate poxvirus species. Examination of the proteins present in vaccinia virus particles identified 63 virally-encoded proteins including two from open reading frames that had not previously been confirmed as being expressed.[25]

Beyond the detection and characterization of viral proteins, MS studies often uncover the presence of host proteins incorporated into viral particles. In the case of influenza, MS detected 36 host-encoded proteins

in addition to nine from viral genes.[28] These proteins included both cytoplasmic and membrane proteins from a range of protein families including cytoskeletal proteins, glycolytic enzymes and annexins, many of which have been reported to be present in virions from other virus families. Proteomics studies of SARS virions identified nearly 200 different host and viral proteins and used previously published interaction maps to identify nonstructural protein 3 as a significant viral factor for virion formation.[29] HIV viral particles derived from monocyte-derived macrophages likewise incorporate a fair number of host factors (more than 30), many of which have been previously found in exosomes, consistent with the notion that HIV utilizes the late endosome/multivesicular body pathway to bud from macrophages.[30] Although these studies may have implications in virus–host interactions, not all of these proteins are necessarily incorporated specifically. Thus, these studies often drive further, more defined experiments to examine the relationship between the virus and the incorporated host proteins.

The interaction between viral and host proteins presents an alternative avenue of viral–host protein studies with MS. Such studies have largely involved characterization of interactions of viral proteins with host proteins within infected cells by utilizing immunoprecipitation techniques.[31–34] One such example identified 14 cellular proteins that bound to the core of Hepatitis C virus (HCV) with MALDI MS.[31] Nine of these proteins were intermediate microfilament proteins, several of which have been implicated in the infection process of other viruses and their detection leads to a better understanding of the infection and pathogenesis process. Often host and viral proteins form functional complexes to enable viral replication and assembly such as in the cucumber necrosis tombusvirus replicase which was observed to consist of two viral proteins (p33 and p92) and four host proteins.[35] These host proteins consist of an Hsp70 homologue, an RNA-binding protein, pyruvate decarboxylase, and a 35-kDa acidic protein of unknown function. The identification of the cell surface receptor utilized by viruses is also possible through MS experiments of host–viral protein interactions. Dengue fever receptors on the surface of liver cells,[36] Japanese encephalitis virus receptors on C6/36 mosquito cells and adenovirus receptors on human ocular cells have been identified with MS experiments.[37,38]

Viral proteins are often co- or post-translationally modified to confer activity to the specific protein. These include phosphorylation,[39]

acetylation,[40,41] lipidation,[42,43] and glycosylation. Many of the studies thus far have identified the site of the modification while several others have also identified the functional role of such modifications. The acetylation of Tat from HIV, for instance, has been shown to affect splicing *in vivo* through preferential interaction with p32, a cofactor to splicing factor ASF/SF-2 resulting in sustained transcriptional activation and virus transcription elongation.[44] Three phosphorylation sites in hepatitis B virus (HBV) Cp protein are implicated in regulating genome encapsidation, capsid localization and maturation. Kang and coworkers found that Cp is phosphorylated by protein kinase A but not protein kinase C *in vitro* and that the phosporylation of Cp at Ser87 facilitates HBV core assembly.[45] This was further confirmed through a mutant that was highly deficient in forming cores although the same quantity of Cp was expressed.

A large number of viral proteins, especially those on the viral surface have been shown to be glycosylated. Glycosylation has been implicated in a wide range of functions including receptor binding, protein folding and immune avoidance. The latter of these functions appears to rely on the inherent heterogeneity of glycan incorporation which has also been shown with MS to evolve in the presence of neutralizing antibodies.[46] Broad surveys of the number of utilized glycosylation sites,[10,47] glycan identity[48] and heterogeneous glycan incorporation have also been performed utilizing MS.[49,50] The development of broadly neutralizing antibodies to HIV has proven difficult, thus thwarting vaccine production efforts. The glycan shield hypothesis posits that heterogeneity and rapid evolution of glycans on the viral surface protein serve as a major mechanism of immune avoidance in HIV. Several studies have examined the heterogenous incorporation of glycans in the gp120 protein of HIV, a protein responsible for binding of CD4 receptors for viral entry.[51,52] Liedtke *et al.* found a characteristic set of gp120 glycoforms which were dependant on the host-cell type as studied with MALDI MS.[50] Studies of gp120 expressed in cell lines have shown that nearly all glycosylation sites (25 of 26 in the HIV-I (SF2) examined) are occupied with glycans[53] with eight of these being sialylated and thirteen containing high mannose glycans.[54] Recent studies by Desaire and co-workers have shown the presence of a large number of heterogeneities in the glycosylation of two forms of recombinant HIV gp120 which can lead to further insight into HIV immune avoidance through the exploitation of glycan heterogeneity.[51,52] Additionally, the use of both ESI and MALDI MS showed added benefit for studies

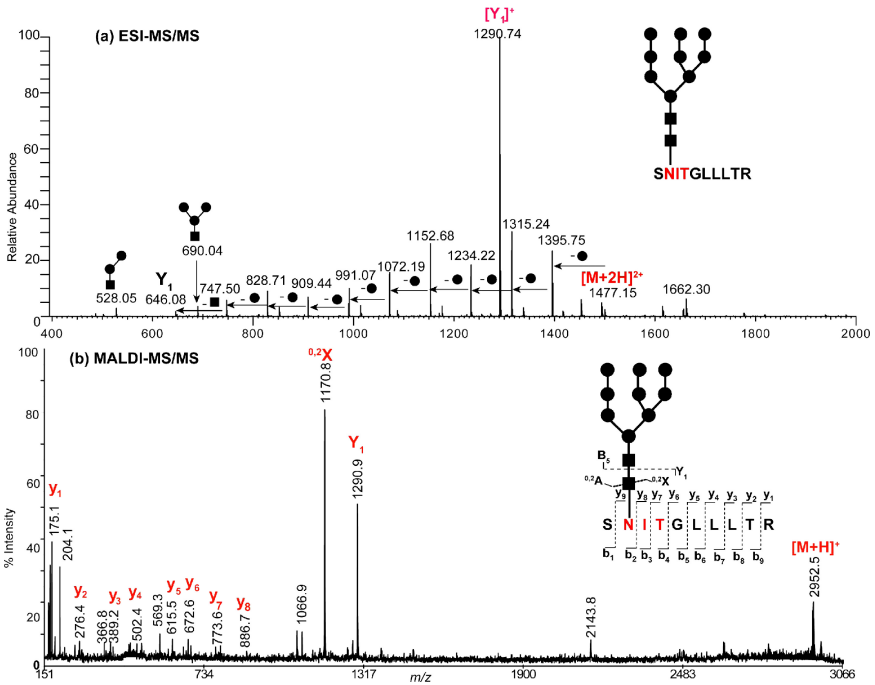


Fig. 2. A representative example of MS/MS data used to confirm the assigned glycopeptide compositions from HIV gp120. (**a**) ESI-MS/MS data for a doubly charged glycopeptide and (**b**) MALDI-MS/MS data of the singly charged form of the same glycopeptide (m/z 2952.55). Reprinted with permission from J. Irungu, E.P. Go, Y. Zhang, D.S. Dalpathado, H.X. Liao, B.F. Haynes and H. Desaire. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass Spectrom.* 1209–20 (2008).[52]

of the glycopeptides as the MS/MS spectra from each methodology were shown to be complementary in nature (Fig. 2). In this regard, ESI MS/MS provided information regarding the glycan moiety while MALDI MS/MS fragmented along the peptide backbone.

During viral infection, the functioning of the host cells is dramatically modulated. Changes in protein and metabolite levels may be examined through the use of quantitative and semi-quantitative methodologies. Utilizing stable isotope labeling, Suizdak and co-workers reported the analysis of 1500 host proteins and 1000 metabolites of which 200 proteins and 30 metabolites were significantly up or down regulated in Drosophila cells in

response to Flock House Virus infection.[55] Studies of SARS-infected Vero cells[56] and Hepatitis C virus infected Huh7 cells[57] also detected expression level variations in 186 (of 355) and 150 (of 1036) proteins respectively. Additional studies have examined the cellular targets of viral immune modulators[58] and protein expression patterns in astrocytes that express HIV Tat protein to examine its role in astrocyte survival.[59] The largest over expression of cellular proteins in Vero cells during African swine fever infection occurred in several redox-related proteins which would suggest functional roles of these proteins during infection in relation to modulation of transcription and apoptosis.[60] The examination of host reactions to viral infection present a means to further study viral–host interactions as well as the means by which viruses take control of the cellular machinery.

These broad-scale proteomic studies represent a large fraction of the -omics based approaches to molecular virology although similar approaches have also been applied to several highly directed studies. Based on biochemical experiments, HIV has been suggested to bud from specific lipid membrane microdomains, lipid rafts. In support of this, Bruuger found that the lipid composition of the HIV envelope is similar in composition to these detergent-resistant membrane microdomains.[61] This finding not only suggests a means and location of viral budding but also provides evidence for the existence of lipid rafts in living cells. The existence of novel or unknown viruses may also be confirmed by MS. Cooper and coworkers identified an uncharacterized infection from tobacco plants after isolating proteins that appeared to be differentially expressed in plant extracts that were infected with the unknown virus and uninfected plants.[62] Following HPLC ESI MS/MS, mass spectra were submitted to a SEAQUEST database. The results allowed for the identification of the causative agent as potato virus X which presents an interesting and possibly powerful tool for biomedical and bioterrorism surveillance once further developed.

4. Macromolecular Mass Spectrometry

Modern ionization strategies are capable of ionizing not only intact proteins but also non-covalent complexes. This enables the study of macromolecular protein complexes present in viral particles and even intact viruses.[3,63] Such experiments can be invaluable in the study of complex subunit stoichiometry as well as protein–protein and protein–ligand interactions. These experiments utilize ESI almost exclusively as

it is a softer technique allowing ionization under near physiological conditions.[3] Furthermore, the multiple charge states obtained from electrospray ionization simplify mass determination of large proteins. Several mass separation/detection strategies including traditional ToF are used to then analyze the highly charged complexes, which often reach into the hundreds of kilodaltons and beyond a megadalton. Beyond ToF, ion mobility and charge reduction MS have also been shown to be remarkably powerful tools for the study of viral macromolecular complexes.

The ionization of macromolecular protein complexes relies on ESI at a reduced voltage, supplemental gas flow and ion source temperature with a relative increase of pressures in the guiding ion optics to further 'soften' the ionization process. As a result of these changes, such native spray experiments do not rely on liquid chromatographic separations and require the use of a volatile buffer system such as ammonium acetate to contain the analytes of interest. These experiments generally require relatively pure and salt-free samples which can be obtained via traditional protein purification strategies and use of spin columns which use size exclusion principles to exchange buffer systems rather efficiently.

Non-covalent MS has shown great utility in examining the native stoichiometry of protein complexes. The native tetrameric, glycosylated form of matrix protein from borna disease virus (BDV) was isolated and detected from the brain material of an infected horse. This protein is the virus-specific component for viral attachment and also shows marked inhibition of BDV infection *in vitro* when the tetrameric assembly was added to culture media.[64] Larger complexes from Hepatitis B virus (HBV) were examined using native mass spectrometry as well. HBV capsids are unusual in that they exist in two distinct icosahedral forms consisting of 90 or 120 capsid dimers. Using mass spectrometry, the molecular weight of complete, intact lattices were measures to within 0.1% of the expected values of ~3 and 4 MDa respectively. No particles with fewer than 90 or more than 120 dimers were detected.[65]

Bacteriophages have attracted a sizable amount of interest in regard to viral mass spectrometry with several groups studying the protein complexes present through the course of the bacteriophage replication cycle. In regard to intact complex studies, several studies have examined the oligomeric states of the portal complex from the phages P22, Phi-29 and SPP1. These complexes range in mass from 430 kDa to one MDa and were unambiguously determined to form exclusively 12-mers in Phi-29

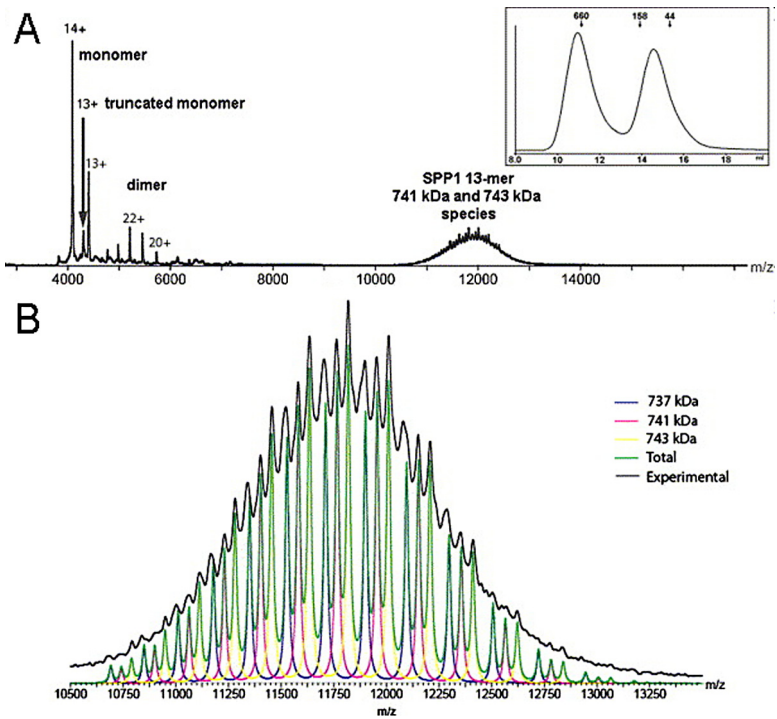


Fig. 3. ESI MS of reconstituted SPP1 portal. (A) The full-range ESI-MS spectrum showing monomer ions (at m/z 4000) and ions originating from the tridecamer at m/z 12,000. (B) Simulated (curve-fitted) mass spectrum of 13-mer signals centered around m/z 12,000. The green line indicates the sum of the three simulated spectra of these three different 13-mers, showing excellent agreement with the experimental spectrum (in black). Adapted with permission from A. Poliakov, E. van Duijn, G. Lander, C.Y. Fu, J.E. Johnson, P.E. Prevelige, Jr. and A.J. Heck. Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. *J. Struct. Biol.* 371–83 (2007).[66]

and 13-mers in SPP1 (Fig. 3).[66] *In vitro* assembled P22 portals, however, have been observed as a mixture of 11- and 12-mer complexes suggesting plasticity in assembly.[66] The phage terminase is used to package DNA through the portal assembly and into the bacteriophage particles. In P22, this terminase consists of a large subunit (gp2), which contains nuclease and ATPase activities, and a small subunit (gp3) which recognizes DNA and self-assembles into a stable olgometric ring shown by native MS to contain nine subunits.[67] A mutant (A112T) was found to form 10-mer

rings despite subunit folding indistinguishable from the wt gp3 again suggesting plasticity in higher order assembly. Following DNA packaging, the terminase complex leaves the phage particle prior to the formation of the portal-tail complex. This complex has also been studied with MS in P22 and was observed to form a 12:12 portal:gp4 complex with a molecular mass of 1.050 MDa.[68] Following the application of tandem MS, gp4 was found to be associated as dimeric products. Binding of gp4 was observed to induce a conformational change in the portal ring leading to the proposal that this conformational shift stabilizes the newly-filled bacteriophage particles thus marking the end of phage morphogenesis.

Many current experiments utilize highly charged ions to study macromolecular complexes with ToF, but applications of charge reduced electrospray mass spectrometry have shown to be beneficial for the determination of the physical size of macromolecular complexes. In these experiments, multiple charged ions produced by ESI are immediately swept by air and CO_2 gas flow into a neutralizing chamber containing a ^{210}Po source of α-radiation which generates charge-reducing agents (complex mixture of ions and radicals of the air components).[69] The charge reduced ions are then separated by gas-phase electrophoretic mobility molecular analysis (GEMMA) by a differential mobility analyzer to which a small electrical field is homogeneously applied. Following the gas phase separation and detection of the ions, the molecular mass and electrophoretic mobility diameter of complexes may be calculated. This technique was used to report the size of intact human rhinovirus as 29.8 ± 0.3 nm, which is consistent with electron microscopy data.[70] The broad peak shape shown in Fig. 4A may be explained by the incomplete desolvation of the virus while heat inactivation of the virus results in the shedding of capsid (VP4) oligomers and slight swelling of the heat-treated capsid is indicative of water intrusion into the viral capsid.[70] Similar methodologies were applied to the comparative study of bacteriophage M2 and T2 and T4 capsid heads.[71] Bacteriophage M2 was found to have a diameter of 24.13 ± 0.06 nm and remained viable following the electrospray process. T2 and T4 capsid heads were observed as having diameters of 87.03 ± 0.18 nm and 88.23 ± 1.02 nm respectively, which is a statistically significant difference. No viable T2 or T4 was detected following electrospray, which is believed to be a result of mechanical deformation within the electrospray droplets whereas those from M2 are small enough to not be subject to such stresses. Additional studies of cowpea chlorotic mottle

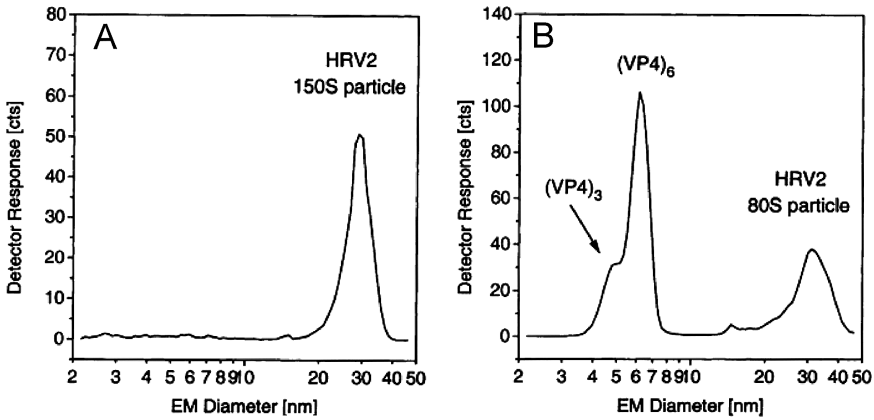


Fig. 4. GEMMA spectra of (A) infectious, intact human rhinovirus and (B) heat-degraded viral particles. (VP4)6 and (VP4)3 are nonspecific oligomers of the viral capsid protein VP4 and are released during the heat treatment. Reprinted with permission from G. Bacher, W.W. Szymanski, S.L. Kaufman, P. Zollner, D. Blaas and G. Allmaier. Charge-reduced nano electrospray ionization combined with differential mobility analysis of peptides, proteins, glycoproteins, noncovalent protein complexes and viruses. *J. Mass Spectrom.* 1038–52 (2001).[70]

virus (CCMV) presented a measured viral diameter of 25.4 nm which is ~10% less than the X-ray crystal structure indicating that desolvation leads to partial viral collapse.[72] A measured density of 0.89 g/mL is larger than typical protein complexes and believed to be due to the density of RNA packaging within the virus over proteins in complexes.

5. Structural Studies

Beyond identifying what biological molecules are present, MS may also be applied to studying structural features of viral proteins. Several methodologies have been applied to the study of viral protein structure including side chain derivatization, crosslinking, limited proteolysis and hydrogen–deuterium exchange (HDX) MS. In addition to these experiments, which are designed to examine protein structure, many of the -omics approaches above often point to structural elements such as glycosylation, protein–protein interactions, *etc.* and results from those experiments may also be used to study viral protein structure, albeit in a more indirect manner.

In HDX experiments, proteins, protein complexes or intact viral particles are diluted or solution exchanged into a deuterated buffer system. Amide protons in the protein backbone then undergo exchange with solvent deuterons at a rate that is dependent on solvent accessibility and protein dynamics.[73,74] This exchange reaction is quenched by the addition of formic acid at prescribed intervals (to pH ~2.4) and rapidly frozen. Samples are thawed and proteolytically digested with pepsin immediately prior to analysis to reduce the opportunity for back exchange (when hydrogen replaces previously exchanged deuterium). By observing the mass shift for each detected peptide from the protein of interest, the structural stability and solvent accessibility of the native protein may be examined.

In studying the molecular motors from bacteriophages responsible for packaging genomic material into a preformed capsid prohead, HDX MS was used to examine the interactions of the hexameric ATPase P4 with the ϕ12 procapsid.[75] P4 was found to associate by its C-terminal face to the capsid prohead. This interaction also stabilized the subunit interfaces, which are prone to ring opening in solution. The proposed model of the structural interface presents a means for both packaging processing and conferring RNA selectivity.

Structural elements of HIV and SIV Nef proteins have been difficult to study as the proteins are not readily amenable to NMR or X-ray crystallography as a result of aggregation of the full-length proteins at high concentrations. A highly solvent-protected core for both HIV and SIV Nef proteins was observed in HDX experiments which is in line with previously demonstrated NMR and X-ray studies of truncated HIV Nef.[76] Many of the regions outside of the core, however, were readily exchanged and are thus highly solvent exposed outside of the N- and C-terminal regions and a large loop region that emanates from the central core which suggest the presence of more structured elements. Structural studies of viral dynamics are also readily accessed with HDX MS as will be discussed in the following section.

An alternative means with which to examine the dynamics or surface presentation of viral proteins involves the use of limited proteolysis. In these experiments, purified viruses or proteins are digested with an endoprotease such as chymotrypsin, Glu-C, pepsin or trypsin. The proteolysis is then quenched, often by the addition of an acid, at selected time points similar to those used in HDX and before complete digestion. The digested samples may then be analyzed via MS to study the regions of the proteins

that are dynamic and exposed to the protein/viral surface as such exposure would be required for protease cleavage.

Intact reovirus, which contains a double-layered capsid core composed of sigma3 and mu1C proteins is proteolytically processed during entry to produce an infectious subviral particle, was examined with several endoproteases to study the proteolytic events involved in this processing.[77] Most of the tested proteases were found to cut the sigma3 protein between amino acids 217 and 238 along with a cleavage of the C-terminal of mu1C. Beyond this C-terminal peptide of mu1C, whose processing had been well described, no other digestion products were observed until sigma3 was entirely digested. This suggests that sigma3 is rapidly cleaved in addition to the C-terminal trimming of mu1C at this hypersensitive region during viral entry, which had not previously been well defined. A combination of limited proteolysis and mutational studies were used to examine the nonstructural glycoprotein NSP4 of rotavirus, which serves as a receptor for inner capsid particle (ICP) budding during virus maturation.[78] Digestion revealed a protease-sensitive region at the C-terminal while mutations demonstrated that 17–20 amino acids at the C-terminal are required for ICP binding which is consistent with models where the protein forms a central coiled tetramer with the C-terminal interacting with the ICP binding sites. Additional limited proteolyis experiments have examined the viral spike glycoprotein of mouse hepatitis virus strain A59 which is responsible for receptor binding and membrane fusion.[79] The heptad repeat regions (HR1 and HR2) were found to occur in the membrane fusion unit complex in an antiparallel fashion where such a conformation would bring the proposed fusion peptide, located in the N-terminal domain of HR1, and the transmembrane anchor into close proximity.

Chemical labeling and cross-linking also can provide valuable insight into viral protein structure. For instance, cysteine residues may be labeled with iodoacetamide as has been applied to the study of E1 and E2 glycoproteins from the Sindbis virus.[80] The modification of these proteins identified those residues which were accessible and those which were involved in disulfide bonds or otherwise inaccessible. In addition, this labeling greatly reduced infectivity but not membrane fusion suggesting that a rearrangement of disulfide bonds is required for infectivity. Lysine residues in the shell glycoproteins E1 and E2 in Sindbis virus were biotinylated by Sharp, *et al.* to observe that a large number of lysines from E2 were modified (and therefore available for modification) while only a single site

in E1 was modified, indicating that E1 is almost completely buried within the virus structure.[81]

Chemical cross-linking utilizes reagents that contain reactive moieties on each end that are capable of covalently attaching to specific functional groups such as lysine residues in proteins. The use of cross-linkers with different arm lengths can provide distance constraints between reactive groups within and between protein subunits. In one study, Kang, *et al.* chemically cross-linked empty procapsid shells from bacteriophage P22 to identify inter- and intra-subunit interactions with MS.[82] A single loop region was observed to cross-link between two subunits in procapsids (K183–K183) with reagents with similar distance constraints (12 Å and 11.4 Å) while a shorter cross-linker (6.4 Å length) linked K175 in one subunit to K183 in another subunit. Coupled with mutational studies, the results were suggestive of intrinsic flexibility within the loop region and close proximity between loop regions of adjacent subunits.

6. Viral Dynamics

Beyond merely observing a static state, MS observations from multiple time points have been utilized to examine dynamic events in the viral life-cycle including dynamic protein movements, structural maturation and chemical modification. These studies observe multiple 'snapshots' using the techniques discussed previously and largely have focused on the structural dynamics of viral proteins.

Limited proteolysis has been applied to the study of several aspects of viral dynamics. Studies by Bothner, *et al.* found that the first digestion products from flock house virus particles were from portions of the capsid subunits that had been shown to be internal to the X-ray structure but had been implicated in RNA neutralization release and delivery.[83] These results suggested that these regions of the protein were transiently exposed to the viral surface, which was not previously predicted. Similarly, free dimer and assembled capsid proteins from hepatitis B virus (HBV) were readily cleaved by thermolysin and trypsin at residues 127–128 which indicated that this region was dynamic and exposed to the surface and allowed the measurement of the conformational equilibria and rate of conversion between protein conformers in the viral complex.[84]

The dynamic nature of human rhinovirus (HRV) was studied under limited proteolysis and isotope labeling conditions in the presence of

several antiviral compounds, a neutralizing antibody and several drug-binding cavity mutations.[85] It was found that the hydrophobic antiviral compounds, neutralizing antibodies and mutations in the canyon region of the capsid protein blocked the normal breathing of the HRV virion thus illustrating that this region is a main determinant in the breathing process.

The application of HDX experiments to protein conformational changes presents an expansion of the structural studies presented above and have covered a relatively wide berth to this point. Capsid swelling in brome mosaic virus and CCMV was found to rely on conformational changes in both the N- and C-terminal arms.[86] HRV HDX experiments produced results in line with the previously reported crystal structure of the capsid proteins (VP1–4) and suggested that the high exchange rates in the N-terminal of VP1 and most of VP4, which suggest high flexibility, may point to potential roles in virus uncoating.[87] In bacteriophage P22, several studies of capsid expansion/maturation have been reported.[88,89] The N-terminal domain of the coat protein was observed to be much more flexible in the empty procapsid shell than the mature capsids while the C-terminal domain maintained a similar exchange rate.[89] The loop region that connected the two domains remained flexible in both forms such that the results point to the global stabilization of the N-terminal domain to be a key component to the P22 maturation process. The relationship between the capsid and the scaffolding protein was examined in bacteriophage Ø29 to reveal a highly dynamic and cooperative opening of the scaffolding molecules in the helix-loop-helix region, where the cooperative opening rate was increased in the procapsid bound form (Fig. 5) suggesting that this region interacts with the capsid protein.[90] The maturation pathway of HIV has also been partially elucidated with HDX MS experiments.[91] By examining the exchange pattern of immature and mature virions, maturation induced interactions between the N- and C-terminal domains in half of the capsid molecules indicated not only the restructuring of intermolecular interactions for the collapsing core but also that only half of the capsid proteins are assembled into the mature, conical core.[92]

The ability to examine the dynamics of viruses with MS does not only involve structural studies. Although such examples are far fewer, the chemical maturation and evolution of viral proteins presents an interesting aspect for future examination. The dynamics of glycan incorporation in SIV was observed to shift the mass of the envelope glycoprotein Env in response to the immune system.[46] The molecular mass of the protein

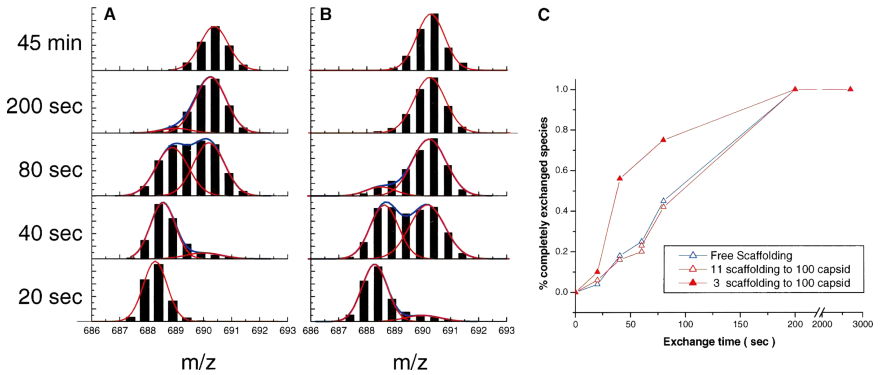


Fig. 5. The opening kinetics of free and procapsid-bound scaffolding protein residues 20–31 in (A) recombinant scaffolding proteins and (B) and procapsid particles with 3:100 scaffolding: capsid protein ratio exchange in a bi-modal manner. (C) The faster exchange for scaffolding molecules when bound with capsid suggests a co-operative interaction with capsid molecules. Reprinted with permission from C.Y. Fu and P.E. Prevelige, Jr. Dynamic motions of free and bound O29 scaffolding protein identified by hydrogen deuterium exchange mass spectrometry. *Protein Sci.* 731–43 (2006).[90]

was observed to increase from 102.3 to 103.5 kDa and was tracked to a change in the glycosylation pattern in variable region 1, which allowed the new virus to escape antibody recognition. Beyond evolution from external stresses such as immune avoidance, virus chemical maturation has also been followed with MS. The core protein from hepadnavirus nucleocapsids has been observed to be phosphorylated on at least six sites in immature virions while the protein is completely dephosphorylated in the mature, secreted virions.[93] This maturation suggests that the dephosphorylation of mature nucleocapsids may trigger envelopment and secretion as phosphorylation is required for the pregenomic RNA.

7. Conclusions

Mass spectrometry provides a toolbox suitable for identifying and characterizing the biological macromolecules required for viral replication. Viral and host proteins may be detected as well as their interactions, structural elements and dynamics as a means to further the understanding of the viral life cycle. Owing to the inherent complexity of biological studies, MS sheds important insight into molecular aspects of viral dynamics and is often used in concert with various molecular and structural biological

approaches including mutational analysis, protein and genetic analysis, electron microscopy, nuclear magnetic resonance (NMR) and crystallography to complement and enhance these other techniques to study the complexities and intricacies of the virus–host interactions.

References

1. D.A. Skoog, F.J. Holler and T.A. Nieman. *Principles of Instrumental Analysis.* Brooks/Cole, New York (1997).
2. M.L. Gross and R.M. Caprioli. Mass analysis and associated instrumentation. In *The Encyclopedia of Mass Spectrometry*, Gross, M.L.,Caprioli, R.M., Eds. Elsevier Science: Oxford (2004).
3. V.L. Morton, P.G. Stockley, N.J. Stonehouse and A.E. Ashcroft. Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass Spectrom Rev.* **27**:575–595 (2008).
4. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**:64–71 (1989).
5. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida and T. Yoshida. Protein and polymer analysis up to *m/z* 100.000 by laser ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**:151–153 (1988).
6. M. Karas and F. Hillenkamp. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**:2299–301 (1988).
7. N.L. Kelleher, Top-down proteomics. *Anal. Chem.* **15**:197A–203A (2004).
8. L. Sleno and D.A. Volmer. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **39**:1091–112 (2004).
9. V.H. Wysocki, K.A. Resing, Q. Zhang and G. Cheng. Mass spectrometry of peptides and proteins. *Methods* **35**:211–22 (2005).
10. Y.J. Kim, A. Freas and C. Fenselau. Analysis of viral glycoproteins by MALDI-TOF mass spectrometry. *Anal. Chem.* **73**:1544–8 (2001).
11. M. Fitzgibbon, Q. Li and M. McIntosh. Modes of inference for evaluating the confidence of peptide identifications. *J. Proteome Res.* **7**:35–9 (2008).
12. G. Lubec, L. Afjehi-Sadat, J.W. Yang and J.P. John. Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog. Neurobiol.* **77**:90–127 (2005).
13. J.R. Chang, A. Poliakov, P.E. Prevelige, J.A. Mobley and T. Dokland. Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology* **370**:352–61 (2008).
14. L. Eyer, R. Pantucek, Z. Zdrahal, H. Konecna, P. Kasparek, V. Ruzickova, L. Hernychova, J. Preisler and J. Doskar. Structural protein analysis of the polyvalent staphylococcal bacteriophage 812. *Proteomics* **7**:64–72 (2007).
15. M. Borriss, T. Lombardot, F.O. Glockner, D. Becher, D. Albrecht and T. Schweder. Genome and proteome characterization of the psychrophilic Flavobacterium bacteriophage 11b. *Extremophiles* **11**:95–104 (2007).

16. X. Xie, L. Xu and F. Yang. Proteomic analysis of the major envelope and nucleocapsid proteins of white spot syndrome virus. *J Virol.* **80**:10615–23 (2006).
17. Z. Li, Q. Lin, J. Chen, J.L. Wu, T.K. Lim, S.S. Loh, X. Tang and C.L. Hew. Shotgun identification of the structural proteome of shrimp white spot syndrome virus and iTRAQ differentiation of envelope and nucleocapsid subproteomes. *Mol. Cell Proteomics* **6**:1609–20 (2007).
18. O. Krokhin, Y. Li, A. Andonov, H. Feldmann, R. Flick, S. Jones, U. Stroeher, N. Bastien, K.V. Dasuri, K. Cheng, J.N. Simonsen, H. Perreault, J. Wilkins, W. Ens, F. Plummer and K.G. Standing. Mass spectrometric characterization of proteins from the SARS virus: a preliminary report. *Mol. Cell Proteomics* **2**:346–56 (2003).
19. W. Song, Q. Lin, S.B. Joshi, T.K. Lim and C.L. Hew. Proteomic studies of the Singapore grouper iridovirus. *Mol. Cell Proteomics* **5**:256–64 (2006).
20. H.C. Liu, E.J. Soderblom and M.B. Goshe. A mass spectrometry-based proteomic approach to study Marek's Disease Virus gene expression. *J. Virol. Methods* **135**:66–75 (2006).
21. C.M. O'Connor and D.H. Kedes. Mass spectrometric analyses of purified rhesus monkey rhadinovirus reveal 33 virion-associated proteins. *J. Virol.* **80**:1574–83 (2006).
22. D. Chelius, A.F. Huhmer, C.H. Shieh, E. Lehmberg, J.A. Traina, T.K. Slattery and E. Pungor Jr. Analysis of the adenovirus type 5 proteome by liquid chromatography and tandem mass spectrometry methods. *J. Proteome Res.* **1**:501–13 (2002).
23. T.G. Senkevich, S. Ojeda, A. Townsley, G.E. Nelson and B. Moss. Poxvirus multiprotein entry-fusion complex. *Proc. Natl. Acad. Sci. USA* 18572–7 (2005).
24. N.P. Manes, R.D. Estep, H.M. Mottaz, R.J. Moore, T.R. Clauss, M.E. Monroe, X. Du, J.N. Adkins, S.W. Wong and R.D. Smith. Comparative proteomics of human monkeypox and vaccinia intracellular mature and extracellular enveloped virions. *J. Proteome Res.* 960–8 (2008).
25. J.D. Yoder, T.S. Chen, C.R. Gagnier, S. Vemulapalli, C.S. Maier and D.E. Hruby. Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. *Virol. J.* 10 (2006).
26. R. Zeng, H.Q. Ruan, X.S. Jiang, H. Zhou, L. Shi, L. Zhang, Q.H. Sheng, Q. Tu, Q.C. Xia and J.R. Wu. Proteomic analysis of SARS associated coronavirus using two-dimensional liquid chromatography mass spectrometry and one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by mass spectroemtric analysis. *J. Proteome Res.* 549–55 (2004).
27. W. Ying, Y. Hao, Y. Zhang, W. Peng, E. Qin, Y. Cai, K. Wei, J. Wang, G. Chang, W. Sun, S. Dai, X. Li, Y. Zhu, J. Li, S. Wu, L. Guo, J. Dai, J. Wang, P. Wan, T. Chen, C. Du, D. Li, J. Wan, X. Kuai, W. Li, R. Shi, H. Wei, C. Cao, M. Yu, H. Liu, F. Dong, D. Wang, X. Zhang, X. Qian, Q. Zhu and F. He. Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus, *Proteomics* 492–504 (2004).
28. M.L. Shaw, K.L. Stone, C.M. Colangelo, E.E. Gulcicek and P. Palese. Cellular proteins in influenza virus particles. *PLoS Pathog* e1000085 (2008).

29. B.W. Neuman, J.S. Joseph, K.S. Saikatendu, P. Serrano, A. Chatterjee, M.A. Johnson, L. Liao, J.P. Klaus, J.R. Yates, 3rd, K. Wuthrich, R.C. Stevens, M.J. Buchmeier and P. Kuhn. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* 5279–94 (2008).
30. E. Chertova, O. Chertov, L.V. Coren, J.D. Roser, C.M. Trubey, J.W. Bess, Jr., R.C. Sowder, 2nd, E. Barsov, B.L. Hood, R.J. Fisher, K. Nagashima, T.P. Conrads, T.D. Veenstra, J.D. Lifson and D.E. Ott. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* 9039–52 (2006).
31. S.M. Kang, M.J. Shin, J.H. Kim and J.W. Oh. Proteomic profiling of cellular proteins interacting with the hepatitis C virus core protein. *Proteomics* 2227–37 (2005).
32. T.J. Taylor and D.M. Knipe. Proteomics of herpes simplex virus replication compartments: association of cellular DNA replication, repair, recombination, and chromatin remodeling proteins with ICP8. *J. Virol.* 5856–66 (2004).
33. Y. Gao, K. Colletti and G.S. Pari. Identification of human cytomegalovirus UL84 virus- and cell-encoded binding partners by using proteomics analysis. *J. Virol.* 96–104 (2008).
34. S. Noisakran, S. Sengsai, V. Thongboonkerd, R. Kanlaya, S. Sinchaikul, S.T. Chen, C. Puttikhunt, W. Kasinrerk, T. Limjindaporn, W. Wongwiwat, P. Malasit and P.T. Yenchitsomanus. Identification of human hnRNP C1/C2 as a dengue virus NS1-interacting protein. *Biochem. Biophys. Res. Commun.* 67–72 (2008).
35. S. Serva and P.D. Nagy. Proteomics analysis of the tombusvirus replicase: Hsp70 molecular chaperone is associated with the replicase and enhances viral RNA replication. *J. Virol.* 2162–9 (2006).
36. C. Thepparit and D.R. Smith. Serotype-specific entry of dengue virus into liver cells: identification of the 37-kilodalton/67-kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor. *J. Virol.* 12647–56 (2004).
37. E. Wu, S.A. Trauger, L. Pache, T.M. Mullen, D.J. von Seggern, G. Siuzdak and G.R. Nemerow. Membrane cofactor protein is a receptor for adenoviruses associated with epidemic keratoconjunctivitis. *J. Virol.* 3897–905 (2004).
38. A. Gaggar, D. Shayakhmetov and A. Lieber. Identifying functional adenovirus-host interactions using tandem mass spectrometry. *Methods Mol. Med.* 141–55 (2007).
39. H. Chen, A. Gill, B.K. Dove, S.R. Emmett, C.F. Kemp, M.A. Ritchie, M. Dee and J.A. Hiscox. Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance. *J. Virol.* 1164–79 (2005).
40. L.V. Kordyukova, M.V. Serebryakova, L.A. Baratova and M. Veit. S acylation of the hemagglutinin of influenza viruses: mass spectrometry reveals site-specific attachment of stearic acid to a transmembrane cysteine. *J. Virol.* 9288–92 (2008).
41. P. Alfonso, J.I. Quetglas, J.M. Escribano and C. Alonso. Protein pE120R of African swine fever virus is post-translationally acetylated as revealed by post-source decay MALDI mass spectrometry. *Virus Genes* 81–5 (2007).
42. J.D. Yoder, T. Chen and D.E. Hruby. Sequence-independent acylation of the vaccinia virus A-type inclusion protein. *Biochemistry* 8297–302 (2004).

43. J. Hensel, M. Hintz, M. Karas, D. Linder, B. Stahl and R. Geyer. Localization of the palmitoylation site in the transmembrane protein p12E of Friend murine leukaemia virus. *Eur. J. Biochem.* 373–80 (1995).
44. R. Berro, K. Kehn, C. de la Fuente, A. Pumfery, R. Adair, J. Wade, A.M. Colberg-Poley, J. Hiscott and F. Kashanchi. Acetylated Tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. *J. Virol.* 3189–204 (2006).
45. H.Y. Kang, S. Lee, S.G. Park, J. Yu, Y. Kim and G. Jung. Phosphorylation of hepatitis B virus Cp at Ser87 facilitates core assembly. *Biochem. J.* 311–7 (2006).
46. B. Chackerian, L.M. Rudensey and J. Overbaugh. Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. *J. Virol.* 7719–27 (1997).
47. J. de Jesus Perez, S. Juarez, D. Chen, C.L. Scott, L.M. Hartweck, N.E. Olszewski and J.A. Garcia. Mapping of two O-GlcNAc modification sites in the capsid protein of the potyvirus Plum pox virus. *FEBS Lett.* 5822–8 (2006).
48. D. Falzarano, O. Krokhin, G. Van Domselaar, K. Wolf, J. Seebach, H. J. Schnittler and H. Feldmann. Ebola sGP — the first viral glycoprotein shown to be C-mannosylated. *Virology* 83–90 (2007).
49. R.E. Iacob, I. Perdivara, M. Przybylski and K.B. Tomer. Mass spectrometric characterization of glycosylation of hepatitis C virus E2 envelope glycoprotein reveals extended microheterogeneity of N-glycans. *J. Am. Soc. Mass Spectrom* 428–44 (2008).
50. S. Liedtke, R. Geyer and H. Geyer. Host-cell-specific glycosylation of HIV-2 envelope glycoprotein. *Glycoconj. J.* 785–93 (1997).
51. E.P. Go, J. Irungu, Y. Zhang, D.S. Dalpathado, H.X. Liao, L.L. Sutherland, S.M. Alam, B.F. Haynes and H. Desaire. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome. Res.* 1660–74 (2008).
52. J. Irungu, E.P. Go, Y. Zhang, D.S. Dalpathado, H.X. Liao, B.F. Haynes and H. Desaire. Comparison of HPLC/ESI-FTICR MS Versus MALDI-TOF/TOF MS for Glycopeptide Analysis of a Highly Glycosylated HIV Envelope Glycoprotein. *J. Am. Soc. Mass Spectrom* 1209–20 (2008).
53. X. Zhu, C. Borchers, R.J. Bienstock and K.B. Tomer. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* 11194–204 (2000).
54. J.M. Cutalo, L.J. Deterding and K.B. Tomer. Characterization of glycopeptides from HIV-I(SF2) gp120 by liquid chromatography mass spectrometry. *J. Am. Soc. Mass Spectrom* 1545–55 (2004).
55. E.P. Go, W.R. Wikoff, Z. Shen, G. O'Maille, H. Morita, T.P. Conrads, A. Nordstrom, S.A. Trauger, W. Uritboonthai, D.A. Lucas, K.C. Chan, T.D. Veenstra, H. Lewicki, M.B. Oldstone, A. Schneemann and G. Siuzdak. Mass spectrometry reveals specific and global molecular transformations during viral infection. *J. Proteome Res.* 2405–16 (2006).

56. X.S. Jiang, L.Y. Tang, J. Dai, H. Zhou, S.J. Li, Q.C. Xia, J.R. Wu and R. Zeng. Quantitative analysis of severe acute respiratory syndrome (SARS)-associated coronavirus-infected cells using proteomic approaches: implications for cellular responses to virus infection. *Mol. Cell Proteomics* 902–13 (2005).
57. P. Mannova, R. Fang, H. Wang, B. Deng, M.W. McIntosh, S.M. Hanash and L. Beretta. Modification of host lipid raft proteome upon hepatitis C virus replication. *Mol. Cell Proteomics* 2319–25 (2006).
58. E. Bartee, A. McCormack and K. Fruh. Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. *PLoS Pathog.* e107 (2006).
59. C.B. Pocernich, D. Boyd-Kimball, H.F. Poon, V. Thongboonkerd, B.C. Lynn, J.B. Klein, V. Calebrese, A. Nath and D.A. Butterfield. Proteomics analysis of human astrocytes expressing the HIV protein Tat. *Brain Res. Mol. Brain Res.* 307–16 (2005).
60. P. Alfonso, J. Rivera, B. Hernaez, C. Alonso and J.M. Escribano. Identification of cellular proteins modified in response to African swine fever virus infection by proteomics. *Proteomics* 2037–46 (2004).
61. B. Brugger, B. Glass, P. Haberkant, I. Leibrecht, F.T. Wieland and H.G. Krausslich. The HIV lipidome: a raft with an unusual composition. *Proc. Natl. Acad. Sci. USA* 2641–6 (2006).
62. B. Cooper, D. Eckert, N.L. Andon, J.R. Yates and P.A. Haynes. Investigative proteomics: identification of an unknown plant virus from infected plants using mass spectrometry. *J. Am. Soc. Mass Spectrom.* 736–41 (2003).
63. S.D. Fuerstenau, W.H. Benner, J.J. Thomas, C. Brugidou, B. Bothner and G. Siuzdak. Mass Spectrometry of an Intact Virus. *Angew Chem. Int. Ed. Engl.* 541–544 (2001).
64. R. Stoyloff, A. Strecker, L. Bode, P. Franke, H. Ludwig and F. Hucho. The glycosylated matrix protein of Borna disease virus is a tetrameric membrane-bound viral component essential for infection. *Eur. J. Biochem.* 252–7 (1997).
65. C. Uetrecht, C. Versluis, N.R. Watts, W.H. Roos, G.J. Wuite, P.T. Wingfield, A.C. Steven and A.J. Heck, High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. USA* 9216–20 (2008).
66. A. Poliakov, E. van Duijn, G. Lander, C.Y. Fu, J.E. Johnson, P.E. Prevelige, Jr. and A.J. Heck. Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. *J. Struct. Biol.* 371–83 (2007).
67. D. Nemecek, E.B. Gilcrease, S. Kang, P.E. Prevelige, Jr., S. Casjens and G.J. Thomas, Jr. Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. *J. Mol. Biol.* 817–36 (2007).
68. K. Lorenzen, A.S. Olia, C. Uetrecht, G. Cingolani and A.J. Heck. Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly. *J. Mol. Biol.* 385–96 (2008).
69. G.P. Reischl, J.M. Mäkelä, R. Karch and J. Necid. Bipolar charging of ultrafine particles in the size range below 10 nm. *Journal of Aerosol Science* 931–949 (1996).

70. G. Bacher, W.W. Szymanski, S.L. Kaufman, P. Zollner, D. Blaas and G. Allmaier. Charge-reduced nano electrospray ionization combined with differential mobility analysis of peptides, proteins, glycoproteins, noncovalent protein complexes and viruses. *J. Mass Spectrom.* 1038–52 (2001).
71. C.J. Hogan, Jr., E.M. Kettleson, B. Ramaswami, D.R. Chen and P. Biswas. Charge reduced electrospray size spectrometry of mega- and gigadalton complexes: whole viruses and virus fragments. *Anal. Chem.* 844–52 (2006).
72. C.S. Kaddis, S.H. Lomeli, S. Yin, B. Berhane, M.I. Apostol, V.A. Kickhoefer, L.H. Rome and J.A. Loo. Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. *J. Am. Soc. Mass Spectrom.* 1206–16 (2007).
73. R.S. Molday, S.W. Englander and R.G. Kallen. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* 150–8 (1972).
74. S.W. Englander and N.R. Kallenbach. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev. Biophys.* 521–655 (1983).
75. J. Lisal, D.E. Kainov, T.T. Lam, M.R. Emmett, H. Wei, P. Gottlieb, A. G. Marshall and R. Tuma. Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. *Virology* 73–9 (2006).
76. J.M. Hochrein, T.E. Wales, E.C. Lerner, A.P. Schiavone, T.E. Smithgall and J.R. Engen. Conformational features of the full-length HIV and SIV Nef proteins determined by mass spectrometry. *Biochemistry* 7733–9 (2006).
77. Mendez, II, Y.M. She, W. Ens and K.M. Coombs. Digestion pattern of reovirus outer capsid protein sigma3 determined by mass spectrometry. *Virology* 289–304 (2003).
78. J.A. O'Brien, J.A. Taylor and A.R. Bellamy. Probing the structure of rotavirus NSP4: a short sequence at the extreme C terminus mediates binding to the inner capsid particle. *J. Virol.* 5388–94 (2000).
79. B.J. Bosch, R. van der Zee, C.A. de Haan and P.J. Rottier. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* 8801–11 (2003).
80. C.B. Whitehurst, E.J. Soderblom, M.L. West, R. Hernandez, M.B. Goshe and D.T. Brown. Location and role of free cysteinyl residues in the Sindbis virus E1 and E2 glycoproteins. *J. Virol.* 6231–40 (2007).
81. J.S. Sharp, S. Nelson, D. Brown and K.B. Tomer. Structural characterization of the E2 glycoprotein from Sindbis by lysine biotinylation and LC-MS/MS. *Virology* 216–23 (2006).
82. S. Kang, A.M. Hawkridge, K.L. Johnson, D.C. Muddiman and P.E. Prevelige, Jr. Identification of subunit-subunit interactions in bacteriophage P22 procapsids by chemical cross-linking and mass spectrometry. *J. Proteome Res.* 370–7 (2006).
83. B. Bothner, X.F. Dong, L. Bibbs, J.E. Johnson and G. Siuzdak. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J. Biol. Chem.* 673–6 (1998).
84. J.K. Hilmer, A. Zlotnick and B. Bothner. Conformational equilibria and rates of localized motion within hepatitis B virus capsids. *J. Mol. Biol.* 581–94 (2008).

85. N. Reisdorph, J.J. Thomas, U. Katpally, E. Chase, K. Harris, G. Siuzdak and T.J. Smith. Human rhinovirus capsid dynamics is controlled by canyon flexibility. *Virology* 34–44 (2003).
86. L. Wang, L.C. Lane and D.L. Smith. Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci.* 1234–43 (2001).
87. L. Wang and D.L. Smith. Capsid structure and dynamics of a human rhinovirus probed by hydrogen exchange mass spectrometry. *Protein Sci.* 1661–72 (2005).
88. R. Tuma, L.U. Coward, M.C. Kirk, S. Barnes and P.E. Prevelige, Jr. Hydrogen-deuterium exchange as a probe of folding and assembly in viral capsids. *J. Mol. Biol.* 389–96 (2001).
89. S. Kang and P.E. Prevelige, Jr. Domain study of bacteriophage p22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange. *J. Mol. Biol.* 935–48 (2005).
90. C.Y. Fu and P.E. Prevelige, Jr. Dynamic motions of free and bound O29 scaffolding protein identified by hydrogen deuterium exchange mass spectrometry. *Protein Sci.* 731–43 (2006).
91. J. Lanman and P.E. Prevelige, Jr. Kinetic and mass spectrometry-based investigation of human immunodeficiency virus type 1 assembly and maturation. *Adv. Virus Res.* 285–309 (2005).
92. J. Lanman, T.T. Lam, M.R. Emmett, A.G. Marshall, M. Sakalian and P. E. Prevelige, Jr. Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat. Struct. Mol. Biol.* 676–7 (2004).
93. D.H. Perlman, E.A. Berg, B. O'Connor P, C.E. Costello and J. Hu. Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. *Proc. Natl. Acad. Sci. USA* 9020–5 (2005).

Chapter 6

An Overview of Capsid Assembly Kinetics

J. Zachary Porterfield* and Adam Zlotnick*,†

1. Introduction

The capsid of an icosahedral virus is a magnificent example of nature's utility and precision. Capsids are self-assembling symmetrical structures. Viral infections can yield hundreds of thousands of correctly assembled virions per cell (Rueckert, 1996). For this to occur, the assembly process must be suitably rapid, robust, and well regulated. In this review we will focus on the kinetics of this assembly, which play a dominant role in determining the outcome of capsid subunit interactions.

The Montreal Biosphere (Fig. 1), designed by Buckminster Fuller, is a well-known example of a geodesic dome with icosahedral symmetry. In nature, icosahedral symmetry offers a selective advantage due to its inherent stability (Zandi *et al.*, 2004) and its ability to enclose a large volume using small subunits (Crick and Watson, 1956). Icosahedral symmetry has convergently evolved in completely unrelated viruses (e.g., α helical hepadnaviruses (Wynne *et al.*, 1999), HK97-fold bacteriophages (Wikoff *et al.*, 2000), and β barrel sobemo-viruses (Abad-Zapatero *et al.*, 1980) and in cellular protein complexes (e.g., human ferritin and bacterial pyruvate dehydrogenase). Geodesic domes naturally inspire questions about their construction. Virus capsid assembly is all the more fascinating as it is spontaneous and relies on physical interactions to guide the overall synthesis.

*University of Oklahoma Health Sciences Center, Department of Biochemistry, Oklahoma City, OK 73104.
†Indiana University, Department of Biology, Bloomington, IN 47405.

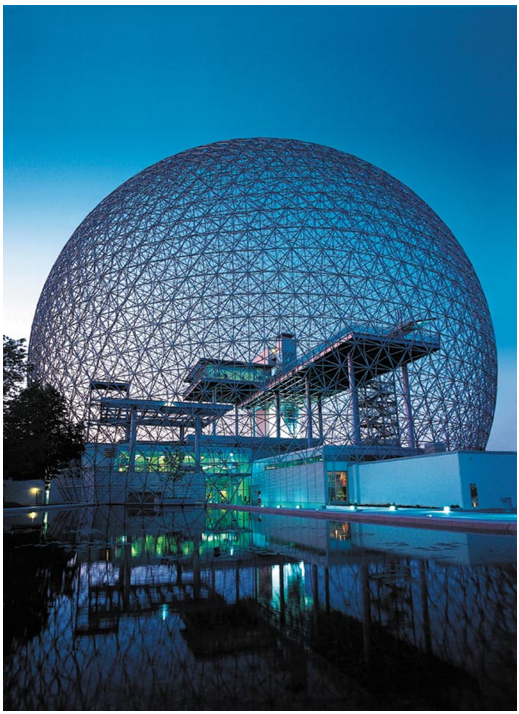


Fig. 1. A Geodesic dome (Courtesy of Environment Canada).

Self-assembly was cleverly demonstrated using a molecular model of poliovirus with intersubunit bonds mimicked by magnets. The model spontaneously assembles using shaking in lieu of thermal energy, as shown in Fig. 2 (Olson *et al.*, 2007). During the reaction, intermediates formed and fell apart until a complete, stable capsid was achieved. This experiment highlights two of the critical features of the self-assembly reaction: the complementarity of the components and the reversibility of association.

Consider one of the earliest *in vitro* observations of the kinetics of spherical capsid assembly: using neutron and small angle X-ray scattering, Cuillel and co-workers observed extremely rapid assembly of Brome Mosaic Virus (BMV) into a complex mixture of intermediates that slowly resolved into capsids (Cuillel *et al.*, 1983). This mirrors some elements of the mechanical study. However, in retrospect the preponderance of intermediates with BMV is likely to reflect kinetic traps that result from the extraordinary concentrations of protein required for these experiments.

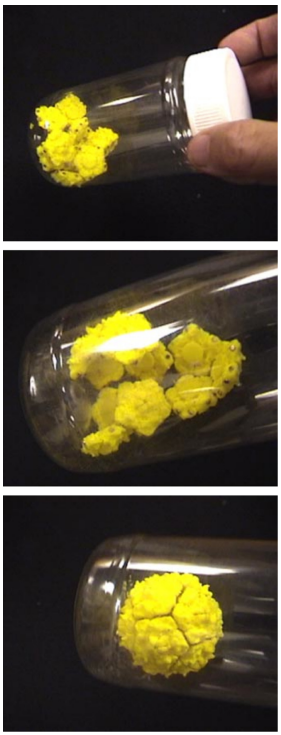


Fig. 2. A self-assembling model of poliovirus (Olson *et al.*, 2007).

2. Modeling Assembly

Interpretation of kinetic data requires a theoretical model. Virus replication is a complex process that can vary greatly from virus to virus. Capsid assembly in the context of the cell is difficult to model. A reductionist approach has led to studies of capsid assembly as a first step towards understanding the physical chemistry of the virus lifecycle. Some studies have expanded to include packaging of the genome during assembly. In all cases, however, simplifications must be made. The goal is to identify salient features of the assembly process without oversimplification. Two broad approaches to describing capsid formation are *stepwise assembly* and *protein condensation*. In stepwise assembly, subunits are added one at a time to a growing multimeric structure, until a completed capsid is achieved. In protein condensation the subunits bind to a template in a loose organization and then reorganize into an ordered structure.

2.1. *Stepwise Assembly*

In stepwise assembly, the formation of capsids is treated as a cascade of bimolecular reactions. Modeling capsid assembly as a cascade of low-order reactions is directly related to the crystallization and the polymerization of filaments. Though the mathematics describing these reactions are also similar, the key difference is that capsids are closed structures of defined size, unlike filaments or crystals. Polymerization of open-structures can theoretically proceed indefinitely given a single nucleation event. Each capsid, however, must be individually nucleated. The physical basis for the sigmoidal kinetics observed for crystallization and capsid assembly are thus very different.

In stepwise capsid assembly, an incoming subunit can be added at many positions along the growing structure. In this way a large number of different intermediate structures can arise. Where a typical virus capsid can be formed of hundreds of subunits, the number of potential intermediates that can form during assembly is combinatorially larger and can be computationally intractable. Two distinctly different techniques have been used to model assembly reactions. Classical kinetics evaluates large populations and has the advantage of relatively small computational overhead but requires that intermediates along the assembly path be explicitly defined. Coarse-grained molecular dynamics does not require definition of intermediates but does require careful definition of the assembly model, a limited pool of intermediates and substantial computer time for calculations. These two approaches, one derived from thermodynamics and the other from statistical mechanics, are complementary. The classical kinetic models simplify examination of bulk properties of assembly reactions for comparison to solution experiments. The molecular dynamic approach promotes a more careful examination of the forms of intermediates and how the dynamics of intermolecular interaction can affect assembly paths.

The simplest exposition of a classical kinetic model of assembly is to consider a single assembly path. The assembly of a dodecahedron from twelve subunits in the form of pentagons was described as a relevant model system (Zlotnick, 1994), see also Fig. 3. This geometrical simplification is analogous to picornavirus assembly from pentamers-of-trimers (Rueckert, 1996). For a dodecahedron composed of 12 pentagonal subunits, a total of 73 intermediates are possible (Endres *et al.*, 2005). As originally modelled, only the most thermodynamically stable intermediates were examined.

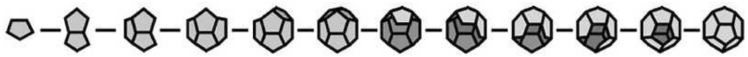


Fig. 3. Assembly of a dodecahedron from pentagons. The single assembly path is based on the addition of one pentagon at a time to form the largest number of new contacts (Zlotnick, 1994).

At each step, the incoming subunit is bound in the position where it makes the most contacts.

A set of equations was developed to describe the concentrations of monomers, capsids and each intermediate that allowed all species to be expressed in terms of the concentration of free subunit. Assembly was treated as a Markov Chain, or random walk along selected intermediates. It was assumed that only one monomer was added at a time. The association constant for addition of one subunit to an intermediate composed of $n-1$ subunits is described in equation 1:

$$K_n = [n]/[n-1][\text{subunit}]. \tag{1}$$

Here K_n is the association constant for the formation of species, n, an intermediate composed of n subunits, $[n]$ the concentration of species n, and $[n-1]$ the concentration of species $n-1$.

Solving this equation for $[n]$, we obtain

$$[n] = K_n[n-1][\text{subunit}]. \tag{2}$$

By iteratively substituting the concentration of each species, the equilibrium concentrations can be expressed as a function of the concentration of free subunit:

$$[n] = \left(\prod_{j=2}^{n} K_j\right)[\text{subunit}]^n. \tag{3}$$

The term K_n can be partitioned into a statistical component and a non-statistical component. The statistical component accounts for the degeneracy of the incoming subunit and of the assembly or disassembly of the intermediate under consideration. The non-statistical term represents the number of new bonds that are formed by binding of the incoming subunit (Zlotnick, 1994).

This model was readily adapted to kinetic simulations by considering reversible assembly of each intermediate as a function of preceding and

succeeding intermediates (equations 4 and 5).

$$d[\mathbf{n}]/dt = -(\text{assembly of } (n+1)\text{mers}) + (\text{assembly of } (n)\text{mers}) + (\text{dissociation of } (n+1)\text{mers}) - (\text{dissociation of } (n)\text{mers}) \quad (4)$$

$$d[\mathbf{n}]/dt = -k_{\text{forward},n}[n][\text{subunit}] + k_{\text{forward},n}[n-1][\text{subunit}] + k_{\text{backward},n}[n+1] - k_{\text{backward},n}[n] \quad (5)$$

For simulations, the microscopic forward rate was an arbitrary value, limited by diffusion, modified by reaction degeneracy (i.e., the number of paths to build an intermediate). As the equilibrium constant is the ratio of forward and backward rates, the backward rate constant for each intermediate is equal to the forward rate divided by the respective K_n.

Several experimentally observed features of capsid assembly were accounted for using this model. Simulated kinetics were sigmoidal and rapidly achieved equilibrium. Intermediates were required for capsid formation but were only found in trace amounts. The outcome of the reaction was also extremely concentration dependent. Below a "pseudo-critical concentration" of subunit, an infinitesimal concentration of capsids was formed. Almost all subunit above this concentration was consumed to form capsid. At very high concentrations of subunit, conditions that resembled those of the early small angle X-ray scattering (SAXS) experiments (Cuillel *et al.*, 1983), the assembly reaction was over-initiated and small intermediates accumulated. This consumed the pool of free subunits at the expense of completed capsids. This phenomenon is referred to as kinetic trapping and is a result of too rapid assembly, in this case brought about by excess concentration of subunit. Perhaps the greatest advantage of this model was its ease of modification to include new elements and/or reflect biology. For example, a nucleation step decreased kinetic trapping, caused by high subunit concentration and/or high association energy, by regulating initiation of assembly (Zlotnick *et al.*, 1999).

2.2. *Contributions of Alternative Pathways to Stepwise Assembly*

Recent efforts to improve these models have focused on the choice of assembly paths. The initial classical simulations were based on a single path of assembly, which is not chemically realistic and has an effect on the shape of the simulation. Virus capsids can contain hundreds to thousands of

copies of capsid protein and the number of possible intermediates increases combinatorially with size (Endres *et al.*, 2005). This increase in the number of potential intermediates is similar to the Levinthal paradox of protein folding, where the available conformational space is too large for it all to be sampled (Levinthal, 1969). By analogy, it is equally unlikely that a capsid samples all of the possible assembly paths. Determining which assembly pathways are dominant has been the focus of several investigations.

A first approximation is that the most likely assembly path follows the most stable intermediates. An early approach to defining a path based on a buried surface was taken to define a most likely assembly path for rhinovirus (Horton and Lewis, 1992). This emphasized the stability of the pentamer-of-trimers assembly unit (Rueckert, 1996). This approach was used to define the most stable assembly paths based on crystal structures for black beetle virus, southern bean virus, and human rhinovirus 14 (Reddy *et al.*, 1998). Molecular interactions were investigated in great detail in this study. By using a build-up procedure, these selections of species define a viable path and suggest interesting constraints on the assembly path and leads to a much smaller library of intermediates than would be predicted for a 180-mer. Using tiling theory as a guide, continuous paths were derived for polyomavirus and papillomavirus structures (i.e., systems of 72 pentamers in a $T = 7$ d lattice) (Keef *et al.*, 2006; Keef *et al.*, 2005). These studies demonstrated a surprising repeat of certain "blocks" of assembly. Such regular features during assembly may be critical for controlling the fidelity of assembly and may represent an underlying mathematical regularity of the structures.

In a brute force attack of a relatively small problem, the assembly of a dodecahedron, the 73 assembly intermediates and the paths connecting them were enumerated (Fig. 4). For example, a dimer of pentagons can be constructed in only one way, a trimer can be made in two ways and a hexamer in twenty. In kinetic simulations based on this tree of species it was shown that: (i) kinetics calculated for a single assembly path do not accurately mirror a complete simulation, (ii) a relatively small number of intermediates in excess of the minimum do a fair job of simulating the complete reaction, (iii) along with the stability of the intermediates, probability (i.e., reaction degeneracy) is an important criterion for inclusion in a model and (iv) the relative populations of intermediates shift during the course of the reaction. The shifting assembly surface rapidly approaches its equilibrium state, far from its condition at the earliest times

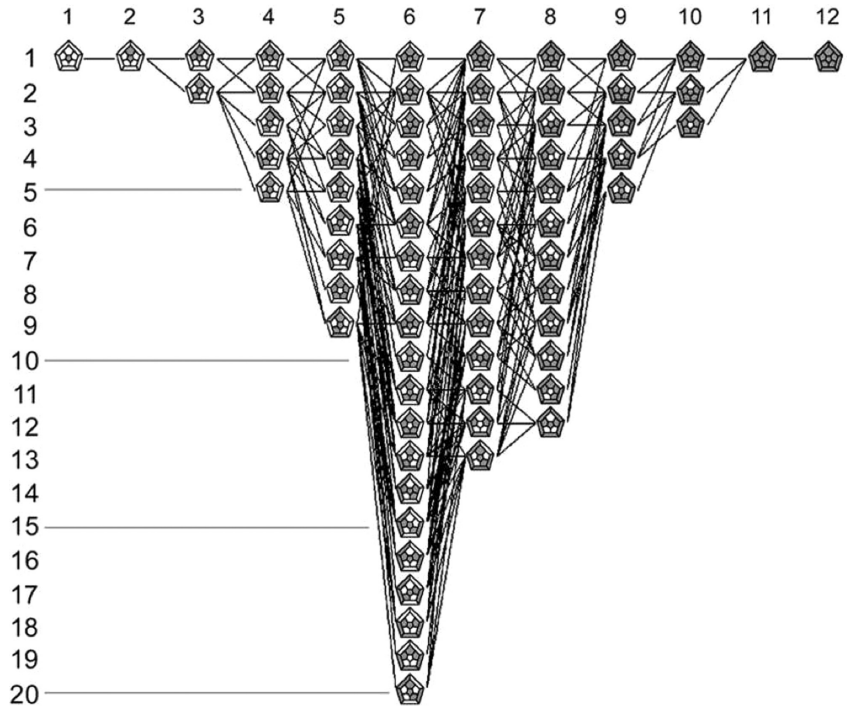


Fig. 4. The paths and intermediates for assembly of a dodecahedral capsid showing all 73 species as two-dimensional projections, Schlegel diagrams (Endres *et al.*, 2005). The segments joining species identify reactions paths. The vertical axis is an arbitrary index ranked by stability. Compare to Fig. 3.

in the reaction, though the reaction may be far from equilibrium. If capsid assembly is indeed a good analogy for protein folding, this shifting reaction surface suggests that predicting protein folding is even harder than it appears.

Many models of stepwise assembly make the assumption that capsids are assembled by addition of monomers to a growing structure. Molecular dynamics simulations support this idea in part because assembly conditions show relatively low concentrations of intermediates and relatively high concentrations of reactant monomers (Nguyen *et al.*, 2007; Rapaport, 2004; Rapaport *et al.*, 1999; Nguyen and Brooks, 2008; Rapaport, 2008). However, we cannot eliminate the possibility that compatible intermediates may join together (Zhang and Schwartz, 2006). This possibility was studied using a discrete event simulation that enumerates each type

of allowable event (Zhang and Schwartz, 2006). In conditions of low protein concentration or low binding rate, the addition of pathways in which intermediates bind together does not significantly change the overall rate of capsid assembly. Under conditions of high subunit concentration or high association rate, however, the contribution of intermediates binding together relieved kinetic trapping (Zhang and Schwartz, 2006).

2.3. *Molecular Dynamics and Stepwise Assembly*

Molecular dynamics (MD) and related approaches can be powerful ways to relate structure and function. Such simulations are limited because of the computational overhead and the assumptions required for coarse graining. However, they have been very successful in identifying how molecular interactions can direct assembly paths.

One of the earliest MD-like analyses probed the precision required of protein–protein interaction for successful assembly. This analysis was based on establishing a set of "local rules" which guide the local interactions of subunits and can explain some of the bulk observations of assembly (Berger *et al.*, 1994; Schwartz *et al.*, 2000). The "local rules" are based upon restriction of the angles and distances between subunits. The local rules for a $T = 7$ particle (Fig. 5) lead exclusively to the expected result. This model assumes a metastable distribution of conformers in solution that exchange with one another at a rate that is slow compared to capsid assembly. However, it is possible that the flexibility allows for protein subunits to adopt the structure dictated by the local environment to fulfill the demands of quasi-equivalence. These rules can be relaxed to surprising degree before aberrant assembly results (Berger *et al.*, 1994; Prevelige, 1998). This demonstrated the viability of such an approach.

Local rules approaches to assembly of single particles were expanded to a population to examine assembly kinetics (Schwartz *et al.*, 1998). These studies emphasized the trade off between fidelity of assembly and assembly rates. Interestingly, based on first principles, it has been estimated that local rules and similar mechanisms can restrict geometry with high fidelity for small T numbers as larger sizes are likely to yield a stochastic distribution of sizes (Lidmar *et al.*, 2003).

Molecular dynamics requires careful parameterization. This is even more critical where coarse graining is required. Rapaport has described assembly of multiple particles from geometric subunits. In initial assembly

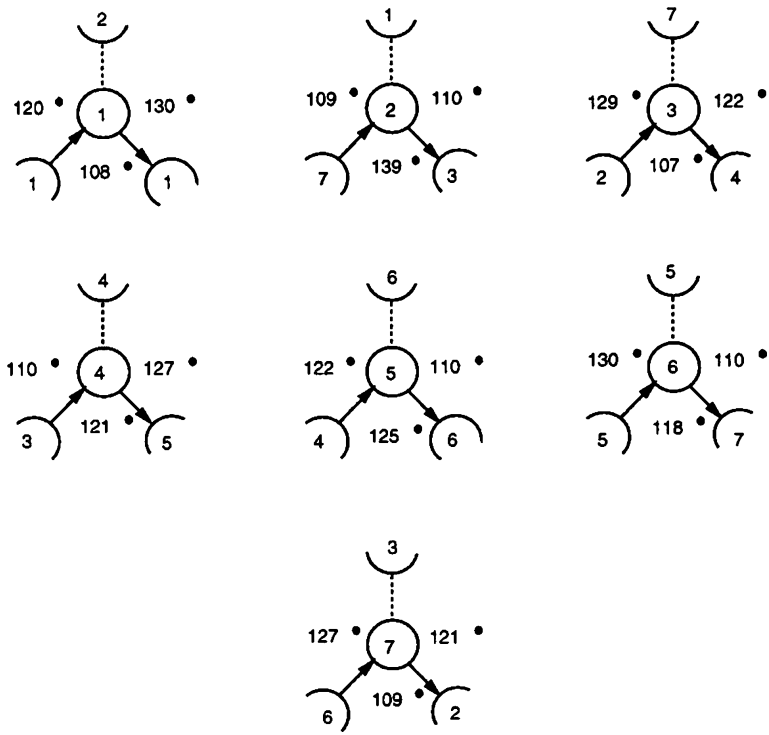


Fig. 5. Local Rules for formation of a $T = 7$ particle. The flexibility of interaction allows for a sub-stoichimetric number of rules (Berger *et al.*, 1994).

studies he observed that reactions tended to go awry unless they were seeded with assembly-active centers (essentially nuclei) and that the rate of addition of the last subunit was only marginally slower than for the earlier subunits (Rapaport *et al.*, 1999). As the models became more sophisticated (Rapaport, 2004; Rapaport, 2008) it became necessary to incorporate reversibility into reactions to prevent kinetic trapping. When reversibility was incorporated, simulations became more robust and more closely resembled actual assembly reactions. Only a small number of highly-bonded, i.e., relatively stable intermediates, were preponderant during assembly (Rapaport, 2008). Examination of dynamics trajectories (Rapaport, 2008) and analysis of shifts in concentrations (Jack *et al.*, 2007) show similar shifting energy surfaces leading to equilibrium, comparable to populations of intermediates seen in classical kinetics simulations (Endres *et al.*, 2005). Remarkably, in order for binding parameters to resemble

those experimentally recorded, it was necessary to include explicit solvent in these simulations (Rapaport, 2008). Similar studies with populations focusing on comparison to experiment have yielded similar results to classical studies, including hysteresis to dissociation (Hagan and Chandler, 2006).

When subunit–subunit interactions are restricted to near ideal values it is not surprising that ideal capsids result. A coarse-grained model with substantial angular freedom assembles normally, again the complete capsid is the lowest energy state, but it is difficult to close (Nguyen *et al.*, 2007). In these simulations insertion of the last subunit is the kinetically limiting step. At present there is little direct evidence to test this hypothesis excepting mass spectrometry of *in vitro* assembled hepatitis B virus capsids, where no incomplete capsids were observed (Uetrecht *et al.*, 2008); however, these may have had time to reach equilibrium. Critically, this MD simulation suggests how flexibility can correlate with assembly path and suggests a need for experimental examination of the last steps of capsid assembly. These results have been extended to look at pleiomorphy in capsid assembly, relating assembly heterogeneity to kinetic paths and the flexibility of intersubunit contacts (Nguyen and Brooks, 2008).

2.4. *Protein Condensation*

McPherson proposed a qualitative model based on micelle formation that offers an explanation for assembly on nucleic acid (McPherson, 2005). This model is also applicable to packaging exogenous structures such as gold particles and quantum dots (Chen *et al.*, 2005; Dixit *et al.*, 2006; Huang *et al.*, 2007; Sun *et al.*, 2007) This model was later quantitatively evaluated in a series of MD simulations (Hagan, 2008). The components of this model are the capsid protein and an assembly template.

Consider virus capsids as analogous to reverse micelles where the hydrophilic nucleic acid binding domain of the capsid protein is oriented toward the center of the virus and the nucleic acid acts as a reservoir of negative charge. As the nucleic acid becomes saturated with protein, the protein-nucleic acid complex begins reorganization into geometric groups. At this stage, the structure remains largely fluid. Rearrangement continues until a global energy minimum is found. Based on a maximization of bonding potential, icosahedral symmetry yields this minimum energy structure (Zandi *et al.*, 2004). It is interesting to note that there

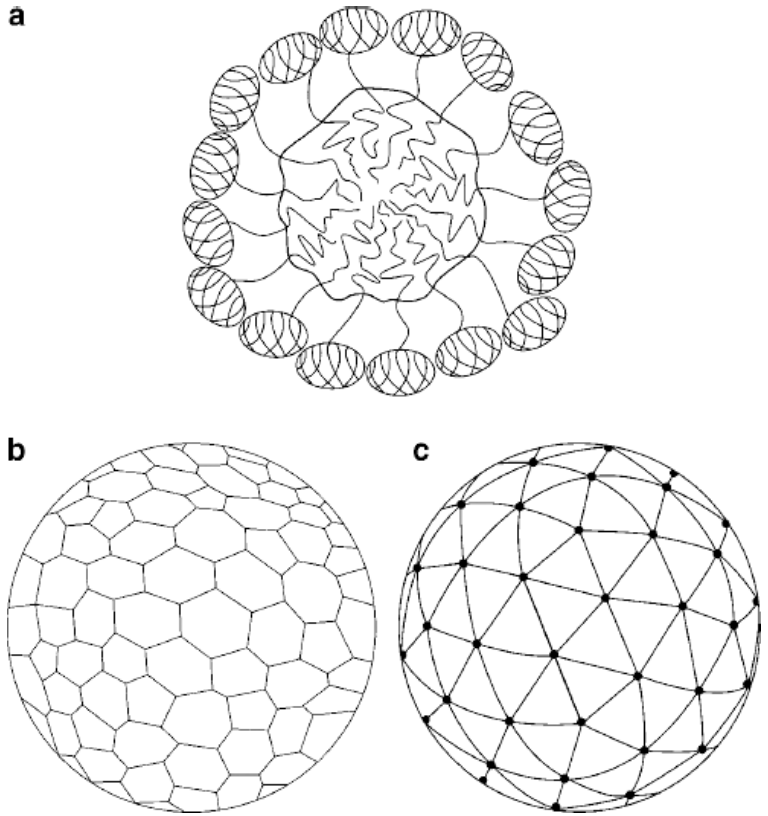


Fig. 6. Metamorphosis of capsid condensation. (a) Capsid proteins associate with a core of nucleic acid or other substrate. (b) The proteins locally condense (b) and eventually achieve an icosahedral global minimum (McPherson, 2005).

was a relatively small global energy difference between an icosahedron and flawed isomers. In contrast, McPherson postulates that hexamers and pentamers of capsid protein form spontaneously and are inter-convertable until they become locked in an icosahedral organization (Fig. 6).

Coarse-grained MD simulations were applied to a model of capsid protein adsorption to a solid surface (Hagan, 2008). *In vitro* capsid proteins can package various inorganic cores (Chen *et al.*, 2005; Dixit *et al.*, 2006; Douglas and Young, 2006; Huang *et al.*, 2007). In the MD simulation, binding between subunits was governed by potential energy functions that included relatively long distance repulsive and attractive terms. Assembly paths depended on assembly parameters. Assembly

of empty capsids by a stepwise mechanism was observed in the case of low concentration of subunits or low free energy of adsorption. When concentration or adsorption energy were high, capsid proteins would bind to the core and then organize upon its surface, as suggested by the micelle model. Thus condensation and stepwise assembly may be thought of as extremes in the continuum of capsid assembly mechanisms.

Condensation on a substrate allowed for more rapid assembly with a higher yield of intact capsids (Hagan, 2008). The rate of adsorption to cores was generally faster than the assembly rate. The greater yield can be attributed to the additional association energy and concentrating effect of the underlying substrate. Under some conditions, this kind of assembly can be more resistant to kinetic trapping. Cores act as a template, which may promote a particular morphology of capsid proteins. Adsorption to a core also increases the local concentration of capsid protein, thereby increasing the rate of assembly.

3. Virus Capsid Assembly in Solution

3.1. *Experimental Systems*

It is generally accepted that most spherical viruses self-assemble. *In vitro* this has been demonstrated with numerous viruses. We will briefly examine just a few systems for which there is rigorous kinetic data. These systems have in common only two important features: ease of handling and a pseudo-critical concentration in the micromolar range. At this time, relatively few viruses have been subjected to kinetic analysis, so serendipity also counts. Virus systems that are likely amenable to a more detailed kinetic analysis of assembly include the sobemovirus sesbania mosaic virus (Lokesh *et al.*, 2002), picornaviruses such as poliovirus (Rombaut *et al.*, 1991b), bacteriophage MS2 (Lago *et al.*, 2001), herpes simplex 1 provirions (Newcomb *et al.*, 1999), sindbis virus (Tellinghuisen *et al.*, 1999), and birnaviruses such as infectious bursal disease virus (Fernandez-Arias *et al.*, 1998). The systems that will be discussed in this review are bacteriophage P22, cowpea chlorotic mottle virus (CCMV), bromegrass mosaic virus (BMV), hepatitis B virus and human papilloma virus (HPV).

Bacteriophage P22 is a tailed bacteriophage with a double-stranded DNA (dsDNA) genome (Fane and Prevelige, 2003). Assembly of the complete virion is complex due its numerous components, including accessory proteins for the phage head and the multi-step assembly of the

tail. However, empty isometric $T = 7$ capsids can be assembled by addition of the dimeric scaffold proteins to a solution of the monomeric HK97-fold coat protein.

HBV is an enveloped DNA virus with an RNA intermediate form (Ganem and Prince, 2004). *In vivo*, the dimeric α helical core protein assembles around a complex of reverse transcriptase and pregenomic RNA. The RNA is reverse transcribed within the core (capsid) in the cytoplasm. *In vitro* assembly studies have focused on Cp149, a 149 residue truncated form of the capsid protein lacking the RNA-binding C-terminus (Zlotnick *et al.*, 1999). Assembly is driven by elevated ionic strength.

CCMV and BMV are closely related bromoviruses (Adolph and Butler, 1976; Bancroft, 1970). These are plant viruses with tripartite RNA genomes. An effort to name this family of viruses *tircornaviridae* stalled when it was realized that *tricorna* is an Italian synonym for cuckold. The β-barrel coat proteins of BMV and CCMV, both dimers in solution, share 70% sequence identity yet notably different assembly properties. Assembly can be driven by addition of a negatively charged substrate (e.g., RNA); empty capsids form in pH < 6.

HPV is a DNA virus with numerous serotypes. HPV16 is responsible for approximately 50% of all human cervical cancer. HPV virus-like particles are an extremely successful, though expensive, vaccine (Koutsky *et al.*, 2002; Schiller and Lowy, 2001). The soluble form of the β-barrel coat protein is a pentamer. These self assemble, by interaction of their C-termini, to generate a $T = 7$ lattice with pentamers at both fivefold and quasi-sixfold positions. Assembly of HPV16 is driven by ionic strength, and may be subsequently stabilized by disulfides (Mukherjee *et al.*, 2008).

3.2. *Assembly of Bacteriophage P22*

Bacteriophage P22 is one of the few well-established assembly systems based on a binary reaction — coat and scaffold proteins — and whose assembly has been subjected to substantial biophysical analysis (Fane and Prevelige, 2003). In the absence of scaffold protein bacteriophage P22 coat protein will spontaneously assemble into aberrant structures (Prevelige *et al.*, 1988). When sub-stoichiometric concentrations of scaffold were added to coat protein, assembly of empty virus-like particles (VLPs) proceeds with sigmoidal kinetics (Prevelige *et al.*, 1993). The overall

rate of capsid appearance was dependent upon both coat and scaffold concentrations (Prevelige *et al.*, 1993). In these experiments, by correlating turbidity and electron microscopy, it was shown that intermediates transiently accumulated and were consumed in subsequent assembly. Nonetheless, based on comparison to crystallization kinetics, it was assumed that the lag of the sigmoidal kinetics was evidence of nucleation; however, the concentration dependence of rate observed in these elegant experiments is in fact the confirming feature. The extent of assembly showed the expected pseudo-critical concentration. While mature virions are extraordinarily stable, provirions (also very stable) can exchange assembled subunits with coat protein in solution, experimentally demonstrating the presence of reversibility in these self-assembling systems, if not the necessity for it (Parent *et al.*, 2006a),

Scaffold protein has at least three roles in assembly: activating coat protein for assembly (Tuma *et al.*, 2008), stabilizing interactions between coat proteins (Parent *et al.*, 2006b), and regulating geometry (Prevelige *et al.*, 1988). *In vivo*, scaffold concentration is closely regulated and too much or too little is strongly inhibitory to provirion synthesis (Casjens *et al.*, 1985). *In vitro* experiments can readily mimic this effect. Increasing scaffold protein binding by increasing its affinity (a function of decreasing ionic strength) or simply raising scaffold protein concentration lead to kinetic trapping of partial capsids (Parent *et al.*, 2005; Parent *et al.*, 2006b). A scaffold protein that does not dimerize is still able to enhance assembly while a chemically crosslinked dimer radically stimulates the assembly reaction, again leading to kinetic traps (Tuma *et al.*, 2008). Recapitulating the early studies of Cuillel (Cuillel *et al.*, 1983), the role of scaffold in regulating assembly was observed in small angle X-ray scattering experiments that revealed an array of kinetically trapped intermediates during assembly (Tuma *et al.*, 2008).

These results demonstrate the importance of balancing assembly nucleation with sufficient subunits for completion of assembly *in vitro* and provide a rationale for regulation of scaffold expression *in vivo*. The demonstration of subunit exchange confirms a role for reversibility of assembly, even in the formation of these extremely stable capsids. The experiments where the scaffold is modified suggest that this may be an avenue for the evolutionary modification of capsid size, as seen with the $P2 - P4$ $T = 7 - T = 4$ switch (Chang *et al.*, 2008; Wang *et al.*, 2000).

3.3. *HBV Assembly*

The kinetics of HBV capsid protein assembly has been the focus of several studies. HBV capsid assembly is a well-characterized experimental system (Birnbaum and Nassal, 1990; Wingfield *et al.*, 1995; Zlotnick *et al.*, 1999). *In vitro* assembly of HBV capsids is dependant on changes in protein concentration, temperature, ionic strength, and pH (Ceres and Zlotnick, 2002; Wingfield *et al.*, 1995). Ionic strength and factors such as zinc (Stray *et al.*, 2004) appear to induce an assembly-activating conformational change. NaCl-induced assembly results in kinetics that closely resemble the predictions of stepwise assembly (Zlotnick *et al.*, 1999) (Fig. 7). Assembly of capsid protein *in vitro* is typically sigmoidal and rapid. Analysis indicates that it proceeds through a nucleus comprised of a trimer of core protein dimers (Zlotnick *et al.*, 1999). The nucleus size determination is based on the dependence of the extent of assembly on subunit concentration (Endres and Zlotnick, 2002). The concentration dependence of subsequent capsid assembly is second order, indicating that elongation proceeds by addition of dimer subunits. HBV assembly data agree well with the stepwise model of assembly, including the paucity of intermediates. HBV capsids are held together by weak subunit-subunit interactions, with association energies of between −3 and −4 kcal/mol (Ceres and Zlotnick, 2002). This corresponds to a millimolar association constant per contact.

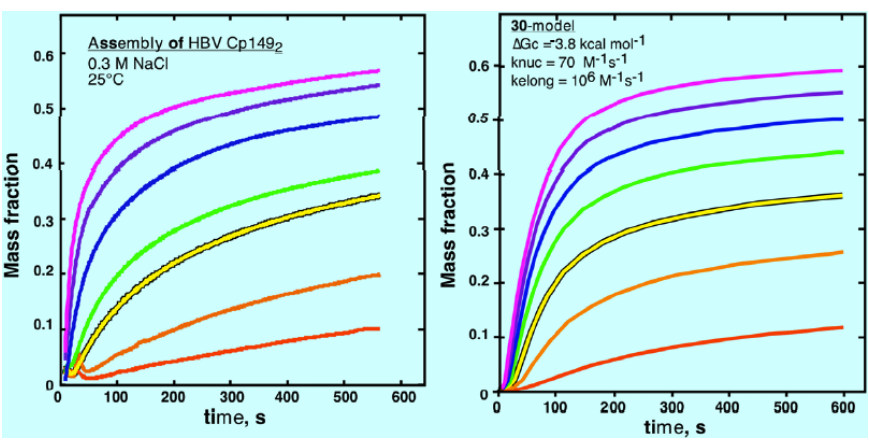


Fig. 7. Assembly of a concentration series of HBV core protein assembly domain, Cp149, compared to assembly simulations of a 30-mer assembly model.

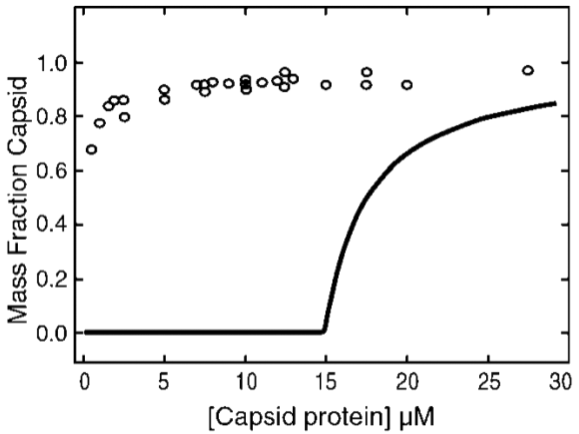


Fig. 8. Dilution of purified HBV capsids reveals hysteresis of dissociation. Solid line shows a capsid assembly isotherm at 21°C and 150 mM NaCl. Open circles show dissociation induced by dilution (Singh and Zlotnick, 2003).

However, the polyvalent subunits result in a globally stable capsid with a micromolar pseudo-critical concentration of assembly (Fig. 7). Capsid stability is further enhanced by hysteresis, a kinetic barrier to dissociation (Singh and Zlotnick, 2003). Hysteresis results in metastable capsids under nominally dissociating conditions, which would be advantageous to any virus in an extracellular environment (Singh and Zlotnick, 2003) (Fig. 8).

Evolutionary forces have selected HBV core protein mutations with altered assembly kinetics. The most frequent mutation of the HBV core protein is Phe or Ile 97 to Leu (Gunther *et al.*, 1999). The 97Leu mutants replicate better than wild-type in some cell lines and even yield an 'immature secretion' phenotype concurrent with enhanced replication in others (Suk *et al.*, 2002; Yuan *et al.*, 1999a; Yuan *et al.*, 1999b). Notably, amino acid 97 is not located at the inter-dimer interface, but rather in a hydrophobic pocket at the center of the dimer subunit and thus strengthens the case for allosteric regulation of assembly. *In vitro*, this mutation increases both the extent of assembly and the assembly rate (Ceres *et al.*, 2004). Furthermore, 97Leu exhibits a more positive enthalpy of assembly than wild-type. These differences, coupled with the location of the mutation suggest that amino acid 97 does not directly strengthen intersubunit contacts, but affects the dynamic behavior of the dimer subunit, encouraging the transition of the capsid protein to the assembly active state.

3.4. *HBV and Small Molecules that Interfere with Assembly*

Capsid assembly is an attractive target for anti-viral therapeutics because it has no cellular homologue. The effects of potential drug molecules on capsid assembly could take the form of traditional competitive or non-competitive inhibitors, preventing assembly as shown with phage P22 (Teschke *et al.*, 1993), HBV (Zlotnick *et al.*, 2002), and in HIV (Kelly 2007, Kelly 2007, Braun 2008, Sticht *et al.*, 2005; Ternois *et al.*, 2005). For therapeutic purposes, complete abrogation of assembly may not be necessary. Another strategy is that the drug molecule could stabilize capsids and prevent disassembly (e.g., the WIN compounds and picornaviruses (Rombaut *et al.*, 1991a; Smith *et al.*, 1986)). As an alternative, small molecules could stabilize binding between subunits while slightly distorting the interaction, resulting in non-capsid structures (Prevelige, 1998; Stray *et al.*, 2005; Zlotnick *et al.*, 2002).

Bayer scientists demonstrated that heteroaryldihydropyrimidines (HAPs) interfered with HBV replication in a core-dependent manner (Deres *et al.*, 2003; Weber *et al.*, 2002). Later it was shown that excess HAP leads to assembly of aberrant particles (Hacker *et al.*, 2003; Stray *et al.*, 2005). More importantly, it was shown that HAPs dramatically enhanced assembly kinetics and the stability of core protein-protein interactions (Stray *et al.*, 2005). High concentrations of HAP were shown to destabilize icosahedral fivefold vertices and stabilize hexagonal lattices of core protein. The effect of HAP-1 on capsid structure was further evaluated in the crystal structures of disulfide crosslinked $T = 4$ HBV capsids with and without the HAP-1 molecule (at low resolution of 5 Å and 4 Å, respectively) (Bourne *et al.*, 2006). The +HAP structure showed no change in tertiary structure though quasi-sixfold vertices were flatter and fivefold vertices were expanded in radius and reduced in buried surface. Though the four classes of core protein were quasi-equivalent, HAP density was only observed associated with the C-D interdimer contact. Presumably the quaternary changes all propagated from this 'burr'.

A particularly revealing set of experiments used isomerically pure BAY41-4109, instead of racemic HAP-1 (Stray and Zlotnick, 2006). BAY41-4109 affected the kinetics of assembly by increasing both the rate of nucleation and elongation. With stoichiometries less than 1 BAY41-4109 per two dimers, assembly yields spherical capsids, suggesting

nucleation continues to regulate assembly. At higher BAY41-4109 concentrations, assembly is misdirected, suggesting roles for kinetic trapping as well as defective geometry. Thus, BAY41-4109 has two functional classes of binding site; it is tempting to equate the first class of site with binding to the C subunit and suggest that subsequent binding events lead to distortion in quaternary structure that cannot be reconciled with icosahedral geometry.

Based on fitting HAP-1 into 5 Å resolution electron density, a series of modified HAP molecules based on the HAP-1 structure were developed and examined for their effects on capsid assembly (Bourne *et al.*, 2008; Stray *et al.*, 2006). Using solution assays, the effect of the HAP's association energy and relative assembly kinetics were evaluated. At their most extreme, HAPs decrease the pseudo-critical concentration by up to 500 (from $\sim$15 μM to $\sim$30 nM) and increase the rate of assembly by more than three orders of magnitude. Effects on *in vitro* assembly were correlated with their effects on HBV replication in a cell culture line that constitutively produces active virus. There was no correlation between the thermodynamic effect and antiviral activity. For HAPs that strengthened pairwise association by 1 ± 0.2 kcal/mol, the effective concentration of HAP that dropped virus replication by 50% (EC50) ranged between 50 nM at best and no detectable effect at worst. There was a remarkable correlation between the HAP effect on assembly kinetics and antiviral effect. The 'fastest' HAP had an EC50 for secreted virus of 6 nM. Notably this antiviral effect occurred at concentrations too low for appreciable assembly misdirection. One explanation for HAP antiviral activity is that it activates assembly at the wrong time for productive replication. It is clear that altering assembly kinetics is therefore a valid approach to the development of anti-viral therapeutics.

3.5. *Bromovirus Assembly*

CCMV was the first spherical virus assembled *in vitro* from purified components (Bancroft and Hiebert, 1967), a decade after re-assembly of rod-like Tobacco Mosaic Virus (Fraenkel-Conrat and Williams, 1955). Unlike HBV (discussed below) and bacteriophage P22, CCMV does not always show simple sigmoidal kinetics (Zlotnick *et al.*, 2000). Instead, there is an initial jump in light scattering indicating the formation of a rapidly assembled intermediate. Based on (i) the concentration

dependence of its formation and (ii) the observation of a 200 kDa complex in size exclusion resolution of reactions shortly after initiation, this early intermediate was identified as a pentamer-of-dimers. This identification was supported by the observation of 'pseudo-$T = 2$' capsids, 120-subunit capsids composed of 12 pentamers-of-dimers by mass determination (Zlotnick *et al.*, 2000) and in image reconstructions of capsids from similar reactions (Johnson *et al.*, 2005; Tan *et al.*, 1999). This assembly path is distinct from that outlined in the discussion of stepwise assembly. But it can be accurately modeled by including an initial fast association reaction and rate limiting steps for (i), which is the addition of a dimer to a pre-existing pentamer-of-dimers, a necessary step for $T = 3$ formation, and for (ii), which is the association of pentamers-of-dimers, a necessary step for formation of pseudo-$T = 2$ capsids (Johnson *et al.*, 2005).

The structure of CCMV (Speir *et al.*, 1995) includes an interleaved series of β-strands, termed β-hexamers, that strongly suggest that the formation of the quasi-sixfold would be critical for virus assembly (see VIPERdb (http://viperdb.scripps.edu/) for an examination of buried surface in this virus (Carrillo-Tripp *et al.*, 2008)), contrary to the results of the kinetic studies. The role of the β-hexamer was tested by deletion (Willits *et al.*, 2003). The resulting coat assembled into $T = 3$ capsids with normal morphology though lacking β-hexamer density. The resulting virus had a much more pathogenic phenotype. But, clearly the β-hexamer was not required for assembly.

Structure can be misleading in other ways. BMV is structurally almost identical to CCMV. Sequence identity of the coat protein is greater than 70%. Yet acid-induced assembly of BMV yields kinetics that are dramatically different from those of CCMV (Chen *et al.*, 2008). There is no evidence of an early assembly phase, i.e. pentamers-of-dimers do not accumulate. There is an early, non-productive lag in assembly that is best modeled as a pre-assembly conformational change into an active form. Based on kinetic models of BMV assembly, this conformational change is visible because it was slow compared to nucleation (Fig. 9).

Pre-assembly conformational change, a form of allosteric control of assembly, may be a common feature in viruses. A very similar pre-assembly lag was observed in an elegant study of the assembly of human papilloma virus HPV11 (Casini *et al.*, 2004). The pre-assembly lag was not concentration dependent and therefore is a unimolecular reaction. As with BMV, this conformational change must be slow compared to nucleation.

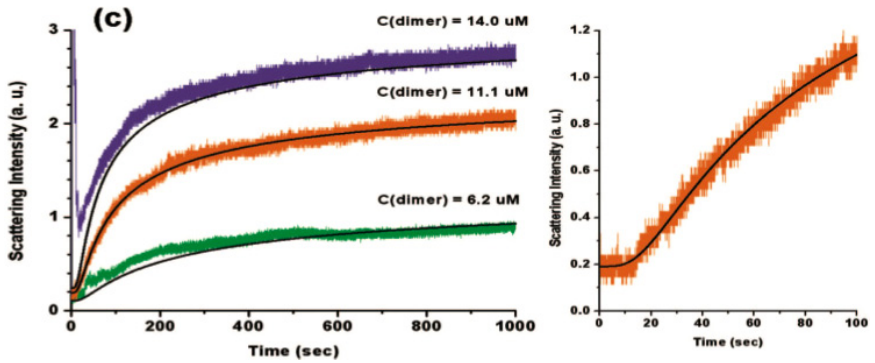


Fig. 9. Assembly of BMV (left) shows a good match to a stepwise assembly model. The right panel shows the pre-assembly lag phase (Chen *et al.*, 2008).

Assembly was concentration dependent with an apparent nucleus of a dimer of coat protein pentamers. Again, these kinetic results contrast with the five-around-one model of assembly that has been hypothesized based on structure (Chen *et al.*, 2001; Stehle *et al.*, 1996). A pre-assembly conformational change has been postulated for HBV based on a variety of evidence (see above and references therein), but, if present, is too fast compared to nucleation to be visible in current kinetic experiments.

4. Concluding Remarks and Future Directions

Some generalizations regarding virus assembly can be made. Reversibility and weak pairwise interactions, two faces of the same coin, are required for high fidelity assembly of regular structures. Nucleation of assembly is critical for prevention of kinetic traps. It appears that 'successful' assembly is tightly regulated. For empty capsids a stepwise assembly model is a safe bet. For assembly on surfaces, a hybrid between stepwise assembly and protein condensation may be more accurate. In spite of their diversity, simulations appear to be converging on a picture of 'successful assembly' and are now exploring the roles of templates and mechanisms of failed assembly reactions.

Some generalizations unfortunately don't hold up. A structure may not be the best starting point for modeling an assembly path; no one in this era would suggest hypothesizing a protein-folding path based on

structure. Just because viruses are closely related does not mean they follow the same assembly path; structure may be evolutionarily conserved but, at least in bromoviruses, assembly paths are not.

As with any field, good science leads to more questions. How can assembly be controlled to generate designed structures? How can small molecules be designed to make assembly-directed therapeutics a reality? How can structures be analyzed to predict assembly? How can we build an accurate model of assembly for rigorous curve-fitting and analysis of data? How can we dissect the behavior of ensembles of particles to predict the behavior of individual capsids? Alternatively, how can we observe the assembly of single particles? And finally, how can we relate the behavior of assembly *in silico* and *in vitro* to assembly *in vivo* in a manner that can be understood by the larger community of virologists.

References

1. C. Abad-Zapatero, S. Abdel-Meguid, J.E. Johnson, A.G.W. Leslie, I. Rayment, M.G. Rossmann, D. Suck and T. Tsukihara. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* **286**:33–39 (1980).
2. K.W. Adolph and P.J.Butler. Assembly of a spherical plant virus. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **276**:113–122 (1976).
3. J.B. Bancroft. The self-assembly of spherical plant viruses. *Adv. Virus Res.* **16**:99–134 (1970).
4. J.B. Bancroft and E. Hiebert. Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. *Virology* **32**:354–356 (1967).
5. B. Berger, P.W. Shor, L. Tucker-Kellogg and J. King. Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci. USA* **91**:7732–7736 (1994).
6. F. Birnbaum and M. Nassal. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* **64**:3319–3330 (1990).
7. C. Bourne, M.G. Finn and A. Zlotnick. Global structural changes in hepatitis B capsids induced by the assembly effector HAP1. *J Virol.* **80**: 11055–11061 (2006).
8. C. Bourne, S. Lee, B. Venkataiah, A. Lee, B. Korba, M.G. Finn and A. Zlotnick. Small-molecule effectors of hepatitis B virus capsid assembly give insight into virus life cycle. *J. Virol.* **82**:10262–10270 (2008).
9. M. Carrillo-Tripp, C.M. Shepherd, I.A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J.E. Johnson, C.L. Brooks, 3rd and V.S. Reddy. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* **37**:D436–D442 (2009).
10. G.L. Casini, D. Graham, D. Heine, R.L. Garcea and D.T. Wu. *In vitro* papillomavirus capsid assembly analyzed by light scattering. *Virology* **325**:320–327 (2004).

11. S. Casjens, M.B. Adams, C. Hall and J. King. Assembly-controlled autogenous modulation of bacteriophage P22 scaffolding protein gene expression. *J. Virol.* **53**:174–179 (1985).
12. P. Ceres, S.J. Stray and A. Zlotnick. Hepatitis B Virus Capsid Assembly is Enhanced by Naturally Occurring Mutation F97L. *J. Virol.* **78**:9538–9543 (2004).
13. P. Ceres and A. Zlotnick. Weak protein–protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**:11525–11531 (2002).
14. J.R. Chang, A. Poliakov, P.E. Prevelige, J.A. Mobley and T. Dokland. Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology* **370**:352–361 (2008).
15. C. Chen, C. Kao and B. Dragnea. Self-Assembly of Brome Mosaic Virus Capsids: Insights from Shorter Time-Scale Experiments. *J. Phys. Chem. A* **112**:9405–9412 (2008).
16. C. Chen, E.S. Kwak, B. Stein, C.C. Kao and B. Dragnea. Packaging of gold particles in viral capsids. *J. Nanosci. Nanotechnol.* **5**:2029–2033 (2005).
17. X.S. Chen, G. Casini, S.C. Harrison and R.L. Garcea. Papillomavirus capsid protein expression in *Escherichia coli*: purification and assembly of HPV11 and HPV16 L1. *J. Mol. Biol.* **307**:173–182 (2001).
18. F.H.C. Crick and J.D. Watson. The structure of small viruses. *Nature* **177**:473–475 (1956).
19. M. Cuillel, C. Berthet-Colominas, B. Krop, A. Tardieu, P. Vachette and B. Jacrot. Self-assembly of brome mosaic virus capsids. Kinetic study using neutron and X-ray solution scattering. *J. Mol. Biol.* **164**:645–650 (1983).
20. K. Deres, C.H. Schroder, A. Paessens, S. Goldmann, H.J. Hacker, O. Weber, T. Kramer, U. Niewohner, U. Pleiss, J. Stoltefuss, E. Graef, D. Koletzki, R.N. Masantschek, A. Reimann, R. Jaeger, R. Gross, B. Beckermann, K.H. Schlemmer, D. Haebich and H. Rubsamen-Waigmann. Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science* **299**:893–896 (2003).
21. S.K. Dixit, N.L. Goicochea, M.C. Daniel, A. Murali, L. Bronstein, M. De, B. Stein, V.M. Rotello, C.C. Kao and B. Dragnea. Quantum dot encapsulation in viral capsids. *Nano Lett.* **6**:1993–1999 (2006).
22. T. Douglas and M. Young. Viruses: making friends with old foes. *Science* **312**:873–875 (2006).
23. D. Endres, M. Miyahara, P. Moisant and A. Zlotnick. A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. *Protein Science* **14**:1518–1525 (2005).
24. D. Endres and A. Zlotnick. Model-based Analysis of Assembly Kinetics for Virus Capsids or Other Spherical Polymers. *Biophys. J.* **83**:1217–1230 (2002).
25. B.A. Fane and P.E. Prevelige, Jr. Mechanism of scaffolding-assisted viral assembly. In W. Chiu and J.E. Johnson (eds.), *Virus Structure*. Academic Press, San Diego, **64**:259–299 (2003).
26. A. Fernandez-Arias, C. Risco, S. Martinez, J.P. Albar and J.F. Rodriguez. Expression of ORF A1 of infectious bursal disease virus results in the formation of virus-like particles. *J. Gen. Virol.* **79(Pt 5)**:1047–1054 (1998).

27. H. Fraenkel-Conrat and R.C. Williams. Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components. *Proc. Natl. Acad. Sci. USA* **41**:690–698 (1955).
28. D. Ganem and A.M. Prince. Hepatitis B virus infection — natural history and clinical consequences. *N. Engl. J. Med.* **350**:1118–1129 (2004).
29. S. Gunther, L. Fischer, I. Pult, M. Sterneck and H. Will. Naturally occurring variants of hepatitis B virus. *Adv. Virus Res.* **52**:25–137 (1999).
30. H.J. Hacker, K. Deres, M. Mildenberger and C.H. Schroder. Antivirals interacting with hepatitis B virus core protein and core mutations may misdirect capsid assembly in a similar fashion. *Biochem. Pharmacol.* **66**:2273–2279 (2003).
31. M.F. Hagan. Controlling viral capsid assembly with templating. *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* 77:051904 (2008).
32. M.F. Hagan and D. Chandler. Dynamic pathways for viral capsid assembly. *Biophys. J.* **91**:42–54 (2006).
33. N. Horton and M. Lewis. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**:169–181 (1992).
34. X. Huang, L.M. Bronstein, J. Retrum, C. Dufort, I. Tsvetkova, S. Aniagyei, B. Stein, G. Stucky, B. McKenna, N. Remmes, D. Baxter, C.C. Kao and B. Dragnea. Self-assembled virus-like particles with magnetic cores. *Nano Lett.* 7:2407–2416 (2007).
35. R.L. Jack, M.F. Hagan and D. Chandler. Fluctuation-dissipation ratios in the dynamics of self-assembly. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**:021119 (2007).
36. J.M. Johnson, J. Tang, Y. Nyame, D. Willits, M.J. Young and A. Zlotnick. Regulating self-assembly of spherical oligomers. *Nano Lett.* **5**:765–770 (2005).
37. T. Keef, C. Micheletti and R. Twarock. Master equation approach to the assembly of viral capsids. *J. Theor. Biol.* **242**:713–721 (2006).
38. T. Keef, A. Taormina and R. Twarock. Assembly models for *Papovaviridae* based on tiling theory. *Phys. Biol.* **2**:175–188 (2005).
39. L.A. Koutsky, K.A. Ault, C.M. Wheeler, D.R. Brown, E. Barr, F.B. Alvarez, L.M. Chiacchierini and K.U. Jansen. A controlled trial of a human papillomavirus type 16 vaccine. *N. Engl. J. Med.* **347**:1645–1651 (2002).
40. H. Lago, A.M. Parrott, T. Moss, N.J. Stonehouse and P.G. Stockley. Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *J. Mol. Biol.* **305**:1131–1144 (2001).
41. C. Levinthal. In P. DeBrunner, J. Tsibris and E. Munck (eds.), *Mossbauer spectroscopy in biological systems.* University of Illinois Press, Allerton House, Monticello, IL, p. 22 (1969).
42. J. Lidmar, L. Mirny and D.R. Nelson. Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**:051910 (2003).
43. G.L. Lokesh, T.D. Gowri, P.S. Satheshkumar, M.R. Murthy and H.S. Savithri. A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. *Virology* **292**:211–223 (2002).
44. A. McPherson. Micelle formation and crystallization as paradigms for virus assembly. *Bioessays* **27**:447–458 (2005).

45. S. Mukherjee, M.V. Thorsteinsson, L.B. Johnston, P. DePhillips and A. Zlotnick. A quantitative description of *in vitro* assembly of human papillomavirus 16 virus-like particles. *J. Mol. Biol.* **381**:229–237 (2008).
46. W.W. Newcomb, F.L. Homa, D.R. Thomsen, B.L. Trus, N. Cheng, A. Steven, F. Booy and J.C. Brown. Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J. Virol.* **73**:4239–4250 (1999).
47. H.D. Nguyen and C.L. Brooks 3rd. Generalized Structural Polymorphism in Self-Assembled Viral Particles. *Nano Lett.* **8**:4574–4581 (2008).
48. H.D. Nguyen, V.S. Reddy and C.L. Brooks, 3rd. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett.* **7**:338–344 (2007).
49. A.J. Olson, Y.H. Hu and E. Keinan. Chemical mimicry of viral capsid self-assembly. *Proc. Natl. Acad. Sci. USA* **104**:20731–20736 (2007).
50. K.N. Parent, S.M. Doyle, E. Anderson and C.M. Teschke. Electrostatic interactions govern both nucleation and elongation during phage P22 procapsid assembly. *Virology* **340**:33–45 (2005).
51. K.N. Parent, M.M. Suhanovsky and C.M. Teschke. Phage P22 procapsids equilibrate with free coat protein subunits. *J. Mol. Biol.* **365**:513–522 (2006a).
52. K.N. Parent, A. Zlotnick and C.M. Teschke. Quantitative analysis of multicomponent spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids. *J. Mol. Biol.* **359**:1097–1106 (2006b).
53. J. Prevelige, D. Thomas and J. King. Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells *in vitro*. *J. Mol. Biol.* **202**:743–757 (1988).
54. P.E. Prevelige, D. Thomas and J. King. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**:824–835 (1993).
55. P.E.J. Prevelige. Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotech.* **16**:61–65 (1998).
56. D.C. Rapaport. Self-assembly of polyhedral shells: A molecular dynamics study. *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* **70**:051905 (2004).
57. D.C. Rapaport. Role of reversibility in viral capsid growth: a paradigm for self-assembly. *Phys. Rev. Lett.* **101**:186101 (2008).
58. D.C. Rapaport, J.E. Johnson and J. Skolnick. Supramolecular self-assembly: Molecular dynamics modeling of polyhedral shell formation. *Comp. Phys. Comm.* **121**:231–235 (1999).
59. V.S. Reddy, H.A. Giesing, R.T. Morton, A. Kumar, C.B. Post, C.L. Brooks, 3rd and J.E. Johnson. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. *Biophys. J.* **74**:546–558 (1998).
60. B. Rombaut, K. Andries and A. Boeye. A comparison of WIN 51711 and R 78206 as stabilizers of poliovirus virions and procapsids. *J. Gen. Virol.* **72**:2153–2157 (1991a).
61. B. Rombaut, A. Foriers and A. Boeye. *In vitro* assembly of poliovirus 14 S subunits: identification of the assembly promoting activity of infected cell extracts. *Virology* **180**:781–787 (1991b).

62. R.R. Rueckert. *Picornaviridae*: The viruses and their replication. In B.N. Fields, D.M. Knipe, P.M. Howley, R.M. Chanock, J.L. Melnick, T.P. Monath, B. Roizman and S.E. Straus (eds.), *Fields Virology*. Lippincott-Raven, Philadelphia, pp. 609–654 (1996).
63. J.T. Schiller and D.R. Lowy. Papillomavirus-like particle vaccines. *J. Natl. Cancer Inst. Monogr.* 50–54 (2001).
64. R. Schwartz, R.L. Garcea and B. Berger. "Local rules" theory applied to polyomavirus polymorphic capsid assemblies. *Virology* **268**:461–470 (2000).
65. R. Schwartz, P.W. Shor, P.E.J. Prevelige and B. Berger. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* **75**:2626–2636 (1998).
66. S. Singh and A. Zlotnick. Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *J. Biol. Chem.* **278**:18249–18255 (2003).
67. T.J. Smith, M.J. Kremer, M. Luo, G. Vriend, E. Arnold, G. Kamer, M.G. Rossmann, M.A. McKinlay, G.D. Diana and M.J. Otto. The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* **233**:1286–1293 (1986).
68. J.A. Speir, S. Munshi, G. Wang, T.S. Baker and J.E. Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**:63–78 (1995).
69. T. Stehle, S.J. Gamblin, Y. Yan and S.C. Harrison. The structure of simian virus 40 refined at 3.1 A resolution. *Structure* **4**:165–182 (1996).
70. J. Sticht, M. Humbert, S. Findlow, J. Bodem, B. Muller, U. Dietrich, J. Werner and H.G. Krausslich. A peptide inhibitor of HIV-1 assembly *in vitro*. *Nat. Struct. Mol. Biol.* **12**:671–677 (2005).
71. S.J. Stray, C.R. Bourne, S. Punna, W.G. Lewis, M.G. Finn and A. Zlotnick. A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc. Natl. Acad. Sci. USA* **102**:8138–8143 (2005).
72. S.J. Stray, P. Ceres and A. Zlotnick. Zinc ions trigger conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry* **43**:9989–9998 (2004).
73. S.J. Stray, J.M. Johnson, B.G. Kopek and A. Zlotnick. An *in vitro* fluorescence screen to identify antivirals that disrupt hepatitis B virus capsid assembly. *Nat. Biotechnol.* **24**:358–362 (2006).
74. S.J. Stray and A. Zlotnick. BAY 41-4109 has multiple effects on Hepatitis B virus capsid assembly. *J. Mol. Recognit.* **19**:542–548 (2006).
75. F.M. Suk, M.H. Lin, M. Newman, S. Pan, S.H. Chen, J.D. Liu and C. Shih. Replication advantage and host factor-independent phenotypes attributable to a common naturally occurring capsid mutation (I97L) in human hepatitis B virus. *J. Virol.* **76**:12069–12077 (2002).
76. J. Sun, C. DuFort, M.C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V.M. Rotello, A. Holzenburg, C.C. Kao and B. Dragnea. Core-controlled polymorphism in virus-like particles. *Proc. Natl. Acad. Sci. USA* **104**:1354–1359 (2007).
77. W.S. Tan, M.R. Dyson and K. Murray. Two distinct segments of the hepatitis B virus surface antigen contribute synergistically to its association with the viral core particles. *J. Mol. Biol.* **286**:797–808 (1999).

78. T.L. Tellinghuisen, A.E. Hamburger, B.R. Fisher, R. Ostendorp and R.J. Kuhn. *In vitro* assembly of alphavirus cores by using nucleocapsid protein expressed in *Escherichia. coli*. *J. Virol.* **73**:5309–5319 (1999).
79. F. Ternois, J. Sticht, S. Duquerroy, H.G. Krausslich and F.A. Rey. The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. *Nat. Struct. Mol. Biol.* **12**:678–682 (2005).
80. C.M. Teschke, J. King and P.E. Prevelige, Jr. Inhibition of viral capsid assembly by 1,1′-bi(4-anilinonaphthalene-5-sulfonic acid). *Biochemistry* **32**:10658–10665 (1993).
81. R. Tuma, H. Tsuruta, K.H. French and P.E. Prevelige. Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid. *J. Mol. Biol.* **381**:1395–1406 (2008).
82. C. Uetrecht, C. Versluis, N.R. Watts, W.H. Roos, G.J. Wuite, P.T. Wingfield, A.C. Steven and A.J. Heck. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. USA* **105**:9216–9220 (2008).
83. S. Wang, P. Palasingam, R.H. Nokling, B.H. Lindqvist and T. Dokland. *In vitro* assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. *Virology* **275**:133–144 (2000).
84. O. Weber, K.H. Schlemmer, E. Hartmann, I. Hagelschuer, A. Paessens, E. Graef, K. Deres, S. Goldmann, U. Niewoehner, J. Stoltefuss, D. Haebich, H. Ruebsamen-Waigmann and S. Wohlfeil. Inhibition of human hepatitis B virus (HBV) by a novel non-nucleosidic compound in a transgenic mouse model. *Antiviral Res.* **54**:69–78 (2002).
85. W.R. Wikoff, L. Liljas, R.L. Duda, H. Tsuruta, R.W. Hendrix and J.E. Johnson. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**:2129–2133 (2000).
86. D. Willits, X. Zhao, N. Olson, T.S. Baker, A. Zlotnick, J.E. Johnson, T. Douglas and M.J. Young. Effects of the cowpea chlorotic mottle bromovirus beta-hexamer structure on virion assembly. *Virology* **306**:280–288 (2003).
87. P.T. Wingfield, S.J. Stahl, R.W. Williams and A.C. Steven. Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry* **34**:4919–4932 (1995).
88. S.A. Wynne, R.A. Crowther and A.G. Leslie. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell* **3**:771–780 (1999).
89. T.T. Yuan, G.K. Sahu, W.E. Whitehead, R. Greenberg and C. Shih. The mechanism of an immature secretion phenotype of a highly frequent naturally occurring missense mutation at codon 97 of human hepatitis B virus core antigen. *J. Virol.* **73**:5731–5740 (1999a).
90. T.T. Yuan, P.C. Tai and C. Shih. Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen. *J. Virol.* **73**:10122–10128 (1999b).
91. R. Zandi, D. Reguera, R.F. Bruinsma, W.M. Gelbart and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA* **101**:15556–15560 (2004).

92. T. Zhang and R. Schwartz. Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys. J.* **90**:57–64 (2006).
93. A. Zlotnick. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**:59–67 (1994).
94. A. Zlotnick, R. Aldrich, J.M. Johnson, P. Ceres and M.J. Young. Mechanism of Capsid Assembly for an Icosahedral Plant Virus. *Virology* **277**:450–456 (2000).
95. A. Zlotnick, P. Ceres, S. Singh and J.M. Johnson. A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J. Virol.* **76**:4848–4854 (2002).
96. A. Zlotnick, J.M. Johnson, P.W. Wingfield, S.J. Stahl and D. Endres. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**:14644–14652 (1999).

Chapter 7

Assembly and Disassembly of Deltahedral Viral Shells

Alexander Yu. Morozov*, Joseph Rudnick*, Robijn F. Bruinsma* and William S. Klug†

We examine the assembly and disassembly of $T = 13$ viral shells using a deformable version of the Caspar-Klug deltahedral model. As the cohesive energy is reduced, the shell releases elastic energy by rupturing. The path of rupture is determined by the intrinsic elastic stress pattern of icosahedral shells. Surprisingly, spherical shells are more prone to rupture than icosahedral shells. Next, assembly of a $T = 13$ deltahedral shell is shown to be *impossible* along the conventional pathway of compact partial shells. Instead, we propose a *non-compact* assembly pathway, with the elastic energy directing the assembly through 'Whiffle-Ball' type assembly intermediates.

1. Introduction

The molecular structure of the protein shell ('capsid') that surrounds the genome of spherical viruses has been elucidated, often in exquisite detail, by X-ray crystallography and cryo-TEM methods. These studies largely confirmed the classical Caspar-Klug (CK) T-Number 'quasi-equivalence' classification, which states that spherical viral shells are composed of twelve pentamers and $10(T-1)$ hexamers with $T = 1, 3, 4, 7, 12, 13, \ldots$ an integer (though some striking exceptions were documented as well[2,3]). These structural studies revealed that even if the capsid is composed of identical proteins — $60T$ in total — the *conformation* of the proteins making up

*Department of Physics and Astronomy, University of California, Los Angeles, CA, 90024.

†Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, 90095.

the twelve pentamers can differ significantly from the conformation of the hexameric proteins, as is reasonable in view of the different local packing symmetry. For large viruses, the structural deformations of the hexamers bordering the pentamers can be productively described by the application of thin-shell elasticity theory.[5] All-atom numerical simulations of small viral shells have been carried out as well.[6]

During viral assembly or disassembly the number of proteins comprising a shell is less than that of the complete shell. Our understanding of the structure and physical properties of *incomplete* viral shells is much more limited than that of complete shells, yet such an understanding is essential for a physical theory of viral assembly and disassembly. First, it is not possible to make crystals of incomplete assembly intermediates, which makes it difficult to reconstruct incomplete shells by diffraction methods. For a given number of proteins, moreover, there can be a wide variety of competing, incomplete shell structures. All-atom numerical simulations of viral shell assembly are challenging, in particular for large shells. Numerical studies of viral assembly have mostly used the 'rigid capsomer' model[6–10] which treats capsomers as rigid units. Growth takes place on a capsomer by capsomer basis. These studies found that the success rate of the assembly process depends sensitively on assumptions about the bonding between the capsomers but few general principles have emerged so far. One of the results of the current study is that rigid capsomer models leaving out the elastic deformations of incomplete viral shells may lead to seriously misleading results. Separately, the classical theory of nucleation and growth has been applied to viral assembly.[12] In that case a *line tension* λ, of the order of the subunit bonding energy divided by the size of a subunit, is assigned to an exposed edge of an incomplete shell. In order to minimize this line energy, an incomplete shell is expected to have a *compact* shape, such as a sphere of radius R with a spherical cap removed. The size of the cap can be defined by the angle θ of a radius vector from the center of the sphere to edge of the cap with an axis through the center perpendicular to the cap. As the cap grows, this angle increases from zero to π while the perimeter varies as $2\pi R \sin\theta$. The line energy reaches a maximum value $\lambda 2\pi R$ when $\theta = \pi/2$ after which it decreases symmetrically back to zero when the capsid is completed. This means that under conditions of thermal equilibrium there would be a *nucleation barrier* $2\pi R\lambda$ preventing assembly — believed to be in the range of hundreds of $k_B T$ — with the half-formed shell playing the role of the critical nucleus. For increasing levels of supersaturation, the

critical nucleus would be progressively smaller. Kinetic experiments on viral assembly indicate that the critical nucleus is in fact the size of a capsomer. The spherical cap model also does not include elastic deformation. One of the results of this study is that it is very likely that growth nuclei of larger shells in fact may not be compact due to the effects of elastic deformation.

In this chapter we discuss the results of a study of capsid assembly and disassembly based on the *deltahedral* model of viral capsids. The deltahedral model was introduced by Caspar and Klug in their original paper to derive the T Number classification of viral shells.[1] Capsid proteins are represented as equilateral triangles in this model, and capsomers as hexamer and pentamer aggregates of such triangles. We will use a *deformable* version of the deltahedral model in which both the shape of the triangles and the bonds linking them are allowed to deviate from their optimal value. The energy cost of these deviations is described by a quadratic elastic energy. In the continuum limit of very large T Numbers, the deformable deltahedral model reduces to thin-shell theory of continuum elasticity theory referred to earlier. Thin-shell theory has been found to provide a good description of certain generic features of larger T Number shells, such as asphericity,[5] shape transitions[13,14] and the response to deformation of viral shells by an Atomic Force Microscope[15,16] (AFM), although it obviously does not describe the internal degrees of freedom of the capsomer proteins.

According to thin-shell theory, an icosahedral viral shell is characterized by *an intrinsic distribution of elastic stress focused on the five-fold symmetry sites* that is not included in the rigid capsomer model and in the nucleation-and-growth theory of viral assembly, nor has the role of elastic stress been examined for the process of viral disassembly. We will show, for the case of a $T = 13$ deformable deltahedral model, that elastic deformation of the pentamers and the surrounding ring of hexamers actually should be expected to play a key role both during assembly and disassembly. In Section 2 of this contribution, we start with a description of the deformable deltahedral model and then compute numerically the shape and elastic stress distribution of a complete $T = 13$ shell as a function of the stretching and bending rigidities of the triangles. We also will examine how the local stress distribution depends on assumptions about the *internal structure* of the pentamers. In Section 3 we study how a $T = 13$ shell *ruptures* as one lowers the binding energy. The pre-existing stress pattern will be shown to *direct* rupture along a very specific template: cracks start along the border of a pentamer, then propagate from pentamer to

pentamer along an icosahedral edge. Separately, rupture also can cut out pentamers from the shell. The elastic energy released by the opening of the crack is computed as a function of the shear rigidity of the triangles and shown to depend sensitively on the assumed structure of the pentamers. Surprisingly, shells with a more spherical shape are much more prone to rupture than shells with a more icosahedral shape. Changes in the ratio of stretching and bending energies can be used as a control parameter to focus elastic energy on the pentamers and thereby initiate disassembly. In Section 4 we study the assembly of a $T = 13$ shell with the following results. Within the deltahedral model, *a $T = 13$ icosahedral capsid cannot grow by self-assembly as a compact shell on a capsomer-by-capsomer basis.* The elastic energy cost of introducing pentamers is so large that it always is more favorable to continue adding hexamers to the growing shells, leading to a 'kinetic trap' in the form of misfolded shells. Within the deltahedral model, the assembly of the shell cannot be described by the classical theory of nucleation and growth, at least for the case of large icosahedral viral shells. Second, we find that as the binding free energy is increased from zero to initiate assembly, large icosahedral shells *can* form within the deltahedral model but only through a *non-compact* assembly pathway with the *elastic energy acting as the directing agent for positioning the capsomers in an icosahedral pattern*. 'Whiffle-Ball' type open icosahedral structures are proposed as key assembly intermediates.

2. Deformable Caspar — Klug Deltahedral Shells

Figure 1 shows a full $T = 13$ CK icosahedral shell as a net of 780 triangles.

The triangles are assigned to 120 hexamers, each with six triangles and twelve pentamers, each with five triangles. In terms of symmetry, the hexamers of a $T = 13$ shell can be divided into two inequivalent groups. The first group comprises the symmetry-equivalent hexamers that surround the three-fold symmetry icosahedral sites (indicated by a black arrow in Fig. 1). The second group comprises the symmetry-equivalent hexamers that surround each pentamer by five equivalent hexamers. We will use and compare two different versions of the deltahedral model. In the first version, hexamer and pentamer triangles are all identical equilateral triangles, as in the original CK model. The minimum energy structure of an isolated

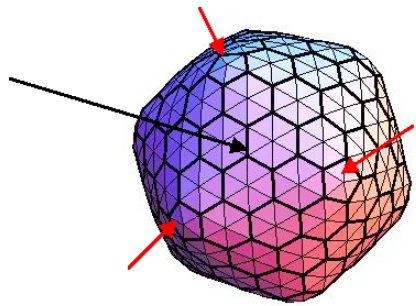


Fig. 1. $T = 13$ Caspar–Klug Icosahedron. The outlines of the 120 hexamers and 12 pentamers are indicated as heavy lines. Pentamers are indicated by red arrows and a three-fold symmetry site, surrounded by three hexamers, is indicated by black arrow. Note that the shell is chiral.

pentamer is then a *pentagonal pyramid*. Note that pentagonal pyramids favor a more facetted icosahedral shape. In the second version, the pentamer triangles are isosceles instead of equilateral and the minimum energy structure of a pentamer is a *regular pentagon*. Regular pentagons favor a more spherical capsid shape. These two versions will be used to test the sensitivity of capsid assembly and disassembly scenarios against assumptions about the internal structure of the pentamers.

The elastic stress distribution will be obtained by minimizing the following quadratic energy:

$$E = \frac{\alpha_3}{2} \sum_{triangles\ j} \left(\sum_{k=1,2,3} \left(\cos\theta_{jk} - \cos\theta_0\right)^2 \right) + \frac{\alpha_2}{2} \sum_{triangles\ j} (A_j - A_0)^2 + \frac{\alpha_1}{2} \sum_{triangle\ neighbors\ <ij>} (\cos\phi_{ij} - \cos\phi_0)^2 - N_e\varepsilon + N_5\mu_5 + N_6\mu_6 \tag{1}$$

The first term, the *shear energy*, is minimized when all hexameric triangles are equilateral and all pentameric triangles either equilateral or isosceles, depending on the variant that is used. The term is a sum over all the triangles j of the net and then, for each triangle, over the three interior angles θ_{jk} of that triangle, with $k = 1, 2$ or 3 (see Fig. 2b).

If the triangle is part of a hexamer then the optimal angle θ_0 is equal to $\pi/3$. If the triangle is part of a pentamer then the value of θ_0 differs

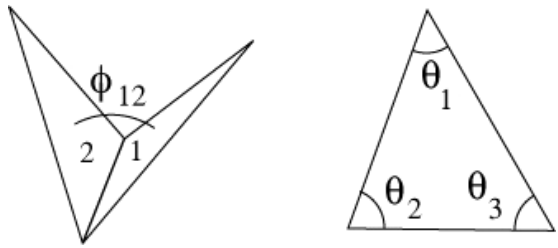


Fig. 2. Bending angle (a) and internal deformation angles (b) as they appear in the energy expression (Eq. (1)).

for the two variants of the model. For the case that isolated pentamers are pentagonal pyramids (as in the CK deltahedral model), θ_0 is equal to $\pi/3$. For the case that isolated pentamers are regular pentagons, θ_0 is equal to $2\pi/5$ for an angle of the triangle pointing to the center of the pentamer and $3\pi/10$ for the two base angles. The parameter α_3 is a measure of the in-plane shear rigidity of the triangular 'proteins'.

The second sum, the *area energy*, is minimized if the area of each triangle is equal to that of the triangles that compose a regular hexamer, a regular pentamer, or regular pentagonal pyramid. The sum is over all triangles with A_j the triangle area. If the triangle is part of a hexamer, then the optimal triangle area A_0 is equal to $(\sqrt{3}/4)l^2$, i.e., the area of an equilateral triangle with edge length l. If the triangle is part of a pentamer, then for the case of isosceles triangles, A_0 is equal to $\frac{1}{4}\tan(3\pi/10)l^2$, i.e., $1/5$th of the area of a flat pentamer with edge length l. The parameter α_2 is a measure of the compressional stiffness of the triangles.

The third sum, the *bending energy*, is minimized if the *normal directions* of two adjacent triangles make an angle θ_0 equal to $\sqrt{3}\frac{l}{R}$ where R is the radius of a completely spherical $T = 13$ shell (see Fig. 2a). The curvature angle between adjacent triangles tessellating a sphere of radius R (with R large compared to the edge length l) equals $\sqrt{3}\frac{l}{R}$. It follows that our choice for the optimal curvature angle is such that the third term is minimized for a $T = 13$ shell with the shape of a sphere of radius R. The sum is over all pairs of adjacent triangles with ϕ_{ij} defined as the angle between the normals of two adjacent triangles (see Fig. 2a). The parameter α_1 measures the *bond stiffness* of the link between two triangles, i.e.,

the energy cost of rotating the relative orientation of the two adjacent triangles away from the relative preferred orientation.

The fourth term represents the *cohesive free energy* gain of shell assembly. Here, ε is the free energy gain per 'bond' shared by two adjacent triangles that are part of different capsomers (hexamers or pentamers) with N_e the number of shared edges of capsomers of the incomplete shell. This term effectively includes the line energy of the classical nucleation-and-growth theory (note that we do not allow break up of capsomers). The fifth and sixth terms represent the free energy cost of removing capsomers from solution with $\mu_{5,6}$ the concentration-dependent chemical potential of, respectively, a pentamer or hexamer in solution, and with $N_{5,6}$ the number of pentamers or hexamers of which the shell is composed. Any structural free energy differences between pentamers and hexamers are included in these two terms.

The last three terms, by themselves, would be the starting point of a conventional nucleation-and-growth theory of capsid assembly/disassembly in the absence of elastic energy effects. The assembly free energy would be $E_0(T = 13) = -390\varepsilon + 12\mu_5 + 120\mu_6$ since a $T = 13$ shell has 390 bonds, 120 hexamers, and 12 pentamers. The bond free energy ε acts here as the key control parameter for assembly. If one increases ε from zero, assembly under conditions of thermal equilibrium would start when E_0 changes sign, i.e., when the bond energy reaches the threshold $390\varepsilon \simeq 12\mu_5 + 120\mu_6$. This form of assembly requires activation over a large barrier, as mentioned. If activated assembly is not permitted, then assembly from a supersaturated solution would require a somewhat larger threshold bonding energy $\varepsilon \sim \mu_{5,6}$.

We now return to the case of where we do include all terms of Eq. (1), starting with assembled $T = 13$ shells. Figure 3 shows the result of a numerical minimization of the energy for the parameter values $\alpha_1 = 400$, $\alpha_2 = 10$, $\alpha_3 = 10$. The bending energy parameter is thus large compared to the area and shear parameters. We are assuming here the model variant that isolated pentamers are flat regular pentagons in equilibrium.

The total energy after minimization is 9.81. If we maintain *fixed ratios* between the three elastic energy parameters — in this example 400:10:10 — then it follows from the general form of the energy expression Eq. (1) that only the overall energy scale would change but not the *shape* of the shell. If one changed the value of the bending energy

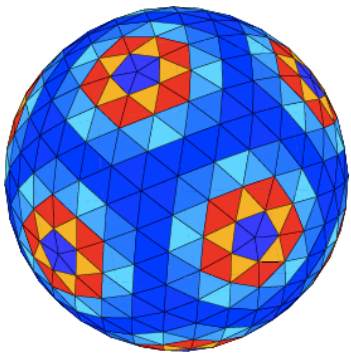


Fig. 3. Elastic energy distribution of a $T = 13$ shell with $\alpha_1 = 400$, $\alpha_2 = 10$, $\alpha_3 = 10$. The color codes for the elastic energy of a triangle with relative energy shown as a spectral change from deep blue (lowest energy) to red (highest energy). The bending energy of two adjacent triangles is assumed shared equally between them. The dominant contribution to the non-uniform stress surrounding the five-fold sites is stretching energy.

α_1, for fixed parameter ratios, then the elastic energy would scale as $(9.81/400)\ \alpha_1$ and the assembly free energy would be $E(T = 13) = 0.024\alpha_1 - 390\varepsilon + 12\mu_5 + 120\mu_6$. The effect of the elastic energy cost is to increase the threshold bonding energy for self-assembly by an amount $\Delta\varepsilon = \alpha_1(0.024/390)$ that is small compared to the bonding energy.

Note that the shape of the shell in Fig. 3 is close to spherical. This is not surprising since the bending energy is minimized for a spherical shape. Note also that each of the twelve pentamers is surrounded by an aureole of elastic stress, indicated by the red triangles. In our units, the energy of the three deep-blue hexamers that surround the three-fold symmetry site is 0.0188 while that of a hexamer bordering a pentamer is 0.144 or about eight times larger. The pentamer triangles themselves are hardly deformed at all, with an energy of only 0.0018 per capsomer. The dominant energy contribution to this stress non-uniformity is, surprisingly, *not* the bending energy but rather the *shear energy*. Physically, this can be understood from the viewpoint of the elasticity theory. A site with five-fold symmetry in an otherwise hexagonal lattice is a five-fold disclination, surrounded by a non-uniform pattern of shear elastic stress distribution. For a disclination located at the center of a flat disc of radius r, the elastic energy is of the order of Υr^2 with $\Upsilon \propto \alpha_1$ the 2D *Young's Modulus* of the lattice.[5] A spherical shell of radius R is thus expected to have a shear elastic energy of the order of ΥR^2.

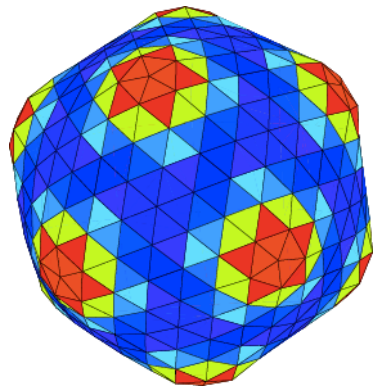


Fig. 4. Elastic energy distribution of a $T = 13$ shell with $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. Note that the pentamers are more deformed than in Fig. 3.

In Fig. 4 we show the result of an energy minimization run for the case of $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. The bending energy parameter is thus small compared to the shear and area energy parameters. Isolated pentamers are again regular pentagons.

The total energy, after relaxation, is 1.102. For a ratio set of 1:10:10, the assembly free energy is thus given by $E(T = 13) = 1.1\alpha_1 - 390\varepsilon + 12\mu_5 + 120\mu_6$. The shell shape is now approximately icosahedral. This can be understood from the fact that the CK construction of T Number icosahedra is *isometric*, meaning that the triangles of a CK shell are not sheared. It is thus reasonable for the shell to adopt the shape of a CK icosahedron if the shear rigidity of the triangles is relatively large. The non-uniform stress distribution is now dominated by the bending and area contributions to the elastic energy. This, too, can be understood from 2D elasticity theory. The twelve five-fold symmetry sites of a CK shell have a *conical* shape and the elastic energy of a cone of radius R is of the order of $\kappa \ln(R/l)$ with κ the 2D bending modulus. The stress non-uniformity is less pronounced than in Fig. 3: the hexamers bordering the three-fold site have an energy of 0.0054 while the hexamers bordering a pentamer have an energy of 0.0104, or about three times larger. The pentamers have an elastic energy of 0.0131 and now the pentamers do carry a significant amount of elastic deformation. This can be understood by noting that if the five-fold symmetry sites of an icosahedron would be capped by perfectly flat pentamers, then this would entail a significant amount of bending energy

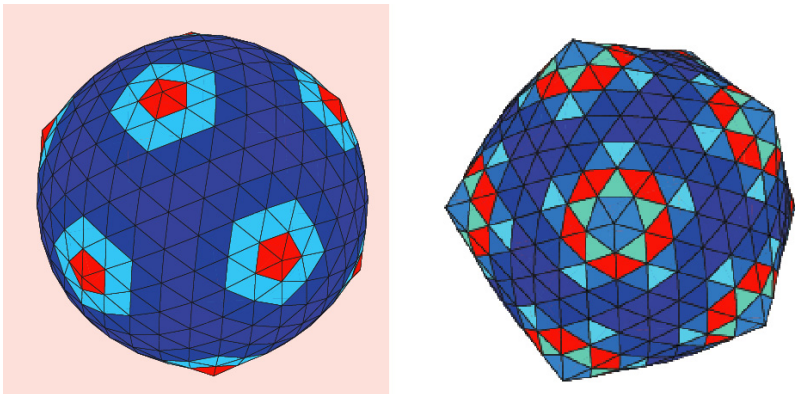


Fig. 5. Elastic energy distribution of a $T = 13$ shell. The left panel shows the case of $\alpha_1 = 400$, $\alpha_2 = 10$, $\alpha_3 = 10$ and the right panel shows the case of $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. Isolated pentamers are pentagonal pyramids. Note that the pentagonal pyramids are more deformed in the left panel.

cost for the bonds between the pentamer and the surrounding hexamers. The pentamers are thus deformed away from their optimal flat shapes.

This energy focusing effect suggests that the elastic energy of the pentamers sites might be sensitive to assumptions that are made about the internal structure of pentamers. So far, we used the variant of the deltahedral model where isolated pentamers had a preferred shape that was a flat equilateral pentagon. Figure 5 shows the minimum energy shell for the same parameters as the ones used in Figs. 3 and 4 but now assuming the case that pentamer triangles are equilateral, so isolated pentamers have the shape of pentagonal pyramids in equilibrium.

In the first case, with $\alpha_1 = 400$, $\alpha_2 = 10$, $\alpha_3 = 10$, the elastic energy is 24.44 so the assembly free energy is $E(T = 13) \approx -390\varepsilon + 0.061\alpha_1 + 12\mu_5 + 120\mu_6$ for this parameter ratio set. In the second case, with $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$, the elastic energy is 0.545 so the assembly free energy is $E(T = 13) \approx -390\varepsilon + 0.545\alpha_1 + 12\mu_5 + 120\mu_6$ for this parameter ratio set. Compared with the previous case, for the 1:10:10 ratio set, the facetted shells with pyramidal pentamers have an elastic energy that is (roughly) a factor of two *smaller* than that of shells with regular pentagon pentamers while for the 400:10:10 ratio set, the elastic energy of the spherical shells is (roughly) a factor of three *larger*. This is a reasonable result since a pyramidal pentamer fits much better

on the vertices of a facetted icosahedron than splayed out on a spherical surface. We conclude that the internal structure of the pentamers rather strongly affects the elastic energy stored in deltahedral $T = 13$ shells. The two model variants do share an important common feature: independent of the internal structure of the pentamers, the pentamer elastic energy focusing is much more pronounced when the shell shape is approximately spherical. For the case of pentagonal pyramidal pentamers, the hexamers bordering the pentamers have an elastic energy that is about 8.62 times larger than that of hexamers in contact with the three-fold symmetry site for the 400:10:10 ratio set while for the case of regular pentagon pentamers this ratio is about 7.68. For the 1:10:10 parameter set, the respective ratios are 2.69 and 1.93. The degree of energy focusing is thus quite similar for the two variants. It is only the *overall elastic energy scale* that is different for the two models.

Does the internal structure of the pentamers exert an influence not only on the energy but also on the *shape* of the capsid? The shape of the shells only depends on parameter ratios as noted, not on the overall scale of the elastic energy. In Fig. 6 we show the *asphericity*, the reduced

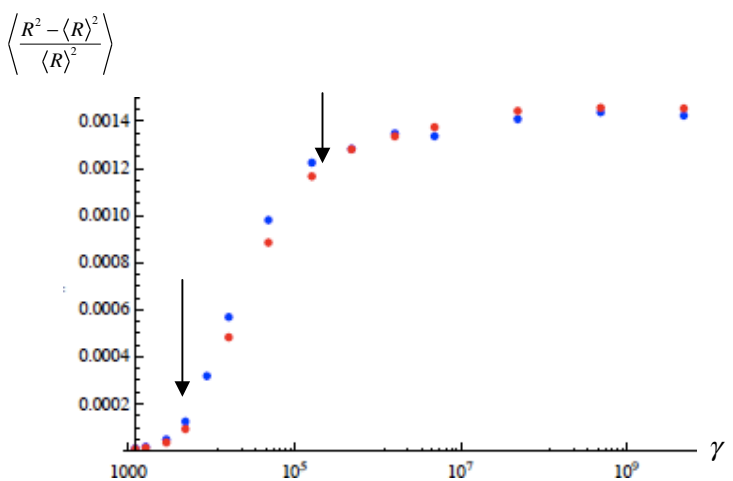


Fig. 6. Asphericity (vertical axis) as a function of the ratio of stretching and bending energy parameters $\frac{\alpha_3}{\alpha_1}$. The horizontal axis is $\gamma = 60\frac{\alpha_3}{\alpha_1}(R/l)^4$ with R the capsid radius and l the edge length of a triangle. Blue dots indicate equilateral pentamer triangles and red dots isosceles pentamer triangles. The two arrows correspond to the two sets of parameter ratios of Figs. 3–5.

second moment of the radius distribution, of a $T = 13$ deltahedral shell as a function of the parameter ratio $\gamma = 60\frac{\alpha_3}{\alpha_1}(R/l)^4$. Blue dots indicate equilateral pentamer triangles and red dots isosceles pentamer triangles.

This choice for the horizontal axis was motivated by the fact that within thin-shell continuum theory there is only one dimensionless parameter ratio that controls the distribution of the elastic energy and the shell shape, namely the ratio $\frac{YR^2}{\kappa}$ of shear and bending elastic energies (the so-called 'Foppl-von Karman (FvK) Number'). If this number exceeds a critical value of order 100 then, within thin-shell theory, the shape of the shell changes in a rather pronounced way from spherical to icosahedral.[12] The parameter $\gamma = 60\frac{\alpha_3}{\alpha_1}(R/l)^4$ roughly corresponds to the FvK parameter.

According to Fig. 6, the asphericity grows in a pronounced way around $\gamma \sim 5{,}000$, though the shape transition is not as well defined as it is in the continuum limit. The two cases shown in Figs. 3, 4, and 5 straddle the transition (arrows in Fig. 6). The two plots are quite similar. We conclude that, even though the internal structure of the pentamers strongly influences the elastic energy, it has no significant effect on the asphericity plot: *the global equilibrium shape of the capsid is not dependent on the internal structure of the pentamers.* In the next section, we will examine the impact of elastic energy focusing on shell rupture.

3. Stress-Driven Capsid Rupture

In this section we will discuss how capsid disassembly is influenced by elastic stress. As before, the bonding energy ε will act as the control parameter. Disassembly will be initiated as follows. A bond linking two triangles will be allowed to dissociate if the elastic energy cost of the bond, as computed in the previous section, exceeds the bonding energy ε. Note that the pentamer and hexamer chemical potentials are not involved in bond rupture at this stage. Bonds with the highest elastic energy cost will dissociate first. We will be assuming a mechanically *adiabatic* rupture process, that is, we will assume that the disassembly process is sufficiently slow so *the partial shell remains in mechanical equilibrium.* Under this assumption, the shape of a shell at any stage of disassembly can be obtained by minimizing the energy of the shell, independent of any kinetic rate parameters.

In order to implement the rupture process, we start from a fully assembled shell, with minimum elastic energy, make a list of bond energies and,

as we reduce ε, allow bonds with the highest energy to fail, replacing the single shared edge of two triangles by *two* edges, one for each triangle. The total capsid energy is minimized after a bond rupture event and a new list is made of bond energies. Next, the highest energy bond *adjacent to a previously failed bond* is allowed to fail. As a result, one or more rupture cracks start to grow along the surface. One of the bonds at the end of one of a cracks is usually the highest energy bond that will fail at the next step.

Figure 7 Shows how a shell with a very large T Number and a large FvK Number ruptures under this scenario.

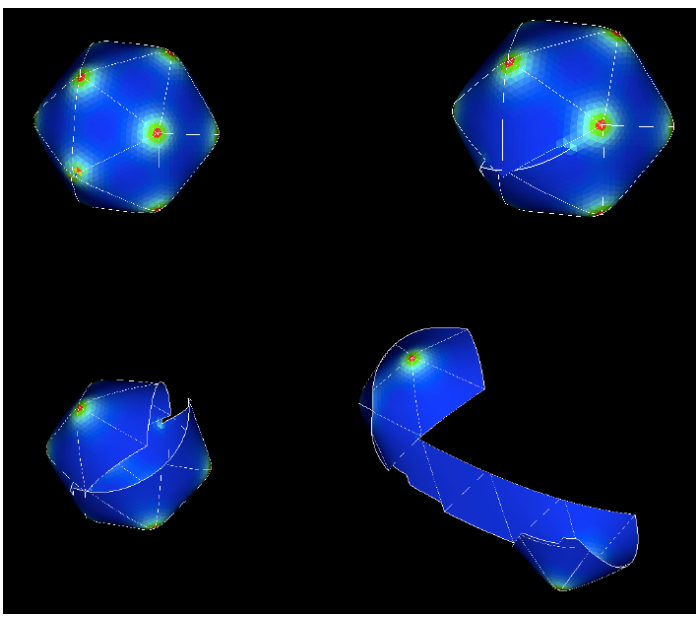


Fig. 7. Rupture of a large T Number CK icosahedron. Upper left-hand panel: disassembly has just started at the five-fold site in the lower left-hand corner of the shell. Upper right-hand panel: two cracks have emerged from this five-fold site propagating along two of the ridges of the icosahedron. The stress distribution of the icosahedron away from the cracks is close to that of the complete shell. Lower left-hand panel: one of the two cracks has stopped propagating while the other crack continues on along icosahedral ridges from five-fold site to five-fold site. Note that it appears to be in the process of cutting out a cap with five pentamers. Lower right-hand panel: at a later stage, both cracks continue to propagate along ridges. Two caps are linked by a flat, stress-free strip leading to a flat, stress-free *template* of an icosahedron.

Rupture starts at a five-fold site. Two connected cracks open up along the ridges of the icosahedron and propagate from five-fold site to five-fold site. The propagation of the crack is driven by the progressive release of the elastic energy stored in the non-uniform stress distribution. Eventually a planar *template* of an icosahedron emerges. A template of an icosahedron is a section of a hexagonal sheet that can be folded into the shape of an icosahedron. The rupture process can be compared with a propagating *Griffiths crack* of classical elasticity theory, except that in our case the elastic stress is not applied externally, but instead is the intrinsic heterogeneous stress of an icosahedral shell. The eventual dissolution of this planar template, with the capsomers being released into solution, was not investigated here except that we note that, unlike the rupture process, this final dissolution of the shell will be dependent on the chemical potential.

Before investigating the energetics of the rupture process of a smaller $T = 13$ shell, it is useful to *guess* how it might depend on the ratio γ of stretching and bending energy parameters. If this ratio is small, so the bending energy is dominant, then the stress energy released by the formation of a crack is expected to be modest. The reason is that in this regime shells have the shape of a sphere, with a radius equal to the preferred curvature radius R. If a spherical shell with radius equal to the preferred curvature radius is cut along an equatorial circle then there is no release of elastic stress (as can be verified by slicing open a tennis ball). On the other hand, in the limit of large γ, the folds of the CK icosahedral shell are a focus of both bending and shear elastic energy, as discussed by Lobkovsky.[17] Cutting open such a shell from a five-fold site to a five-fold site *is* expected to release a significant amount of elastic energy. A spherical shell is thus expected to be more resistant against rupture than an icosahedral shell.

In order to check this guess, we evaluated the elastic energy of ruptured $T = 13$ shells for different FvK Numbers. An example of a cracked shell is shown in Fig. 8: the crack starts at a pentamer, reaches a second pentamer and is in the process of continuing on to a third pentamer (by assumption, a crack does not fracture pentamers or hexamers).

The crack breaks one of the bonds linking the second pentamer to its neighboring hexamers. Curiously, the *highest* concentration of elastic stress on the shell — indicated by the two red triangles — is on the opposite side of this same pentamer so the progression of a crack generates additional stress non-uniformities that could be the starting point of new cracks.

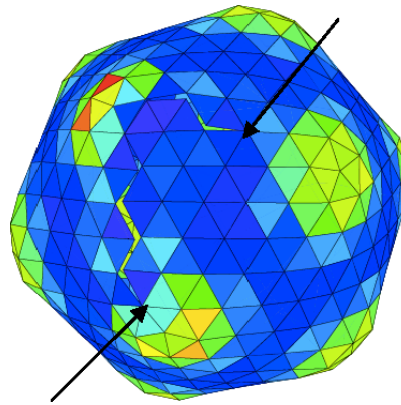


Fig. 8. $T = 13$ shell with a crack for $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. The crack starts from the tip of the lower arrow at the edge of the lower center pentamer and ends between two pentamers (upper arrow) after releasing one link between the pentamer at the upper left and a hexamer. Note the increased elastic energy on the side of this pentamer opposite to the crack.

Figure 9 shows how the elastic energy of the shell decreases as this crack elongates.

The elastic energy is released in *bursts* by the elongating crack. Between link numbers $n = 2$ and $n = 4$, and between $n = 7$ and $n = 10$, the fraction of released energy is rather small. These values correspond to dissociation of links between hexamers along the ridge that separates pentamers. The larger fraction of released energy between $n = 4$ and $n = 5$, between $n = 6$ and $n = 7$, and between $n = 10$ and $n = 11$ is associated with the dissociation of a bond between two hexamers immediately adjacent to a pentamer. The amount of energy released between $n = 5$ and $n = 6$ is again modest and corresponds to bond dissociation between a pentamer and a hexamer: *most of the elastic energy release during rupture takes place close to the pentamers*. We encountered a very similar plot when we studied the *release* of a pentamer by allowing the crack to circle a pentamer, instead of linking two pentamers: the elastic energy is again released in bursts.

We can use Fig. 9 to graphically analyze the rupture process. The total change in free energy when n links are broken is $\Delta F(n) = \varepsilon n + E(n) - E(0)$ where the last two terms are purely elastic. If this quantity is negative then it is energetically favorable to release n links. A simple graphical analysis,

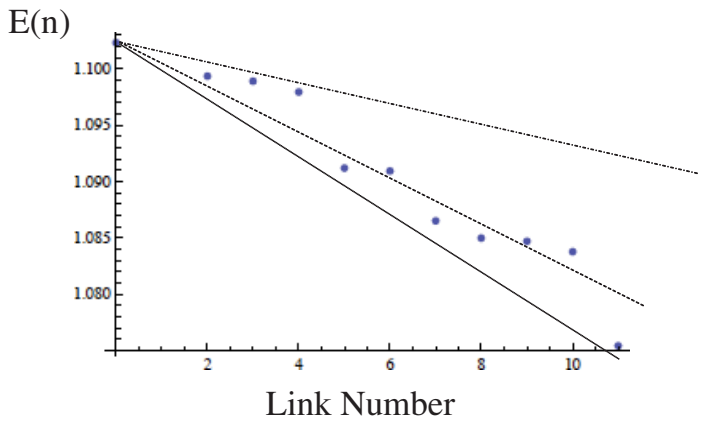


Fig. 9. Elastic energy $E(n)$ of the shell as a crack progresses along the path of Fig. 8. The energy parameters are $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. The energy release is most pronounced when a bond is broken that linked two hexamers that both were adjacent to a pentamer, such as $n = 4$ to $n = 5$, $n = 6$ to $n = 7$ and $n = 10$ to $n = 11$ (isolated pentamers are regular pentagons). The geometrical construction involving the straight lines is described in the text.

shown in Fig. 9, shows what happens as one reduces the bonding energy ε. The change in free energy $\Delta F(n)$ is the distance between $E(n)$ in the graph and the straight lines representing different values of ε. The dotted line (with steepest slope) is representative of the case of large ε. All $\Delta F(n)$ values are positive so there is no rupture. The dash-dotted line is representative of the case of small ε. All $\Delta F(n)$ values are now negative so the rupture process is 'downhill' and spontaneous. Rupture continues until the capsid resembles the last stage of Fig. 8. Finally, for the dashed line, some points are above and some below the curve. In fact, only bonds adjacent to pentamers can rupture spontaneously: the capsid is *metastable* for ε values in this range. It follows from the graphical analysis of Fig. 9, and some trivial scaling arguments, that for a fixed 1:10:10 ratio set, spontaneous rupture starts when the bond energy drops below a critical value $\varepsilon \sim 0.003\alpha_1$.

Percentage of released elastic energy

In Fig. 10, we show the percentage of elastic energy released by a crack of five broken bonds linking two pentamers as a function of $\gamma = 60\frac{\alpha_3}{\alpha_1}(R/l)^4$, the dimensionless ratio of stretching and bending energies.

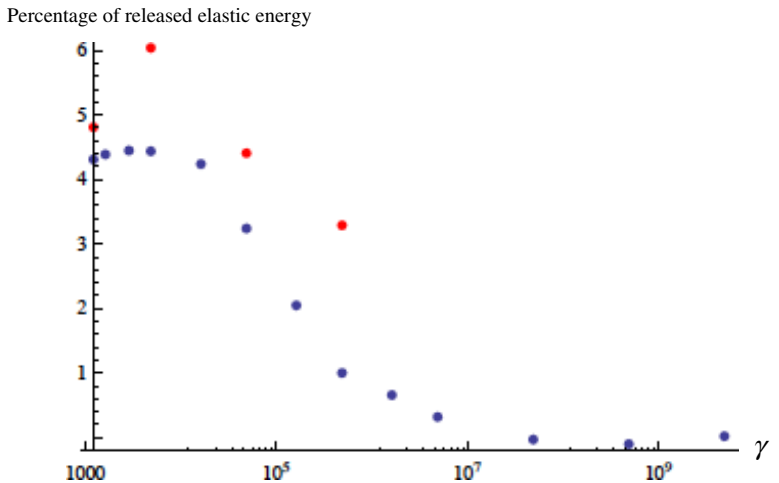


Fig. 10. Percentage of released elastic energy of a crack linking two pentamers for different values of the dimensionless ratio γ of stretching and bending elastic energies. The energy parameters are those of Fig. 8. Blue dots: isolated pentamers are flat pentagons. Red dots: isolated pentamers are pentagonal pyramids.

For the case that isolated pentamers are flat pentagons, the amount of released elastic energy decreases in a pronounced way as a function of γ (blue dots). For values of γ small compared to 10^4, about 4% of the stored elastic energy is released by the crack while for γ large compared to 10^4, the percentage of released elastic energy is negligible. The drop in released energy takes place over the same range of γ values where the shape transition took place in Fig. 6. It follows that if two deltahedral shells have the same total elastic energy, but a different value of γ and hence a different shape, then the shell that is more spherical is more prone to rupture. This is *exactly* opposite to our original guess: for the same level of total stored elastic energy, *spherical shells are energetically less stable than icosahedral shells.*

The most obvious interpretation is that the release of elastic energy upon rupture for small γ values is due to the release of the large amount of shear elastic energy surrounding the immediate neighborhood of a five-fold symmetry site of a flattened hexagonal lattice. For larger γ values, buckling of the shells reduces the energy focusing effect and this apparently stabilizes the shell against rupture. This interpretation can be tested by considering the case in which isolated pentamers are regular pentagonal

pyramids. In this case, even more energy is stored in the five-fold sites for small γ values, since the pyramids have to splay out to fit on a shell. Indeed, the decrease of the percentage of released energy is significantly more pronounced. For very small γ values the percentage of released energy appears to come back down to approximately the same value as in the previous case, presumably because the shear energy eventually has to go to zero in that limit. One conclusion that follows from Fig. 10 is that, unlike the asphericity, the percentage of elastic energy released by a crack is not a universal quantity, depending only on the FvK Number. Instead, it appears to be strongly affected by the internal structure of the pentamers so the application of continuum elasticity theory to study capsid rupture may be questionable.

4. Capsid Assembly and Elastic Stress

In this final section, we look at the effects of elastic energy on the assembly process of a $T = 13$ deltahedral shell. As before, the control parameter is the cohesion free energy per bond ε except that now we increase ε from $\varepsilon = 0$ and try to grow a shell, capsomer by capsomer, from solution. Assume a shell is growing, capsomer-by-capsomer, driven by the cohesive free energy gain. How does an incomplete shell 'know' when to add a hexamer and when a pentamer? Pentamers, and the hexamers that surround them, in general are in a state of significant elastic strain. Within continuum theory, a pentamer corresponds to a five-fold disclination, and an isolated disclination is *attracted* to a free boundary, such as the edge of an incomplete shell, by an 'image-type' force. How can the introduction of a five-fold disclination at specific locations inside a growing shell ever be energetically favorable? For a defect-free hexagonal sheet with no preferred curvature, the introduction of a pentamer should only raise the energy. However, this need not be the case in the presence of preferred curvature because the bending free energy of a *flat* circular hexagonal sheet of radius R increases as R^2 if there is a non-zero preferred curvature. By introducing pentamers into a flat hexagonal sheet, it is provided with intrinsic ('Gaussian') curvature so the bending energy can be reduced, albeit at the cost of introducing a defect. For large enough R, it must be energetically favorable to include pentamers in a growing hexagonal shell.[18] One thus expects on intuitive grounds that spontaneous curvature must be the mechanism that guides pentamer introduction.

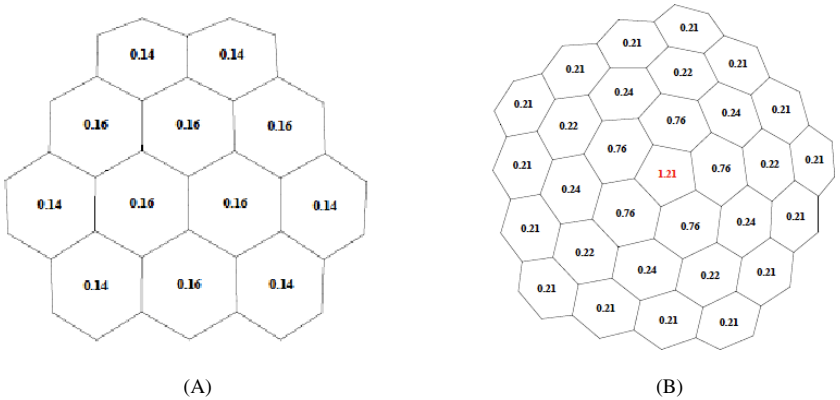


Fig. 11. Compact incomplete shells of a $T = 13$ icosahedron with three-fold symmetry (A) and five-fold symmetry (B). The energy parameters were $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$. The elastic energy (times 100) of each capsomer is indicated. Isolated pentamers are flat pentagons.

To verify this guess, we start by examining the elastic energy of spherical caps of the classical theory of nucleation and growth discussed in Section 1. We will consider incomplete shells of this form having three-fold and five-fold symmetry, respectively. Figure 11 shows the elastic energy distribution of a three-fold and of a five-fold symmetric incomplete shell (isosceles pentamer triangles) following energy minimization. The three-fold shell has two coordination rings of hexamers and the five-fold shell three coordination rings. In either case, additional pentamers would have to be introduced in the next coordination ring if one wanted to grow a CK shell.

In the first — pentamer-free — case, the elastic energy of a patch of twelve hexamers was 0.018 for $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$, mainly due to the bending energy imposed by the preferred curvature. In the second case, the central pentamer introduced a local stress field that tapered off as a function of distance. Here, the shear stress introduced by the pentamer is the main contribution to the energy. Note that even along the third coordination ring of hexamers surrounding the pentamer, the energy per hexamer remains significantly above that of the elastic energy imposed by the spontaneous curvature in the first figure. This indicates that the shear energy cost of introducing a pentamer at the center of a hexagonal shell exceeds any reduction of the curvature energy, at least for these elastic energy parameters.

Fig. 12. Same as Fig. 11, with $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$, but now adding the next set of pentamers as imposed by icosahedral symmetry for the full capsid. The elastic energy (times 100) of each capsomer is indicated. Isolated pentamers are flat pentagons.

In Fig. 12 we show what happens when we introduce pentamers at the appropriate locations in the next coordination ring to continue building a CK icosahedron.

In each case, the elastic energy of the hexamers that border the added pentamers is increased dramatically. The deformation energy of the pentamers is significant as well. If *hexamers* were introduced at these same locations, they would have an elastic energy of only about 0.0015 for

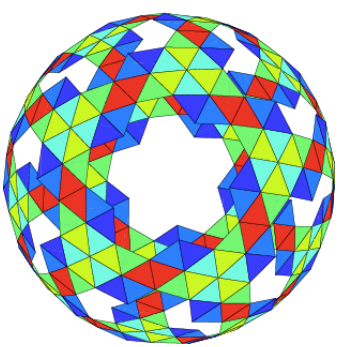


Fig. 13. *Whiffle-Ball* composed of 60 hexamers. The ball has icosahedral symmetry and all hexamers are located at symmetry-equivalent locations.

Fig. 12A and about 0.0020 for Fig. 12B, while the bonding energy gain of course would be the same as for pentamers. The introduction of pentamers is in this case not an energetically favorable assembly step, even in the presence of spontaneous curvature. The same conclusion was reached for different values of the stretching and bending energies as well as for a variety of different compact shell geometries. Also, we increased the magnitude of the preferred curvature beyond the optimal value of a $T = 13$ shell. This does eventually favor the introduction of pentamers, but it first leads to a mechanical instability. Finally, it might be thought that reducing the chemical potential of pentamers could make their introduction a favorable step. Indeed, it is easy to see that if $\mu_6 - \mu_5 > (0.0024 - 0.0015)\alpha_1$ then it would be favorable to introduce a pentamer (for the present parameter ratio). However, if one assumes $\mu_6 > \mu_5$ — as is necessary for this condition to hold — then the growing shell would be *entirely* composed of pentamers leading to a $T = 1$ dodecahedral shell. It appears that introducing pentamers at the locations required for the construction of an icosahedral shell simply is not a low energy assembly 'pathway' for the growth of a $T = 13$ shell even in the presence of a spontaneous curvature term in the energy. Within the deformable deltahedral model, it would not seem to be possible to grow a $T = 13$ icosahedral shell, at least not under conditions of local mechanical equilibrium.

This undesirable conclusion can be avoided *if one relaxes the condition of compact partial shells.* Consider somewhat bizarre 'Whiffle-Ball' (shown in Fig. 13): It is composed of 60 symmetry-equivalent hexamers, with each hexamer linked to three other hexamers, so it has 90 bonds in total. The structure has full icosahedral symmetry and can be constructed by removing from a $T = 13$ shell first all pentamers and then all hexamers that were in contact with a pentamer. The twelve resulting holes remove the loci of large elastic stress encountered in Section 2, so we expect this structure to have a rather low elastic energy. After energy minimization, the Whiffle-Ball was found to have an elastic energy of about 0.0317 for $\alpha_1 = 1.0$, $\alpha_2 = 10$, $\alpha_3 = 10$, about 35 times less than that of the $T = 13$ shell for the same parameters. The elastic energy of a compact patch of twelve hexamers is less than that, about 0.018 (again for the same parameter set, see Fig. 11A). However, the elastic energy of *five* compact incomplete shells of twelve hexamers, having the same total number of hexamers, is about 0.09 and that is higher than that of the Whiffle-Ball. It follows that by fusing five compact shells of twelve hexamers into a

single Whiffle-Ball one lowers the elastic energy. A different way to view this is as a stress-driven instability of the boundary edge of a compact shell. Stress-driven interfacial instabilities are familiar from the theory of epitaxial growth of solid films where they are known as 'Asaro–Tiller–Grinfeld instabilities'.[19,20] The difference with the present case is that the elastic stress is introduced by the mismatch between the substrate and the overlying elastic film, whereas here the elastic stress is introduced by the twelve pentamers.

It does *not* follow that compact shells are necessarily unstable. Five compact shells collectively share 120 bonds, 30 more than the Whiffle-Ball and that obviously favors growth of compact shells. This, effectively, is the line-tension mechanism discussed in the introduction. The bonding energy in general favors formation of compact partial shells while the elastic energy in general favors non-compact shells. More quantitatively, if one maintains a 1:10:10 ratio for α parameters, the sum of the elastic plus bond energies of the Whiffle-Ball equals $E(WB) \approx -90\varepsilon + 0.032\alpha_1$ while that of five compact hexagonal shells equals $E(5) \approx -120\varepsilon + 0.09\alpha_1$. The free energy difference between these two states is then $\Delta F \approx 30\varepsilon - 0.058\alpha_1 + k_B T \ln 5$ (the last term is the difference in entropy between five caps in solution versus a single Whiffle-Ball). The Whiffle-Ball has the lower free energy of the two states if ΔF is negative, that is if $\varepsilon < 0.0019\alpha_1 - k_B T(\ln 5)/30$. On the other hand, $E(WB)$ must be negative for the aggregation process to be energetically favorable in the first place, which means that we also must demand that $\varepsilon/\alpha_1 > 0.00035$.

Assume one increases the ratio ε/α_1 from zero at the onset of the assembly process. A Whiffle-Ball type structure with icosahedral symmetry should grow spontaneously for a range of ε/α_1 values between 0.00035 and $0.0019 - (k_B T/\alpha_1)(\ln 5)/30$. For this range of bonding energies, a fully assembled icosahedral shell indeed would not be stable. Recall that when the bonding energy is less than $\varepsilon \sim 0.003\alpha_1$, a completed shell is ruptured by internal elastic stresses. Continue to increase the ratio ε/α_1. If one adds additional hexamers to fill in the holes of the Whiffle-Ball then the elastic energy of this set of 'second generation' hexamers is significantly higher than that of the 'first generation', as shown in Fig. 14.

However, as ε/α_1 increases, the gain in bonding energy will overcome the increased elastic energy. Finally, in order for the free energy $E(T = 13) \approx -390\varepsilon + 1.1\alpha_1 + 12\mu_5 + 120\mu_6$ of a completed icosahedral shell to be less than the free energy of a Whiffle-Ball, which is about

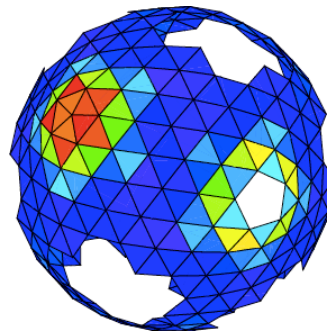


Fig. 14. Stress distribution of a Whiffle-Ball with partially completed holes. The stress distribution of the completed pentamer is nearly the same as that of the completed shell.

$E(WB) \approx -90\varepsilon + 0.032\alpha_1 + 60\mu_6$, the bonding energy obeys the inequality $\varepsilon/\alpha_1 > 0.0035 + (\mu_5 + 5\mu_6)/25\alpha_1$.

This suggests a possible scenario for error-free assembly of $T = 13$ icosahedral capsids. Assume that hexamers have a lower chemical potential than pentamers and that assembly is initiated by the aggregation of hexamers on a spherical surface, the 'scaffold'. Next, allow the ratio ε/α_1 of cohesion and elastic energies to slowly increase from zero. For sufficiently low values of ε/α_1 the non-compact, icosahedral Whiffle-Ball type structures form in preference over compact shells. Elastic energy provides no significant barrier and so the Whiffle-Ball structure establishes the icosahedral symmetry. As ε/α_1 continues to grow, the twelve holes in the structure start to fill in with additional rings of five hexamers. No more hexamers can be added after this. As we increase ε/α_1 even further, it becomes favorable to introduce the energetically costly pentamers, which close off the holes to form the completed capsid. Effectively, the Whiffle-Ball acts as a 'growth intermediate' along the assembly pathway of a CK shell.

The assembly of $T = 13$ shells often takes place in the presence of such spherical scaffolds, as for example in the Herpes virus.[21] Interestingly, Whiffle-Ball type viral shells have been engineered as mutants of the HK97 shell,[22] which means that even without a support scaffold a Whiffle-Ball shell can be mechanically stable. We thus propose the Whiffle-Ball structures as key growth intermediates on the pathway of the formation of the growth of icosahedral shells, with the elastic energy directing the assembly process.

Acknowledgement

This work was supported by the NSF under DMR Grant 04-04507.

References

1. D.L.D. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Spring Harbor Symp. Quant. Biol.* **27**:1–22 (1962).
2. R.C. Liddington, Y. Yan, J. Moulai, R. Sahli, T.L. Benjamin and S.C. Harrison. Structure of simian virus 40 at 3.8-A resolution. *Nature* **354**:278–284 (1991).
3. J.M. Grimes, J.N. Burroughs, P. Gouet, J.M. Diprose, R. Malby, S. Zientara, P.P. Mertens and D.I. Stuart. The atomic structure of the bluetongue virus core. *Nature* **395**:470–478 (1998).
4. J.A. Speir, S. Munshi, G.S. Wang, T.S. Baker and J.E. Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **15;3(1)**:63–78 (1995).
5. J. Lidmar, L. Mirny and D.R. Nelson. Virus shapes and buckling transitions in spherical shells. *Phys. Rev. E* **68**:051910 (2003).
6. P.L. Freddolino, A.S. Arkhipov, S.B. Larson, A. McPherson and K. Schulten. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* **14(3)**:437–449 (2006).
7. D. Endres, M. Miyahara, P. Moisant and A. Zlotnick. A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. *Protein Sci.* **14(6)**:1518–1525 (2005).
8. T. Zhang and R. Schwartz. Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys. J.* **90(1)**:57–64 (2006).
9. M. Hemberg, S.N. Yaliraki and M. Barahona. Stochastic kinetics of viral capsid assembly based on detailed protein structures. *Biophys. J.* **90(9)**:3029–3042 (2006).
10. M.F. Hagan and D. Chandler. Dynamic pathways for viral capsid assembly. *Biophys. J.* **91(1)**:42–54 (2006).
11. S.D. Hicks and C.L. Henley. Irreversible growth model for virus capsid assembly. *Phys. Rev. E* **74(3 Pt 1)**:031912 (2006).
12. R. Zandi, P. van der Schoot, D. Reguera, W. Kegel and H. Reiss. Classical nucleation theory of virus capsids. *Biophys. J.* **90(6)**:1939–1948 (2006).
13. T. Guerin and R. Bruinsma. Theory of conformational transitions of viral shells. *Phys. Rev. E* **76**:061911 (2007).
14. T.T. Nguyen, R.F. Bruinsma and W.M. Gelbart. Continuum theory of retroviral capsids. *Phys. Rev. Lett.* **96**:078102 (2006).
15. I.L. Ivanovska, P.J. de Pablo, B. Ibarra, G. Sgalari, F.C. MacKintosh, J.L. Carrascosa, C.F. Schmidt and G.J. Wuite. Bacteriophage capsids: tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci. USA* **101**:7600 (2004).

16. J.P. Michel, I.L. Ivanovska, M.M. Gibbons, W.S. Klug, C.M. Knobler, G.J.L. Wuite and C.F. Schmidt. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc. Nat. Acad. Sci. USA* **103**:6184–6189 (2006).
17. A.E. Lobkovsky. Boundary layer analysis of the ridge singularity in a thin plate. *Phys. Rev. E* **53**:3750 (1996).
18. R.J. Mashl and R.F. Bruinsma. Spontaneous-curvature theory of clathrin-coated membranes. *Biophys. J.* **74(6)**:2862–2875 (1998).
19. R.J. Asaro and W.A. Tiller. Interface morphology development during stress corrosion cracking: Part I. Via surface diffusion. *Metall. Trans.* **3**:1789 (1972).
20. M.A. Grinfeld. The stress driven instabilities in crystals: mathematical models and physical manifestations. *J. Nonlinear Sci.* **3**:35 (1993).
21. J.B. Heymann, N. Cheng, W.W. Newcomb, B.L. Trus, J.C. Brown and A.C. Steven. Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nat. Struct. Biol.* **10(5)**:334–341 (2003).
22. Y. Li, J.F. Conway, N. Cheng, A.C. Steven, R.W. Hendrix and R.L. Duda. Control of virus assembly: HK97 'Whiffleball' mutant capsids without pentons. *J. Mol. Biol.* **348(1)**:167–182 (2005).

Chapter 8

What Determines the Size of an RNA Virus?

Charles M. Knobler*,† and William M. Gelbart*,‡

1. Introduction

The vast majority of viruses have icosahedral capsids that can be classified according to their Caspar–Klug (CK) triangulation number T (with $T = h^2 + k^2 + hk$, h and k positive integers). Their capsids are composed of $60T$ proteins and organized into 12 pentamers and $10(T-1)$ hexamers. Thus, the smallest of the viral capsids, those with $T = 1$, are composed of 60 proteins, organized as 12 pentamers. The *largest* icosahedral virus that has been structurally characterized has a T number of 219, with a capsid composed of 13,140 proteins, organized into 12 pentamers and 2180 hexamers. It is remarkable that the same principles of organization that account for the icosahedral ordering of fewer than 100 protein subunits are still operative in the case of more than 10,000 subunits. Clearly the protein–protein interactions involved are at once specific and robust enough to give rise in each instance — and for each of the many intervening T values — to icosahedrally symmetric shells consisting of 12 pentamers and unique 'magic' numbers of hexamers, up to at least 2180!

What about the corresponding capsid *sizes*? A typical $T = 1$ capsid, say that of the Satellite panicum mosaic virus (SPMV), is 17 nm in diameter.[1] In contrast, the *Phaeocystis pouchetii* virus (PpV01) that infects marine algae has a T number of 219 and a capsid whose diameter[2] is approximately 200 nm. Thus, the *volumes* of their capsids differ by more than three orders of magnitude. This difference in size is necessitated by the

*Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA 90095-1569. E-mails: †knobler@chem.ucla.edu; ‡gelbart@chem.ucla.edu

very large differences in the genetic information the viruses contain. In the case of SPMV, it is in the form of a single molecule of single-stranded RNA (ssRNA), 1059 nt in length, not sufficiently long to convey all of the information needed for the replication of the virus in its plant host; indeed, this RNA encodes for only a single gene product. The rest of the genetic information is supplied by the panicum virus, which must *co*infect the plant (hence the classification of SPMV as a *satellite* virus). The genome of PpV01, on the other hand, consists of a 485 kbp double-stranded DNA (dsDNA), containing almost 1000 times the number of nucleotides (and hundreds of times as many genes) as the SPMV RNA.

While it is straightforward to compare the *contour lengths* of the genomes — 170 μm for PpV01 and 0.4 μm for SPMV — we cannot easily compare their effective *volumes*. For example, it is possible to characterize dsDNA simply as a semiflexible polymer with a 50 nm persistence length. Accordingly, any linear measure of its 3D size — e.g., its radius of gyration or root-mean-square end-to-end distance — is expected to scale with the square-root of its contour length (or with a slightly higher power when excluded volume effects are taken into account), and its volume will scale with the cube of this linear measure. But the volume of ssRNA *cannot* be described in this way because single-stranded nucleic acids are known to develop a large degree of secondary structure arising from base-pairing between short stretches of widely-separated nucleotides distributed throughout their length. The resulting duplexes and the associated internal 'loops' and 'junctions' between them give rise in turn to a complicated secondary and tertiary structure, the details of which determine the effective volume of the molecule in ways that are known only for a few, very short (< 400 nt) RNA molecules.

We confine ourselves in the present discussion to viruses whose genome is ssRNA and whose capsids are icosahedral. A crude picture of the relation between genome length and capsid size for these viruses is shown in Fig. 1, where the outside diameters of viral capsids are plotted against the lengths of their ssRNA genomes. (While there are ssRNA viruses with longer genomes, e.g., the coronaviruses with 33,000 nt, their capsids are not icosahedral.) In general, as one would expect, the capsid diameters increase with genome length. Similarly the T numbers increase with genome length, from 1 to 4; correspondingly, the number of capsid proteins involved increases from 60 to 240. With one exception, the smallest capsids, those indicated by the filled green circles, with diameters

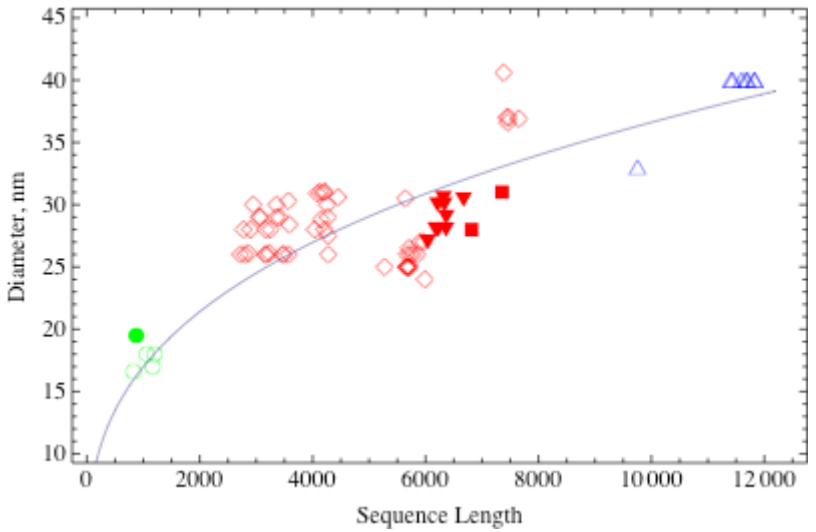


Fig. 1. The relation between capsid diameter and number of nucleotides in icosahedral ssRNA viruses. The green symbols represent viruses with $T = 1$ triangulation numbers, the red symbols $T = 3$ and those in blue $T = 4$. The line is a fit to a cubic equation; its significance is discussed in Sec. 6.

of about 17 nm, are satellite viruses and all are $T = 1$. The largest viral capsids shown, those of the animal alphaviruses, have a diameter of 40 nm with a $T = 4$ triangulation number and genomes almost 12,000 nt long. All the other examples involve $T = 3$ capsids, composed of 180 proteins, and yet display a significant range of diameters. Some of the variations within this group can be ascribed to special properties.

For example, the group of viruses represented by the red squares (from the *Astroviridae* family) stand out amongst the $T = 3$ examples because they involve a relatively high density of nucleotides, corresponding to capsid diameters as small as 30 nm with packaged genomes longer than 7000 nt. On the other hand, the bromoviruses have the same capsid diameter but contain only about 3000 nt of packaged RNA length. As we will see in Sec. 4 (cf., Fig. 8) this special compactness of the *Astroviridae* is related to the compactness of their secondary structures. In other cases of high nucleotide density, say the *Tymoviridae* (indicated by the solid inverted triangles), cationic polyamines are known to be incorporated into the capsids,[3] accounting for a 6500 nt genome being contained in a

30 nm capsid. The amines neutralize some of the charge on the phosphate backbone of the RNA, reduce the electrostatic repulsions, and effectively reduce the genome size. A particularly striking dependence on genome length is the solid green circle for $T = 1$ representing a Brome mosaic virus capsid that has *in vitro* assembled around its capsid protein messenger RNA, which is 876 nt long, rather than the 3000 nt viral RNA which is found in the normal 28 nm diameter capsids.[4]

No quantitative information can be derived from the relationships shown in Fig. 1. The capsids are icosahedral, not round, so in several cases there is not a unique diameter. Here we have used the geometric mean of the largest and smallest diameter. Moreover, the internal volume of the capsids is the more relevant quantity[5] but this is difficult to define because parts of the capsid proteins occupy space in the capsid interior. As we will show in Sec. 5, however, there is a rationale for fitting the data with a cubic curve such as the line shown.

In this chapter we discuss — for ssRNA icosahedral viruses — some of the fundamental physical factors determining the relationship between the sizes of viral capsids and the sizes of their genomes. The first such factor (see Sec. 2) is the total charge of the genome, or number of nucleotides (since there is one phosphate charge per nucleotide), and how it correlates with the total (and opposite-sign) charge associated with the interior of the capsid. A second factor (see Sec. 3) is the local radius of curvature preferred by the capsid proteins as they aggregate to form a one-molecule-thick 2D crystal that is closed on itself. This driving force for a preferred capsid size can be conveniently addressed in studies of shell formation by capsid protein in the absence of RNA. A third factor is the preferred 3D size of the genome itself, i.e., the intrinsic 3D volume of an ssRNA molecule of a given nucleotide length and sequence. To explore this issue free of uncertainties about not-yet-determined secondary and tertiary structures of viral RNAs, we first discuss this factor (see Sec. 4) for the case of a simple linear polyanion with the same flexibility and linear charge density as ssRNA; here the volume of the molecule is simply determined by its contour and persistence lengths. Then, in Sec. 5, we confront the problem of ssRNA molecules, where the 3D size of the molecule can be shown to depend on sequence; two molecules with the same nucleotide length (and hence same charge) will have different volumes because of different secondary and tertiary structures. Finally, in Sec. 6 we review various theoretical treatments

of the packaging of ssRNA genomes in their self-assembling capsids and discuss the prospects for and importance of systematic measurements of packaging efficiencies for these systems.

Before beginning the discussion of physical determinants of viral capsid size, we mention some biological factors. There are many reasons why a virus particle should be as small as possible, for a given size of genome. Because a minimum of *thousands* of viruses are produced in a typical generation and the success/virulence of the infection depends on this number being as large as possible, each of the particles should be as small as possible. Also, the smaller they are the easier it is for each of them to get out of the host cell and infect neighboring cells. This is especially true in the case of plant viruses like Cowpea chlorotic mottle virus (CCMV), where the self-assembled, fully mature virions are constrained to move to neighboring cells through narrow channels — the plasmodesmata — in the rigid plant cell walls. The inner diameters of these channels are comparable to the outer diameters of the viral capsids, making passage through them difficult. In CCMV, this translocation is facilitated by a virally-encoded 'movement' protein. Clearly then, the rigid capsids should be no bigger than absolutely necessary. Accordingly, for a given nucleotide length, the viral RNA should be as small as possible so that its packaging by capsid protein is as efficient as possible.

2. Effect of Charge

The capsid proteins of many viruses have N-termini that are rich in basic residues and these positively charged tails project into the capsid interior where they can interact with the negatively charged RNA. This electrostatic interaction is known to play a role in virus assembly. For example, Cowpea chlorotic mottle virus (CCMV) assembles into infectious virus when its pure capsid protein and RNA are mixed *in vitro* under appropriate buffer and ionic strength conditions.[6] (We use the term *in vitro* in the sense that the assembly involves purified capsid protein and RNA.) *In vitro* assembly with RNA does not occur with mutant protein that lacks the positive charge on the N-terminus, but *in vitro* self-assembly of empty capsids from wild-type protein in the absence of RNA *does* occur.[7]

Belyi and Muthukumar[8] examined the relation between the total charge on the N-termini and genome length by plotting genome length

(effectively the number of negative charges) against the total positive charge (the product of the charge per tail and the number of proteins) for 24 viruses ranging in genome length from 1000 to 12,000 nucleotides and including $T = 1$, 3 and 4 triangulation numbers. Their plot, shown in Fig. 2, shows a remarkable linear correlation. The slope, 1.6, corresponds to slightly less than two negative charges for each positive charge. This experimental correlation shows clearly that charge plays a role in determining the size of many viruses. But charge cannot be the only determinant. For example, there are ssRNA viruses such as the Norwalk virus, which do not have an N-terminal basic region or contain polyamines.[9] Also, for a given capsid protein (e.g., that of wild-type CCMV) and hence cationic charge, we shall see below that a broad range of anionic charge can be efficiently packaged. The nature of the capsid protein also must play a role. Consider, for example, Turnip yellow mosaic virus (TYMV) and Tobacco mosaic virus (TMV) both of which have RNA genomes about 6000 nt in length. TMV viruses are rod-like but TYMV viruses are icosahedral. Moreover, if TMV RNA is mixed with CCMV capsid protein, icosahedral virions form, but the capsids are larger than those of wild-type CCMV. On the other hand, CCMV RNA molecules mixed with TMV capsid proteins

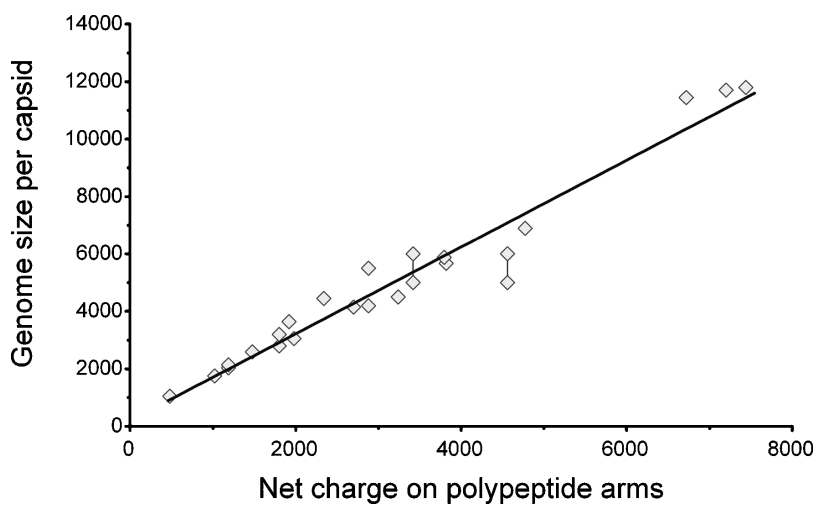


Fig. 2. Relation between genome size and the net charge on the capsid protein tails. From Belyi and Muthukumar.[8] Copyright *Proceedings of the National Academy of Sciences USA*.

form hollow tubes as in TMV virions, but their lengths are shorter. Clearly each of these two different capsid proteins prefers to aggregate into shells of different curvature.

3. Effect of Preferred Curvature

If the top and bottom surfaces of a membrane differ, it may be curved at equilibrium, rather than flat. Departures from this spontaneous curvature can occur only at a cost in free energy. Applying a coarse-grained approach, we can think of the protein shell that constitutes a capsid as a homogeneous membrane, but one that is asymmetric. Such a description has been shown to provide a remarkably good first-order account of the mechanical properties of viral capsids[10] and also provides an understanding of the polymorphism exhibited by CCMV capsid protein in the absence of RNA.[11] Indeed, a most direct way to learn about the local radius of curvature preferred by capsid proteins is to study the structures they form when present alone (i.e., without RNA) under different solution conditions.

In an impressive and systematic series of experimental studies completed over 30 years ago, Bancroft and coworkers investigated the range of polymorphs that can be obtained by self-assembly of capsid protein alone[6] (see also Adolph and Butler[12]). The domains of stability of the various polymorphs are shown on the ionic strength-pH phase diagram in Fig. 3, which comes from a more recent study.[13]

The existence of spontaneous curvature and its role in the formation of capsids is evident from the appearance of multishell structures that are found at low ionic strength. At low pH, single-walled 28 nm diameter capsids identical to those in wild-type CCMV are observed. A transition to shells with two concentric protein layers begins near the isoelectric point of the capsids, $pH = 3.2$, and the number of layers increases with increasing pH. This progression can be understood in terms of a balance between the electrostatic interaction between successive layers of capsid protein and the elastic energy related to the departure from the preferred shell curvature (i.e., a diameter of 28 nm).[13,14] The N-terminus ('inside surface') of the CCMV protein has basic residues that remain fully (positively) charged until high pH, but the anionic residues on the capsid exterior have lower acid dissociation constants and their degree of ionization is much more sensitive to the pH under assembly conditions.

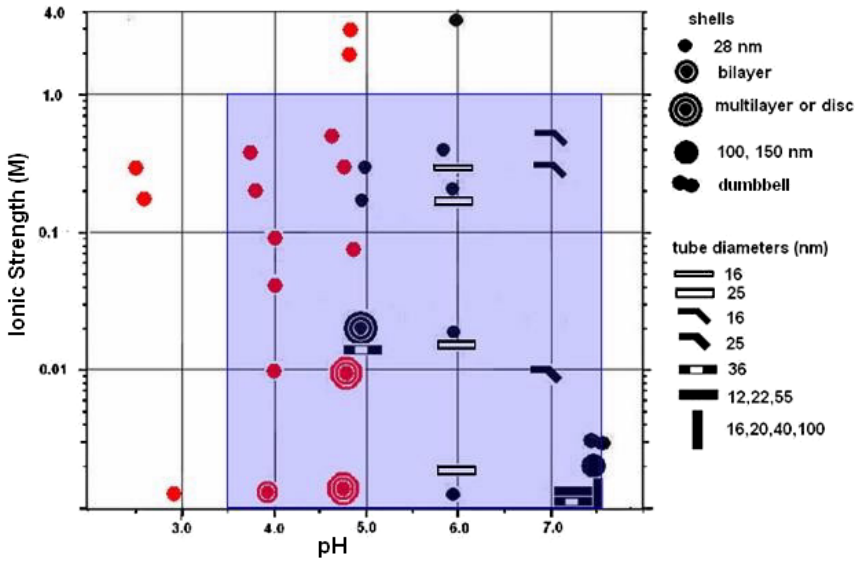


Fig. 3. Ionic strength-*pH* phase diagram for self-assembled CCMV capsid protein polymorphs. The blue rectangle shows the range of conditions examined by Bancroft and Adolph and Butler. From Lavelle *et al.*[13] Copyright 2009 American Chemical Society.

Electrophoretic mobility measurements of empty and full capsids show that the external charge on the assembled protein becomes increasingly negative from *pH* 3.2 to about *pH* 5, where it levels off.[15] As a result, the difference in charge across the membrane (i.e., between the inside and outside surfaces) increases with increasing *pH*. Theory[16] shows that asymmetrically charged membranes will stack with a separation that depends on the charge differential, with the equilibrium separation controlled by the release of counterions as the membranes interact. In the case of concentric layers around a core capsid, however, there is an additional effect, the increasingly lower curvature (larger diameter) with each successive shell. There is a free energy cost associated with the change in shell curvature from the equilibrium spontaneous curvature of the 28 nm diameter capsid.

The role of spontaneous curvature is also directly seen in the formation of 'rosette' structures, Fig. 4. Here the bending energy cost is so high that no more complete shells can form but the electrostatic interaction is

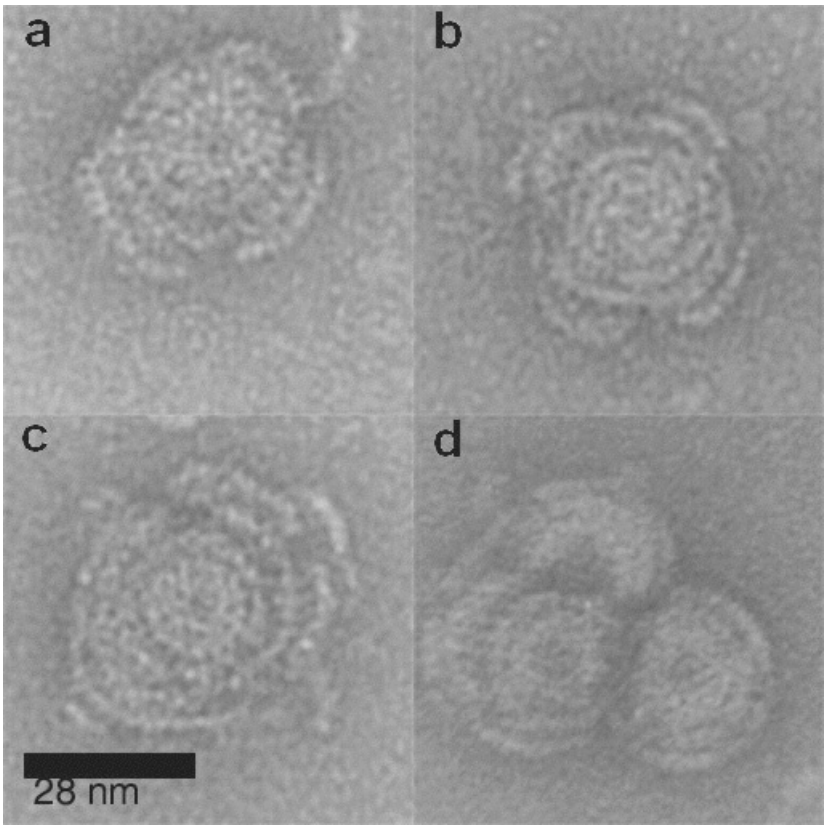


Fig. 4. Rosette structures formed by self-assembly of CCMV capsid protein. The outer incomplete shells have higher curvatures than those of the largest complete shell and similar to that of the innermost shell (adapted from Lavelle *et al.*[13]). Copyright 2009 American Chemical Society.

sufficient to stabilize partial shells with higher curvature, hence a smaller expenditure of bending energy.

4. Effect of Polyanion Size

To examine the interplay become genome size, charge and spontaneous curvature, Hu *et al.*[17] studied the *in vitro* packaging of poly(styrene sulfonate) (PSS) by CCMV capsid protein. CCMV protein has been shown[6] to be capable of assembling into capsids around a variety of anionically

charged polymers. As already mentioned in the Introduction, the general relationship between the length of viral RNA in terms of the number of nucleotides and its physical size, e.g., its radius of gyration in solution, is problematic because ssRNA has a complex structure that depends on the details of its self-complementarity, which is determined by its primary structure. In contrast, PSS is a flexible polymer with a radius of gyration that depends in a regular way on its molecular weight[18]: $R_g \propto M^{1/2}$. The charge density of the fully sulfonated polymer, 1 e^-/3 Å, is close to that of ssRNA.

Hu *et al.* employed electron microscopy to examine the products of self-assembly of CCMV protein with PSS samples that ranged in molecular mass from 400 kDa to 3.4 MDa. The size distributions of the products are shown in Fig. 5. The virus-like particles (VLPs) formed have unimodal size-distribution histograms based in each case on at least 100 particles that can be represented by Gaussians. When the measurements for all of the particles are combined it is evident that they form a bimodal distribution (Fig. 5(f)) for which a fit with two Gaussians gives maxima at 21.5 and 27 nm. For comparison, a distribution of sizes obtained for assembly of CCMV capsid protein with RNA under the same conditions is also shown.

In the absence of structural information about the two types of capsids, we can deduce the T numbers from their relative sizes. The number of capsid proteins is $60T$, so the capsid surface area is proportional to $60T$. Thus, if we assume that the area for a given protein is independent of T, the ratio of the diameters of two capsids composed of the same protein will be given by

$$\frac{D_1}{D_2} = \sqrt{\frac{T_1}{T_2}},$$

a relation that has been shown to be reliable by computing the diameter ratio for capsids whose T numbers are known from structure determinations.[17] The ratio of the diameters for the two populations of capsids corresponds closely to $\sqrt{3/2}$, which identifies them as $T = 2$ and $T = 3$.

For molecular weights of about 1 MDa and less, PSS assembles into $T = 2$ capsids, as shown in Fig. 6. $T = 2$ shells (which might more accurately be characterized as *pseudo*- $T = 2$) are non-Caspar–Klug structures; they consist of 120 proteins arranged into 12 pentamers of dimers. In contrast, polymers with MW above 2 MDa assemble into $T = 3$ capsids. The

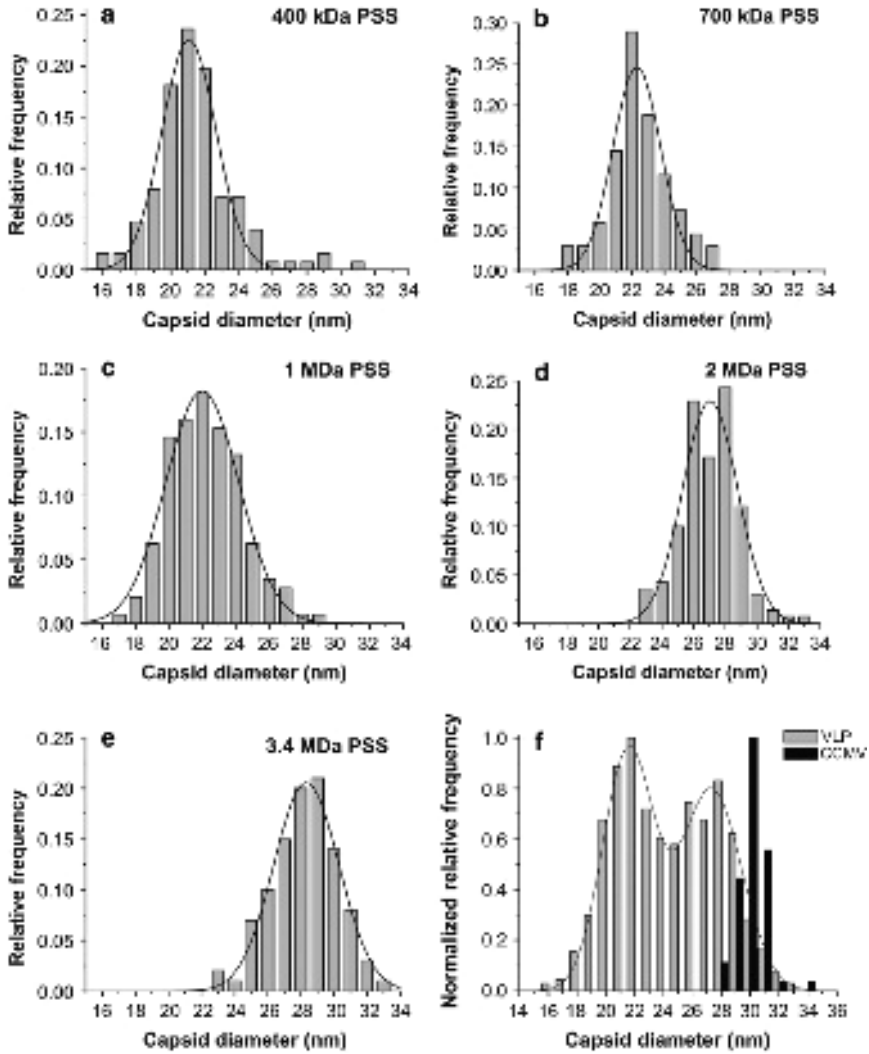


Fig. 5. Histograms showing diameters of VLPs formed by self assembly of CCMC capsid protein around poly(styrene sulfonate) polymers of different molecular weights (from Hu *et al.*[17]). Copyright Biophysical Society 2008.

hydrodynamic radius of the largest polymer examined, that with molecular weight 3.4 MDa, is 42.8 nm as measured under assembly conditions by dynamic light scattering; the radius of gyration is nearly twice as large.[18] Thus, in the assembly process, interaction with the CCMV capsid protein

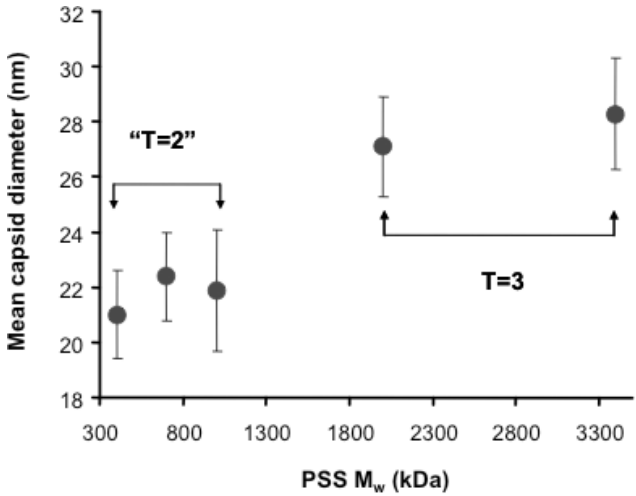


Fig. 6. Diameters of VLPs as a function of the molecular weight of the poly(styrene sulfonate) polymer (from Hu *et al.*[17]). Copyright Biophysical Society 2008.

is able to compact PSS by a factor of around 5. In experiments in which CCMV RNA is titrated with its capsid protein, an initial compaction of the RNA has been observed at a ratio of 10 CPs to one RNA, but the degree of compaction is unknown.[19] (In Sec. 5, we will see that the intrinsic size of the RNA — free in solution under *in vitro* assembly buffer conditions — is only slightly larger than the inner volume of the CCMV capsid.) Each of the CCMV proteins has an N-terminus with 10 positive charges, so the total charge on the interior of a $T = 2$ capsid is 1200, while it is 1800 for $T = 3$. The smallest PSS polymer studied, which has a molecular weight of 408 kDa, has nearly 2000 negative charges, and the ratio of charge on the polymer to that in the interior of the capsid is about 1.6, in agreement with the correlation found by Belyi and Muthukumar.[8] However, the ratio exceeds this value for all of the other polymers studied and in the case of the 3.4 MDa polymer it is greater than 9.

5. Dependence of RNA Size on Sequence

5.1. *Theory*

As mentioned in the Introduction, the 3D size of linear polymers is completely determined by their contour length and persistence length. More

explicitly, the average radius of gyration is given by $R_g \cong \xi^{1-\nu} L^{\nu}$, where L and ξ are the contour and persistence length, respectively, and ν is approximately $1/2$. In particular, ν is equal to: $1/3$ for poor solvents where the polymer excludes solvent and effectively collapses on itself; $3/5$ in good solvents where excluded volume effects are important; and $1/2$ for ideal solvents where these effects are essentially cancelled. These 'details' aside, the important point for our present purposes is that any two linear polymers with the same contour and persistence lengths will have the same 3D size under the same solution conditions. Consider the case of dsDNA. It is a heterogeneous linear polymer whose persistence length varies locally with its sequence of base pairs. But because the persistence length varies by at most a factor of two and is always small (of the order of 50 nm) compared with the contour lengths (typically 10,000–50,000 nm) of dsDNA viral genomes, the dependence of radius of gyration on overall sequence is negligible. Accordingly, any two viral genomes consisting of tens of thousands of base pairs will have the same 3D size if they have the same number of base pairs.

The situation with ssRNA genomes is completely different, because the configurational statistics of ssRNA are qualitatively different from those of a linear homopolymer. The key point, as already mentioned several times in earlier sections, is that ssRNA develops a complicated secondary (and hence tertiary) structure determined by the self-complementarity of its particular primary sequence of nucleotides. And, as we discuss below, the spatial extent and 'branchedness' of each of these secondary structures give rise to a distinctly different overall 3D size for the molecule, even for a fixed number of nucleotides. The issue we explore here is the relationship between nucleotide sequence and overall molecular size, for a given nucleotide *length*. Nucleotide length, of course, is determined by the number of genes, and so it is natural to ask whether there is evolutionary pressure on a genome to be as compact as possible (e.g., have as small as possible a radius of gyration), for a given amount of genetic information.

Figure 7(a) shows the predicted secondary structures[20] for two different ssRNA sequences of equal nucleotide length (2117 nt). It is important to note that basically no experimental information whatsoever is available for the secondary (and tertiary) structures of sequences this long. The structures shown here are predicted by a particular RNA folding algorithm — RNAsubopt, in the Vienna RNA Package, Version 1.7[21] — but for purposes of the present discussion the results are essentially the same

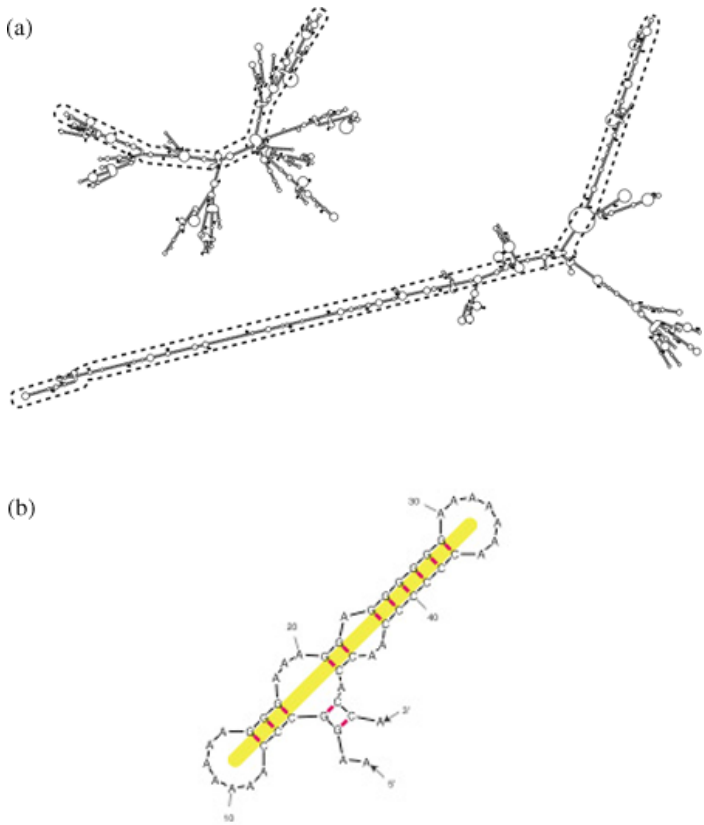


Fig. 7. (a) MFE secondary structures, drawn to the same scale, calculated for two different 2117 nt-long RNA sequences chosen at random from the yeast genome; the corresponding MLDs (see text) are 148 (top) and 368 (bottom), respectively, and are depicted by dotted-line overlays. (b) The definition of MLD for the secondary structure of an arbitrary (here 50 nt) RNA sequence. The MLD is the number of the duplex links along the path shown in yellow leading from the upper-right-hand hairpin to the hairpin at the lower left.

when computed using any of several alternatives. The upper figure and lower figures show the secondary structures predicted for two different sequences pulled at random from the yeast genome. It is immediately striking that the secondary structure associated with the first, upper, sequence is much more compact than that of the lower sequence.

A convenient measure of the extendedness of a secondary structure is given by its 'maximum ladder distance' (MLD) — illustrated by

Fig. 7(b) — which is essentially the longest path across the structure. More precisely, for an arbitrary pair of nucleotides i and j, consider all possible paths between them, measuring the length of each as follows. If nucleotide i is in a single-stranded loop, go to one of the duplexes associated with it and proceed towards j, counting the number of 'rungs' (base pairs) crossed (and if i is in a duplex, start counting rungs immediately). Upon encountering an internal loop, jump to one of the duplexes coming off it and continue counting rungs, and so on, until nucleotide j is reached. The minimum total number of rungs traversed in this way is the ladder distance between i and j, a quantity first introduced by Bundschuh and Hwa.[22] Note that a ladder distance, LD_{ij}, can be calculated in this way for each pair ij in a secondary structure. The MLD associated with an arbitrary, overall, secondary structure is defined by us as the maximum of LD_{ij} over all pairs ij, i.e., it is the ladder distance associated with the longest direct path across the structure. For the secondary structure of the 50 nt-long sequence shown in Fig. 7(b), for example, the MLD is 11. It is the ladder distance associated with, i.e. the sum of duplex lengths separating — the pairs of nucleotides in the two hairpin loops (see upper right and lower right) at the 'ends' of the structure.

The molecules depicted in Figs. 7(a) are 2117 nt long, with very different MLDs. A similar analysis of the secondary structures associated with hundreds of other 2117 nt-long sequences taken at random from a yeast genome finds a broad distribution of MLDs ranging from just below to just above the two shown in the figure, i.e., ranging from about 140 to 370; the average MLD is 240, *just what is found for a large number of mathematically random sequences of the same length.*[20] This suggests that — while the yeast sequences, both coding and noncoding, have evolved to have special features of various kinds — the extendedness/compactness of their RNA secondary structures is no different from that of random sequences.

In contrast, we find that viral sequences have significantly more compact secondary structures, i.e., smaller MLDs. Consider, for example, the dicistronic RNA3 molecule of BMV, which contains the genes coding for the capsid and movement proteins and which is 2117 nt long. The MLD for this viral sequence is 183, 24% smaller than the value typical for equal-length random (or yeast) sequences. We have shown[20] that this holds in general, i.e., the MLD associated with the secondary structure of a viral sequence is significantly smaller than the MLD of random (and nonviral) sequences of the same length. This fact is demonstrated in Fig. 8, which

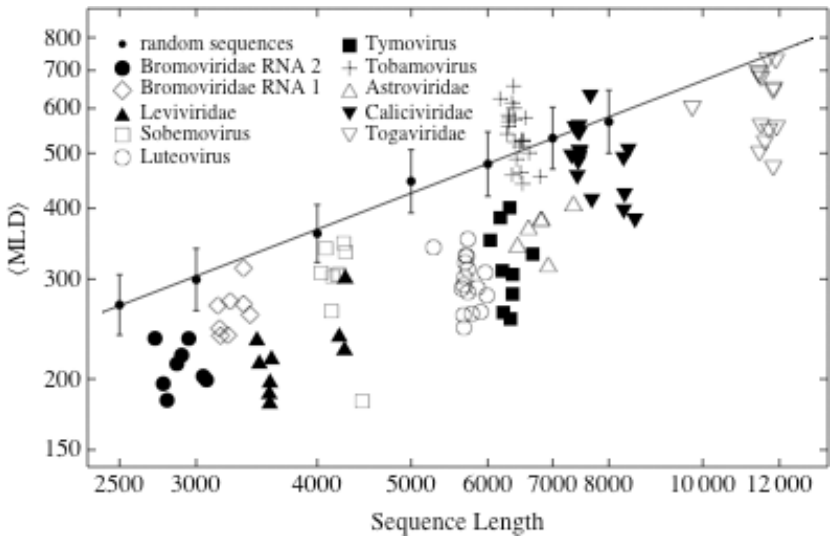


Fig. 8. Average MLDs for viral and random sequences over the range of 2500–8000 nt. Adapted from Yoffe *et al.*[20] Copyright *Proceedings of the National Academy of Sciences USA*.

compares the MLDs of viral *vs.* random sequences for a large number of RNAs with lengths ranging from 2500 to 8000 nt. Actually what is plotted here is the *ensemble-averaged* MLD associated with each of the sequences involved. This is because the minimum free energy (MFE) secondary structure of a sequence as long as thousands of nucleotides is one of a huge number of secondary structures that are thermally accessible to the molecule. Accordingly, we define the average maximum ladder distance ($\langle \mathrm{MLD} \rangle$) of an arbitrary sequence by the Boltzmann-weighted ensemble average of the MLDs associated with the secondary structures of that sequence.

Nine different families of viruses — the *Bromoviridae*, *Leviviridae*, *Sobemoviridae*, *Luteoviridae*, *Tymoviridae*, *Astroviridae*, *Tobamoviridae*, *Caliciviridae* and *Togaviridae* — are included in the data shown in Fig. 8, all corresponding to virions consisting of a ssRNA molecule enclosed within a rigid protein shell. The shells of all but the togaviruses (see ∇ symbols) and tobamoviruses (see + symbols) are $T = 3$ capsids, with outer diameters of 26–28 nm, made up of 180 copies of a single gene product, the capsid protein; the lengths of packaged RNA molecules range from just below 3000 nt to just above 8000 nt. The capsids of the *Togaviridae* are

$T = 4$ structures with 40 nm diameters. The virions of the *Tobamoviridae*, on the other hand, involve hollow, open, cylindrical shells of fixed diameter (about 15 nm), but variable length (averaging 300 nm), depending on the length of the ssRNA enclosed. Based on our comments in the Introduction, we expect that only the spherical viruses will be under evolutionary pressure for their genomes to be compact. Indeed, this is consistent with the results shown in Fig. 8 and discussed below. More explicitly, the ⟨MLD⟩ values of the spherical virus RNAs are distinctly and systematically smaller than those of random sequence RNAs of the same length. The tobamovirus RNAs, however, form a notable exception.

The straight line in Fig. 8, with respect to which viral sequence ⟨MLD⟩s are compared, is a fit through the data calculated for random sequences with virus-like nucleotide composition. For each of several nucleotide lengths — 2500, 3000, 4000, 5000, 6000, 7000 and 8000 — Yoffe *et al.*[20] have generated several hundred randomly permuted sequences, calculated an ⟨MLD⟩ for each one of the sequences and then averaged over all of the ⟨MLD⟩s for those sequences, to obtain an average ⟨MLD⟩ value for each of the nucleotide lengths ranging from 2500 to 8000. These seven values, with their standard deviations depicted by vertical lines with horizontal end-bars, are plotted in Fig. 8. The straight line fit through them turns out to have a slope consistent with average ⟨MLD⟩ scaling as $N^{0.67}$.

In comparing extendedness/compactness of random-sequence versus viral-sequence RNA secondary structures, it is interesting to expand the discussion to include the case of non-random but non-viral sequences. This has been done[20] by analyzing a large number of ssRNA sequences corresponding to transcripts of 3000 bp-long sections chosen at random from chromosomes XI and XII of the *S. cerevisiae* yeast genome. These sequences represent evolved biological RNAs that are not expected to have been under pressure to be compact or otherwise have a special overall size or shape. Indeed, it is found (as in the case of the 2117 nt sequence mentioned earlier) that the average ⟨MLD⟩s for these sequences are indistinguishable from those of randomly permuted sequences of the same length (3000 nt), i.e., 300 ± 46, as compared with values $2/3$ that size for viral sequences (see Fig. 8).

Recall that the motivation for examining the compactness/extendness of RNA secondary structures was the idea that the RNA genomes of spherical viruses should be as small as possible so that their associated capsids could be as small as possible. Accordingly, we need to correlate

compactness as measured by ⟨MLD⟩ with appropriate measures of 3D size, like the radius of gyration. One way to do this is to map secondary structures onto the 3D configurational statistics of linear polymers. More explicitly, consider replacing the secondary structures shown in Fig. 7(a) by linear chains whose effective *contour* lengths are given by their ⟨MLD⟩ values (see dotted overlay lines). The effective *persistence* length of these linear chains is determined by the average length of a duplex because the single-stranded loops are relatively flexible (dsRNA has a persistence length almost 100 times larger than that of single-stranded portions). But because the *average* duplex length is virtually constant (about 5 bp) and independent of overall sequence, we can approximate the radius of gyration of a large RNA molecule by[20]

$$R_g \cong \xi_{eff}^{1-\nu} L_{eff}^{\nu} \cong \xi_{eff}^{1-\nu} \langle MLD \rangle^{\nu} \propto \langle MLD \rangle^{\nu}.$$

This result allows us to correlate secondary structure compactness/extendedness with 3D size, and to account qualitatively for the preliminary measurements discussed below for radii of gyration and hydrodynamic radii of viral *vs.* nonviral sequences of long RNA molecules. By combining this result with the relation $\langle MLD \rangle \propto N^{0.67}$ that follows from the scaling analysis of Fig. 8, we have

$$R_g \propto \langle MLD \rangle^{\nu} \propto N^{0.67\nu}.$$

For a non-self-avoiding chain $\nu = 1/2$, which then gives $R_g \propto N^{1/3}$; a relation obtained by Hyeon *et al.*[23] by fitting structural data for a variety of RNAs. If we assume that the size of a viral capsid is proportional to the volume of its genome, we recover the cubic relation between capsid diameter and genome length that was shown in Fig. 1.

It is interesting to note that general relations can be derived for ideal polymers that provide added insight into the effective 3D sizes of branched RNA molecules and their dependence on secondary structure statistics. Let L_{ij} denote the distance along the backbone between monomers i and j in an arbitrary ideal polymer, *linear* or *branched*. Then the average radius of gyration of this N-monomer polymer can be written as

$$R_g = \left\langle \frac{b}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} L_{ij} \right\rangle^{1/2},$$

where b is the effective persistence length (say, the Kuhn length in a freely-jointed chain model). Substituting ξ_{eff} for b and *ladder* distance LD_{ij} for

the effective distance between monomers i and j, it follows (since ξ_{eff}, the average duplex length, is a constant, approximately independent of the overall sequence) that

$$R_g \propto \left\langle \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} LD_{ij} \right\rangle^{1/2} \equiv \langle ALD \rangle^{1/2}.$$

Note that the average ladder distance (ALD) appearing here is the average over ladder distances LD_{ij} in a secondary structure, and $\langle$ALD$\rangle$ denotes the thermal ensemble average of this quantity. A calculation of $\langle$ALD$\rangle$s for the same random and viral sequences included in Fig. 8 gives a plot essentially identical to Fig. 8, i.e., the $\langle$ALD$\rangle$s for viral RNA molecules are systematically smaller than those for nonviral sequences, with the difference increasing with increasing nucleotide length. From the last equation, this implies in turn that the radii of gyration — the 3D sizes — of viral RNAs should be systematically smaller than those of nonviral sequences of the same length. Preliminary experimental results suggest that these differences are measurable and significant.

5.2. *Experiment*

In principle, the simplest way to measure the size of a large ($1 >$nm) biomolecule is by small angle X-ray scattering (SAXS). The classic experiment involves collecting data from a dilute sample at angles sufficiently small so that the log of intensity scattered at angle ϑ is a linear function of q^2, with a slope of $R_g^2/3$;[24] here q is related to the scattering angle by $q = \frac{2\pi}{\lambda} \sin(\vartheta/2)$, with λ the wavelength of the X-rays. In this way one determines the radius of gyration of the particles of interest.

The radius of gyration — and additional information about the shape of the RNA molecule — can also be obtained from the Fourier inversion of the q-dependent scattering intensity. More explicitly, the distribution of distances, $p(r)$, is given by

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty dq\ I(q)(qr) \sin(qr),$$

where $I(q)$ is the intensity in units of $I(0)$. R_g is given by the ratio of the second and zeroth moments of $p(r)$ according to

$$R_g^2 = \int_0^\infty r^2 p(r) dr / 2 \int_0^\infty p(r) dr.$$

By analyzing SAXS data in these alternative ways, Gopal *et al.*[25] have deduced radii of gyration for different-sequence RNA molecules of equal nucleotide length. They have also shown for sequences of *different* lengths that a *non*viral-sequence molecule has a *significantly* larger size than a significantly longer viral-sequence molecule. Figure 9, for example, shows distance distribution curves, $p(r)$s, for a 3171 nt-viral RNA molecule (RNA1 of CCMV), a 2117 nt-viral RNA molecule (RNA3 of BMV), and a 2117 nt-RNA transcribed from the *S. cerevisiae* chromosomes XII. Also shown is the distance distribution for the full virion (RNA-containing capsid).

Note that the non-viral sequence 2117 nt molecule has a size significantly larger than that for the equal-length viral molecule, and comparable to that of the viral molecule whose nucleotide length (3171) is 50% larger. In particular the $p(r)$ distributions for the two viral RNAs fit more comfortably in the capsid than does that of the small non-viral sequence molecule.

A special power of angle-dependent scattering experiments is that one can obtain important information about the shape and anisotropy of the scatterer, i.e., large RNA molecules in this case. In particular, it is possible

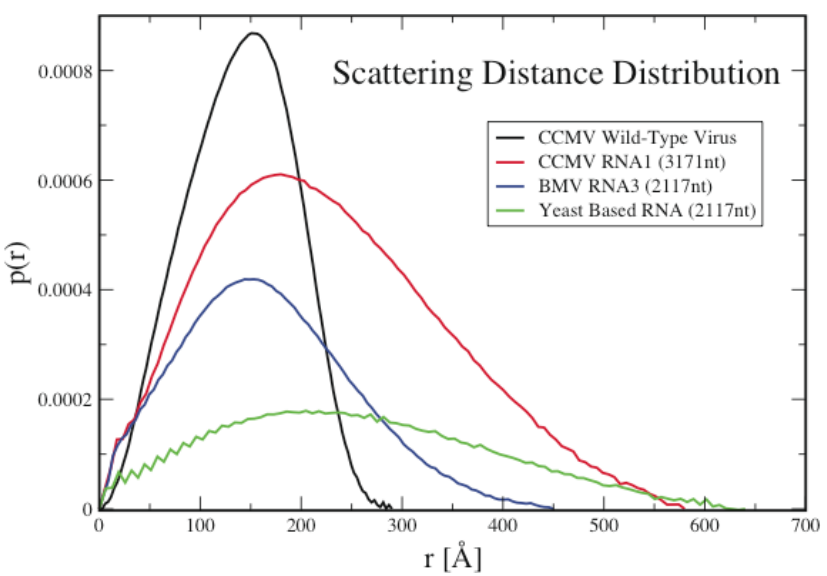


Fig. 9. The pair distance distributions, $p(r)$, for CCMV and several RNAs.

from the 'one-dimensional' $I(q)$ plots to deduce *three*-dimensional real-space images of the molecules, in addition to R_gs and distance distributions. More explicitly, computational procedures developed by Svergun[26] and by Doniach[27] involve Monte Carlo calculations of the $I(q)$ curve from real-space distributions of scattering-length density. Starting with a random distribution of density over a volume characterized by the largest distance in the molecule, one successively generates new distributions — by randomly switching scattering contrast locally from that for solvent to that for RNA — by Monte Carlo moves weighted by a least-squares difference between the measured $I(q)$ and the one calculated for the new distribution. The size of the local regions ('beads'), whose scattering densities are repeatedly reassigned by this procedure, is determined by the spatial resolution of the experiment, i.e., the maximum value of q. Iteration proceeds until a least-squares minimum difference with the measured scattering curve is obtained. The structures to which these calculations converge are not unique, but are found to differ very little from one another, and to depend only weakly on initial conditions. The middle figure in Fig. 10 shows a superposition of five such structures obtained from measurements on 2774 nt-long RNA2 of CCMV in virus storage buffer ($pH = 4.8$). Each bead has a radius of 1.45 nm; a solid surface reflecting the volume of the bead distribution is shown on the left. Finally, the figure on the right is a scaled cut-away view of the CCMV capsid (based

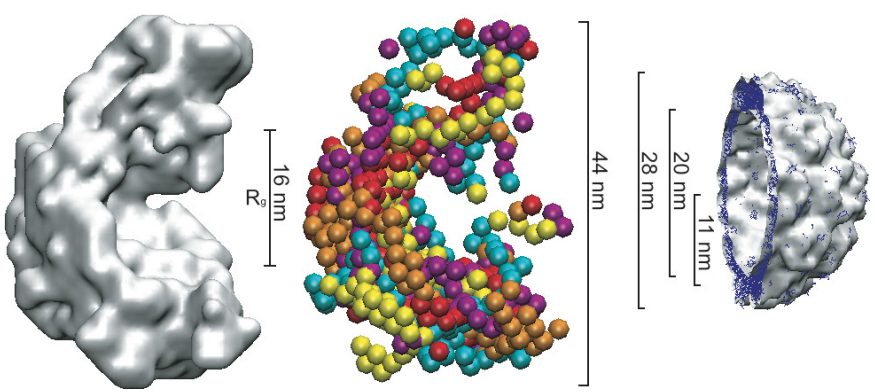


Fig. 10. A real-space 3D image reconstructions of the scattering-length distribution corresponding to the 3171 nt-long RNA2 of CCMV, as deduced from SAXS measurements.

on PDB structure 1CWP) into which this molecule is spontaneously packaged under assembly buffer conditions.

Several qualitative features of the reconstructed image stand out. First, the molecule occupies a volume comparable in size to, but somewhat *larger* than, the volume available to it inside the capsid. This is consistent with what one would expect for the optimum size of a polymer packaged in a shell with which it interacts attractively[17]; by contracting slightly, the density of chain monomers available for interacting with the inside surface of the shell is enhanced, with minimal configurational cost. Second, and perhaps most significantly, the reconstructed image (see central figure in Fig. 10) is both anisotropic and 'hollow'. The anisotropy is typical of any single trajectory of a random walk, or of appropriately averaged properties like the moment of inertia tensor; see, for example, the discussion of RNA anisotropy by Hyeon, Dima and Thirumalai.[23] 'Hollowness' refers to the fact that — unlike, say, a linear polymer whose mass density as a function of distance from the center-of-mass decreases monotonically, corresponding to a 'compact' object — here the mass (scattering) density is concentrated at the 'surface' of the molecule. We conjecture that this is not true for nonviral sequences, and that the 'hollowness' — as much as the overall size – is important for enhancing the interaction with oppositely charged capsid protein and thereby optimizing the packaging efficiency.

While small-angle X-ray scattering is the canonical technique for determining sizes of biomolecules like proteins and nucleic acids, its quantitative power is limited to small molecules. If the molecules are too large (> 10 nm) much of the scattered intensity lies at very small angles, where it is experimentally difficult to collect data. For this reason, while one is able to obtain reliable R_g values from SAXS experiments for viral sequences as long as 3171 nt, say (see above), this is no longer possible for (the larger) non-viral sequences of the same length.

Alternatively, one can determine the sizes of RNAs by measuring their diffusion coefficients. This can be accomplished conveniently by Fluorescence Correlation Spectroscopy (FCS) of molecules labeled with a fluorescent probe. The fluctuations in the number density as they diffuse into and out of a small well-defined volume defined by a laser beam in a confocal microscope lead to fluctuations in the fluorescence, and the time

autocorrelation function of the fluorescence is measured. This function decays on a time scale τ_D equal to the time it takes for a molecule to diffuse out of the illuminated volume, namely l^2/D, where l is the dimension of the volume (on the order of 1μm) and D is the diffusion coefficient of the molecule. Assuming the Stokes–Einstein relation, $D = \frac{kT}{6\pi\eta R_h}$ (with η the solution viscosity), gives the 'equivalent sphere size' — hydrodynamic radius, R_h — of the molecule. Thus, the measured correlation time is proportional to R_h, from which the relative sizes of our RNA molecules can be directly inferred.

Figure 11 shows results of this kind, comparing Brome Mosaic Virus RNA3 and non-viral (yeast) RNA sequence, each 2117 nt long. Note that the correlation time for the yeast RNA (3.4 ms), l^2/D, is almost twice that for RNA3 (1.8 ms), implying a hydrodynamic radius almost twice that of the viral sequence. Recall (from above) that $\tau_D \propto (1/D) \propto R_h$, by the Stokes–Einstein relation.

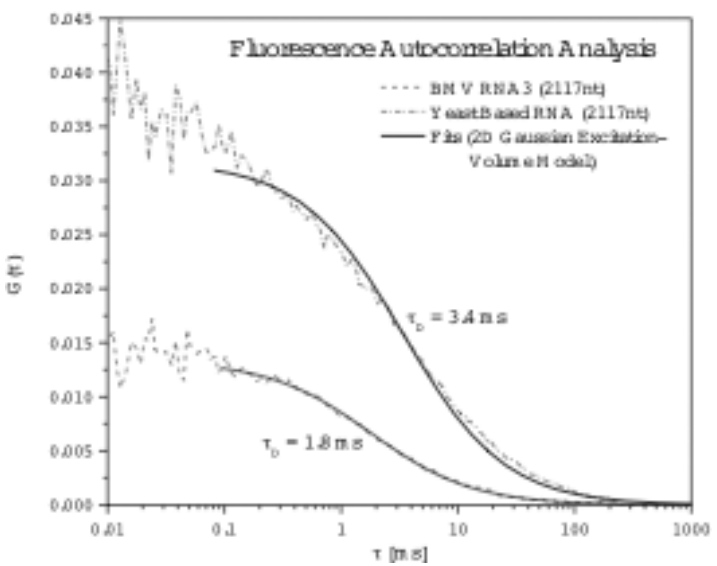


Fig. 11. Intensity autocorrelation functions obtained from FCS measurements of viral and nonviral RNA sequences of length 2,117 nt. τ_D is the time it takes for each molecule to diffuse through the illuminated volume, and is proportional to its effective size (hydrodynamic radius).

6. So … What Determines the Size of an RNA Virus?

We have seen that understanding the relationship between the size of a viral capsid and the size of RNA or other polyelectrolyte that it contains is a formidable problem. The structure of a long charged polymer in solution is in itself very difficult to predict, even more so when, as in the case of ssRNA, self-complementarity gives it a complex secondary and tertiary structure that depends on its primary sequence. If, as in many cases, the capsid proteins have charged tails that extend into the capsid interior, their structure will depend on the number of charged residues, their position and the backbone hydrogen bonding that can lead to helices. The structure of both the packaged polyelectrolyte and the tails will be altered by the strong electrostatic interactions between them, which depend as well on the concentration of the surrounding electrolyte.

In this last Section we return to this chapter's eponymous question: What determines the size of the virus — the size preferred by the capsid protein, or the size preferred by the RNA being packaged? We discuss some of the theory underlying the several size-determining factors treated in previous sections, and conclude with an outline of experiments that we believe can begin to discriminate between the competing roles played by these factors.

In the approach adopted by Belyi and Muthukumar[8] the basic N-termini of the capsid protein are modeled as a polyelectrolyte brush, assumed to be strongly stretched, which interacts only electrostatically with other peptide chains, i.e., there are no excluded-volume interactions. They evaluate the electrostatic potential due to the chains, which has the form of a parabolic well. The conformation of the genome in this potential is then determined, with the genome treated as a long flexible polyelectrolyte chain. The brush potential is affected by the genome charge, and the changes in the brush thickness (and therefore the electrostatic potential) are determined self-consistently. The total length of the genome (the number of negative charges) is then taken as the amount trapped in the well, which is energetically the most favorable situation. A direct proportionality is found between the total charge on the RNA and on the N-termini. The theory does not allow the proportionality factor to be determined, but Belyi and Muthukumar give some simple arguments that suggest that it has a value close to the 1.6 found empirically (cf. Fig. 2).

The tails and RNA are treated as semi-flexible polymers and the Coulomb energy is estimated by treating the RNA and tails as oppositely charged rigid cylinders. They conclude that the free energy minimum corresponds to overcharging by a factor of two, i.e., that the total charge on the RNA is twice that on the tails.

Belyi and Muthukumar also calculated the distribution of the viral RNA within the shell. They found that the RNA is contained in a spherical shell, with a gap between it and the capsid interior. The spatial distribution of the RNA in CCMV has also been determined in a Monte Carlo simulation.[28] In this case, the capsid was modeled in full atomic detail, both with and without the protein tails. In contrast to the detailed description of the capsid, a coarse-grained model was employed for the RNA in which each nucleotide was treated as an isolated sphere. Each sphere has one quarter of an electronic charge rather than a full charge in order to take into account the effects of counterions and solvent. About 1/5 of the spheres were fixed on the capsid interior; they represent the density of nucleotide identified from cryo-electron density maps. The remaining nucleotides were treated as unconnected and were displaced independently in the Monte Carlo moves; those closest to the capsid organize into an icosahedral arrangement. A better approximation for the RNA based on a mean-field treatment of a *branched* polymer has recently been used by Lee and Nguyen[29] to determine the radial distribution of the RNA inside a spherical capsid with a uniformly charged internal surface, which is appropriate for viruses such as Dengue and the MS2 bacteriophage, as they do not have extended protein tails. The RNA interacts with the capsid only at its interior surface. If the adsorption is high, the RNA forms a band near the surface and the interior is empty; this is the profile observed in MS2 (where there is actually a double-shell RNA organization in the particle, with connectors along the five-fold axes[30]). For weak adsorption, the thermodynamically stable profile has a maximum at the capsid center, as was observed in Dengue virus.

The way in which protein curvature and polyelectrolyte size determine the size of viral capsids can be understood by considering the free energy of such assemblies, which can be written as a sum:

$$E = E_{capsid} + E_{pol} + E_{pol-entropy} + E_{\mathrm{int}}.$$

The self-energy of the capsid, E_{capsid}, represents the energy per capsomer of the aggregate of the protein pentamers and hexamers that constitute

the capsid, i.e., the free energy of an empty capsid. This has been evaluated by Zandi *et al.*[31] from a model in which the hexamers and pentamers were treated as attracting disks confined to a spherical surface of radius R. The energy and the equilibrium spatial distributions of the disks as a function of R and N, the total number of capsomers, were determined by Monte Carlo simulations. Minima in the free energy were observed at values of N corresponding to the Caspar–Klug T numbers and can be represented by a curve qualitatively similar to the green line shown in Fig. 12, shown for convenience as a continuous function of N rather than a set of discrete values. It is important to note that these calculations

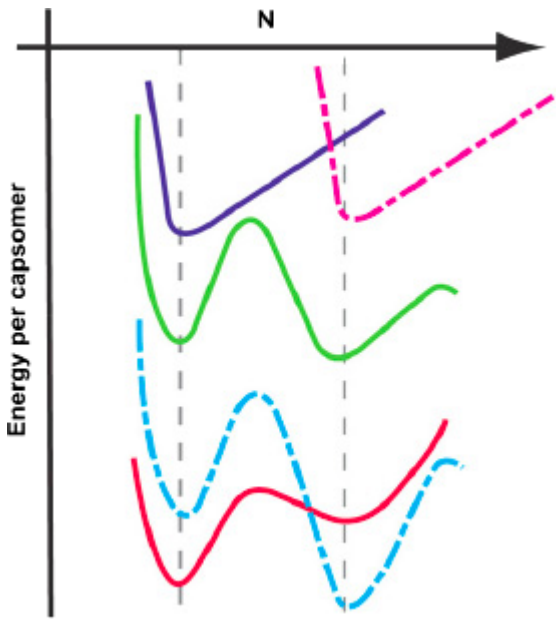


Fig. 12. Schematic illustration of the VLP size-selection mechanism. The green curve represents the self-energy of an empty capsid. For convenience the curve is shown as continuous, but it has meaning only for integral values of N, the total number of capsomers in the shell. Energy minima are found for capsids with specific T numbers. The upper curves represent the interaction energy per capsomer for each of two different molecular weights of encapsidated polymer. The lower two curves arise from the superimposition of these curves with the green curve, stabilizing a smaller capsid (red curve) in the case of the purple curve on top, and a larger capsid (dashed blue curve) in the case of the dashed pink curve (adapted from Hu *et al.*[17]). Copyright Biophysical Society 2008.

do not include the effects of spontaneous curvature, i.e., the preference for capsomers to bind to each other with special angles (smaller than 2π) between their normals. Instead the minima arise from the magic-number nature of particles attracting each other on the surface of a sphere, i.e., from the special advantage of icosahedrally symmetric distributions of nearest and non-nearest neighbors. As we have seen, however, from the discussion of capsid formation in solutions of pure CCMV protein, for example, each capsid protein must be expected to show a preferred radius of curvature upon aggregating into shells. It is these preferences that determine the relative positions of the minima shown at different N-values (hence T-numbers) in the green curves of Fig. 12.

The other contributions to the free energy are those associated with the confinement of a polyelectrolyte inside the capsid. E_{pol} is the self-energy of the polymer chain, $E_{pol-entropy}$ is the chain entropy, and $E_{\rm int}$ represents the interaction between the polymer and the capsid interior. Here, as in the various theoretical treatments described above, these terms can be evaluated only by making approximations, some of them severe. However, poly(styrene sulfonate), in which there is no self-complementarity, can appropriately be treated as a flexible linear polymer. The interaction with the capsid interior can be represented by adsorption on a wall.[32,33]

For a given value of the capsid-monomer contact energy and chain length, there is a minimum in the interaction energy per capsid unit area as a function of capsid radius. The upper curves in Fig. 12 represent such curves for two different polymer lengths (L), the purple curve representing the shorter chain and the dashed pink curve the longer. The total energy is the sum of this energy and the capsid energy (green curve). Thus, in the case of the shorter polymer, the smaller capsid would be preferred, while for the longer polymer the minimum energy corresponds to the larger capsid. In essence, the chain contribution selects the smaller or larger capsid size. More detailed Monte Carlo simulations of viral assembly[34] lead to similar conclusions. Note that while there is an optimal polymer length for a given capsid size and the energy of interaction with the surface, the free energy change allows the packaging of polymers of other lengths as well, although it is thermodynamically less favorable and probably less efficient.

The composite free energy curves in Fig. 12 make clear that there is not a single answer to the question 'What is the maximum (or minimum) length of RNA that can be packaged in the capsid of a specific

virus?' Suppose, for example, that the left-most minimum corresponds to a $T = 1$ capsid and the minimum in the curve corresponds to a polymer of length L. As the length of the polymer is increased, the packaging becomes less favorable. But still longer polymers are able to stabilize the $T = 3$ capsid and packaging again becomes favorable. A still longer polymer may be able to 'select' a $T = 7$ capsid. This is what has been observed by Verduin and Bancroft[35] who found that $T = 7$ capsids of CCMV self-assembled around TMV RNA, which is about twice as long as the CCMV RNA. This behavior is analogous to that discussed for the PSS packaging experiments, in which CCMV capsid protein is induced by progressively higher-molecular-weight PSS molecules to self-assemble into larger (in this case $T = 3$ instead of $T = 2$) capsids.

To help resolve some of the current open questions and controversies regarding the relationship between RNA size and capsid size, it will be important to carry out experiments that measure quantitatively the relative packaging efficiencies of different RNA molecules by the same capsid protein. CCMV capsid protein will be the protein of choice for these experiments, because of its 'promiscuity'. More explicitly, while it prefers a particular T number (and hence capsid size) under a given set of buffer conditions, it can be persuaded to form other structures — as just described above — when presented with different sizes of anionic polymer or RNA. Also, unlike BMV, whose capsid protein can only package RNA molecules containing a specific transfer-RNA-like sequence at their 3′ ends, CCMV capsid protein can efficiently package a broad range of anionic polymers. For example, it can package TMV RNA whose nucleotide length (6500 nt) is twice that packaged by the wild-type CCMV virions (each containing either 3200 nt [RNA1], 2800 nt [RNA2], or 2900 nt [RNA3 + RNA4]). But in these instances the capsid is much larger than wild-type, and essentially no information is available on the relative efficiencies of packaging.

Consider several different RNA molecules of the same nucleotide length — and hence carrying the same total charge, Q — but with distinctly different 3D sizes (R_gs). Suppose, in particular, that $Q = Q^*$, the optimum charge for a CCMV capsid; we can take Q^* to be the value on the plot in Fig. 2, i.e., around -3000e for CCMV. We then ask the following question: If equal numbers of each of these several different RNA molecules are added to equal numbers of CCMV capsid proteins, how will their packaging efficiencies — the fractions of them ending up in capsids — depend on their 3D size (R_g)? In particular, will the viral sequence

correspond to the optimum 3D size for a capsid this size? Or will there be essentially no observable dependence of packaging efficiency on size, consistent with the idea that total charge (which is the same for all of the molecules) is the only important determining factor?

Alternatively, consider several different RNA molecules with the same R_gs but with distinctly different total charges (nucleotide lengths). Suppose further that this common R_g is the optimum one, i.e., the one corresponding to the viral value. In the preceding Section we saw, for example, that a 1000 nt-long yeast-derived RNA sequence is found to have the same radius of gyration as a 3000 nt-long viral RNA molecule, with their total charges differing by as much as a factor of 3. Will a plot of packaging efficiency *vs.* Q show a sharp maximum at -3000e — or will there be only a weak dependence on Q, consistent with the 3D size being the dominant determining factor?

We believe that capsid efficiency measurements of the kind discussed here, along with further high-resolution structural measurements — both X-ray and cryo-electron microscopy — will soon elucidate these issues of RNA viral packaging that are likely to play an important role in determining *in vivo* infectivity.

Acknowledgement

This work was supported by grant CHE-0714411from the U.S. National Science Foundation.

References

1. N. Ban and A. McPherson. The structure of satellite panicum mosaic virus at 1.0 A resolution. *Nature Struct. Biol.* **2**:882–90 (1995).
2. X. Yan, P.R. Chipman, T. Castberg, G. Bratbak and T.S. Baker. The marine algal virus PpV01 has an icosahedral capsid with T = 219 quasisymmetry. *J. Virol.* **79**:9236–43 (2005).
3. S.V. Beer and T. Kosuge. Spermidine and spermine-polyamine components of turnip yellow mosaic virus. *Virol.* **40**:930–8 (1970).
4. M.A. Krol, N.H. Olson, J. Tate, J.E. Johnson, T.S. Baker and P. Ahlquist. RNA-controlled polymorphism in the *in vivo* assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proc. Natl. Acad. Sci. USA* **96**:13650–5 (1999).

5. P. Plevka, K. Tars, A. Zeltins, I. Balke, E. Truve and L. Liljas. The three-dimensional structure of ryegrass mottle virus at 2.9 A resolution. *Virol.* **369**:364–74 (2007).
6. J.B. Bancroft. The self-assembly of spherical plant viruses. *Adv. Virus Res.* **16**:99–134 (1970).
7. S. Brumfield, D. Willits, L. Tang, J.E. Johnson, T. Douglas and M. Young. Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *J. Gen. Virol.* **85**:1049–53 (2004).
8. V.A. Belyi and M. Muthukumar. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. USA* **103**:17174–8 (2006).
9. V. Prasad, R. Rothnagel, X. Jiang and M.K. Estes. Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *J. Virol.* **68**:5117–25 (1994).
10. Michel, I.L. Ivanovska, M.M. Gibbons, W.S. Klug and C.M. Knobler. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc. Natl. Acad. Sci. USA* **103**:6184–9 (2006).
11. R.F. Bruinsma, W.M. Gelbart, D. Reguera, J. Rudnick and R. Zandi. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* **90**:248101 (2003).
12. K.W. Adolph and P.J.G. Butler. Studies on the assembly of a spherical plant virus. I. States of aggregation of the isolated protein. *J. Mol. Biol.* **88**:327–41 (1974).
13. L. Lavelle, M. Gingery, M. Phillips, W.M. Gelbart, C.M. Knobler, R. Cadena-Nava, J.R. Vega-Acosta, L.A. Pinedo-Torres and J. Ruiz-Garcia. Phase diagram of self-assembled viral capsid protein polymorphs. *J. Phys. Chem. B* **113**:3813–9 (2009).
14. P. Prinsen, P. van der Schoot, W.M. Gelbart and C.M. Knobler, in preparation.
15. M.W. Johnson, G. Wagner and J.B. Bancroft. A trimetric and electrophoretic study of cowpea chlorotic mottle virus and its protein. *J. Gen. Virol.* **19**:263–73 (1973).
16. D. Ben Yaakov, Y. Burak, D. Andelman and S.A. Safran. Electrostatic interactions of asymmetrically charged membranes. *Euro. Phys. Lett.* **79**:48002 (2007).
17. Y. Hu, R. Zandi, A. Anavitarte, C.M. Knobler and W.M. Gelbart. Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. *Biophys. J.* **94**:1428–36 (2008).
18. L. Wang and Y. Yu. Chain conformation of linear polyelectrolyte in salt solutions: sodium poly(styrenesulfonate) in potassium chloride and sodium chloride. *Macromolecules* **21**:3498–501 (1988).
19. J.M. Johnson, D.A. Willits, M.J. Young and A. Zlotnick. Interaction with capsid protein alters RNA structure and the pathway for *in vitro* assembly of Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* **335**:455–64 (2004).
20. A.M. Yoffe, P. Prinsen, A.J. Gopal, C.M. Knobler, W.M. Gelbart and A. Ben-Shaul. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. USA* **105**:16153–8 (2008).
21. I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125**:167–88 (1994).

22. R. Bundschuh and T. Hwa. Statistical mechanics of secondary structures formed by RNA sequences. *Phys. Rev. E* **65**:031903 (2002).
23. C. Hyeon, R.I. Dima and D. Thirumalai. Size, shape and flexibility of RNA structures. *J. Chem. Phys.* **125**:194905 (2000).
24. A. Guinier and G. Fournet. *Small angle scattering of X-rays.* Trans. by C.B. Wilson, Wiley, New York (1955).
25. A. Gopal, D. Egecioglu, M. Niebuhr, L.T. Fang, A.L.N. Rao, W.M. Gelbart and C.M. Knobler, in preparation.
26. D.L. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**:2879–86 (1999).
27. S. Doniach. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.* **101**:1763–78 (2001).
28. D. Zhang, W. Konecny, N.A. Baker and J.A. McCammon. Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. *Biopolymers* **75**:325–37 (2004).
29. S.I. Lee and T.T. Nguyen. Radial distribution of RNA genomes packaged inside spherical viruses. *Phys. Rev. Lett.* **100**:198102 (2008).
30. K. Toropova, G.Basnak, R. Twarock, P.G. Stockley and N.A. Ranson. The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**:824–36 (2008).
31. R. Zandi, D. Reguera, R.F. Bruinsma, W.M. Gelbart and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA* **101**, 15556–60 (2004).
32. H. Ji and D. Hone. Polymer adsorption on rough surfaces. 2. Good solvent conditions. *Macromolecules* **21**:2600–5 (1998).
33. P. van der Schoot and R.F. Bruinsma. Electrostatics and the assembly of an RNA virus. *Phys. Rev. E* **71**:061928 (2005).
34. D.G. Angelescu, R. Bruinsma and P. Linse. Monte Carlo simulations of polyelectrolytes inside viral capsids. *Phys, Rev. E* **73**:041921 (2000).
35. B.J.M. Verduin and J.B. Bancroft. The infectivity of tobacco mosaic virus RNA in coat proteins from spherical viruses. *Virology* **37**:501–6 (1969).

Chapter 9

Physics of Viral Infectivity: Matching Genome Length with Capsid Size

Alex Evilevitch[*,†,§] and Martin Castelnovo[‡]

In this work, we review recent advances in the field of physical virology, presenting both experimental and theoretical studies on the physical properties of viruses. We focus on the double-stranded DNA (dsDNA) bacteriophages as model systems for all of the dsDNA viruses both prokaryotic and eukaryotic. Recent studies demonstrate that the DNA packaged into many dsDNA viral capsids is highly pressurized, which provides a force for the first step of passive injection of viral DNA into either bacterial or eukaryotic cells. Moreover, specific studies on capsid strength show a strong correlation between genome length and capsid size and robustness. The implications of these newly appreciated physical properties of a viral particle with respect to the infection process are discussed.

1. Introduction

Viruses are simple lifeless entities that cannot reproduce on their own and therefore depend on host cells to provide them with the necessary life support mechanisms. Viruses are categorized as either prokaryotic (also called bacteriophage, or 'phage' for short, originating from the Greek word '*phagein*' [to eat]) or eukaryotic (i.e. viruses that infect plant and animal cells). Simplified, all viruses consist of a protein shell (capsid) that

[*]Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA.
[†]Department of Biochemistry, Center for Chemistry and Chemical Engineering, Lund University, Box 124, S-221 00, Lund, Sweden.
[‡]Laboratoire de Physique, Ecole Normale Superieure de Lyon, 46 Allee d'Italie, 69364 Lyon Cedex 07, France.
[§]For Correspondence: Alex.Evilevitch@biochemistry.lu.se or alexe@andrew.cmu.edu

protects the viral genome (DNA or RNA). To infect, the viral genome must enter the cell, either alone or as part of a viral nucleoprotein complex. This entry is assisted by receptors that are present on the host cell surface (Flint *et al.*, 2004b; Hendrix *et al.*, 1983; Ptashne, 2004). Many phages are equipped with a tubular 'tail' structure through which the viral genome passes into the cell; the capsid is not internalized. In contrast, most animal viruses are internalized into the host cell essentially intact through endocytosis or membrane fusion mechanisms. In the latter case, the capsid is covered with a lipid envelope that fuses with the host cell plasma membrane. Plant viruses utilize insects as vectors that aid in mechanical penetration of the plant cell wall (Flint *et al.*, 2004b; Hendrix *et al.*, 1983; Ptashne, 2004). Once the viral genome has entered the cell, the virus hijacks the host cell's machinery and synthesizes multiple copies of the viral genome. Transcription and translation of viral genes yields proteins that assemble into new viral particles that exit the host cell. This may occur by a budding process that is not necessarily detrimental to the cell or by cell lysis, which is a lethal event (Flint *et al.*, 2004a; Hendrix *et al.*, 1983; Ptashne, 2004).

Viruses present a major threat to human health and welfare. The extent of this threat becomes obvious if one considers the morbidity and mortality caused by the human immunodeficiency virus (HIV), hepatitis virus, or influenza as well as threats from SARS (severe acute respiratory syndrome) and H5N1 (avian influenza virus) (WHO, 2007). The potential for global spread of lethal viral infections is accentuated by the modern aviation network and traveling habits, thus putting millions of people at risk (Brockmann *et al.*, 2006). Aside from vaccination, current antiviral strategies either target the receptors through which viruses enter host cells, or target specific enzymes that are critical to viral replication. For instance, abacavir and retrovir inhibit HIV reverse transcriptase, acyclovir inhibits DNA polymerase of herpes simplex virus, and oseltamivir and zanamivir inhibit neuraminidase of the influenza virus (Flint *et al.*, 2004c). The main limitation of these kinds of anti-viral agents is that they are highly 'specialized', as they target specific events in individual viral replication cycles. Thus, their use is limited to a specific virus or a small group of viruses. Furthermore, viruses frequently undergo mutations that alter the target sites. Therefore, there is a growing need to develop less specialized viral intervention strategies that will enable targeting of a broad range of viruses with little susceptibility to mutation. For instance, drugs that directly interfere

with the assembly of the viral capsid and/or packaging of the viral DNA into the capsid could provide a 'broad spectrum' anti-viral agent.

Biophysical virology is a rather new field that seeks to define the physical mechanisms controlling virus development. This knowledge can provide information essential to the rational design of new antiviral strategies with less restriction to a limited number of viruses. Biological and physical simplicity relative to other viral systems have made the bacteriophages an attractive model system in biophysical virology. Notwithstanding, the fundamental mechanistic and structural features of phages can be directly applied to the eukaryotic viruses. For many years phages have served as reliable experimental models that have significantly contributed to the advent of modern molecular biology. To mention a few examples, initial studies on gene regulation and protein synthesis were conducted using phage (Echols, 2001; Hendrix *et al.*, 1983; Ptashne, 2004). Further, Hershey and Chase used phages for one of the first demonstrations that DNA is the hereditary material (Hershey & Chase, 1952). Today, phages provide ideal systems to study biophysical virology and this review focuses on recent advances in this field.

A common feature of many phages is that the microns-long genome is confined within a nanometre-size capsid by means of an ATP-driven motor (Riemer & Bloomfield, 1978). In the case of phage λ and many other double-stranded DNA (dsDNA) phages, this confinement creates internal pressures of tens of atmospheres, which is exerted on the inner capsid walls (Evilevitch *et al.*, 2004; Kindt *et al.*, 2001; Purohit *et al.*, 2003b; Smith *et al.*, 2001; Tzlil *et al.*, 2003). DNA is packed to nearly crystalline densities inside the capsid; given the limited volume of the rigid viral capsid, such levels of compression lead to strong bending of the DNA and DNA–DNA repulsion between its negatively charged strands (Kindt *et al.*, 2001; Tzlil *et al.*, 2003). The high force generated by the pressurized genome is in turn necessary to initiate the infection process through DNA ejection into the cell, which is energetically a downhill process that does not require ATP (Evilevitch *et al.*, 2005; Evilevitch *et al.*, 2003). This common physical property of many dsDNA phages is likely applicable to eukaryotic dsDNA viruses that use motor proteins to package their genome, e.g. the herpesviruses and possibly adenovirus (Catalano, 2005). The physical properties of the DNA-filled capsid suggest that high internal pressure induced by tightly packed DNA must be physically matched by the mechanical strength of the capsid, which must withstand

this pressure (Ivanovska *et al.*, 2007; Ivanovska *et al.*, 2004). This raises several fundamental questions that will be discussed in this review:

- What is the physical basis for the internal pressure generated by packaged DNA? What internal and external factors control this pressure?
- How can phage capsids, which are held together by non-covalent hydrophilic/hydrophobic interactions between capsid-protein subunits, withstand this high pressure and how does it affect survival of viruses outside of the host cell?
- How is DNA packaged and ejected from phage? Is the internal pressure required for infection *in vivo*?

This review is an interdisciplinary account of theoretical and experimental aspects of biophysical virology, where most of the experiments have been performed on phages λ, ϕ29, T4, T5, T7 and SPP1. We discuss recent work on the mechanism of DNA packaging and ejection in *Section 2* and the mechanical properties of viral capsids are reviewed in *Section 3*. We also discuss some of the eukaryotic viruses for which relevant measurements have been made. Viral infectivity and virion stability are discussed from the perspective of the internal genome stress and the consequences of it for the virus. We believe that a deeper understanding of this common physical parameter for many dsDNA motor-packaged viruses will lead to new possibilities for development of antiviral strategies as well as novel tools for gene therapy and nanotechnology.

2. DNA Ejection

2.1. *Theoretical Considerations*

In this section, we review the physical concepts underlying the current understanding of DNA packaging properties. Indeed it has long been a challenging theme for physicists, because it implies basic concepts of polymer physics that can be explicitly addressed in the case of bacteriophages. Early systematic studies traced back to work of Earnshaw and Harrison (Earnshaw *et al.*, 1976; Earnshaw & Casjens, 1980; Earnshaw & Harrison, 1977; Earnshaw *et al.*, 1978) highlighted a concentric layered structure of the DNA within the phage λ capsid, and a strong correlation between genome length and spacing between consecutive DNA layers. This study led to the proposal of several packaging models: the DNA genome has

been envisaged as being wound into a spool (Earnshaw & Casjens, 1980; Earnshaw & Harrison, 1977; Gabashvili & Grosberg, 1991; Gabashvili & Grosberg, 1992; Harrison, 1983; Hendrix, 1978; Riemer & Bloomfield, 1978), a ball (Earnshaw & Harrison, 1977; Richards *et al.*, 1973), and a liquid crystal with hairpins or folds (Black *et al.*, 1985; Earnshaw & Harrison, 1977; Serwer, 1986; Sun *et al.*, 1999). The first quantitative estimate of the energetics of DNA packaging was proposed by Riemer and Bloomfield (Riemer & Bloomfield, 1978) on theoretical grounds. However, the lack of precise information regarding the true conformation of packaged DNA has hampered direct comparison with experimental data. It was nevertheless realized that the Gibbs free energy of the system is dominated by the balance of two main contributions associated with both the small size of viral capsid and high DNA density: these are (i) the bending energy that tends to crowd DNA on the capsid wall in order to minimize DNA curvatures, and (ii) DNA–DNA self-interactions that tend to oppose this self-crowding.

This is nicely illustrated by bacteriophage λ (Fig. 1): in this case, the internal capsid diameter is approximately 55 nm, i.e. of the order of the DNA persistence length (50 nm), which represents the typical length scale of spontaneous bending due to thermal fluctuations. Packaging of DNA lengths longer than the persistence length inside such a small volume will therefore cost a significant amount of mechanical work. Similarly, the volume fraction of DNA inside the capsid for wild-type λ is roughly 0.59, implying small distances between consecutive DNA layers, on the order of 2.7 nm in the case of an inverse spool conformation to be described below. At these very short separations, dense hexagonal phases of DNA exhibit strong electrostatic and hydration self-interactions, as shown using the osmotic stress method by Rau and Parsegian (Rau & Parsegian, 1992; Strey *et al.*, 1998). As discussed below, the buffer conditions, particularly the multivalent ion content, strongly influence these self-interactions, which can be either repulsive or attractive (Bloomfield, 1997; Evilevitch *et al.*, 2008a). The balance between these two characteristic energies leads to an internal 'pressure' on the order of several tens of atm, equivalently to an internal force of roughly tens of pN ($1\,\text{atm} = 0.1\,\text{pN/nm}^2$). We note that in addition to these two major energetic contributions to DNA packaging, the entropy cost associated with DNA packaging inside the capsid has not been thoroughly investigated analytically and this might also play

some non-negligible role in the total Gibbs free energy, as is suggested by recent simulations (Forrey & Muthukumar, 2006).

Using an inverse spool conformation, Odijk was able to cast analytically these two dominant energies into a single predictive model for bacteriophage T7 (Odijk, 1998). In this proposed conformation, DNA is wound in the interior of the capsid in an ordered way, filling first layers close to the walls followed by layers of decreasing radius, finally leaving a cylindrical hole inside the capsid devoid of any DNA (Fig. 1). It has been well established by Cerritelli (Cerritelli *et al.*, 1997) that this conformation is the most likely and is the result of a central protein connector that imposes internal cylindrical symmetry to the DNA condensate. In Odijk's model, total bending energy is calculated as the sum of elementary bending energies of circles filling the inverse spool region associated with the DNA condensate. Similarly, DNA self-interactions are assumed to be of purely electrostatic origin and are evaluated explicitly using a classical Poisson–Boltzmann approach. The balance between these two energies specifies the optimal spacing between neighboring turns of DNA in the condensate and the main predictions of this model are (i) the spacing between DNAs in the hexagonal packing within the inverse spool and (ii) the inner radius of the condensate. The resulting theoretical values are in quantitative agreement with Cerritelli's experimental data. While this model was the first to address the structural properties of packaged DNA,

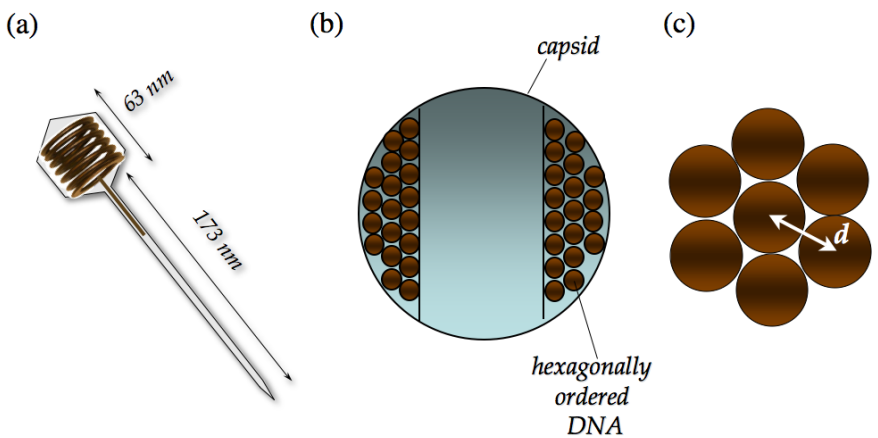


Fig. 1. (a) Schematic illustration of phage. (b) Cross-section of the capsid with hexagonally ordered DNA, (c) d is interaxial DNA–DNA spacing.

a quantitative discussion of the energetics in connection with a particular structure was still missing at that time.

The next step towards this direction came from single molecule experiments. Indeed, original experiments performed by Smith *et al.* (Smith *et al.*, 2001) renewed the interest of the physics community on the problem of viral DNA packaging, by providing direct single-molecule measurements of the *in vitro* packaging rates in phage $\phi 29$. From these studies, an internal force resisting packaging on the order of tens of pN can be inferred, thus providing quantitative estimates of the energy associated with packaging (as will be discussed in the next section). Based on these experimental results, several theoretical models of DNA packaging have been proposed in recent years, addressing a variety of issues by both analytical models and numerical simulations.

Tzlil *et al.* (Tzlil *et al.*, 2003) extended Odijk's original model to quantitatively include the energetics. First they allowed for shape variations of the inverse spool condensate ranging from ideal spool to toroid, and second they used more realistic DNA–DNA interactions based on osmotic stress experiments (Rau & Parsegian, 1992; Strey *et al.*, 1998). The outcome of this extended model is to describe the relative contribution of each of the energies (bending, self-interaction, and surface tension in the case of attractive self-interactions induced by multivalent ions, etc.). A similar model was proposed by Purohit *et al.* (Purohit *et al.*, 2005; Purohit *et al.*, 2003a; Purohit *et al.*, 2003b), further extending the theoretical discussion on experimental quantities affecting the amplitude of internal force: capsid size and shape, genome length, discreteness of DNA condensate and buffer conditions. This work showed a strong reduction of internal force resisting packaging if DNA–DNA interactions are switched from repulsive to attractive conditions. Indeed in the former case, the repulsive self-interactions result in a high cost for self-crowding as soon as a few persistence lengths of DNA are packaged. In the latter case, attractive interactions induce the condensation of DNA into a toroid in bulk solution, whose dimensions are slightly larger than the capsid size in the case of bacteriophage λ. This means that early stages of DNA packaging cost much less crowding energy, because the capsid influence on the toroid is negligible. Since more DNA is internalized in the capsid, the DNA toroidal condensate becomes compressed, its shape changes from a toroid to an inverse spool, and consequently free energy increases sharply.

One of the first simulations was performed by Kindt *et al.* (Kindt *et al.*, 2001). In this work, DNA packaging is mimicked by Brownian molecular dynamics by inserting a semi-flexible chain made of beads into a sphere in a stepwise fashion. By an appropriate choice of physical parameters (e.g. persistence length, capsid size, and interactions between beads), these simulations are able to quantitatively reproduce variations of the internal force observed in single molecule experiments as nucleic acid is translocated into the capsid. They showed for example that if attractive DNA–DNA self-interactions are assumed due to the presence of multivalent ions in the buffer, the packaging proceeds mainly in two steps. During the first two third of genome packaging, the force increases smoothly, the attractive interactions helping to reduce the cost of packaging, as discussed above. Then the force increases sharply during the last third of packaging, in qualitative agreement with the experiments by Smith *et al.* (Smith *et al.*, 2001).

Many more numerical simulations of semi-flexible packaging inside a container were subsequently proposed (Ali *et al.*, 2006; Arsuaga *et al.*, 2002; Forrey & Muthukumar, 2006; Klug & Ortiz, 2003 ; LaMarque *et al.*, 2004; Locker & Harvey, 2006; Petrov *et al.*, 2007a; Petrov *et al.*, 2007b; Spakowitz & Wang, 2005). Unlike experimental studies to date, the use of computer simulations allows changing at will specific parameters of the system, such as capsid shape, twisting DNA during packaging, constant rate or constant force packaging, strength of electrostatic interactions, *etc.*, which allows theoretical testing of the relevance of these parameters. Consequently these works have provided important structural insights on DNA conformation inside the capsid. However, precise experimental results on DNA structure within viral capsids are not yet available for direct comparison. Most of the simulations nevertheless showed that semi-flexible chain organization inside cavities might not be as well ordered as it is assumed to be in previous analytical models, unless attractive self-interactions are included (Forrey & Muthukumar, 2006) or elongated capsid shapes are assumed (Petrov *et al.*, 2007b).

Another complementary way of providing quantitative data on DNA packing energetics is to investigate DNA ejection from the capsid. Since DNA packing properties are expected to produce mechanical ‘pressure’ on the order of several tens of atm within the capsid, it was proposed by Kindt *et al.* (Kindt *et al.*, 2001; Tzlil *et al.*, 2003) that this pressure could

be balanced during *in vitro* triggered DNA ejection by osmotic pressure in the solution bathing the capsids. This was confirmed by Evilevitch *et al.* (Evilevitch *et al.*, 2005; Evilevitch *et al.*, 2003), where incomplete DNA ejection from the bacteriophage λ capsid is observed (see Section 2.2 below). Using these data together with an appropriate modeling of the osmotic force resisting ejection from the bacteriophage (Castelnovo *et al.*, 2003; Evilevitch *et al.*, 2004), it was possible to compare the analytical models mentioned above with experimental data. This osmotic force is associated with the thermodynamic PV work of creating a cavity large enough to accommodate DNA, where P is the osmotic pressure outside the capsid and V is DNA volume. An excellent quantitative agreement was observed with λ deletion mutants (Grayson *et al.*, 2006), thus validating the current theoretical description.

As mentioned above, buffer conditions strongly influence the amplitude of internal force resisting packaging. This was shown recently using the osmotic suppression technique with bacteriophage λ (Evilevitch *et al.*, 2008a) by increasing the concentration of various multivalent ions (magnesium, spermine) in solution. The experimental data were successfully interpreted within the framework of previous theoretical models by changing the parameters of DNA self-interactions, providing further credibility to the inverse spool model; however, recent studies indicate that our understanding of DNA packaging properties is still incomplete. First, a striking feature of all the simulations mentioned so far is that the final conformation of packaged DNA might deviate significantly from the inverse spool, which is the basis of the successful model described above (Forrey & Muthukumar, 2006). Further, it appears that the order of magnitude of internal force resisting packaging is not very sensitive to the final conformation. These facts suggest that some rather general formulation of this force beyond the inverse spool models is needed. Indeed, Castelnovo *et al.* (Castelnovo & Evilevitch, 2007) showed that osmotic suppression data on bacteriophages λ and T5 is equally well described using a structureless model (i.e. no hexagonal packing is assumed) of the DNA condensate. In this simple model, DNA self-interactions are assumed to be of excluded volume or hard-core type, and the only length dependent contribution to the internal force comes from bending penalty. Therefore this work questions the reason for the agreement of structured and structureless models, which remains an open question.

2.2. *DNA Ejection Experiments*

In order to infect a cell, viruses must release the nucleic content of their capsids into the host cell. Tailed bacteriophages attach to specific receptors on the surface of the host, which triggers DNA ejection into the cell. Extracting and purifying such receptors provides an important tool to study phage ejection mechanism *in vitro*; only a few phage receptors have been successfully purified, however. For example, the LamB receptor of λ (Roa & Scandella, 1976), the FhuA receptor of T5 (Boulanger *et al.*, 1996) and the YueB780 receptor of SPP1 (Sao-Jose *et al.*, 2006) can be purified in functional form. Although the mechanism of complete DNA ejection from phage *in vivo* is not entirely clarified, there is experimental and theoretical evidence that the DNA internal force is responsible for initiating the ejection process and for the translocation of the first DNA fraction into the cell. Some of the main questions related to the DNA ejection process are:

- Is the internal force stored in the capsid responsible for DNA release from phage *in vivo*?
- How do osmotic pressure, salt and DNA length affect the extent and rate of DNA ejection?

Capsids are permeable to water and smaller ions but are impermeable to larger molecules like polyethylene glycol (PEG) (Evilevitch *et al.*, 2003). DNA packaging experiments, described above, demonstrate that forces inside a phage capsid can reach values greater than 50 pN (Fuller *et al.*, 2007; Smith *et al.*, 2001) (corresponding to ~50 atm internal pressure) when a genome is fully packaged. Binding of the cellular receptor to the phage tail opens the tube and the internal force of the pressurized DNA, F_{eject}, promotes full ejection of the genome. When the ejection is triggered in the presence of added osmolytes, which generate an external osmotic pressure, this ejection force will be balanced by the osmotic force resisting ejection of DNA from the capsid. The osmotic force resisting ejection has been attributed to a decrease of the hydration properties of DNA inside the capsid resulting from a change in the chemical activity of the water hydrating DNA induced by osmolytes (Evilevitch *et al.*, 2008a). DNA ejection will thus proceed until the forces are equal, $F_{eject} = F_{resist}$. PEG is an inert osmotic agent that is widely used to set osmotic pressure in biological systems and to mimic bacterial cytoplasm (Zimmerman, 1993). In order to address the question as to whether phage DNA ejection can be suppressed

by osmotic pressure, we have used PEG 8000 (MW = 8000 Da) which is too large to enter the capsid and for which the osmotic pressure has been well calibrated (Evilevitch *et al.*, 2003; Parsegian *et al.*, 1986; Rau & Parsegian, 1992). In our experimental setup, purified phage λ particles were incubated with increasing concentrations of PEG 8000 in the presence of DNase I (Evilevitch *et al.*, 2003). Once the tail tube is opened with the LamB receptor, DNA is ejected into solution and digested into nucleotides by DNase I, leaving the unejected DNA fraction intact (undigested) inside the capsid. The fraction of DNA ejected is determined by measuring the concentration of nucleotides (digested DNA) by UV absorbance after removing DNA-containing capsids by centrifugation (see Fig. 2a). We found that the fraction of DNA ejected from phage λ (capsid outer diameter 63 nm containing 41.5 kb DNA, 86% the length of the wt 48.5 kb genome) decreases monotonically with increasing osmolyte concentration and is completely inhibited at osmotic pressure corresponding to 20 atm (see Fig. 2b) (Evilevitch *et al.*, 2003).

Analysis of DNA length remaining in the capsid, using pulsed-field gel electrophoresis, nicely confirmed these data (Evilevitch *et al.*, 2005). Osmotic suppression measurements were also performed with

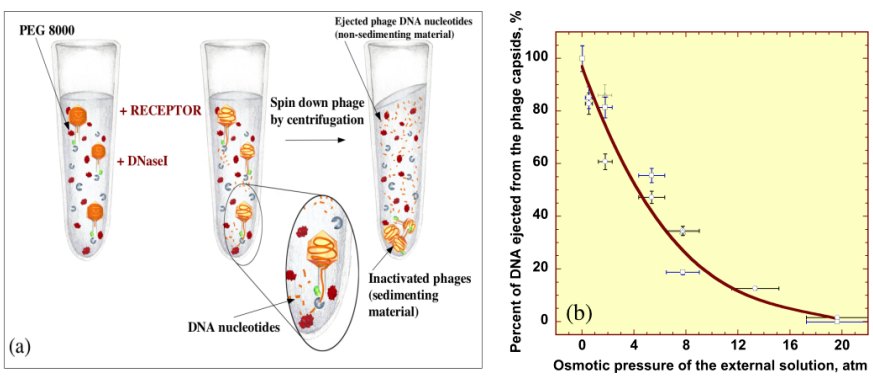


Fig. 2. (a) Illustration of osmotic suppression experiment, (b) Fraction of DNA ejected as a function of the osmotic pressure (calculated from the concentration of PEG 8000). Data are presented for samples taken from three different phage stocks. Horizontal and vertical error bars represent one standard deviation and were determined by propagation of errors in the PEG concentration and the absorbance. Adapted from: E. Nurmemmedov, M. Castelnovo, C.E. Catalano, A. Evilevitch. Biophysics of viral infectivity: Matching genome length with capsid size. *Q. Rev. Biophys.* **40**(4): 327–356 (2007). Copyright Cambridge University Press.

bacteriophage T5 (capsid outer diameter 90 nm containing 113 kb DNA), where the fraction of ejected DNA was determined as a function of osmotic pressure set by PEG 8000 (Castelnovo & Evilevitch, 2007). Similar to phage λ, the fraction of DNA ejected progressively decreased with increase in osmotic pressure exerted by the osmolyte; however, the pressure required to completely inhibit DNA ejection was only 6 atm (compared to 20 atm for λ) (Castelnovo & Evilevitch, 2007). Modeling internal ejection forces for phages λ and T5 as a function of DNA remaining inside the capsid and rescaling these data with regard to the capsid size and DNA filling density produces a 'master plot' showing that despite differences in internal pressure the osmotic suppression data for these two different phages is superimposable (Castelnovo & Evilevitch, 2007). This nicely illustrates that the same physical parameters control DNA ejection from two distinctly different bacteriophages and suggests that there is a general mechanism of DNA ejection that applies to many dsDNA viruses that use ATP-driven motors to package their genomes. These osmotic suppression experiments also partially recapitulate ejection behavior *in vivo* by mimicking the colloidal milieu of the host cell cytoplasm, which also exerts osmotic force on phage DNA during ejection and packaging processes. For instance, osmotic pressure inside *E.coli* cells under normal conditions is at least 2–4 atm (Neidhardt, 1996). We have demonstrated that wild-type phage λ can passively eject only 60% of its genome against this pressure (Grayson *et al.*, 2006). Hence, the pressure induced by the tightly packaged DNA inside the phage capsid is insufficient to complete DNA ejection *in vivo*.

Various two-step *in vivo* DNA internalization scenarios have been suggested for phage (Alcorlo *et al.*, 2007; de Frutos *et al.*, 2005b; Evilevitch, 2006; Gonzalez-Huici *et al.*, 2004; Inamdar *et al.*, 2006; Kemp *et al.*, 2004; Lof *et al.*, 2007b; Molineux, 2001). For example, genome transfer from ϕ29 is described as a push-pull mechanism such that the first ~65% of the genome is 'pushed' into the cell by pressure, while the remaining 35% is 'pulled' into the cell by enzyme-assisted processes (Alcorlo *et al.*, 2007; Gonzalez-Huici *et al.*, 2004). Similarly, it was shown that only 8% of T5 genome is injected into cell in the first step, allowing production of phage-encoded proteins that are necessary for the internalization of the remaining DNA (Lanni, 1968). Interestingly, the *in vitro* osmotic suppression measurements on phage T5 described above indicate only ~9% of the wild-type T5 genome (121.75 kbp) is ejected at an osmotic pressure of 3–4 atm (corresponding to cytoplasmic pressure),

consistent with the *in vivo* observations (Castelnovo & Evilevitch, 2007). For bacteriophage T7, experimental evidence suggests that its entire genome is actively pulled into the cell by RNA polymerases, similar to the situation observed for T5 (Kemp *et al.*, 2004; Molineux, 2001). In addition, there is now evidence that the portion of DNA that is not passively ejected into the cell by pressure can be pulled out of the phage capsid by adsorbing and ratcheting effects of non-specific DNA-binding proteins, such as HU protein in *E. coli* (Lof et al., 2007b), as well as by the condensing effects of polyamines (*i.e.* spermidine) also present in cells (de Frutos *et al.*, 2005a; Evilevitch, 2006). For instance, the addition of a non-specific DNA-binding protein, HU, to a solution of phage λ has been shown to significantly increase the rate of ejection (Lof *et al.*, 2007b).

As has been discussed above, the length of the packaged DNA has a pronounced effect on the internal pressure. The ability to accurately predict the force required to stop DNA ejection or to stop packaging of the DNA into a capsid is clearly important to our fundamental understanding of the viral replication cycle and has many practical applications. For instance, for gene therapy purposes non-viral DNA vectors are packaged into adenovirus capsids that are transfected into mammalian cells (Amalfitano & Parks, 2002). Therefore, it is important to have models that can predict whether vector packaging is feasible. As described above, such direct comparisons and rescaling based on the DNA filling density and capsid size has been made with phages λ and T5 (Castelnovo & Evilevitch, 2007). Furthermore, one can directly verify the effect of genome length on internal force by osmotic suppression studies on the same phage but with different genome lengths packaged into it. The effect of genome length was verified for phage λ mutants containing 48.5 (wild-type phage), 41.5 and 37.7 kb DNA (Grayson *et al.*, 2006). Ejection of DNA from all three λ mutants rapidly decreases up to ~50% ejected fraction and then slowly progresses to completion. It was evident that the length of ejected DNA of the shortest DNA λ mutant (37.7 kb) was always smaller than that of wild-type λ (48.5 kb) at every PEG concentration. The external osmotic pressure required to completely inhibit DNA ejection was found to be strongly dependent on the genome length. For λ (37.7 kb) it was found to be 10–15 atm while that for λ (48.5 kb) it was 20–25 atm. Thus, a 20% longer genome (37.7 kb vs 48.5 kb) doubles the pressure suppressing ejection. A nice agreement between these data and the theoretical model description (Kindt *et al.*, 2001; Purohit *et al.*, 2003b; Tzlil *et al.*, 2003) was found (Grayson *et al.*, 2006).

Considering that viral capsids are permeable to ions and realizing that electrostatic interactions between DNA strands is the major determining parameter of internal force, salt conditions become equally important, on a par with packaged genome length. The effect of mono-, di- and multivalent DNA counter ions on internal pressure has been investigated. The extent of ejection from phage λ at a fixed osmotic pressure of 4 atm (mimicking cytoplasmic pressure) has been measured as a function of concentration of salts with 50 mM Tris-HCl and 10 mM $MgSO_4$ as standard buffer present in all solutions (Evilevitch *et al.*, 2008a). While Na^+ ions had no effect on the fraction of DNA ejected, adding 30 mM Mg^{2+} ions halved the ejected DNA fraction (which initially was around 60%) and with 1 mM tetravalent spermine the ejection was nearly completely suppressed with only 5% DNA ejected (Evilevitch *et al.*, 2008a). This is explained by the fact that Mg^{2+} ions screen DNA–DNA repulsive interactions while spermine also induces attraction between DNA strands (Arscott *et al.*, 1990; Bloomfield, 1991; Raspaud *et al.*, 2005; Raspaud *et al.*, 1998; Schellman & Parthasarathy, 1984; Sikorav *et al.*, 1994; Suwalsky *et al.*, 1969), leading to a significant decrease in the internal ejecting force. In fact, cryo-electron micrographs have shown that a stable DNA toroid remains inside the capsid when λ is incubated with LamB receptor *in vitro* in the presence of 1 mM spermine (4+) and DNase I (Evilevitch, 2006) (see Fig. 3).

This cryo-electron microscopy (cryoEM) micrograph shows a relaxed DNA toroid inside the phage capsid when 60% of its genome has been ejected allowing the interaxial spacing between neighboring DNA strands in the capsid to be equal to that found in the a DNA toroid condensed by spermine in the bulk (Arscott *et al.*, 1990; Bloomfield, 1991; Raspaud *et al.*, 2005; Raspaud *et al.*, 1998; Schellman & Parthasarathy, 1984; Sikorav *et al.*, 1994; Suwalsky *et al.*, 1969). Interestingly, if DNase is not present in the sample, all of the DNA will be pulled out of the capsid by the larger DNA condensate (condensed by spermine) formed outside phage particles. This is similar to an Ostwald ripening process, with the larger toroid growing at the expense of the smaller one (Evilevitch, 2006). Coincidently, early studies found that infection of phage λ *in vivo* is inhibited in the presence of several polyamines (Harrison & Bode, 1975). The fact that spermine significantly decreases the DNA ejection force within the capsid and that this force is required to eject DNA into an osmotic pressure gradient of a few atmospheres in the cell can explain this data. It should be noted, however, that the presence of polyamines can also affect

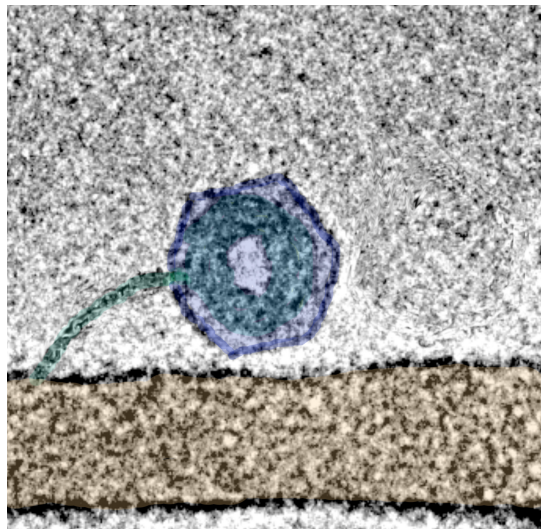


Fig. 3. Cryoelectron micrograph of toroidally condensed DNA remaining unstressed in the capsid (wild-type phage, 48.5 kbp), following ejection at zero "external" osmotic pressure in the presence of spermine. Adapted from: E. Nurmemmedov, M. Castelnovo, C.E. Catalano, A. Evilevitch. Biophysics of viral infectivity: Matching genome length with capsid size. *Q. Rev. Biophys.* 40(4):327–356 (2007). Copyright Cambridge University Press.

the transmembrane potential which has been shown to be important for phage T7 infection (Kemp *et al.*, 2004; Molineux, 2001).

Internal and external forces acting on viral DNA can also be probed by measuring the *rate* of DNA ejection. The rate of ejection *in vitro* from phage λ was determined using time-resolved dynamic light scattering as a function of DNA length and temperature (Lof *et al.*, 2007a; Lof *et al.*, 2007b). Comparing two genome length mutants (37.7 and 45.7 kbp) with wild-type λ (48.5 kbp), it was found that the initial ejection rate increases in almost direct relationship with the initially packaged genome length, while the total time to completely eject DNA was nearly constant for all the genome lengths. Assuming that the net force of ejection is $F = F_{eject} - fv$, where F_{eject} is the internal DNA ejection force, f is the friction coefficient and v is the rate of ejection, $F_{eject} - fv = 0$ at steady-state and the rate of ejection is therefore $v = F_{eject}/f$. Assuming that the friction coefficient is the same for all three phages (since capsid and tail dimensions, where friction occurs with the DNA, are constant for all three viruses), then the rate

of ejection will depend directly on F_{eject}, which has indeed been observed (Lof *et al.*, 2007b). This again confirms that initial DNA ejection from phage is entirely dependent on the internal DNA force. In accord with earlier observations of salt effects on internal genome force (Evilevitch *et al.*, 2004; Evilevitch *et al.*, 2008a), single molecule fluorescence measurements (Grayson *et al.*, 2007) of DNA ejection kinetics from λ also show strong ejection rate dependence on salt conditions. It was found that replacing 10 mM Na^+ with 10 mM Mg^{2+} ions increased ejection time nearly 10 fold, suggesting that internal force inside phage decreases by a factor of 2. Maximum measured DNA translocation rates for DNA ejection from λ were up to 60 kb/s (with 10 mM Na^+ ions in 10 mM Tris-HCl buffer), comparable to translocation rates found for phage T5 (lower bound of 75 kb/s) in an analogous single molecule fluorescence study (Mangenot *et al.*, 2005). Similar speeds were found for the opposite process of DNA packaging in ϕ29 (10 kb/s) (Rickgauer *et al.*, 2008). The kinetics of T5 DNA ejection was also investigated using static light scattering (de Frutos *et al.*, 2005b); however, direct comparison with monotonic ejection from λ cannot be made since step-wise ejection behavior was observed in T5 due to the presence of nicks in the DNA. While most of the genome is ejected by internal pressure *in vitro* in the absence of external osmotic pressure, the last unconstrained DNA part will remain in the tail for a long time before it will diffuse out, since it is not affected by DNA–DNA repulsion and bending forces (Evilevitch, 2006). This makes it difficult to determine the total time required for complete ejection in both single molecule and bulk measurements.

Current internal force models accurately account for changes in ejection rate associated with changes in genome length and salt conditions. They fail, however, to explain the temperature effect on ejection kinetics. While internal force predicts weak temperature dependence, we found a nearly exponential increase in ejection rate with increasing temperature (Lof *et al.*, 2007b). Comparing the ejection times for λ, we see that there is a nearly ten times rate difference, with most DNA released in 11 min at 15°C compared to 1 min at 37 °C. A suggested mechanism that explains such ejection behavior is due to the temperature sensitive conformational changes induced in the tail–receptor complex forming the pore (including the portal) through which DNA passes. It is likely that the tail-receptor portal can adopt at least two conformations — an 'open' conformation through which DNA can freely pass or a 'closed' conformation blocking

the DNA exit (*vide supra*). We propose that the probability of occurrence of the 'open' conformation increases with increasing temperature, keeping the portal-tail pore open longer and thus speeding up the ejection of DNA from phage. Such a mechanism will be coupled to an energy of activation. This also explains why no ejection occurs *in vitro* from λ at 4°C with receptor present (Lof *et al.*, 2007b). Similarly, DNA ejection from phage T5 is completed within minutes at 30–41°C, while at lower temperatures it takes hours to days for completion (de Frutos *et al.*, 2005b); the free energy of activation barrier for T5 was estimated to 2.9×10^{-19} J (de Frutos *et al.*, 2005b).

In contrast to the phages, the entire viral particle enters the host cell in many eukaryotic viruses. In many cases, however, the viral genome still needs to be ejected into the cell nucleolus (Ojala *et al.*, 2000). For some dsDNA eukaryotic viruses, like herpes simplex virus type 1 (HSV-1), it was shown that it injects its DNA into the nucleus in a manner analogous to bacteriophage DNA ejection into the host cytoplasm (Cardone *et al.*, 2007; Newcomb *et al.*, 2007). After the HSV-1 nucleocapsid has entered the host cell by a membrane fusion event, it is transported towards the nucleus where it docks to the nuclear pore and ejects its genome into the nucleus, leaving the empty capsid in the cytosol. Thus, HSV-1 strongly resembles many aspects of phage biology in that it also packages its DNA into a pre-assembled capsid using an ATP driven motor and ejects the genome through a portal located in one of its vertices (Cardone *et al.*, 2007; Newcomb *et al.*, 2007).

The experiments described above suggest that internal pressure resulting from DNA confinement within the capsid may provide a common mechanism for genome ejection in many dsDNA viruses, specifically those that package their genomes with an ATP driven motor (the small dsDNA viruses that spontaneously self-assemble a nucleocapsid e.g. polyoma virus, do not utilize a packaging motor and thus are not under pressure (Oppenheim & Peleg, 1989)). Thus the ability of these viruses to infect by injecting their genome into the cell or cell nucleus is determined by DNA internal force set by the packaged DNA density and capsid size. This force also strongly depends on ambient salt conditions. However, this pressure alone appears to be insufficient to deliver the entire genome into the cell. Once passive pressure-driven ejection step has ceased, cellular DNA-binding proteins and polyamines can complete internalization of the viral genome by ratcheting, adsorption and/or condensation mechanisms.

3. Capsid Strength

3.1. *Theoretical Considerations*

In this section, we introduce briefly the physical concepts underlying capsid stability. Viral capsids are formed by the self-assembly of one or several proteins (two in the case of phage λ excluding portal protein) to form a closed shell structure with high symmetry. As outlined by the pioneering work of Caspar and Klug (Caspar & Klug, 1962), most viral shells exhibit icosahedral symmetry composed of pentameric (i.e. five proteins) and hexameric (i.e. six proteins) structures at the shell's faces. Accordingly, viral capsids are classified using the so-called T-numbers: an icosahedral structure can be constructed from a hexagonal lattice by inserting pentagons at sites of the hexagonal lattice that are separated by h steps along one lattice direction and k steps along the other, with h and k non-negative integers. The integer $T = h^2 + k^2 + hk$ is then the number of inequivalent sites in the resulting shell, which is made of 12 pentagons and $10(T - 1)$ hexagons. So for instance, phage λ head is therefore classified as an icosahedral $T = 7$ capsid ($h = 2$ *and* $k = 1$) (Dokland & Murialdo, 1993).

From a physical point of view, the capsid must withstand several tens of atmospheres internal pressure due to packaged DNA, as discussed above. The interactions between proteins at the capsid surface are mediated by non-covalent hydrophobic, hydrophilic and electrostatic bonding interactions that provide the robustness and stability of the capsid. A variety of approaches have been taken to gain theoretical insight into the question of capsid stability (reviewed in Gibbons & Klug, 2007a).

First, capsids were investigated using Normal Mode Analysis starting with atomic information on capsid structure available from crystallographic data in the Protein Data Bank (Tama & Brooks, 2005). This method allows the main directions of motion of the whole capsid to be determined and has proven a useful method to analyze large-scale conformational changes in some plant viruses (e.g. CCMV, which is an RNA virus). However, this method does not allow precise quantification of capsid robustness. A second approach to investigate capsid stability is to start from a macroscopic continuous description of closed shells. We restrict ourselves in this section to the particular case of thin shells, which is effectively the most relevant one for interpretation of nanoindentation studies on capsids (see Gibbons & Klug, 2007b). In this approach, the capsid is described by two elastic constants characterizing the in-plane stretching of

the shell, and its bending out of plane. These elastic constants are conveniently combined into the Foppl-von Karman (FvK) dimensionless parameter, which for a spherical shell of radius R:

$$\gamma = \frac{ER^2}{\kappa}$$

where E is the 2D Young Modulus and κ is the out-of-plane bending modulus. This number represents the ratio between stretching and bending stiffness. Several works were devoted to the analysis of these continuous models with different geometries by varying the FvK number (Lidmar *et al.*, 2003; Nguyen *et al.*, 2005; Widom *et al.*, 2007). In particular, it was shown that FvK number controls the general morphology of the thin shell: at low FvK numbers, the shell is nearly spherical while for larger FvK numbers a significant faceting of the shell is observed. Furthermore, the models constitute a theoretical framework for the interpretation of capsid nano-indentation experiments, since it predicts a nearly linear relationship between force and indentation (Ivanovska *et al.*, 2007; Ivanovska *et al.*, 2004; Klug *et al.*, 2006; Michel *et al.*, 2006). More precisely, the effective spring constant associated with this linear behavior is given by:

$$k_{cap} = \frac{\alpha h^2 E}{R}$$

where h is the capsid wall thickness and α is a factor depending on indentation geometry.

Complementary insights on capsid stability are coming from the use of triangulated surfaces, as described by Buenemann *et al.* (Buenemann & Lenz, 2007). These authors reproduced a number of experimental observations using triangulated surfaces with elastic (stretching) and bending energy, where each capsomer is represented by several vertices. Moreover different behavior is observed when the indentation is performed on pentamers or hexamers, the former being softer than the latter. This weakness of pentamers in the icosahedral capsid is predicted to result in a buckling instability, where the main curvature of the shell has been inverted at the pentamers. This type of model allows prediction of the critical breaking force of empty and full capsids. It was theoretically found that pressurized capsids, like in DNA-filled capsids, increase the critical breaking force upon indentation, the pressurized genome helping to resist the imposed deformation. Finally, critical rupture by internal stress (no external force applied), as one provoked by osmotic shock, experiments

was also investigated. In such an experiment, the internal pressure of a capsid is sharply increased due to a sudden change in the environment, and forces larger than critical breaking force are generated. Interestingly, Buenemann *et al.* found that robustness of capsid as determined by critical rupture force correlates well with the presence of a pressurized or a non-pressurized genome.

3.2. *Measuring Elasticity and Strength of Viral Capsids*

Survival success of viruses hinges in their ability to take control over host's cellular machineries and to withstand physical conditions to which they are exposed. The latter is critically dependent on factors such as genome density and capsid strength (De Paepe & Taddei, 2006). More explicitly, safekeeping of viral genetic material, successful delivery of the genome into the host cell, and survival between infection events, during which viruses are susceptible to a variety of mechanical damage, are critical. As discussed above, phages (and quite likely many eukaryotic dsDNA viruses) are exposed to internal genome pressures of tens of atmospheres. While high internal pressure is required for successful genome injection into a host cell, it was also shown that viability of 12 different phages, defined by capsid rupture, was significantly increased with increasing density of packaged DNA (De Paepe & Taddei, 2006). Notwithstanding, phages in nature are well suited to host-to-host replication cycles by virtue of their capacity to withstand environmental stresses, including pH, temperature, osmotic pressure, salt concentration and mechanical shear forces. For example, *E. coli*, which is host to many tailed phages, are subjected to high shear forces in nature (Thomas *et al.*, 2002). Therefore, phages that in turn adsorb to the surface of *E. coli* cells during infection are also exposed to high external sheer forces (Ivanovska *et al.*, 2007). As discussed above, viral capsids are typically held together by non-covalent bonding interactions (hydrophobic, electrostatic, *etc.*) and the ensemble of interactions must withstand the internal capsid pressures and in addition a variety of external forces. All of these factors raise several conceptual questions:

- What are the mechanical properties of capsids and how do they depend on capsid dimensions?
- How is packaged genome length correlated with the capsid strength?

Capsids of many viruses can be self-assembled *in vitro* from their purified proteins (the morphologically defined units of which are also called 'capsomers'), given the proper buffer conditions (Caspar & Klug, 1962). So for instance, cowpea chlorotic mottle virus (CCMV, a plant virus), can self-assemble around its RNA genome without involvement of auxiliary factors (Bancroft *et al.*, 1968a; Bancroft *et al.*, 1968b; Finch & Bancroft, 1968; Hiebert *et al.*, 1968; Wagner & Bancroft, 1968). In contrast, many dsDNA viruses actively package their genomes into a pre-assembled capsid. In these cases, capsid assembly requires the transient association of chaperones, scaffolding proteins and maturation proteases (Dokland, 1999; Gaussier *et al.*, 2006; Hendrix *et al.*, 1983; Ptashne, 2004). Scaffolding proteins bind to capsid shell proteins inducing co-polymerization into an icosahedral protein shell. Regardless of the mode of assembly, the resultant capsids are comparable in structure (typically icosahedral symmetry) and in non-covalent capsomer–capsomer interactions. (Gan *et al.*, 2004; Gan *et al.*, 2006; Lata *et al.*, 2000; Lee *et al.*, 2004; Wikoff *et al.*, 2003; Wikoff *et al.*, 2006; Wikoff *et al.*, 1998; Wikoff *et al.*, 1999; Wikoff *et al.*, 2000). The assembly process is often followed by a maturation step (typically induced by DNA packaging), in which the capsid proteins undergo rearrangement (sometimes capsid proteins leave while new ones are added) resulting in capsid expansion and shell wall thinning.

The use of Atomic Force Microscopy (AFM) allows quantitative study of the mechanical properties of viral particles by manipulating individual capsids at the nanometer scale (Ivanovska *et al.*, 2004). In addition to providing high resolution imaging of viral particles in liquid (see Fig. 4a and b), the AFM tip can apply force on single viral particles in order to obtain real-time force-distance (F-z) curves as the nanometer size tip of the cantilever scans the specimen surface (Ivanovska *et al.*, 2004), see Fig. 4c. Probing phage capsids with an AFM tip generally results in two regimes: reversible deformations are usually observed when force below a certain value is applied, while irreversible rupturing occurs at higher applied force values (Ivanovska *et al.*, 2007; Ivanovska *et al.*, 2004). These studies provide three parameters describing the capsid mechanical properties: i) information about the capsid's mechanical limits and strength of capsomer–capsomer bonds is provided by the threshold force beyond which the interactions between capsid proteins collapse, referred to as the 'breaking force' F_{break}, ii) due to an essentially linear response to deformation, the spring constant, k, can be obtained by linear fit to the

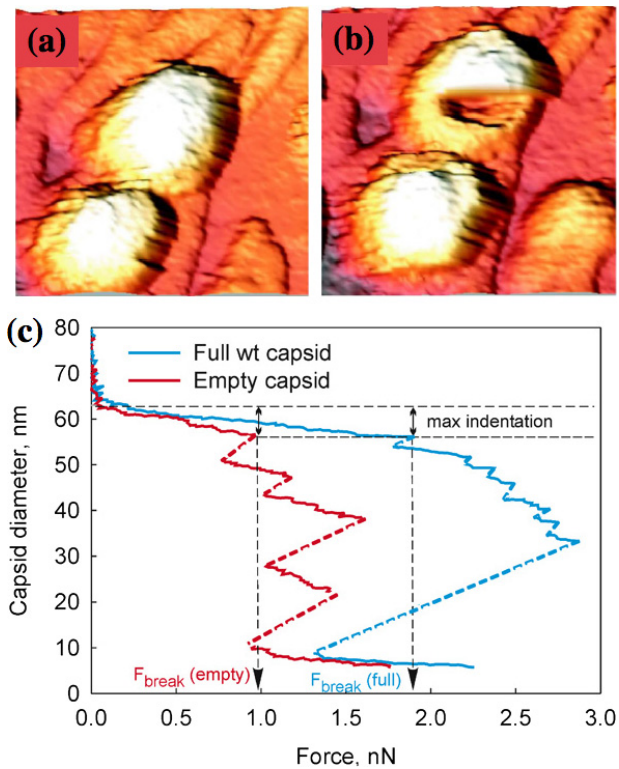


Fig. 4. High-resolution AFM imaging of λ phage capsids (3-D with shadow): (a) intact phage λ, (b) same phage as in panel (a) but after breakage induced by an AFM tip, (c) FZ curves of full wild-type versus empty phage λ. The data are displayed as observed capsid height vs. force. The deformation of the capsid is linear until a critical deformation is reached, after which a sudden decrease in capsid height can be observed. The dashed line is inserted to guide the eye, because there are no data recorded between these gaps. Adapted from Ivanovska *et al.* (2007).[60] Copyright 2007 National Academy of Sciences, U.S.A.

force-distance curves, iii) using a homogeneous thin-shell theory approximation (Ivanovska *et al.*, 2004) or finite element simulations (Klug *et al.*, 2006), Young's modulus, E, can also be obtained. The spring constant and Young's modulus are measures of a capsid's elasticity, and whilst k is dependent on the material geometry, E is an intrinsic geometry-independent material property.

AFM nanoindentation studies of empty viral capsids reveal unique mechanical properties for different classes of viruses, which can in turn

be directly related to a virion's replication cycle. For example, we have stressed throughout this review that the presence of the internal genome pressure in the complex dsDNA viruses, which is required for DNA ejection into the host cell, is distinct from many eukaryotic single stranded RNA (ssRNA) viruses, which do not appear to require internal pressure for genome release. Indeed, it was found that Young's modulus of the empty ϕ29 procapsid is ~1.8 GPa (Ivanovska *et al.*, 2004) (comparable to hard plastics) which is close to the value obtained for empty phage λ capsid, ~1 GPa (Ivanovska *et al.*, 2007). For comparison, Young's modulus for the ssRNA plant virus CCMV was measured to be 0.14 GPa (Michel *et al.*, 2006) and in the animal virus murine leukemia virus (MLV) it was 0.23 GPa (Kol *et al.*, 2006). The fact that Young's modulus for these two ssRNA viruses is one order of magnitude smaller than for dsDNA bacteriophages reflects the fact that the phage capsids are significantly more robust and can tolerate much higher internal force exerted by their genomes. At the same time, the maximum indentation is similar for all four viruses above, about 20–30% of capsid height. Above this indentation, the AFM cantilever tip causes irreversible capsid rupture with catastrophic drop of force. However, the force required to break an empty λ capsid is 0.8 nN (Ivanovska *et al.*, 2007), similar to the force found for CCMV (0.6 nN) (Michel *et al.*, 2006) but two times lower than the breaking force found for empty ϕ29 (~1.5 nN) (Ivanovska *et al.*, 2004). Taking into account capsid thickness, this provides us with information about capsid rigidity and capsomer–capsomer interactions. The ϕ29 procapsid, which has a thickness similar to that of λ (about 1.8 nm), is able to withstand higher force by being more elastic (also confirmed by measured spring constants: 0.31 N/m for ϕ29 and 0.13 N/m for λ). This can be related to the fact that λ has slightly higher DNA packaging density and thus higher internal pressure compared to ϕ29 (Purohit *et al.*, 2003b) and therefore needs to be more rigid in order to maintain its pressure constant without expansion.

Thus, AFM probing of viral shells reveals fundamental physical properties controlling the viral replication cycle. The recent discovery of genome pressures in several phages raises the question about the correlation between capsid strength and genome pressure. Many phages exhibit lower and upper packaged DNA length limits; for instance, it has been shown that phage λ can package only between 78 and 106% of its genome *in vitro* (where 100% corresponds to wild-type λ-DNA 48.5 kb) (Feiss *et al.*, 1977). With the help of AFM one can learn whether capsid strength

limits the extent to which a viral shell can be pressurized, which must in turn be correlated with the maximum force that the DNA packaging motor can exert. With jumping mode AFM, we have measured force-distance deformation of phage λ capsids as a function of packaged genome length for 78–100% of wild-type λ-DNA length (37.7, 45.7 and 48.5 kb DNA) (Ivanovska *et al.*, 2007). All capsids behaved elastically at small deformations, and spring constants could be obtained from the force-distance curve slopes. The maximum capsid breaking force was also determined. These data indicate that phages with wild-type DNA are twice as strong as shorter genome mutants, which behave like empty capsids ($F_{break} \sim 1.6$ nN *vs.* ~ 0.8 nN and $k \sim 0.23$ N/m *vs.* ~ 0.13 N/m), as shown in Fig. 4c. This is regardless of the high internal pressure of tens of atm present also in shorter DNA length phage mutants (Grayson *et al.*, 2006). This is due to the fact that DNA-hydrating water molecules exert an osmotic pressure inside capsids that increases exponentially when the packaged DNA density is close to wild-type phage. This osmotic pressure raises the wild-type DNA capsid strength and is approximately equal to the maximum breaking force of empty shells. This result confirms that the strength of the shells limits the maximal packaged genome length.

Upon AFM tip induced deformation of the capsid, water molecules hydrating tightly packed DNA are displaced through the capsid pores. The DNA hydration force is the force required to displace water molecules from DNA out to the bulk solution and is described by the osmotic pressure. Using DNA osmotic pressure data (Parsegian *et al.*, 1986; Rau & Parsegian, 1992; Strey *et al.*, 1998) it is evident that the osmotic pressure inside the capsid is sufficient to provide extra strength against external deformation only at 100% wild-type DNA packaging density, analogous to the osmotic pressure in biological cells. Thus it is not the DNA itself but water molecules hydrating it that provide significant contribution to the force resisting capsid deformation and breaking. Consistently, if the DNA hydration forces are decreased by incubating wild-type DNA (wt-DNA) phage λ in higher salt concentration, the capsid changes its mechanical properties and behaves like an empty capsid (Ivanovska *et al.*, 2007). This demonstration provides a stringent test of the hydration force interpretation above. We have proposed that such evolutionary energy optimization presumably determines the genome length of wt-DNA phages, since they can survive twice the external mechanical stress compared to its shorter genome mutants that are abundant in nature (Ivanovska *et al.*, 2007). As

mentioned above, external mechanical stresses comparable to the stress induced on the capsid by the AFM cantilever is present in the natural phage environment due to the fact that its host (*E. coli*) forms colonies under high shear flows (Thomas *et al.*, 2002). Thus, the capsid's strength is related to the phage's ability to infect (by optimizing genome packaging density that must provide sufficient pressure) and also survive internal and external stresses to which it is exposed between infection events.

Similar to phage λ, the genome-filled capsids of CCMV virus (Michel *et al.*, 2006) and ssDNA minute virus of mice (MVM) virus (Carrasco *et al.*, 2006) have greater resistance to breaking with an AFM tip, indicating greater capsid strength in comparison to empty capsids. However, the physical basis for this behavior may differ from that found in λ since these viruses do not have internal pressure and the genome hydration force is therefore low.

Internal force from higher genome packaging density in phages leads to increased lateral stress on the capsid walls due to the large outward pressure. Because of the thermal fluctuations of capsid proteins, this results in more frequent rupture of the capsids in direct correlation with the genome density (De Paepe & Taddei, 2006). In contrast, osmotic outward pressure also caused by high genome confinement stabilizes phage particles against external deformations, as discussed above. Thus, in nature, there is precise balance of the forces acting externally and internally on a viral particle that are determined by the viral genome length and capsid strength, that are in turn controlled by salt ions in the viral milieu.

4. Concluding Remarks

The majority of viruses possess spherical, icosahedral protein shells with radii varying between 10 and 100 nm and with thicknesses of a few nanometers corresponding to a single protein layer. Viral capsids protect genomes that are tens of microns in contour length. Complete genome encapsidation implies that the virus must be stable enough to withstand the internal forces exerted by its packaged genome and external forces from its environment. Yet, it must be unstable enough to rapidly release its genome in the cell during infection. Thus, there must exist a unique match between the viral genome length and capsid size and strength that is adjusted to the biological and physical properties of the host cell. An understanding of the fundamental physical principles that control viral

encapsidation and genome release provides tools for the rational design of therapeutic agents that selectively interfere with the encapsidation process (e.g. charged molecules that specifically bind to viral capsid proteins and slightly change the interior charge of the capsid will interfere directly with the genome charge and thus affect the internal pressure), and in addition tools to improve encapsidation *in vitro* in order to make stable vectors for gene delivery.

There are two main pathways of viral assembly. In the first case, viral capsid proteins spontaneously self-assemble around the genome to make an infectious particle. Electrostatic interactions between capsid proteins and genome have been shown to play a central role in the self-assembly process, which can therefore be controlled by ionic strength and pH (Kegel & Schoot Pv, 2004). In the second case exemplified by the herpesviruses and many bacteriophages, the capsid proteins first assemble into an empty procapsid and the viral genome is then actively packaged into the capsid by a terminase motor complex fuelled by ATP hydrolysis. In both cases, a precise match between the capsid size and the genome length is of great importance for efficient encapsidation. In this review we have focused on the latter genome encapsidation process, with dsDNA bacteriophage as a main experimental system.

The primary purpose of this review is to describe biophysical approaches to learn about the mechanisms of viral infectivity, specifically focusing on the observation of high internal capsid pressures due to strongly confined DNA. Phage provide ideal experimental systems to define the correlation between genome and capsid dimensions since they can be genetically modified and assembled both *in vivo* and *in vitro* to form mutants with varying capsid size (e.g. phage T4) as well as packaged DNA length (e.g. phage λ) (Hendrix *et al.*, 1983; Ptashne, 2004). Driven mainly by physical models (see Section 2.1), experimental studies have combined single molecule measurements, which allow imaging and manipulation at nanometer resolutions, with bulk measurements that provide direct comparison between different phages and between wild-type vs. mutant viruses, to study the systems in great detail.

The studies described herein illustrate key physical requirements for viral infectivity. Internal genome pressure is required for phage and many other dsDNA viruses to be able to infect by passive ejection of its genome. The same pressure also provides additional support to the strength of viral capsid helping the virus survive external deformations imposed on it

between infections. Our aim was to show that three parameters — genome length, capsid size and capsid strength — are precisely correlated in order to produce an infectious virion. Experiments and theoretical models made for viruses with distinctly different dimensions suggest that generalizations of genome packaging and release mechanisms can be made for whole classes of viruses. By answering the question — *what determines the upper and lower limit for an encapsidated genome?* — one learns about a virus's replication cycle and its host.

So for instance, the phage λ capsid can be filled with DNA lengths within the range of 78–106% of its wild-type DNA length and remain infectious (Feiss *et al.*, 1977). If the genome is shorter than 78%, DNA is still packaged but the phage is non-infectious. Conversely, genomes over 106% DNA are not packaged. The fact that DNA ejection is driven by internal pressure can in part explain this observation. If DNA is too short the internal pressure is too low and there is not enough force to eject sufficient DNA length in the cell against the osmotic pressure of its cytoplasm, in order to initiate the infection. If on the other hand the DNA is too long, the force required to package the entire duplex exceeds the force that can be generated by the DNA packaging motor and/or the force that the capsid can withstand, as we have shown with AFM (Ivanovska *et al.*, 2007).

Yet the volume fraction of nucleic acid is not the only determining parameter in the equilibrium DNA packaging energetics. Although the dimensions of phages T5 and λ are different (capsid size inner radii $R_{T5} = 39$ nm vs $R_\lambda = 27.5$ nm and wild-type DNA length $L_{T5} = 121.75$ kb *vs.* $L_\lambda = 48.5$ kb), their DNA volume filling fractions are similar (T5 = 0.52 and λ = 0.59). Thus, one would expect similar forces ejecting DNA from these phages. However, osmotic suppression measurements on phages T5 and λ, showed that the force driving DNA ejection is significantly lower in T5 than in λ and the osmotic pressure required to completely inhibit ejection is 7 *vs.* 15 atm (Castelnovo & Evilevitch, 2007). Therefore, at least one more relevant variable is necessary to describe equilibrium DNA packaging energetics. In the framework of the simple model that we derived (Castelnovo & Evilevitch, 2007), this variable is the characteristic ejection force amplitude $F = kTl_p/2R^2$, which can be modulated by the capsid size R considering that DNA persistence length l_p is unchanged. Rescaling internal force in phages with these two parameters (genome density and force amplitude) provides a master plot describing ejected genome length

versus osmotic pressure required to stop the ejection. This demonstrates how a general model can be applied to successfully describe energetics in different viruses.

According to this model, smaller phages are typically expected to require higher osmotic pressure in order to inhibit completely the ejection of their DNA. The difference in ejecting force for the fully packaged bacteriophages T5 and λ($F_{T5} \times 3$ pN *vs.* $F_{\lambda} \times 7$ pN) is likely to reflect different maximal capsid strength, and also different mechanisms of DNA transport into the host. In the case of phage λ, the build-up of large internal stress upon packaging as compared to phage T5 requires a stronger capsid. Similarly, the large internal stress allows the phage λ to rely mostly on passive ejection for DNA to be transported into the host. In the case of the phage T5, the genome is longer and more complex than in the phage λ. Nevertheless its large capsid size allows it to have less elastic energy stored within the DNA condensate. This lower energy is compensated by the larger genetic information encoded into T5 DNA that allows it to have both passive and 'active' DNA transport into the host. Here, it should be noted however, that internal force is also strongly dependent on the salt conditions due to the capsid permeability to water and smaller ions (Evilevitch *et al.*, 2008a). Different phages require different buffer conditions in order to stay stable, which mirrors differences in their natural environment. Therefore, osmotic suppression measurements on T5 and λ were also performed in slightly different buffers, which can in turn affect the observed differences in forces.

Thus, both capsid size and strength limit the extent to which a capsid can be pressurized. Capsid walls are of similar thickness for most viruses (between 2 and 4 nm) and since lateral stress of an elastic shell is directly proportional to its radius, capsid walls should indeed become weaker as the capsid radius is increased. This suggests an explanation for why the smaller λ capsid can withstand a higher internal pressure that that of T5. Notwithstanding, the DNA packaging density in T5 is almost as high as in λ and it is likely that additional geometrical constraints are associated with the packaging of DNA inside the capsid; for instance, T5 DNA has nicks which are absent in λ genome. It is noteworthy that while the capsid size for phage T4 can vary widely, packaging its DNA follows a head full packaging mechanism, which always results in the same DNA packaging density and suggests similar internal pressures in all mutants (Earnshaw *et al.*, 1978).

An additional consideration to understanding the consequences of optimizing viral dimensions and strength appears when one asks why genetically modified phage mutants are infectious but not abundant in nature, as for example shorter genome mutants of λ. We provide a feasible explanation to this evolutionary development by comparing the mechanical properties of shorter genome λ mutants with empty λ capsids and wild-type DNA phage λ measured by AFM (Ivanovska *et al.*, 2007). Since capsids are permeable to water, it is important to consider hydration force of tightly packed DNA not only to describe internal DNA interactions and force but also since water molecules are displaced from the capsid when it is deformed with an AFM tip. As discussed above, the striking feature of λ is that *only* full-length wild-type λ-DNA provides sufficient osmotic pressure to additionally support the capsid against external deformation induced by the AFM tip with forces comparable to stresses that the phage experiences in nature (Ivanovska *et al.*, 2007; Thomas *et al.*, 2002)). This makes wild-type phage λ nearly two times stronger than any of the its partially filled mutants, which have much lower survival rates in nature.

In conclusion, internal genome pressure is one of the key determining factors for phage infectivity and capsid stability. The future challenges for biophysical virology will be to develop systematic biophysical studies that will test and verify analytical and numerical models for capsid assembly, structure, stability and function. Further, these studies must be extended to the eukaryotic dsDNA viruses, the physical dimensions of which suggest that genome pressure is also a crucial factor for their development.

Acknowledgements

We would like to thank Elmar Nurmemmedov and Carlos E. Catalano for help with editing and valuable advice and discussions.

References

1. M. Alcorlo, V. Gonzalez-Huici, J.M. Hermoso, W.J. Meijer and M. Salas. The phage phi29 membrane protein p16.7, involved in DNA replication, is required for efficient ejection of the viral genome. *Journal of Bacteriology* **189**:5542–5549 (2007).
2. I. Ali, D. Marenduzzo and J.M. Yeomans. Polymer packaging and ejection in viral capsids: shape matters. *Physical Review Letters* **96**:208102 (2006).

3. A. Amalfitano and R.J. Parks. Separating fact from fiction: assessing the potential of modified adenovirus vectors for use in human gene therapy. *Current Gene Therapy* **2**: 111–133 (2002).
4. P.G. Arscott, A Z. Li and V.A. Bloomfield. Condensation of DNA by trivalent cations. 1. Effects of DNA length and topology on the size and shape of condensed particles. *Biopolymers* **30**:619–630 (1990).
5. J. Arsuaga, R.K. Tan, M. Vazquez and D.W. Sumners and S.C. Harvey. Investigation of viral DNA packaging using molecular mechanics models. *Biophysical Chemistry* **101–102**:475–484 (2002).
6. J.B. Bancroft, E. Hiebert, M.W. Rees and R. Markham. Properties of cowpea chlorotic mottle virus, its protein and nucleic acid. *Virology* **34**:224–239 (1968a).
7. J.B. Bancroft, G.W. Wagner and C.E. Bracker. The self-assembly of a nucleic-acid free pseudo-top component for a small spherical virus. *Virology* **36**:146–149 (1968b).
8. L.W. Black, W.W. Newcomb, J.W. Boring and J.C. Brown. Ion etching bacteriophage T4: support for a spiral-fold model of packaged DNA. *Proceedings of the National Academy of Sciences USA* **82**:7960–7964 (1985).
9. V.A. Bloomfield. Condensation of DNA by multivalent cations: considerations on mechanism. *Biopolymers* **31**: 1471–1481 (1991).
10. V.A. Bloomfield. DNA condensation by multivalent cations. *Biopolymers* **44**:269–282 (1997).
11. P. Boulanger, M. Le Maire, M. Bonhivers, S. Dubois, M. Desmadril and L. Letellier. Purification and structural and functional characterization of FhuA, a transporter of the Escherichia coli outer membranel. *Biochemistry* **35**:14216–14224 (1996).
12. D. Brockmann, L. Hufnagel and T. Geisel. The scaling laws of human travel. *Nature* **439**:7075, 462–465 (2006).
13. M. Buenemann and P. Lenz. Mechanical limits of viral capsids. *Proceedings of the National Academy of Sciences USA* **104**: 9925–9930 (2007).
14. G. Cardone, D.C. Winkler, B.L. Trus, N. Cheng, J.E. Heuser, W.W. Newcomb, J.C. Brown and A.C. Steven. Visualization of the herpes simplex virus portal in situ by cryo-electron tomography. *Virology* **361**:426–434 (2007).
15. C. Carrasco, A. Carreira, I.A. Schaap, P.A. Serena, J. Gomez-Herrero, M.G. Mateu, and P.J. De Pablo. DNA-mediated anisotropic mechanical reinforcement of a virus. *Proceedings of the National Academy of Sciences USA* **103**:13706–13711 (2006).
16. D.L. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbour Symposia on Quantitative Biology* **27**:1–24 (1962).
17. M. Castelnovo, R.K. Bowles, H. Reiss and W.M. Gelbart. Osmotic force resisting chain insertion in a colloidal suspension. *The European Physical Journal E Soft Matter* **10**:191–197 (2003).
18. M. Castelnovo and A. Evilevitch. DNA ejection from bacteriophage: towards a general behavior for osmotic-suppression experiments. *The European Physical Journal E Soft Matter* **24**:9–18 (2007).

19. C.E. Catalano. Viral Genome Packaging Machines: Genetics, Structure and Mechanism: Landes Bioscience/Eurekah.com (Georgetown, TX, USA) Kluwer Academic/Plenum Publishers (New York, NY, USA) (2005).
20. M.E. Cerritelli, N. Cheng, A.H. Rosenberg, C.E. Mcpherson, F.P. Booy and A.C Steven. Encapsidated conformation of bacteriophage T7 DNA. *Cell* **91**:271–280 (1997).
21. M. De Frutos, S. Brasiles, P. Tavares and E. Raspaud, Effect of spermine and DNase on DNA release from bacteriophage T5, *The European Physical Journal E Soft Matter* **17**:429–434 (2005a).
22. M. De Frutos, L. Letellier and E. Raspaud. DNA ejection from bacteriophage T5: analysis of the kinetics and energetics. *Biophysical Journal* **88**:1364–1370 (2005b).
23. M. De Paepe and F. Taddei. Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages. *PLoS Biol* **4**:e193 (2006).
24. T. Dokland. Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences* **56**:580–603 (1999).
25. T. Dokland and H. Murialdo. Structural transitions during maturation of bacteriophage lambda capsids. *Journal of Molecular Biology* **233**:682–694 (1993).
26. W. Earnshaw, S. Casjens and S.C. Harrison. Assembly of the head of bacteriophage P22: x-ray diffraction from heads, proheads and related structures. *Journal of Molecular Biology* **104**:387–410 (1976).
27. W.C. Earnshaw and S.R. Casjens, DNA packaging by the double-stranded DNA bacteriophages, *Cell* **21**: 319–331 (1980).
28. W.C. Earnshaw and S.C. Harrison. DNA arrangement in isometric phage heads. *Nature* **268**:598–602 (1977).
29. W.C. Earnshaw, J. King, S.C. Harrison and F.A. Eiserling. The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages. *Cell* **14**:559–568 (1978).
30. H. Echols. *Operators and promoters: The story of molecular biology and its creators.* Berkeley, CA: University of California Press (2001).
31. A. Evilevitch. Effects of condensing agent and nuclease on the extent of ejection from phage lambda. *Journal of Physical Chemistry B* **110**:22261–22265 (2006).
32. A. Evilevitch, M. Castelnovo, C.M. Knobler and W.M. Gelbart. Measuring the force ejecting DNA from phage. *Journal of Physical Chemistry B* **108**:6838–6843 (2004).
33. A. Evilevitch, L.T. Fang, A.M. Yoffe, M. Castelnovo, D.C. Rau, V.A. Parsegian, W.M. Gelbart and C.M. Knobler. Effects of salt concentrations and bending energy on the extent of ejection of phage genomes. *Biophysical Journal* **94**:1110–1120 (2008a).
34. A. Evilevitch, J.W. Gober, M. Phillips, C.M. Knobler and W.M. Gelbart. Measurements of DNA lengths remaining in a viral capsid after osmotically suppressed partial ejection. *Biophysical Journal* **88**:751–756 (2005).
35. A. Evilevitch, L. Lavelle, C.M. Knobler, E. Raspaud and W.M. Gelbart. Osmotic pressure inhibition of DNA ejection from phage. *Proceedings of the National Academy of Sciences USA* **100**:9292–9295 (2003).

36. M. Feiss, R.A. Fisher, M.A. Crayton and C. Egner. Packaging of the bacteriophage lambda chromosome: effect of chromosome length. *Virology* **77**:281–293 (1977).
37. J.T. Finch and J.B. Bancroft. Structure of the reaggregted protein shells of 2 spherical viruses. *Nature* **220**:815–816 (1968).
38. S.J. Flint, L.W. Enquist, V.R. Racaniello and A.M. Skalka. Assembly, Exit, and Maturation. In *Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses* (ed. S. J. Flint), pp. 451–491. Washington D.C., USA: ASM Press (2004a).
39. S.J. Flint, L.W. Enquist, V.R. Racaniello and A.M. Skalka. Attachment and Entry. In *Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses* (ed. S. J. Flint), pp. 127–180. Washington D.C., USA: ASM Press (2004b).
40. S.J. Flint, L.W. Enquist, V.R. Racaniello and A.M. Skalka. Prevention and Control of viral Diseases. In *Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses* (ed. S. J. Flint), pp. 703–757. Washington D.C., USA: ASM Press (2004c).
41. C. Forrey and M. Muthukumar. Langevin dynamics simulations of genome packing in bacteriophage. *Biophysical Journal* **91**: 25–41 (2006).
42. D.N. Fuller, D.M. Raymer, J.P. Rickgauer, R.M. Robertson, C.E. Catalano, D.L. Anderson, S. Grimes and D.E. Smith. Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *Journal of Molecular Biology* **373**: 1113–1122 (2007).
43. I.S. Gabashvili and A. Grosberg. [Bacteriophage DNA reptation]. *Biofizika* **36**: 788–793 (1991).
44. I.S. Gabashvili and A. Grosberg. Dynamics of double stranded DNA reptation from bacteriophage. *Journal of Biomolecular Structure and Dynamics* **9**:911–920 (1992).
45. L. Gan, J.F. Conway, B.A. Firek, N. Cheng, R.W. Hendrix, A.C. Steven, J.E. Johnson and R.L. Duda. Control of crosslinking by quaternary structure changes during bacteriophage HK97 maturation. *Molecular Cell* **145**:559–569 (2004).
46. L. Gan, J.A. Speir, J.F. Conway, G. Lander, N. Cheng, B.A. Firek, R.W. Hendrix, R.L. Duda, L. Liljas and J.E. Johnson. Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure* **14**:1655–1665 (2006).
47. H. Gaussier, Q. Yang and C.E. Catalano. Building a virus from scratch: assembly of an infectious virus using purified components in a rigorously defined biochemical assay system. *Journal of Molecular Biology* **357**:1154–1166 (2006).
48. M.M. Gibbons and W.S. Klug. Mechanical modeling of viral capsids. *Journal of Materials Science* **42**:8995 (2007a).
49. M.M. Gibbons and W.S. Klug. Nonlinear finite-element analysis of nanoindentation of viral capsids. *Physical Review E* **75**:031901 (2007b).
50. V. Gonzalez-Huici, M. Salas and J.M. Hermoso. The push-pull mechanism of bacteriophage O29 DNA injection. *Molecular Microbiology* **52**:529–540 (2004).

51. P. Grayson, A. Evilevitch, M.M. Inamdar, P.K. Purohit, W.M. Gelbart, C.M. Knobler and R. Phillips. The effect of genome length on ejection forces in bacteriophage lambda. *Virology* **348**:430–436 (2006).
52. P. Grayson, L. Han, T. Winther and R. Phillips. Real-time observations of single bacteriophage lambda DNA ejections in vitro. *Proceedings of the National Academy of Sciences USA* **104**:14652–14657 (2007).
53. D.P. Harrison and V.C. Bode. Putrescine and certain polyamines can inhibit DNA injection from bacteriophage lambda. *Journal of Molecular Biology* **96**:461–470 (1975).
54. S.C. Harrison. Packaging of DNA into bacteriophage heads: a model. *Journal of Molecular Biology* **171**:577–580 (1983).
55. R.W. Hendrix. Symmetry mismatch and DNA packaging in large bacteriophages. *Proceedings of the National Academy of Sciences USA* **75**:4779–4783 (1978).
56. R.X. Hendrix, J.W. Roberts, F.W. Stahl and R.A. Weisberg. *Lambda II*. New York: Cold Spring Harbour Laboratory (1983).
57. A.D. Hershey and M. Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology* **36**:39–56 (1952).
58. E. Hiebert, J.B. Bancroft and C.E. Bracker. The assembly in vitro of some small spherical viruses, hybrid viruses, and other nucleoproteins. *Virology* **34**:492–508 (1968).
59. M.M. Inamdar, W.M. Gelbart and R. Phillips. Dynamics of DNA ejection from bacteriophage. *Biophysical Journal* **91**:411–420 (2006).
60. I. Ivanovska, G. Wuite, B. Jonsson and A. Evilevitch. Internal DNA pressure modifies stability of WT phage. *Proceedings of the National Academy of Sciences USA* **104**:9603–9608 (2007).
61. I.L. Ivanovska, P.J. De Pablo, B. Ibarra, G. Sgalari, F.C. Mackintosh, J.L. Carrascosa, C.F. Schmidt and G.J. Wuite. Bacteriophage capsids: tough nanoshells with complex elastic properties. *Proceedings of the National Academy of Sciences USA* **101**:7600–7605 (2004).
62. W.K. Kegel and P. Schoot Pv. Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophysical Journal* **86**:3905–3913 (2004).
63. P. Kemp, M. Gupta and I.J. Molineux. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Molecular Microbiology* **53**:1251–1265 (2004).
64. J. Kindt, S. Tzlil, A. Ben-Shaul and W.M. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proceedings of the National Academy of Sciences USA* **98**:13671–13674 (2001).
65. W.S. Klug, R.F. Bruinsma, J.P. Michel, C.M. Knobler, I.L. Ivanovska, C.F. Schmidt and G.J. Wuite. Failure of viral shells. *Physical Review Letters* **97**:228101 (2006).
66. W.S. Klug and M. Ortiz. A director field model of DNA packaging in viral capsids. *Journal of the Mechanics and Physics of Solids* **51**:1815–1847 (2003).

67. N. Kol, M. Gladnikoff, D. Barlam, R Z. Shneck, A. Rein and I. Rousso. Mechanical properties of murine leukemia virus particles: effect of maturation. *Biophysical Journal* **91**:767–774 (2006).
68. D.J. Kushner. Self-assembly of biological structures. *Bacteriological Reviews* **33**:302–345 (1969).
69. J.C. Lamarque, T.V. Le and S.C. Harvey. Packaging double-helical DNA into viral capsids. *Biopolymers* **73**:348–355 (2004).
70. Y.T. Lanni. First-step-transfer deoxyribonucleic acid of bacteriophage T5. *Bacteriological Reviews* **32**:227–242 (1968).
71. R. Lata, J.F. Conway, N. Cheng, R.L. Duda, R.W. Hendrix, W.R. Wikoff, J.E. Johnson, H. Tsuruta and A.C. Steven. Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell* **100**:253–263 (2000).
72. K.K. Lee, L. Gan, H. Tsuruta, R.W. Hendrix, R.L. Duda and J.E. Johnson. Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid. *Journal of Molecular Biology* **340**:419–433 (2004).
73. J. Lidmar, L. Mirny and D.R. Nelson. Virus shapes and buckling transitions in spherical shells. *Physical Review E* **68**:051910 (2003).
74. C.R. Locker and S.C. Harvey. A model for viral genome packing. *Multiscale Modeling and Simulation* **5**:1264–1279 (2006).
75. D. Lof, K. Schillen, B. Jonsson and A. Evilevitch. Dynamic and static light scattering analysis of DNA ejection from the phage lambda. *Physical Review E Statistical Nonlinear and Soft Matter Physics* **76**:011914 (2007a).
76. D. Lof, K. Schillen, B. Jonsson and A. Evilevitch. Forces controlling the rate of DNA ejection from phage lambda. *Journal of Molecular Biology* **368**:55–65 (2007b).
77. S. Mangenot, M. Hochrein, J. Radler and L. Letellier. Real-time imaging of DNA ejection from single phage particles. *Current Biology* **15**:430–435 (2005).
78. J.P. Michel, I.L. Ivanovska, M.M. Gibbons, W.S. Klug, C.M. Knobler, G.J. Wuite and C.F. Schmidt. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proceedings of the National Academy of Sciences USA* **103**:6184–6189 (2006).
79. I.J. Molineux. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular Microbiology* **40**:1–8 (2001).
80. F. Neidhardt. *Escherichia coli* and *Salmonella typhimirium*. p. 1211: ASM Press (1996).
81. W.W. Newcomb, F.P. Booy and J.C. Brown. Uncoating the herpes simplex virus genome. *Journal of Molecular Biology* **370**:633–642 (2007).
82. T.T. Nguyen, R.F. Bruinsma and W.M. Gelbart. Elasticity theory and shape transitions of viral shells. *Physical Review E* **72**:051923 (2005).
83. T. Odijk. Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophysical Journal* **75**:1223–1227 (1998).
84. P.M. Ojala, B. Sodeik, M.W. Ebersold, U. Kutay and A. Helenius. Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. *Molecular and Cellular Biology* **20**:4922–4931 (2000).

85. A. Oppenheim and A. Peleg. Helpers for efficient encapsidation of SV40 pseudovirions. *Gene* **77**:79–86 (1989).
86. V.A. Parsegian, R.P. Rand, N.L. Fuller and D.C. Rau. Osmotic stress for the direct measurement of intermolecular forces. *Methods in Enzymology* **127**:400–416 (1986).
87. A.S. Petrov, M.B. Boz and S.C. Harvey. The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. *Journal of Structural Biology* **160**:241–248 (2007a).
88. A.S. Petrov, K. Lim-Hing and S.C. Harvey. Packaging of DNA by bacteriophage epsilon15: structure, forces, and thermodynamics. *Structure* **15**:807–812 (2007b).
89. M. Ptashne. *Genetic Switch: Phage Lambda Revisited*. New York: Cold Spring Harbour Laboratory (2004).
90. P.K. Purohit, M.M. Inamdar, P.D. Grayson, T.M. Squires, J. Kondev and R. Phillips. Forces during bacteriophage DNA packaging and ejection. *Biophysical Journal* **88**:851–866 (2005).
91. P.K. Purohit, J. Kondev and M. Phillips. Force steps during viral DNA packaging? *Journal of the Mechanics and Physics of Solids* **51**:2239–2257 (2003a).
92. P.K. Purohit, J. Kondev and R. Phillips. Mechanics of DNA packaging in viruses. *Proceedings of the National Academy of Sciences USA* **100**:3173–3178 (2003b).
93. E. Raspaud, D. Durand and F. Livolant. Interhelical spacing in liquid crystalline spermine and spermidine-DNA precipitates. *Biophysical Journal* **88**:392–403 (2005).
94. E. Raspaud, M. Olvera De La Cruz, J.L. Sikorav and F. Livolant. Precipitation of DNA by polyamines: a polyelectrolyte behavior. *Biophysical Journal* **74**:381–393 (1998).
95. D.C. Rau and V.A. Parsegian. Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. *Biophysical Journal* **61**:246–259 (1992).
96. K.E. Richards, R.C. Williams and R. Calendar. Mode of DNA packing within bacteriophage heads. *Journal of Molecular Biology* **78**:255–259 (1973).
97. J.P. Rickgauer, D.N. Fuller, S. Grimes, P.J. Jardine, D.L. Anderson and D.E. Smith. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage {phi}29. *Biophysical Journal* **94**:159–167 (2008).
98. S.C. Riemer and V.A. Bloomfield. Packaging of DNA in bacteriophage heads: some considerations on energetics. *Biopolymers* **17**:785–794 (1978).
99. M. Roa and D. Scandella. Multiple steps during the interaction between coliphage lambda and its receptor protein in vitro. *Virology* **72**:182–194 (1976).
100. C. Sao-Jose, S. Lhuillier, R. Lurz, R. Melki, J. Lepault, M.A. Santos and P. Tavares. The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. *Journal of Biological Chemistry* **281**:11464–11470 (2006).
101. J.A. Schellman and N. Parthasarathy. X-ray diffraction studies on cation-collapsed DNA. *Journal of Molecular Biology* **175**:313–329 (1984).

102. P. Serwer. Arrangement of double-stranded DNA packaged in bacteriophage capsids. An alternative model. *Journal of Molecular Biology* **190**:509–512 (1986).
103. J.L. Sikorav, J. Pelta and F. Livolant. A liquid crystalline phase in spermidine-condensed DNA. *Biophysical Journal* **67**:1387–1392 (1994).
104. D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson and C. Bustamante. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413**:748–752 (2001).
105. A.J. Spakowitz and Z.G. Wang. DNA packaging in bacteriophage: is twist important? *Biophysical Journal* **88**:3912–3923 (2005).
106. H.H. Strey, R. Podgornik, D.C. Rau and V.A. Parsegian. DNA–DNA interactions. *Current Opinion in Structural Biology* **8**:309–313 (1998).
107. M. Sun, D. Louie and P. Serwer. Single-event analysis of the packaging of bacteriophage T7 DNA concatemers in vitro. *Biophysical Journal* **77**:1627–1637 (1999).
108. M. Suwalsky, W. Traub, U. Shmueli and J.A. Subirana. An X-ray study of the interaction of DNA with spermine. *Journal of Molecular Biology* **42**:363–373 (1969).
109. F. Tama and C.L. Brooks, 3rd. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *Journal of Molecular Biology* **345**:299–314 (2005).
110. W.E. Thomas, E. Trintchina, M. Forero, V. Vogel and E.V. Sokurenko. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**:913–923 (2002).
111. S. Tzlil, J.T. Kindt, W.M. Gelbart and A. Ben-Shaul. Forces and pressures in DNA packaging and release from viral capsids. *Biophysical Journal* **84**:1616–1627 (2003).
112. G.W. Wagner and J.B. Bancroft. The self-assembly of spherical viruses with mixed coat proteins. *Virology* **34**:748–756 (1968).
113. W.H.O. Who. The world health report 2007: A safer future: Global public health security in the 21st century (2007).
114. M. Widom, J. Lidmar and D.R. Nelson. Soft modes near the buckling transition of icosahedral shells. *Physical Review E* **76**:031911 (2007).
115. W.R. Wikoff, Z. Che, R.L. Duda, R.W. Hendrix and J.E. Johnson. Crystallization and preliminary analysis of a dsDNA bacteriophage capsid intermediate: Prohead II of HK97. *Acta Crystallographica D Biological Crystallography* **59**:2060–2064 (2003).
116. W.R. Wikoff, J.F. Conway, J. Tang, K.K. Lee, L. Gan, N. Cheng, R.L. Duda, R.W. Hendrix, A.C. Steven and J.E. Johnson. Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. *Journal of Structural Biology* **153**:300–306 (2006).
117. W.R. Wikoff, R.L. Duda, R.W. Hendrix and J.E. Johnson. Crystallization and preliminary X-ray analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Virology* **243**:113–118 (1998).

118. W.R. Wikoff, R.L. Duda, R.W. Hendrix and J.E. Johnson. Crystallographic analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Acta Crystallographica D Biological Crystallography* **55**:763–771 (1999).
119. S.B. Zimmerman. Macromolecular crowding effects on macromolecular interactions: some implications for genome structure and function. *Biochimica et Biophysica Acta* **1216**:175–185 (1993).

Chapter 10

Topology of Viral DNA

Javier Arsuaga*, Joaquim Roca† and De Witt Sumners‡

1. Introduction: The Organization of Double-Stranded DNA in Bacteriophage Capsids

Chromosome organization plays a key role in many biological processes. In viruses genome organization is essential for the packaging and releasing of the genome as well as for maintaining the stability of the viral capsid. This organization varies across different families and is highly dependent on the virus morphogenetic pathway (Casjens, 1997). In this review we will focus on the packing of double-stranded DNA (dsDNA) in bacteriophages. Understanding how dsDNA is packed in bacteriophage capsids is important because it arguably yields the simplest example of a full genome organization in a biological system. Furthermore bacteriophages are believed to pack their DNA similarly to some animal viruses such as herpes- and adeno-viruses (Casjens, 1997). Therefore understanding the basic principles of DNA packing in bacteriophages may provide essential insights into the viral assembly pathways for a wide range of bacterial and animal viruses.

During bacteriophage morphogenesis a proteinic procapsid is first assembled (reviewed in Johnson and Chiu, 2007; Earnshaw and Casjens, 1980; Dockland, 1992; and in Wang *et al.*, 2006). This assembly is followed by the active transfer of one copy of the viral genome from the host cytoplasm to the procapsid through a molecular motor called the

*Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA.
†Instituto de Biologia Molecular de Barcelona, CSIC, Barcelona, Spain.
‡Department of Mathematics, Florida State University, Tallahassee, FL, USA.

connector (Fuller *et al.*, 2007; Ibarra *et al.*, 2000; Rishov *et al.*, 1998). This packaging reaction is performed against strong repulsive forces due to the excessive bending and the self-repulsive nature of the double helix (Tzlil *et al.*, 2003; Kindt *et al.*, 2001; Smith *et al.*, 2001). Inside the capsid, the viral DNA reaches near-crystalline density (Earnshaw and Casjens, 1980) with concentrations of DNA fibers of up to 800 mg/ml (Kellenberger *et al.*, 1986) and an osmotic pressure of 50 atmospheres (Evilevitch *et al.*, 2003). Once the DNA is packed inside the capsid, the phage tail is assembled at the connector site and a mature virus is completed. How the DNA is folded under the extreme physical conditions described above is unknown. Biochemical and structural analyses have shown that DNA inside the capsid is kept in its B-form (Aubrey *et al.*, 1992), that the DNA forms local domains of parallel fibers that are 25 Å away from each other (Earnshaw and Harrison, 1977; Earnshaw *et al.*, 1978; Lepault *et al.*, 1986). Moreover, they show that there is no correlation between DNA sequences and their spatial location inside the capsid, with the exception of the DNA ends in some viruses (Chattoraj and Inman, 1974). The question whether there are contacts between the double helix and the interior of the capsid remains open. Work by Serwer and colleagues showed that such contacts were inexistent or minimal (Serwer *et al.*, 1992). On the other hand, new asymmetrical reconstruction techniques of frozen samples have revealed frequent contact points. Whether these contact points have any role in the packing or if they affect the overall geometry of the chromosome remains to be explored (Jiang *et al.*, 2006). Most studies agree on the existence of concentric layers of DNA fibers organized in spooling, toroidal or parallel fashion. Next we describe the proposed models for the long-range organization of the DNA inside the viral capsid.

1.1. *Spooling Models*

There are two basic spooling models: coaxial and concentric. The coaxially spooled model (Fig. 1) was initially proposed by Richards *et al.*, 1973 after performing electron micrographs in bacteriophages P2, λ, T4, T7 and T5. In this model, the DNA is organized in shells that are wound about an axis. The orientation of the axis has been a matter of debate and may be dependent on the virus. It has been proposed that the spooling axis can be perpendicular (Cerritelli *et al.*, 1997; Geller and Davis, 1964), parallel (Earnshaw and Harrison, 1977; Jiang *et al.*, 2006), or tilted a few degrees

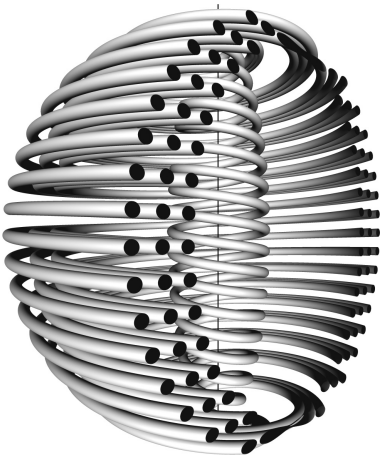


Fig. 1. Coaxial spooling model. Recent experimental results in bacteriophage phi 29 have estimated that there are about five concentric shells of DNA fibers containing up to 50% of the entire genome. This arrangement is more disordered in the center of the capsid where the radius is much shorter than the persistence length of the DNA molecule (Comolli *et al.*, 2007). The idealized model in the figure shows four coaxial shells.

(Sewer, 1986) with respect to the axis defined by the tail. In concentric models, each DNA shell is organized around an axis whose orientation is different from the previous shells. This configuration has been observed in molecular dynamics studies and represents energy minima in the absence of long-range interactions (LaMarque *et al.*, 2004).

1.2. *Toroidal Models*

Toroidal models have been proposed not only for bacteriophages but also for DNA molecules exposed to condensing agents (reviewed in Hud and Vilfan, 2005). In these models DNA travels along the surface of a torus (Fig. 2). Several toroidal models have been proposed. These include the interwound toroid (Earnshaw *et al.*, 1978), toroidal winding (Kosturko *et al.*, 1979), folded toroid (Hud, 1995) and twisted toroid (Petrov *et al.*, 2007). In the folded toroid model the DNA first assumes a conformation similar to the one illustrated in Fig. 2. Since the DNA molecule is too large to fit in the capsid by this arrangement alone the toroid needs to fold in half (or quarters) in a next step. The twisted toroid has been recently proposed

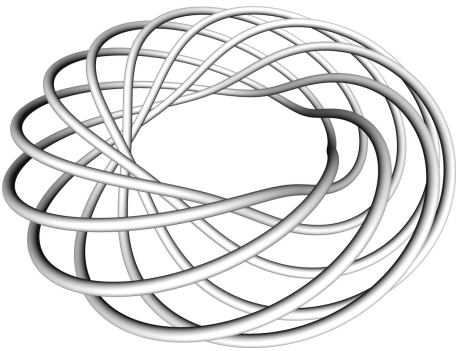


Fig. 2. Toroidal trajectory. The figure shows an idealized toroidal trajectory in which the DNA molecule is represented by a tube and whose axis lies on the surface of a toroid.

using computational models (Petrov *et al.*, 2007). In these studies it is suggested the toroidal shape may be in part a consequence of the elongated shape of some capsids.

1.3. *Spiral-Fold Model*

This model was initially proposed by Earnshaw and Harrison in 1977 and was later supported by eroding experiments on T4 phages (Black *et al.*, 1985). In these experiments partially packed DNA molecules were eroded using ion etching techniques. The illustration in Fig. 3 shows an idealized image of a spiral-fold model. DNA strands run parallel to each other making sharp turns due to the interaction between the genome and the viral capsid. In elongated capsids and in order to avoid unnecessary bending the DNA remains straight for as long as possible. This forces the DNA to eventually take a sharp turn. Current molecular dynamics simulations show configurations that are more consistent with other models. However, sharp turns are also observed in these models (Petrov *et al.*, 2007).

1.4. *Liquid Crystalline Model*

Extreme concentrations of DNA fibers form liquid crystalline phases (Strzelecka *et al.*, 1988; Sikorav *et al.*, 1994). In a liquid crystal state the DNA fibers are locally organized in stacks of parallel planes (with possibly some twisting between the planes) but there is no long-range

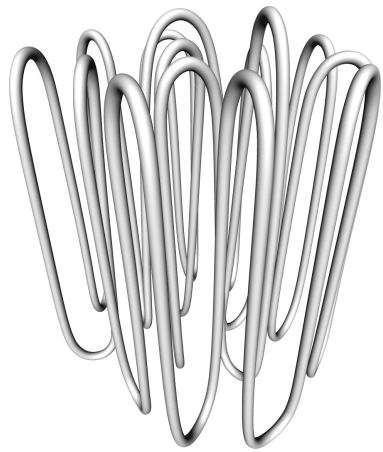


Fig. 3. Spiral-fold model. Black *et al.* proposed that in a T4 capsid the DNA would fold into 700 segments parallel to each other containing between 160 bp and 300 bp forming a total of about 15 shells.

order between the stacks or the fibers in the stacks. Lepault and colleagues (Lepault *et al.*, 1987) proposed this model after analyzing vitrified samples of phages Giant T4, T3 and Lambda phages by X ray diffraction and direct imaging. Similar findings have been previously reported for herpesviruses (Booy *et al.*, 1991) and also in a later report for DNA in lipid capsules used for gene therapy (Schmutz *et al.*, 1999). Figure 4 shows an idealized image of this model where only the organized fibers are shown.

1.5. *Other Models*

Ball of string model. This model was described in (Richards, 1973) as "a sphere generated by winding the DNA in series of concentric circles, each being approximately a great circle, with the planes containing successive circles oriented at random" and was quickly rejected by experimental evidence (Earnshaw and Harrison, 1977).

Random model. Due to repulsion between DNA fibers and the extreme bending of the DNA inside the capsid it is possible that some of the features described above could be obtained by simply randomly packing the DNA chain in the capsid. We will review some of the properties of the random

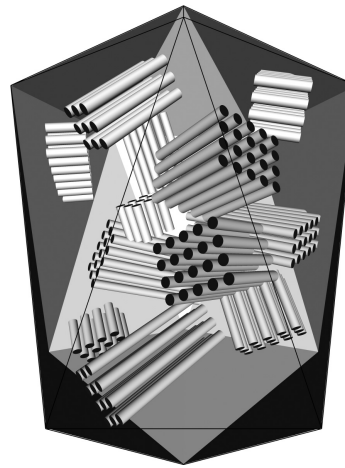


Fig. 4. Liquid crystalline model. In liquid crystalline models, DNA fibers are organized locally but not globally. Sharp turns of the DNA, believed to characteristic of liquid crystalline phases of long molecules, occur in the space between domains. This liquid crystalline phase is a consequence of the extreme density and pressure of the DNA inside the capsid. It is important to highlight that such a model would not provide information on the DNA packaging reaction.

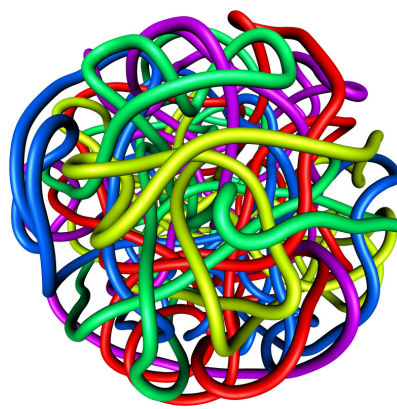


Fig. 5. Random Packing. In the case of random packing the vertices along the trajectory follow a distribution of points in space that is usually difficult to express analytically.

model in detail. Investigating a random model is useful because it provides a reference for random fluctuations in the packing models described above. An illustration of random packing is shown in Fig. 5. The essential feature of a random model is that the points that define the trajectory of

the DNA molecule follow a specific random distribution. This distribution is usually difficult to obtain analytically except for very simple cases (Millet, 1993; Arsuaga *et al.*, 2007). Therefore researchers are performing computer simulations to approach these distributions.

In this review we introduce a new approach to investigating chromosome organization in bacteriophage capsids. Our studies are based on experimental work initiated by Liu and colleagues on P2 and P4 bacteriophages. In Liu *et al.*, 1981a and 1981b it was found that DNA molecules extracted from P2 and P4 tailless mutants were mostly knotted circles. Further studies (Wolfson *et al.*, 1983) found that knots are also observed in wild type P4 phages and that the number of knotted circles (i.e. knotting probability) depends on the length of the DNA being packed. In this review we will argue that these molecules become cyclic inside the capsid, or soon after the disruption of the capsid, and therefore can be used ofto retrieve information about the DNA organization inside the viral capsid. Furthermore we will explain how the DNA knot spectrum can provide new insights into the spatial organization and geometry of DNA inside the viral capsid.

2. Knot Theory and Its Applications to Molecular Biology

2.1. *Mathematical Knots*

In this section a brief introduction to mathematical knot theory is given. For a more detailed exposition of the basic concepts in Knot Theory we refer the reader to standard texts (Adams, 2004; Murasugui, 2007).

A *knot* K is a simple (i.e. without self-intersections) closed curve in R^3. One studies knots by projecting them into a plane; a projection in which each crossing involves two strings in transverse intersection is called a *regular projection*. If the over/under information for the strands at each intersection in a regular projection is kept, represented by a break in the undercrossing strand, then the projection is called a *knot diagram*. Figure 6 shows some examples of knot diagrams.

Two knots K_1 and K_2 are *equivalent* if it is possible to elastically move K_1 into K_2 without introducing self-crossings. Equivalent knots are representatives of the same *knot type*. Two knots K_1 and K_2 are equivalent if and only if any regular diagram for K_1 can be converted to a regular diagram for K_2 via a sequence of *Reidemeister moves*, as illustrated in Fig. 7.

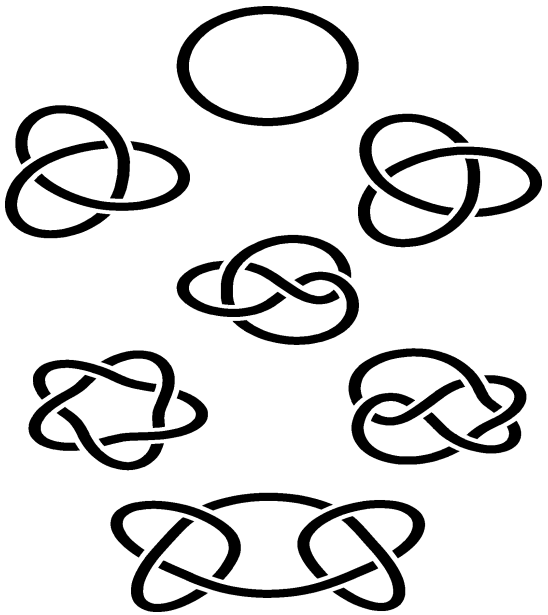


Fig. 6. Examples of knots. The figure shows (from top to bottom) the trivial knot (unknot), the positive (left, denoted here by 3_1) and negative (right, referred here by the mirror image of 3_1 or 3_{1*}) trefoil knots with three crossings each, the four crossing figure-eight knot (center- denoted here by 4_1), the five crossing negative twist knot (right, denoted here 5_{2*}) and the positive torus knot (left, denoted here by 5_1). The knot on the bottom is a six crossing composite knot composed of two trefoils of opposite sign (square knot). From this collection of knots, the trivial knot, the four and six crossing knots have mirror images that are equivalent (achiral), whilst the rest are all chiral. For the trefoil knots, each member of the chiral pair is shown.

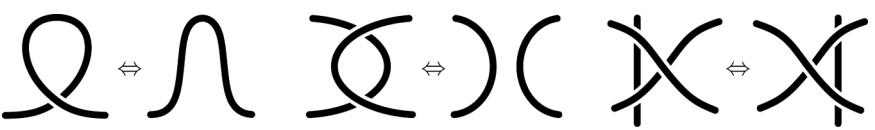


Fig. 7. The three Reidemeister moves. Performing any of these moves in a knot diagram is equivalent to elastically deforming the knot. These moves are used to simplify the projection of the knot without changing the knot type, and therefore facilitate the identification of the knot.

Given two knots, one can construct another knot by cutting a small segment from each of the knots and pasting their endpoints together. The resulting knot is called a *composite knot*. For instance in Fig. 6 the square knot is a composite knot made of two trefoils of opposite sign. Knots that are not composite are called *prime* knots. In Fig. 6, all of the knots except the trivial knot and square knot are prime knots.

The minimum number of crossings among all possible knot diagrams for K is called the *crossing number* of K. A knot diagram that realizes that minimum crossing number for K is called a minimal diagram for K. Figure 6 shows minimal diagrams. A concept that has proven to be very useful in biology is the *average crossing number*. It is given by the average number of crossings over all possible regular projections of K(as a fixed space curve). A second important concept is that of *chirality*. A knot K is said to be *achiral* if it is topologically equivalent to its mirror image, otherwise it is said to be *chiral*. In order to obtain the mirror image of a knot, take any diagram for the knot, and reverse all the crossings. In Fig. 6, the right-handed trefoil is topologically distinct from its mirror image, the left-handed trefoil. The figure-eight knot and square knots are achiral since one can find a sequence of Reidemeister moves that can transform the diagrams shown to their mirror images. A geometrical concept that helps quantify the chirality of a knot is the *writhe*. Given a knot K (as a fixed curve in space) and an orientation on the knot (i.e. a direction of travel along the knotted chain) one can define the *projected writhe* of the knot by assigning an orientation to the knot, and at each crossing in a regular diagram of K either a $+1$ or a -1 is assigned following the right hand rule crossing sign convention (Fig. 8) for oriented skew lines in space. Note that the crossing signs are independent of knot orientation,

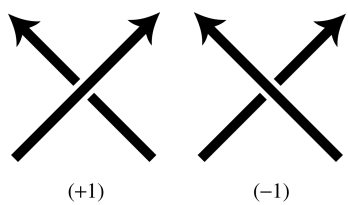


Fig. 8. Crossing Sign Convention.

because changing the orientation of the knot changes both of the arrows at a crossing, leaving the crossing sign unchanged.

If one averages the projected writhe over all projections of a fixed spatial curve K, one obtains the *writhe* of the knot K. The writhe is a measure of non-planarity of the spatial knot. For the knot diagrams in Fig. 6, the projected writhe of the unknot, figure-eight knot and square knots are each zero; the projected writhe of the positive trefoil is $+3$, the projected writhe of the negative trefoil is -3, the projected writhe of the negative twist knot is -5, and the projected writhe of the positive torus knot is $+5$. If a knot K is achiral the average of the writhe of a random sample of curves of fixed knot type K tends to zero as the number of knots in the sample goes to infinity.

The computational identification (classification) of knots is done by using algebraic knot invariants that can be calculated from any regular projection of the knot. If a given invariant is different for two knots, then the knots are different. If a given invariant is the same for two knots, the knots may or may not be different; that invariant fails to distinguish them. In computational studies, polynomial invariants such as the Alexander, Jones and HOMFLY polynomials are commonly used. The Alexander polynomial, denoted by $\Delta(t)$, is easier and faster to compute than the others, but it is not as powerful as the HOMFLY or the Jones polynomials in classifying knots. An important drawback of the Alexander polynomial is that it cannot detect knot chirality. It is known that the computation of the Jones polynomial is NP-hard (i.e. cannot be computed in time, which is polynomial in the number of knot crossings) (Jaeger *et al.*, 1990). In order to make computations more efficient, the Alexander polynomial evaluated at $t = -1$ (denoted by $\Delta(-1)$) is commonly used. However theoretical studies have shown that there exist many non-trivial knots with trivial Alexander polynomials ($\Delta(t) = 1$) and it is not clear that $\Delta(-1)$ can be used to estimate knot probability for arbitrarily knot ensembles (Garoufalidis and Teichner, 2004). Another approach is to use look-up tables of algebraic descriptors of knots called Dowker codes (Hoste and Thistlethwaite, 1999; Micheletti *et al.*, 2006). The disadvantage of look — up tables is that the exact tabulation of knots excludes knots of more than 16 crossings.

2.2. *Knots and DNA*

Knots and links in circular DNA molecules are of biological interest because they can detect and preserve information about DNA folding and

about enzymes that manipulate DNA. DNA knots can be obtained in the test tube (*in vitro*) as products of biochemical reactions or can be designed as building blocks for nanotechnology purposes (e.g. Mueller *et al.*, 1991; Seeman 2003). Figure 9 shows a DNA five-crossing torus knot produced *in vitro* by reacting a plasmid with the site-specific recombinase mutant Gin.

DNA knots obtained as the product of biochemical reactions help analyze the process under study. For instance, knots formed by random cyclization of linear molecules have been used to quantify the effective diameter of the DNA (Rybenkov *et al.*, 1993; Shaw and Wang, 1993) and the effects of confinement on the cyclization of DNA molecules (Arsuaga *et al.*, 2002). DNA knots have been also key to unveiling the mechanism of site-specific recombination enzymes (e.g. Buck and Flapan, 2007; Darcy *et al.*, 2006; Ernst and Sumners, 1990; Wasserman and Cozzarelli, 1986; Grainge *et al.*, 2007; Vazquez and Sumners, 2004; Vazquez *et al.*, 2005; Vetcher *et al.*, 2006), the structure of the Mu transpososome (Darcy *et al.*, 2007; Pathania *et al.*, 2002), the mechanism of type-2 topoisomerases (Flammini *et al.*, 2004; Hua *et al.*, 2001; Liu *et al.*, 2006; Roca, 2001; Rybenkov *et al.*, 1997; Vologodskii *et al.*, 2001), the mechanism of condensins (e.g. Kimura *et al.*, 1999), and the structure of replication bubbles (e.g. Olavarrieta *et al.*, 2002). In most of these cases, the development of mathematical and computational tools has helped greatly in the

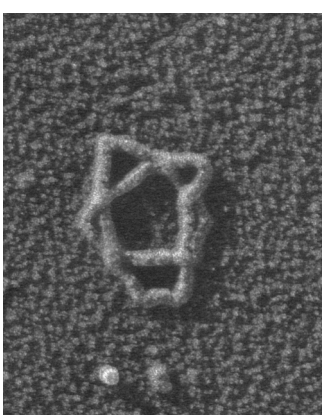


Fig. 9. Electron micrograph of a +DNA five-crossing torus knot obtained after the action of a site-specific recombinase on a dsDNA plasmid. (Figure kindly provided by Nancy Crisona.)

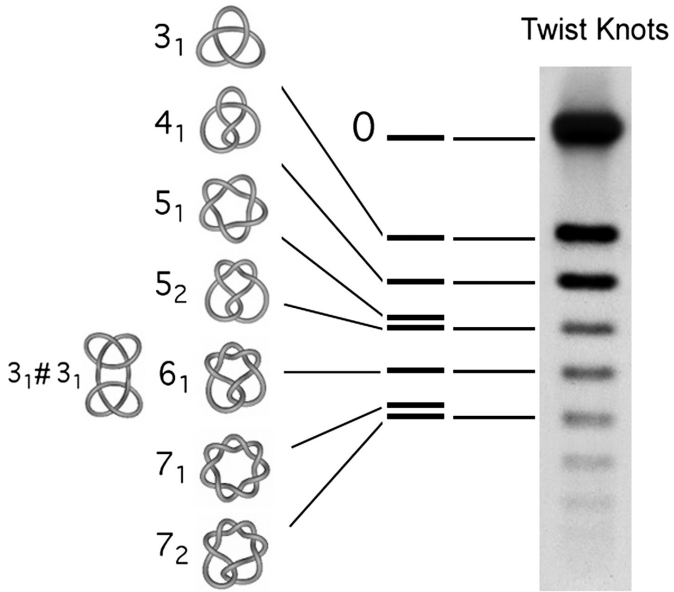


Fig. 10. Electrophoretic velocity of DNA knots. Figure adapted from Arsuaga *et al.*, 2005. DNA knots migrate approximately proportional to their average crossing number. The figure shows the migration pattern of twist knots generated by the action of type II topoisomerase from bacteriophage T4 observed first in Wasserman and Cozzarelli, 1991.

analysis of the experimental results. Yet, the origin of some DNA knots found in other biological systems remains mostly unknown. Examples include *Escherichia coli* cells harboring mutations in the GyrB or GyrA genes (Shishido, 1987), and the cauliflower mosaic virus (Menissier *et al.*, 1983). A recent *in vivo* study (Diebler *et al.*, 2007) shows that knotting of a DNA plasmid that contains a gene for drug resistance leads to an increased mutation rate of the bacterial host, inhibition of replication and cell death when the knot prevents expression of the drug resistance gene.

Knotted and linked (catenated) DNA molecules are stable in the laboratory and can be characterized experimentally by gel electrophoresis and electron microscopy. The simplest method to detect DNA knots in the laboratory, and the one used in the experiments described below, is agarose gel electrophoresis. In this assay, DNA molecules are run at low voltage and knots migrate approximately according to their average crossing number (Vologodskii *et al.*, 1998). Figure 10 shows the twist knot ladder.

A second method to identify knots in the laboratory is by microscopy, using both transmission electron and atomic force microscopes. Since we have not used these assays in our studies we will not provide details here and instead refer the reader to Zechiedrich and Crisona, 1999. However it is important to highlight that the advantage of microscopy over gel electrophoresis is that it allows detection of DNA knot chirality. For example, in Fig. 9, the + five torus DNA knot has been coated with the RecA protein to enhance resolution of DNA crossings in the electron micrograph.

2.3. *Computer Models of DNA Knotting*

Both Molecular Dynamics (MD) and Monte-Carlo (MC) methods have been used to study circular DNA in solution and in confinement. Here we will review only those studies where Monte-Carlo methods were used.

Monte-Carlo simulations of off-lattice (continuum) and on-lattice models have been used to generate ensembles of simulated closed chains. Initial computational efforts as well as some more recent studies have used different lattices such as the tetrahedral lattice (Frank-Kamenetskii *et al.*, 1975) or the simple cubic lattice (Tesi *et al.*, 1994; Hua *et al.*, 2005). However off-lattice models are usually preferred to study DNA properties. These models include the equilateral random polygon model (ERP) (also known as the freely-jointed chain model), the Gaussian random polygon model (GRP), and the wormlike chain model (Cantor and Schimmel, 1980). Among the off-lattice models, the worm-like chain is believed to be the one that best approximates DNA in dilute solution. For the scales considered in the work reviewed here and for comparison with bulk experiments both the wormlike chain and the ERP give very similar results. One advantage of the ERP is that a number of properties, such as the knotting probability and average crossing number, have been characterized not only computationally but also analytically. These results provide a solid foundation for the computational studies (e.g. Diao *et al.*, 1994; Diao 1995).

2.4. *Algorithms for Generating Ensembles of Random Knots*

There are a number of algorithms that generate *equilateral random polygons.* These include the crankshaft (Klenin *et al.*, 1988), the hedgehog (Klenin *et al.*, 1988), and the generalized hedgehog algorithm (Varela *et al.*, 2009). The crankshaft algorithm generates a Markov chain in the

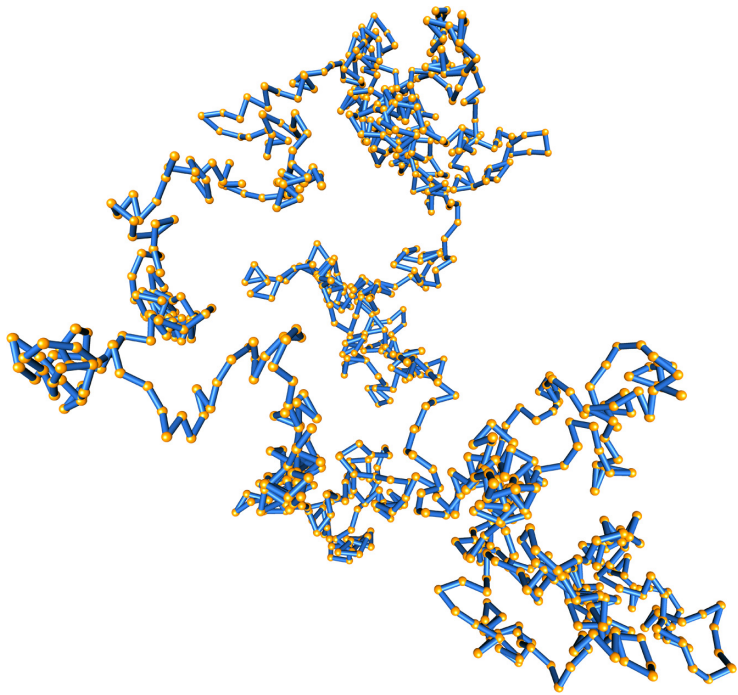


Fig. 11. An equilateral random polygon. An equilateral random walk with 1000 independent segments is shown. The equilateral walk was generated using the hedgehog algorithm and rendered using Knotplot (Scharein, 1998).

space of all ERPs (or wormlike polygons) of fixed length. This Markov chain is known to be ergodic (Millet, 2000). However this algorithm is computationally very inefficient due to the high correlation between consecutive samples. The hedgehog algorithm on the other hand is more efficient than the crankshaft; however it is unknown whether it is ergodic or not. The generalized hedgehog is the only algorithm known to be ergodic and fast.

Once a polygon has been generated by any of the methods described above, a projection is taken and the knot type computed. This process however may be very costly because the knot projection may have many extraneous crossings and the computation of any knot invariant depends on the number of crossings. In order to lower the computational burden one can smooth the polygon before computing its projection (e.g. Micheletti *et al.*, 2006). This smoothing is done by perturbing the knot

across those triangles that are spanned by adjacent edges and that do not intersect other segments of the knot. For instance, in Micheletti *et al.*, 2006, it was observed that unknots of length $N = 400$ edges confined to a sphere of $R = 10$ typically have approximately 400 crossings! By applying these simplification methods, polygons were simplified to lengths of $N' = 30$ with approximately 30 crossings. Next a projection of the knot is taken and Reidemeister moves (Fig. 7) performed to further remove extraneous crossings. Despite these efforts the knot diagrams may still have many extraneous crossings. In order to determine if a configuration is knotted or not, it is necessary to compute an algebraic invariant. A quick and reasonably efficient determination of curve knotting is computing $\Delta(-1)$. Computing the knot type is a more challenging problem and researchers have either computed the Alexander polynomial $\Delta(t)$ evaluated at different values of t (e.g. $t = -2, -3$, see Arsuaga *et al.*, 2005; Mansfield, 1994) or have used more sophisticated methods such as the HOMFLY polynomial (Millet, 1994; Micheletti *et al.*, 2006; 2008).

3. DNA Knots in Bacteriophage P4

3.1. *Bacteriophage P4*

Bacteriophage P4 is a satellite of bacteriophage P2. The P4 capsid has icosahedral symmetry ($T = 4$) (Dockland *et al.*, 1992) and its genome is an 11.7 kb linear double-stranded DNA molecule with two 16 bp long cohesive ends (Wang *et al.*, 1973) called *cos* sites. In the mature phage capsid, one of the DNA ends is attached near the connector region (Chattorraj and Inman, 1974), while the position of the other end is unknown. The attachment of one of the ends to the entry region prevents the two *cos* ends from annealing to each other and forming a circle inside the capsid. It is likely that this anchoring facilitates the DNA in leaving the capsid as a linear molecule through the tail at the time of infection.

However, most DNA molecules extracted from bacteriophage P4, as detailed in Section III.2, are non-covalently closed circular molecules and have a very high percentage of nontrivial and highly complex knots ($\sim$50%) (Arsuaga *et al.*, 2002; Wolfson *et al.*, 1985). The percentage of knots is higher ($\sim$95%) for DNA extracted from tailless mutants of the bacteriophage P4 (Arsuaga *et al.*, 2002; Liu *et al.*, 1981a; Liu *et al.*, 1981b; Wolfson *et al.*, 1985). Interestingly, the percentage of knots is also higher in

P4 strains with shorter genomes (~10 kb) than in those strains with full genomes (Wolfson *et al.*, 1985).

What is the origin of these knots? In order to form a knot, both ends of a linear DNA chain need to meet and anneal through their *cos* ends. This cyclization reaction results in non-covalently closed circles (since the DNA backbones remain unsealed). The cyclization reactions of purified linear P4 DNA molecules in free solution have been extensively studied. In these experiments, DNA molecules were first extracted and linearized by thermal denaturation of their *cos* ends. Upon cyclization in dilute free solution and at physiologic ionic conditions, it was found that the knotting probability for P4 DNA molecules is about 3% and that most of the knots (> 90%) had only three crossings (Rybenkov *et al.*, 1993). Therefore DNA knots have a non-zero probability of occurrence during any cyclization reaction. We ask then why the knotting probability and complexity are so high in molecules extracted from P4 phages. Another important question is why this probability reaches saturation levels in tailless mutants or in the genome deletion strains.

3.2. *Experimental Methods: P4 Production and DNA Extraction*

Bacteriophage P4 vir1 del22, which carries a 1.7 Kb deletion compared with the wild type P4, are used in the laboratory as source of knotted DNA. These phages with a shorter DNA (~10 Kb) show higher DNA knotting percentages than those with full genomes (Wolfson *et al.*, 1985). In order to obtain a maximum yield of knotted molecules, P4 vir1 del22 is first amplified in an *E. coli* strain (for example, C-1895) lysogenic for the helper prophage P2. Stocks of P4 vir1 del22 are then used to infect an *E. coli* strain lysogenic for P2 (for example, C-8001) that carries the mutation amH13 in the H gene. Because gene H encodes part of the P4 phage tail, such strain produces then P4 vir1 del22 particles without tail (tailless mutants).

After bacterial lysis, bacteriophages are collected as described in Isaksen *et al.*, 1999. Briefly, phages are precipitated with polyethylene glycol 8000 and then dissolved in phage buffer (10 mM MgCl2/10 mM Tris·HCl, pH 7.2/130 mM ammonium acetate). Phage particles can then be purified by equilibrium density centrifugation across a cesium chloride

gradient (14 h at 45,000 rpm in an NVT65 Beckman rotor). Banded viral particles are dialyzed against phage buffer and their DNA is extracted with phenol and phenol/chloroform. The viral DNA is then dialyzed against TEN buffer (10 mM Tris·HCl, pH 8/1 mM EDTA/100 mM NaCl) and kept at 4°C.

An optimal way to qualitatively and quantitatively analyze knotted DNA is by means of a high-resolution two-dimensional gel electrophoresis (Trigueros *et al.*, 2001). In this technique, the first gel dimension is run at low voltage, and DNA knots migrate according to their compactness. The second gel dimension is run at high voltage, and DNA knots migrate according to other physical parameters, such as shape and flexibility. In comparison to one-dimensional gel electrophoresis, this procedure segregates the knotted DNA molecules from other unknotted forms of DNA, and partially resolves populations of knots that have the same number of crossings. For the analysis of P4 DNA knots (about 10 Kb size), these gels are prepared as follows. Gel slabs of 0.4% agarose concentration are equilibrated with TBE buffer (100 mM Tris-borate, pH 8.3/2 mM EDTA). In the first dimension, DNA samples run at 0.8 V/cm for 40 h at room temperature. After a 90° rotation of the gel, a second dimension is run in the same electrophoresis buffer at 3.4 V/cm for 4 h at room temperature. After ethidium bromide staining of DNA and photography, gels can be blotted to nylon membranes, and DNA bands radioprobed for phosphorimaging quantification (Fig. 12).

3.3. *Experimental Results*

The quantitative analyses of DNA species resolved by high-resolution two-dimensional gel electrophoresis, like that seen in Fig. 12, indicated the following:

— The 10-kb DNA from the tailless mutant of phage P4 vir1 del22 produces 95% knotted molecules (*i.e.* a knotting probability of 0.95).
— The average estimated number of knot crossings was about 26.7, and the largest estimated number of knot crossings detected was about 40.
— Of the total amount of knotted molecules extracted from P4 tailless capsids, only 2.2% had crossing numbers between three and nine. These populations are of main interest since they can be individually quantified.

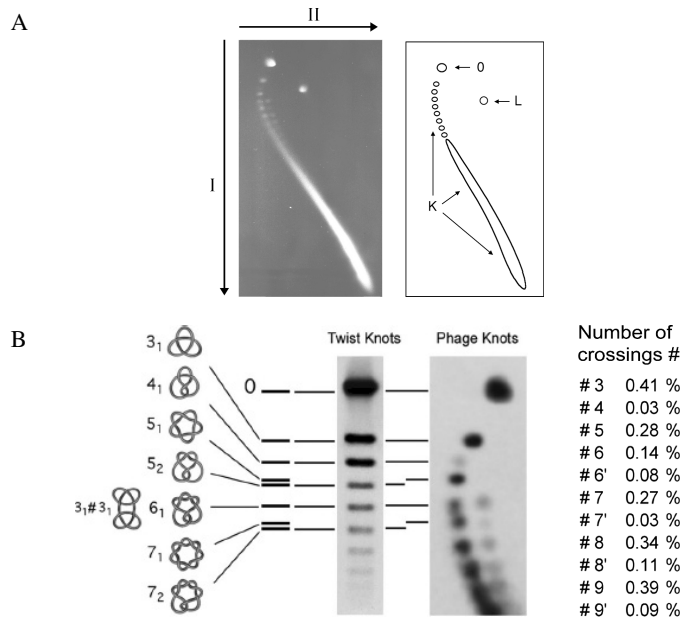


Fig. 12. Qualitative and quantitative analyses of DNA knots resolved by gel electrophoresis (from Arsuaga *et al.*, 2005). A) Knots are identified by 2D gel electrophoresis. In the first dimension (vertically) knots were run for 40 hours at 25 V and were run parallel to a control of twist knots (Wasserman and Cozzarelli, 1991). Under these conditions knots migrate proportionally to their average crossing number. The second dimension (horizontally) is run for under 2 hours at 100 V. Knots migrate differently creating the characteristic bell shape. For details see Trigueros *et al.*, 2001. B) The gel-blot region of less complex knot populations (knots of 3 to 9 crossings) is shown in detail. Gel positions of DNA bands are compared to a marker ladder for twist-type knots that was generated by incubating a supercoiled 10-kb plasmid with equimolar amounts of T4 topoisomerase II (Wasserman and Cozzarelli, 1991). After the knotting reaction, supercoils were removed from DNA by limited nicking in a reaction containing 50 mM Tris (pH 7.5), 5 mM MgC12, 100 pg/ml each DNA and ethidium bromide, and 2 pg/ml DNase I.

— Some of the bands observed in the 1D gel split into a secondary arch of faster velocity in the second gel dimension (denoted as 6–9 and 6'–9' in Fig. 12).
— Densitometer readings of individual knot populations of three to nine crossings revealed that knots of four crossings are severely reduced relative to the other knot populations.

— At low voltage, torus knots (such as 5_1 and 7_1) migrate slightly slower than their corresponding twist knots (5_2 and 7_2) (Vologodskii *et al.*, 1988); this knowledge in conjunction with a marker ladder for twist knots (Fig. 10 and Wasserman and Cozzarelli, 1991) allows the identification of several gel bands of the phage DNA matching the migration of known knot types.
— In the main arch of the gel, in addition to the unambiguous knots 3_1 and 4_1, the knot population of five crossings matched the migration of the torus knot 5_1. The other possible five-crossing knot, the twist knot 5_2 that migrates between and is equidistant to the four- and six-crossing knot populations, appeared to be negligible or absent.
— The knot population of seven crossings matched the migration of the torus knot 7_1 rather than the twist knot 7_2, which has slightly higher gel velocity. Although this visible gel band is likely populated by seven-crossing torus knots, other knot types of seven crossings cannot be excluded.
— There is an evident shortage of the knot subpopulation of seven crossings in the second arch of the gel (denoted by 7' in Fig. 12).

When comparing these knot distributions with those obtained in free solution, we conclude that the DNA particles become cyclic inside the viral capsid.

4. Models of Random Packing

In this section we consider the computational aspects of the problem of DNA knotting in bacteriophages. As discussed above we mainly work with the equilateral random polygon (ERP) which in this case is confined inside a spherical volume that represents the viral capsid. Before we go in a detailed discussion of the models and the results, a few assumptions need to be highlighted.

The first assumption deals with the flexibility of the chain. The flexibility of the dsDNA chain is given by the persistence length which is 50 nm in free solution. However the radius of the P4 capsid is 45 nm. This strong confinement indicates that there is an effective persistence length inside the capsid that is different from that in solution. Due to these restrictions researchers are forced to make approximations (Arsuaga *et al.*, 2002; LaMarque *et al.*, 2004; Arsuaga *et al.*, 2005; Arsuaga *et al.*, 2007;

Micheletti, 2008). In our studies and those of others, the flexibility of the chain is considered by expressing capsid radius R as a multiple of the effective persistence length. The second assumption has to do with the generation of random samples of closed curves inside a sphere of fixed radius. The sample of knots obtained by randomizing polygons in confinement, using the algorithms described above, is believed to be comparable to the sampling process that the DNA undergoes during random cyclization. This assumption has been confirmed when analyzing DNA cyclization experiments in free solution (Rybenkov *et al.*, 1993).

4.1. *Computational Methods*

Modeling DNA under spatial constraints remains a challenge and many of the techniques explained above (Section II.3) become very expensive computationally. The problem of knotting in confined volumes was first addressed by Michels and Wiegel (Michels and Wiegel, 1986) using molecular dynamics simulations. Today, the most common method to generate statistical ensembles of polygons confined to spheres is Monte-Carlo. The simplest approximation is to generate polygons by one of the methods described earlier and reject those that are outside the confining sphere. However when one tries to sample polygons in confinement one finds that the vast majority of rings will be "swollen" and accumulating reliable statistics for highly confined conformations leads to impractically long computing times. A more efficient computational alternative is to generate a succession of conformations chosen using importance sampling criteria (Arsuaga *et al.*, 2002). However this method is still not sufficient for long polygons confined to small volumes. In Micheletti *et al.*, 2006 and 2008 a dramatic improvement on these calculations was accomplished. First, calculations were done in the conjugated ensemble rather than in the ensemble of a sphere of fixed radius. Second, a faster sampling along the Markov Chain was achieved by a multiple Markov chain scheme, in which all "replicas" of the system were run in parallel, running each chain at a different temperature. Third, the data were collected and analyzed using histogram reweighting techniques.

A second approach consists of initially defining points that are physically within the confining region. Both continuous and discrete models have been implemented (Arsuaga *et al.*, 2007; Mansfield, 1994; Millet, 2000). Current continuum models are known to be non-realistic for

simulating DNA since the flexibility of the chain is not well defined. However these studies have shown that, qualitatively speaking, results are similar to those obtained by other methods. The advantage of these methods is that they are drastically more efficient and also more amenable for the development of analytical results than the methods described above.

4.2. *Results*

4.2.1. *Knotting probability, average crossing number and writhe in confined random polygons*

In Arsuaga *et al.*, 2002, 2005 and in Micheletti *et al.*, 2006 the effects of confinement on the ERPs knotting probability were analyzed by Markov-chain Monte Carlo simulations as explained above. Figure 13 (from Micheletti *et al.*, 2006) shows the knotting probability of closed ideal chains confined to spherical volumes of various radii (in multiples of edge length) as a function of the total chain length (number of edges in the equilateral polygon). As one would expect, and in agreement with

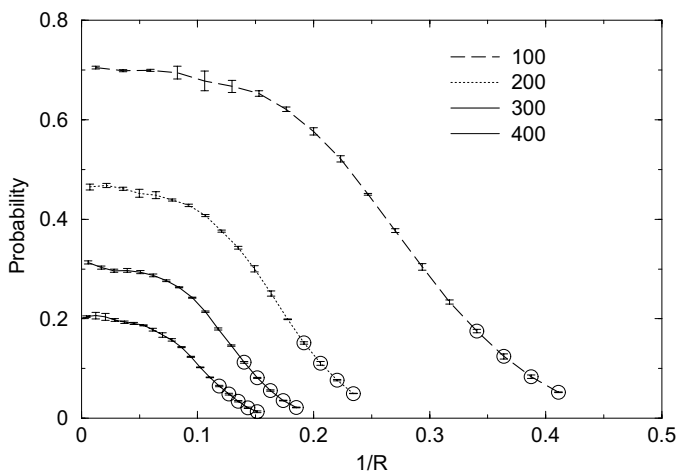


Fig. 13. Knotting probability for chains of length N with $50 \leq N \leq 450$ as a function of $1/R$. Open symbols denote the region where the fraction of unknown knots is between 10% and 50%. No data is plotted where the fraction of unknown knots exceeds 50% (Figure Adapted from Micheletti *et al.*, 2006).

previous models, the knotting probability decreases with increasing radii and decreasing polygon length (since the effects of the confining sphere are gradually removed under these two conditions). In Micheletti *et al.* (2006), it was found that the ratio of probabilities for polygons confined with respect to those that are not confined is given by $P(N,1/R)/P(N,$ no confinement$) \sim \exp(N^a/R^3)$ where $a = 2.15$.

In addition, the complexity of the knots was estimated in Arsuaga *et al.*, 2002 by computing the average number of crossings of the knot. These results are summarized in Fig. 14. As expected, it was observed that for a fixed confining volume the complexity increased with the length of the polygon. For fixed length, the complexity decreased when the confining radius was increased.

In Micheletti *et al.* (2006), a preliminary study of the average writhe of polygons confined to spheres was performed. Since the average value of the writhe is zero one needs to compute the absolute value of the writhe or the square of the writhe to estimate its variation with the confining volume.

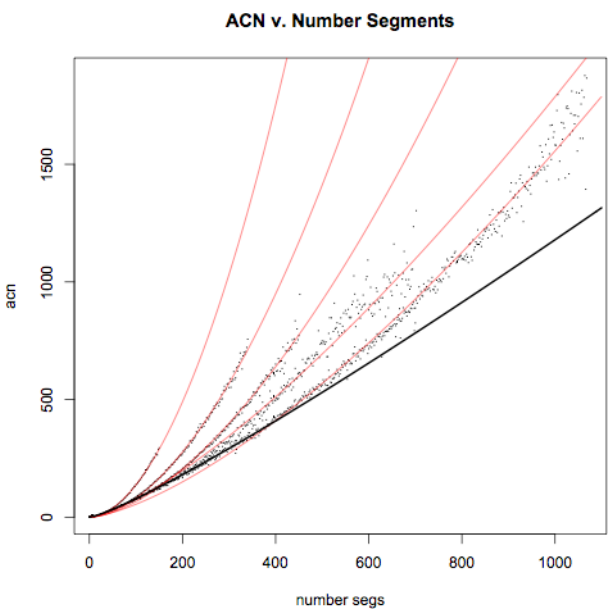


Fig. 14. Average Crossing Number of ERPs in Confined Volumes. The graph shows results for chains of length N with $20 \leq N \leq 1000$ confined to radii ($R = 2, 3, 4, 5, 7$) Results for non confinement are shown by the black line.

It has been rigorously shown that the absolute values of the writhe of polygons in the simple cubic lattice without confinement increase as N^a with $a = 0.52 \pm 0.04$. In Micheletti *et al.*, 2006 it was found that the value of the exponent a changes when in confinement.

4.3. *Distributions of Knots for Low Crossing Numbers: Non-Random Packing of DNA in Bacteriophage P4*

In Arsuaga *et al.* (2005) and in Micheletti *et al.* (2006) and 2008 the distributions of knots with low crossing number, that is knots with fewer than 6 crossings, were computed and plotted as functions of the inverse of the radius or of the length of the polygon.

Figure 15 (left) shows the R dependence of the probability to observe a trefoil knot $P(3_1)$ for various lengths of the polymer ring. As for the unknotting probability there is a range of R ($R > Rc$) for which $P(3_1)$ does not change too much with R. This "plateau" is more visible for small values of N when, for sufficiently small confining radii, $P(3_1)$ is a monotonic-nonmonotonic curve with one maximum. As the confinement radius R is further reduced, $P(3_1)$ decreases in favor of more compact conformations. For longer polymers ($N > 125$) the maximum becomes progressively less evident and we observe a shoulder for small values of $1/R$ that disappears for $N = 400$.

Figure 15 (right) show plots analogous to the one in Fig. 13 but now refer to the probability of forming four crossing knots. The trend

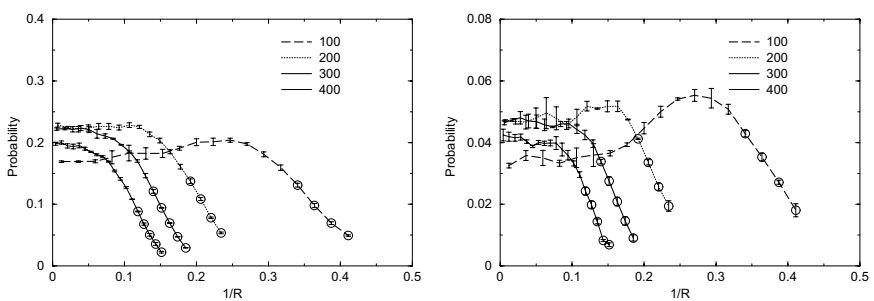


Fig. 15. Knot probabilities as a function of N and $1/R$. The plots are for the 3_1 and 4_1 knots.(Figure adapted from Micheletti *et al.*, 2006)

observed for the trefoil knot is also observed for the four crossing knot and also other prime knots with less than 7 crossings. Indeed, all the curves for short polymers (small N) have a maximum as a function of the inverse radius while those for longer polymers decrease monotonically with $1/R$.

In Arsuaga *et al.* (2005), a comparison between the distributions of knots for the knot families $3_1, 4_1, 5_1$ and 5_2 was made for different lengths and radii. It was found that in all tested cases the probability of the four crossing knot was higher than that of each of the fives (but not combined). Furthermore, as it is the case for knots without confinement, the twist five (5_2) was more probable than the toroidal five (5_1). This result has been observed for different polymer models in confined volumes such as the Uniform Random Polygon (Millet, 2000) and the wormlike chain (Micheletti *et al.*, 2008). Importantly, Fig. 16 shows the simulated

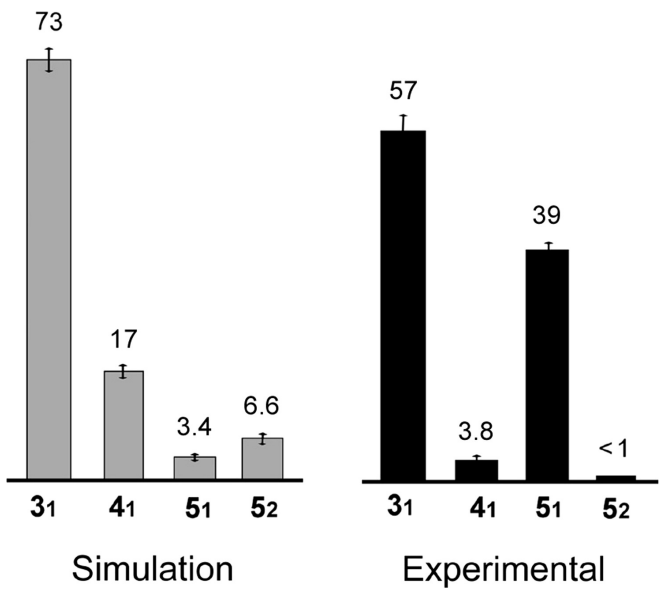


Fig. 16. DNA packing in bacteriophage P4 is not random. Comparison of the computed probabilities of the knots 3_1, 4_1, 5_1, and 5_2 (for polymers of length $n = 90$ randomly embedded into a sphere of radius $R = 4$) with the experimental distribution of knots. The relative amount of each knot type is plotted. Note that the fractions of knots 3_1 and 4_1 plausibly formed in free solution are not subtracted from the experimental distribution. If these corrections are considered, the relative amount of knot 4_1 is further reduced. (Figure adapted from Arsuaga *et al.*, 2005)

distribution of knots that differs strongly from the one observed experimentally (see Fig. 12). Since the simulated case represents the distribution of knots obtained for random embeddings of closed curves in confined volumes it was concluded that the packing in bacteriophage capsids cannot be random (see Fig. 16).

4.4. *Knotting Probability and Complexity in Biased Models*

How can we explain the differences between the knot distributions obtained from random embeddings and those observed experimentally?

Chiral models

In Arsuaga *et al.*, 2005 it was argued that elevated writhe was responsible for these differences in the knot distribution. One interesting observation is that the 4_1 knot is achiral, that is it is equivalent to its mirror image. This property is somehow captured by the writhe of the molecule since achiral knots have writhe 0. In fact in (Rensburg *et al.*, 1993) it was found that random polygonal realizations of the 4_1 knot in free space produce a family of polygons whose writhe distribution for any polygonal length is a Gaussian curve with zero mean and whose variance grows as the square root of the length. As discussed above the confinement alone increases the writhe (Micheletti *et al.*, 2006). In order to obtain conformations with elevated writhe, Arsuaga and colleagues sampled polygons whose writhe was above the average writhe induced by the confinement. It was found that the distribution of knots changes dramatically and produces results that are comparable to those obtained experimentally. Figure 17 shows how the knot distribution for the 4 and 5 crossing knots changes with increasing writhe. In Arsuaga *et al.* it was proposed that the main reason for the scarcity of the knot 4_1 is a writhe bias imposed on the DNA inside the phage capsid. A drop of the probability of the knot 4_1, as well as a rapid increase of the probability of the torus knot 5_1 but not of the twist knot 5_2, readily emerged by increasing the writhe rejection value. These writhe-induced changes in the knot probability distribution are independent of the number of edges in the equilateral polygon and the sphere radius length. Accordingly, previous studies had shown that the mean writhe value of random conformations of a given knot does not

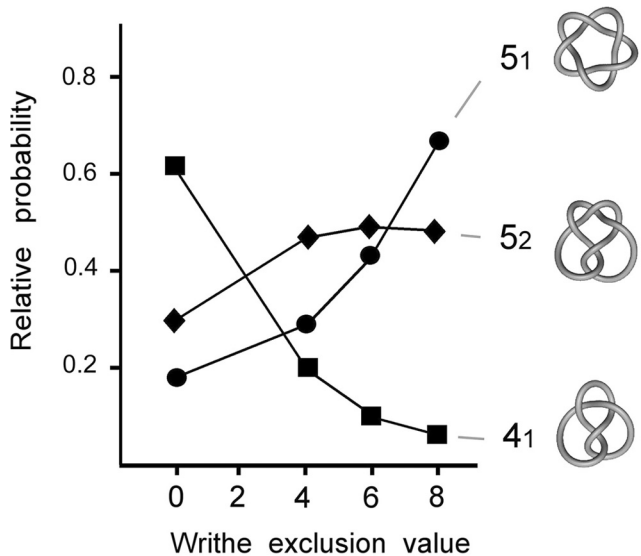


Fig. 17. Effect of writhe-biased sampling on the probability of knots 4_1, 5_1, and 5_2. Figure adapted from Arsuaga *et al.*, 2005. Increasing the value of the writhe modulates the distributions of observed knots. Importantly it decreases the probability of the four crossing knot as well as the five twist knot while at the same time increasing the five toroidal knot.

depend on the length of the chain but only on the knot type and that these values are model-independent (Rensburg *et al.*, 1993).

5. Conclusions and Future Challenges

Here we have reviewed some of our topological approaches to the problem of genome organization in bacteriophage P4. There are two main conclusions: First, P4 DNA knots are formed either inside the bacteriophage capsid or soon after its disruption. This observation indicates that the knotting observed is driven by the confinement of the capsid. It also suggests that P4 knots can be used as reporters for DNA packing in P4 phages (Arsuaga *et al.*, 2002). Second, analysis of P4 knot distributions suggests that the viral genome is chirally organized (Arsuaga *et al.*, 2005). This property had not been previously observed using other experimental techniques. Next we briefly describe some of the latest advances that will require further exploration.

i) *Excess writhe is consistent with the knotting probabilities observed in deletion mutants.*

Results obtained by Wolfson *et al.* that show that shorter DNA chains extracted from bacteriophage P4 have higher knotting probabilities than long ones have long puzzled researchers. A possible explanation was first provided by Blackstone *et al.* (2007), where it was shown that the knotting probability of simulated ERPs with high writhe values decreases for increasing chain length until a minimum is reached. After this point the knotting probabilities increase as expected.

ii) *Currently proposed DNA packing models are not consistent with P4 data.*

Simulations of random closure of the DNA molecules packed as described in Section 1 show very little knotting. This observation suggests that current models for DNA organization inside phage capsids are simplistic and do not completely reflect the complexity of the packing. A new model called "random spooling model" has been proposed by Arsuaga and Diao to address this problem (Arsuaga and Diao, 2008). This model simulates concentric spools in which the DNA strands from different concentric layers are allowed to interweave.

iii) *Dealing with knot complexity.*

As described in the sections above, only about 2% of the knots can be identified using current experimental methods. One way to estimate the complexity of those knots is by creating deletion mutants that can be better characterized by gel electrophoresis or microscopy. A second option is to treat samples with type II topoisomerases. Topoisomerases can unknot knots and therefore by studying possible unknotting pathways one may infer the complexity of the initial population. In Hua *et al.* (2005), a computational study simulating this approach was proposed and equilibrium distributions of knots were found.

iv) *New deletion mutant strains may help better identify the knots observed as well as the packing arrangement.*

One drawback of this natural system is that knot formation is restricted to the viral DNA. Hence, we asked whether other DNA molecules of length and sequence different than P4 DNA could be packaged inside P4 capsids and recovered also as highly knotted forms. Accordingly, we envisaged that a bacterial plasmid containing the P4 *cos* sequence (i.e., a P4 cosmid) could be cleaved and threaded into a viral capsid in the course of a bacterial infection by phage P4. Therefore, we constructed

different P4 cosmids and introduced them in bacteria lysogenic for P2. These bacteria were then infected with P4 phage and the DNA in newly formed viral particles was analyzed. We found that cosmids as small as 5 kb were packaged inside P4 capsids. More interestingly, as well as P4 DNA, such cosmids were recovered in the form of highly knotted DNA circles (Trigueros and Roca, 2006). These results further indicate that P4 knots are not a consequence of some property of the viral DNA, and also demonstrate that complete filling of the capsid is not essential for the packing of individual DNA molecules. With this biological system, DNA molecules of varying length and sequence will be shaped into very complex and heterogeneous knotted forms. These molecules could be produced in preparative amounts suitable for systematic studies and novel applications.

Acknowledgements

We thank R. Scharein for his help in generating the DNA packing figures shown in Section 1 as well as Figures 6–8. Also we want to thank R. Varela and B. Borgo for their help in generating Figs. 11 and 14 respectively. J. Arsuaga is supported by NIH grant 2S06GM52588–12 and NSF grant DMS 0920887.

References

1. C. Adams. *The Knot Book: An elementary introduction to the mathematical theory of knots, knots.* Revised reprint of the 1994 original, American Mathematical Society, Providence, RI (2004).
2. J. Arsuaga and Y. Diao. DNA knotting in spooling like conformations in bacteriophages. *Journal Computational and Mathematical Methods in Medicine* **9(3–4)**:303–316 (2008).
3. J. Arsuaga, R.K. Tan, M. Vazquez, D.W. Sumners and S.C. Harvey. Investigation of viral DNA packaging using molecular mechanics models. *Biophys. Chem.* **101**:475–484 (2002).
4. J. Arsuaga, M. Vazquez, S. Trigueros, D.W. Sumners and J. Roca. Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proc. Natl. Acad. Sci. USA* **99**:5373–5377 (2002).
5. J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, D.W. Sumners and J. Roca. DNA knots reveal a charal organization of DNA in phage capsids. *Proc. Natl. Acad. Sci. USA* **102**:9165–9169 (2005).

6. J. Arsuaga, T. Blackstone, Y. Diao, E. Karadayi and Y. Saito. Knotting of uniform random polygons in confined spaces. *J. Phys. A: Math Gen.* **40**:11697–11711 (2007).
7. K. Aubrey, S. Casjens and G. Thomas. Secondary structure and interactions of the packaged dsDNA genome of bacteriophage P22 investigated by Raman difference spectroscopy. *Biochemistry* **31**:11835–11842 (1992).
8. L. Black, W. Newcomb, J. Boring and J. Brown. Ion etching bacteriophage T4: support for a spiral-fold model of packaged DNA. *Proc. Natl. Acad. Sci. USA* **82**:7960–7964 (1985).
9. T. Blackstone, P. McGuirk, C. Laing, M. Vazquez, J. Roca and J. Arsuaga. The role of writhe in DNA condensation, Proceedings of International Workshop on Knot Theory for Scientific Objects. *OCAMI Studies* Osaka Municipal Universities Press, **1**:239–250 (2007).
10. F.P. Booy, W.W. Newcomb, B.L. Truss, J.C. Brown, T.S. Baker and A.C. Steven. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* **64**:1007–1015 (1991).
11. D. Buck and E. Flapan. Predicting knot or catenane type of site-specific recombination products. *J. Mol. Biol.* **374**:1186–1199 (2007).
12. C.R. Cantor and P.R. Schimel. *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules.* Freeman, San Francisco (1980).
13. S. Casjens. Principles of virion structure, function and assembly, 1–3. In:*Structural Biology of Viruses.* Eds. W. Chiu, R. M. Burnett and R. Garcea. Oxford University Press (1997).
14. M. Cerritelli, N. Cheng, A. Rosenberg, C. McPherson, F. Booy and A. Steven. Encapsidated conformation of bacteriophage T7 DNA. *Cell* **91**:271–280 (1997).
15. D.K. Chattoraj and R.B. Inman. Location of DNA ends in P2, 186, P4 and lambda bacteriophage heads. *J. Mol. Biol.* **87**:11–22 (1974).
16. L.R. Comolli, A.J. Comolli, C.E. Spakowitz, P.J. Siegerist, P.J. Jardine, S. Grimes, D.L. Anderson, C. Bustamante and K.H. Downing. Three-dimensional architecture of the bacteriophage phi29 packaged genome and elucidation of its packaging process. *Virology* **371**:267–277 (2008).
17. I.K. Darcy, J. Chang, N. Druivenga, C. McKinney, R.K. Medikonduri, S. Mills, J. Navarra-Madsen, A. Ponnusamy, J. Sweet and T. Thompson. Coloring the Mu transpososome. *BMC Bioinformatics* **7**:435 (2006).
18. Y. Diao. The knotting of equilateral polygons in R^3. *J. Knot Theory and its Ramifications* **4**:189–196 (1995).
19. Y. Diao, A. Dobay, R.B. Kusner and A. Stasiak. The average crossing number of equilateral random polygons. *J. Phys. A* **36**:11561–11574 (2003).
20. R.W. Diebler, J.K. Mann, D.W. Sumners and L. Zechiedrich. Hin-mediated DNA knotting and recombining promote replicon dysfunction and mutation. *BMC Mol. Biol.* **8**:44 (2007).
21. T. Dokland, B.H. Lindqvist and S.D. Fuller. Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2–P4 bacteriophage system. *EMBO J.* **3**:839–846 (1992).
22. W.C. Earnshaw and S.R. Casjens. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**:319–331 (1980).

23. W.C. Earhshaw and S.C. Harrison. DNA arrangement in isometric phage heads. *Nature* **268**:598–602 (1977).
24. C. Ernst and D.W. Sumners. A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Camb. Phil. Soc.* **108**:489–515 (1990).
25. A. Evilevitch, L. Lavelle, C.M. Knobler, E. Raspaud and W.M. Gelbart. Osmotic pressure inhibition of DNA ejection from phage. *Proc. Natl. Acad. Sci. USA* **100**: 9292–9295 (2003).
26. A. Flammini, A. Maritan and A. Stasiak. Simulations of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots. *Biophys. J.* **87**:2968–2975 (2004).
27. M.D. Frank-Kamenetskii, A.V. Lukashin and A.V. Vologodskii, AV. Statistical mechanics and topology of polymer chains. *Nature* **258**:398–402 (1975).
28. D.N. Fuller, J.P. Rickgauer, P.J. Jardine, S. Grimes, D.L. Anderson and D.E. Smith. Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi 29. *Proc. Natl. Acad. Sci. USA* **104**:11245–11250 (2007).
29. S. Garoufalidis and P. Teichner. On knots with trivial Alexander polynomial. *J. Differential Geom.* **67**:167–193 (2004).
30. M. Gellert and D.R. Davies. Organization of DNA in bacteriophage P4. *J. Mol. Biol.* **8**:341–347 (1964).
31. I. Grainge, M. Bregu, M. Vazquez, V Sivanathan, S.C. Ip and D.J. Sherratt. Unlinking chromosomes catenated in vivo by site-specific recombination. *EMBO J.* **126**:4228–4238 (2007).
32. J. Hoste and M. Thistlethwaite. KNOTFIND, (1999), www.math.utk.edu/morwen/knotscape.html.
33. X. Hua, B. Baghavan, D. Nguyen, J. Arsuaga and M. Vazquez. Random state transitions of knots: a first step towards modeling unknotting by type II topoisomerases. *Topology and its Applications* **154**:1381–1397 (2007).
34. N. Hud. Double-stranded DNA organization in bacteriophage heads: an alternative toroid-based model. *Biophys. J.* **69**:1355–1362 (1995).
35. N.V. Hud and I.D. Vilfan. Toroidal condensates: unraveling the fine structure and the role of nucleation in determining size. *Annual Reviews* **34**:295–318 (2005).
36. B. Ibarra, J.R. Caston, O. Llorca, M. Valle, J.M. Valpuesta and J.L. Carrascosa. Topology of the components of the DNA packaging machinery in the phage Phi29 prohead. *J. Mol. Biol.* **298**:807–815 (2000).
37. M. Isaksen, B. Julien, R. Calendar and B.H. Lindquist. DNA topoisomerase protocols. In: *DNA Topology and Enzymes.* Eds. M.A. Bjornsti, N. Osheroff. (Humana, Totowa, NJ), **94**:69–74 (1999).
38. F. Jaeger, D. Welsh and D. Vertigan. On the computational complexity of the Jones and Tutte polynomials. *Proc. Cambridge Phil. Soc.* **108**:5–53 (1990).
39. W. Jiang, J. Chang, J. Jakana, P. Weigele, J. King and W. Chiu. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**:612–616 (2006).
40. J.E. Johnson and W. Chiu. Structures of virus and virus-like particles. *Curr. Op. in Str. Biol.* **10**:229–235 (2000).

41. J. Kindt, S. Tzlil, A. Ben-Shaul and W.M. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proc. Natl. Acad. Sci. USA* **98**:13671–13674 (2001).
42. E. Kellenberger, E. Carlemalm, J. Sechaud, A. Ryter and G. Haller. Considerations on the condensation and the degree of compactness in non-eukaryotic DNA-containing plasmas, In: *Bacterial Chromatin: Proceedings of the Symposium "Selected Topics on Chromatin Structure and Function"*. Eds. C. Gualerzi and C.L. Pon. Springer, Berlin, 11–25 (1986).
43. L.D. Kosturko, M. Hogan and N. Dattagupta. Structure of DNA within three isometric bacteriophages. *Cell* **16**:515–522 (1979).
44. K. Kimura, V.V. Rybenkov, N.J. Crisona, T. Hirano and N.R. Cozzarelli. 13S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. *Cell* **98**:239–248 (1999).
45. K.V. Klenin, A.V. Vologodskii, V.V. Anshelevich, A.M. Dykhne and M.D. Frank-Kamenetskii. Effect of excluded volume on topological properties of circular DNA. *J. Biomol. Struct. Dyn.* **5**:1173–1185 (1988).
46. J.C. LaMarque, T.L. Le and S.C. Harvey. Packaging double-helical DNA into viral capsids. *Biopolymers* **73**:348–355 (2004).
47. J. Lepault, J. Dubochet, W. Baschong and E. Kellenberger. Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. *EMBO J.* **6**:1507–1512 (1987).
48. L.F. Liu, J.L. Davis and R. Calendar. Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases. *Nucleic Acids Res.* **9**:3979–3989 (1981a).
49. L.F. Liu, L. Perkocha, R. Calendar and J.C. Wang. Knotted DNA from bacteriophage capsids. *Proc. Natl. Acad. Sci. USA* **78**:5498–5502 (1981b).
50. Z. Liu, E.L. Zechiedrich and H.S. Chan. Inferring global topology from local juxtaposition geometry: interlinking polymer rings and ramifications for topoisomerase action. *Biophys. J.* **90**:2344–2355 (2006).
51. C.R. Locker, S.D. Fuller and S.C. Harvey. DNA organization and thermodynamics during viral packing. *Biophys J.* **93**:2861–2869 (2007).
52. M.L. Mansfield. Knots in Hamilton cycles. *Macromolecules* **27**:5924–5926 (1994).
53. J. Menissier, G. de Murcia, G. Lebeurier and L. Hirth. Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome. *EMBO J.* **2**: 1067–1071 (1983).
54. J.P.J. Michels and F.W. Wiegel. On the topology of a polymer ring. *Proc. R. Soc. London Ser. A* **403**:269–284 (1986).
55. C. Micheletti, D. Marenduzzo, E. Orlandini and D.W. Sumners. Knotting of random ring polymers in confined spaces. *J. Chem. Phys.* **124**:64903 (2006).
56. C. Micheletti, D. Marenduzzo, E. Orlandini and D.W. Sumners. Simulations of knotting in confined DNA rings. *Biophys J.* **95**:3591–3599 (2008).

57. K. Millett. Knotting of regular polygons in 3-space. In: *Random Knotting and Linking*. Eds. K. Millet and D. Sumners. World Scientific, Singapore, 31–46 (1994).
58. K. Millett. Monte Carlo explorations of polygonal knot spaces. In: *Knots in Hellas '98*. World Scientific, Singapore, 306–334 (2000).
59. J.E. Mueller, S.M. Du and N.C. Seeman. The design and synthesis of a knot from single-stranded DNA. *J. Am. Chem. Soc.* **113**:6306–6308 (1991).
60. K. Murasugi. *Knot Theory and its Applications*. Birkhauser (2007).
61. L. Olavarrieta, M.L. Martinez-Robles, P. Hernandez, D.B. Krimer and J.B. Schvartzman. Knotting dynamics during DNA replication. *Mol. Microbiol.* **46**:699–707 (2002).
62. S. Pathania, M. Jayaram and R.M. Harshey. Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. *Cell* **109**:425–36 (2002).
63. A.S. Petrov, M.B. Boz and S.C. Harvey. The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. *J. Struct. Biol.* **160**:241–248 (2007).
64. G.R. Pruss, J.C. Wang and R. Calendar. In vitro packaging of satellite phage P4 DNA. *Proc. Natl. Acad. Sci. USA* **71**:2367–2371 (1974).
65. E.J. van Rensburg, E. Orlandini, D.W. Sumners, M.C. Tesi and S.G. Whittington. The writhe of a self-avoiding polygon. *J. Phys A* **26**:L981–L986 (1994).
66. K. Richards, R. Williams and R. Calendar. Model of DNA packing within bacteriophage heads. *J. Mol. Biol.* **78**:255–259 (1973).
67. S. Rishovd, A. Holzenburg, B.V. Johansen and B.H. Lindquist. Bacteriophage P2 and P4 morphogenesis: structure and function of the connector. *Virology* **245**:11–17 (1998).
68. V.V. Rybenkov, N.R. Cozzarelli and A.V. Vologodskii. Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. USA* **90**:5307–5311 (1993).
69. V.V. Rybenkov, C. Ullsperger, A.V. Vologodskii and N.R. Cozarelli. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **27**:690–693 (1997).
70. J. Roca. Varying levels of positive and negative supercoiling differently a?ect the efficiency with which topoisomerase II catenates and decatenates DNA. *J. Mol. Biol.* **305**:441–450 (2001).
71. R. Scharein. *Interactive Topological Drawing*. PhD Thesis, Department of Computer Science, University of British Columbia (1998).
72. M. Schmutz, D. Durand, A. Debin, Y. Palavadeau, E.R. Etienne and A.R. Thierry. DNA packing in stable lipid complexes designed for gene transfer imitates DNA compaction in bacteriophage. *Proc. Natl. Acad. Sci. USA* **96**:12293–12298 (1999).
73. N.C. Seeman. Biochemistry and structural DNA nanotechnology: an evolving symbiotic relationship. *Biochemistry* **42**:7259–7269 (2003).
74. P. Serwer. Arrangement of double-stranded DNA packaged in bacteriophage capsids: an alternative model. *J. Mol. Biol.* **190**:509–512 (1986).

75. P. Serwer, S.J. Hayes and R.H. Watson. Conformation of DNA packaged in bacteriophage T7. Analysis by use of ultraviolet light-induced DNA-capsid cross-linking. *J. Mol. Biol.* **223**:999–1011 (1992).
76. S.Y. Shaw and J.C. Wang. Knotting of a DNA chain during ring closure. *Science* **260**:533–536 (1993).
77. K. Shishido, N. Komiyama and S. Ikawa. Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants. *J. Mol. Biol.* **195**:215–218 (1987).
78. J.L. Sikorav, J. Pelta and F. Livolant. A liquid crystalline phase in spermidine-condensed DNA. *Biophys. J.* **67**:1387–1392 (1994).
79. D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson and C. Bustamante. The bacteriophage phi29 portal motor can package DNA against a large internal force. *Nature* **413**:748–752 (2001).
80. T.E. Strzelecka, M.W. Davidson and R.L. Rill. Multiple liquid crystal phases of DNA at high concentrations. *Nature* **331**:457–60 (1988).
81. M.C. Tesi, E.J. Janse van Rensburg, E. Orlandini, D.W. Sumners and S.G. Whittington. Knotting and supercoiling in circular DNA: a model incorporating the effect of added salt. *Phys. Rev. E* **49**:868–872 (1994).
82. S. Trigueros, J. Arsuaga, M. Vazquez, D.W. Sumners and J. Roca. Novel display of knotted DNA molecules by two dimensional gel electrophoresis. *Nucleic Acids Research* **29**:E67 (2001).
83. S. Trigueros and J. Roca. Production of highly knotted DNA by means of cosmid circularization inside phage capsids. *BMC Biotechnol.* **7**:94 (2007).
84. S. Tzlil, J.K. Kindt, W.M. Gelbart and A. Ben-Shaul. Forces and pressures in DNA packaging and release from viral capsids. *Biophys. J.* **84**:1616–1627 (2003).
85. R. Varela, K. Hinson, J. Arsuaga and Y. Diao. A fast ergodic algorithm for generating ensembles of random polygons. *J. Phys A: Math Gen* **42**:095204 (2009).
86. M. Vazquez and D.W. Sumners. Tangle analysis of gin site-specific recombination. *Math. Proc. Camb. Phil. Soc.* **136**:565–582 (2004).
87. A.A. Vetcher, A.Y. Lushnikov, J. Navarra-Madsen, R.G. Scharein, Y.L. Lyubchenko, I.K. Darcy and S.D. Levene. DNA topology and geometry in Flp and Cre recombination. *J. Mol. Biol.* **357**:1089–1104 (2006).
88. M. Vazquez, S.D. Colloms and D.W. Sumners. Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway. *J. Mol. Biol.* **346**:493–504 (2005).
89. A.V. Vologodskii, N.J. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet and A. Stasiak. Sedimentation and electrophoretic migration of DNA knots and catenanes. *J. Mol. Biol.* **278**:1–3 (1998).
90. A.V. Vologodskii, W. Zhang, V.V. Rybenkov, A.A. Podtelezhnikov, D. Subramanian, J.D. Griffith and N.R. Cozzarelli. Mechanism of topology simplification by type II DNA topoisomerases. *Proc. Natl. Acad. Sci. USA* **98**: 3045–3049 (2001).
91. S. Wang, J.R. Chang and T. Dokland. Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein. *Virology* **348**:133–140 (2006).

92. J.C. Wang, K.V. Martin and R. Calendar. On the sequence similarity of the cohesive ends of coliphage P4, P2, and 186 deoxyribonucleic acid. *Biochemistry* **12**:2119–2123 (1973).
93. S.A. Wasserman and N.R. Cozzarelli. Biochemical topology: applications to DNA recombination and replication. *Science* **232**:951–960 (1986).
94. S.A. Wasserman and N.R. Cozzarelli. Supercoiled DNA-directed knotting by T4 topoisomerase. *J. Biol. Chem.* **266**:20567–20573 (1991).
95. C. Weber, A. Stasiak, P. de los Rios and G. Dietler. Numerical simulation of gel electrophoresis of DNA knots in weak and strong electric fields. *Biophys. J.* **90**:3100–3105 (2006).
96. J.S. Wolfson, G.L. McHugh, D.C. Hooper and M.N. Schwartz. Knotting of DNA molecules isolated from deletion mutants of intact bacteriophage P4. *Nucleic Acids Res.* **13**:6695–6702 (1985).
97. E.L. Zichiedrich and N.J. Crisona. Coating DNA with RecA protein to distinguish DNA path by electron microscopy. In: *DNA Topoisomerase Protocols, DNA Topology and Enzymes*. Eds. M.A. Bjornsti, N. Osheroff. (Humana, Totowa, NJ), **94**:99 (1999).

Chapter 11

The Use of Viruses in Biomedical Nanotechnology

Kristopher J. Koudelka* and Marianne Manchester†

There is a growing need for medical treatments on the nanoscale in the fields of imaging, vaccine development and targeted drug delivery. Viruses are an ideal platform for these technologies because of their size, structure and ease of genetic or chemical modification. This chapter will highlight the use of viruses to achieve these goals, and will focus on the major contributions of a single member, cowpea mosaic virus. In the pursuit of advancing biomedical nanotechnology, significant contributions have been made to both medicinal applications and the understanding of the viruses themselves.

1. Introduction

Before describing the strengths of using viruses for biomedical nanotechnology, one must first appreciate why developing therapies on the nanoscale are so critical. Recent breakthroughs in structural, molecular, cellular and systems biology has for the first time enabled utilization of precise three-dimensional molecular controls to mimic, block or elicit novel protein interactions. The creation of functional machines on this nanoscale level takes advantage of the ability to interact with individual bio-molecules or their sub-domains, allowing for higher specificity and control to induce a desired therapeutic response. In addition, one can also impart layers of

*Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. koudelka@scripps.edu

†Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, MC 0749, La Jolla, CA 92093, USA. mmanchester@ucsd.edu

different functions into a single unit, resulting in higher efficacy and less invasive methodology.

For example, current chemotherapeutics used to fight cancer not only kill many healthy cell types in a non-specific manner, but also very little drug actually reaches the cancer cells.[1] Several recently developed nanoplatforms have achieved longer circulation times, selective tumor targeting and delivery of more potent payloads in model systems.[2] Nanoscale products ultimately provide the opportunity to achieve unambiguous targeting, better pharmacokinetics, increased protection from metabolism and clearance, environmental sensing and real-time imaging in a single formulation.

As a testament to the success of nanoscale products, there are already over 96 subcategories of nanomedicine classifications including nanomaterials, regulated artificial surfaces, cell diagnostics, 'smart drugs', nanotherapeutics, and nanorobotics.[3] Viruses have played a key role in nanoproducts establishing their initial foothold in the biomedical field, and continue to contribute to future applications as well.

2. The Uses of Viruses in Biomedical Nanotechnology

There are many materials, or platforms, used for nanotechnology each with its own individual strengths and weaknesses. The major players include antibodies,[4,5] dendrimers,[6–8] liposomes,[9–11] nanoshells,[12–14] iron oxide nanoparticles,[15–17] quantum dots[18,19] and viruses (see Fig. 1). We will highlight the major contributions of viruses to this field.

The initial applications of virus-based therapeutics have been in the area of gene therapy. One of the most famous cases is the use of a modified Moloney murine leukemia virus to treat X-linked severe combined immunodeficiency through retroviral gene replacement, which was initially celebrated.[20] However, several patients later developed adverse complications.[21] More recently a retrovirus vector has been used to treat X-linked chronic granulomatous disease and appears successful at least in the short term.[22] A localized gene therapy using adenoviruses has also been effective using its native cell-surface receptors; however, a better understanding of this virus's tropism is required before systemic use is clinically feasible.[23]

Several mammalian viruses have been studied as viral therapy vectors. Promising success has been seen in Phase II clinical trial of ONYX-015, a

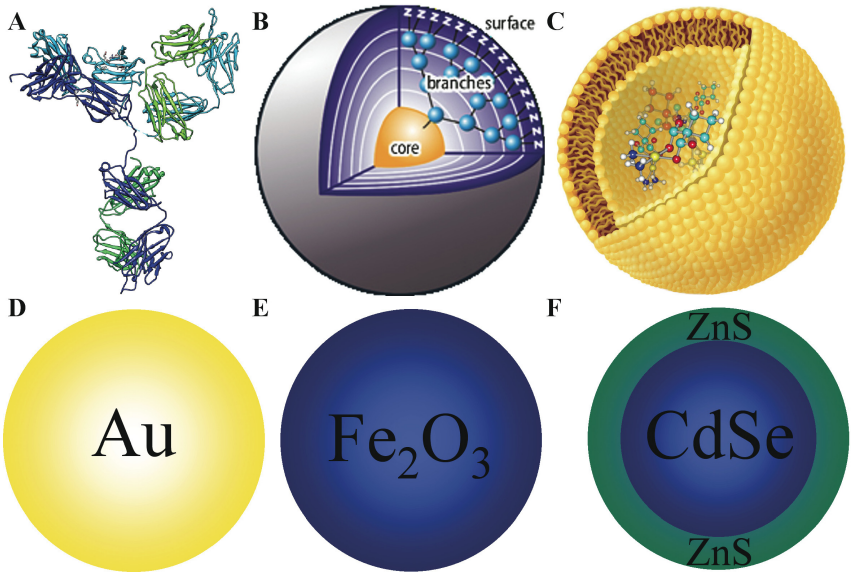


Fig. 1. Types of nanomaterials. (A) Antibodies, antibody 1HZH[24] rendered using Chimera.[25] (B) Dendrimers, image reprinted with permission of Dendritic Nanotechnologies, Inc. (C) Liposomes, image reprinted with permission of Encapsula NanoSciences LLC. (D) Gold nanoshells. (E) Iron oxide nanoparticles. (F) Quantum dots. Not drawn to scale.

replication selective adenovirus, for treatment via specific lysis of squamous cell carcinoma of the head and neck region.[26] In addition, retargeting strategies utilizing adenovirus serotype 5 have also been successful for cancer treatment.[27] Oncolytic herpes simplex virus has been used for anti-cancer therapeutics.[28] Through the addition of an antibody-binding domain Z to polyomavirus, antibodies were attached that directed targeting to ErbB2-positive tumor cell lines.[29] Hepatitis B has been harnessed for targeted delivery of drugs and genes as well.[30] The live attenuated Edmonston B stain of measles virus (MV) has potent oncolytic properties and has been used with[31] and without[32] additional targeting moieties. A modified MV is now in Phase I clinical trials to treat multiple myeloma.[33]

In addition to these natural human pathogens, there exists a subset of non-mammalian viruses, as well as endogenous proteins that form cage-like structures reminiscent of viruses, which are particularly intriguing to nanotechnologists because of several key features. First, they are in the nanometer size range and vary in size and shape ranging from 10 to 100s of nanometers (nm). Second, they self-assemble and therefore do not

require a complex synthesis pathway. Third, each member is monodisperse with equal size and shape within a population. Fourth, known genome sequences and expression systems allow for genetic modification, and many are defined to atomic resolution through X-ray crystallography or cryo-electron microscopy allowing for precise chemical manipulation. Fifth, they exhibit repeated surface and interior patterns that allow for multivalent displays and attachments on multiple levels. Sixth, they may be produced in much higher quantities than the human viruses discussed above, and are non-pathogenic for humans in general. These collective properties have led to their use in intravital imaging, vaccine development, and targeted drug delivery (see Fig. 2).

Fully assembled virus members of this group without nucleic acid are referred to as virus-like particles (VLPs). Modified virions and VLPs used for nanotechnology applications are called viral nanoparticles (VNPs). This chapter will first provide an overview of the use of VLPs and VNPs in nanotechnology, and then focus on cowpea mosaic virus (CPMV) and its use in biomedical applications.

2.1. *VLP and VNP Uses in Imaging*

Intravital imaging is an essential tool for non-invasive diagnosis and management of disease. Toxicity and signal intensity issues hinder currently available imaging agents including small gadolinium conjugates and dye-labeled dextrans or lectins. Multivalent displays of fluorophores

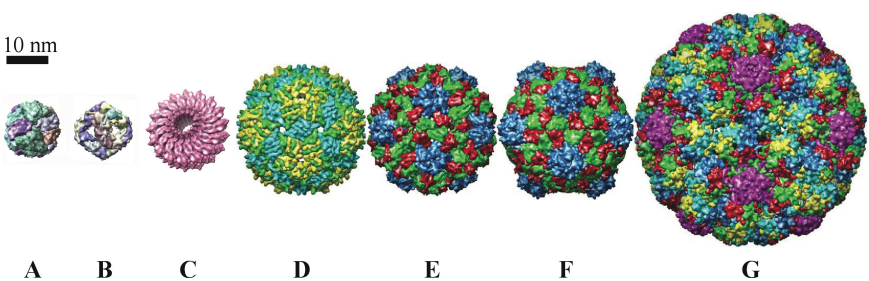


Fig. 2. Examples of viruses and protein cages used in nanotechnology. (A) Ferritin cages 12–14 nm.[34] (B) Heat shock protein cages 12–14 nm.[35] (C) Tobacco mosaic virus intermediates 18 nm wide and variable in length up to 300 nm.[36] (D) Bacteriophage MS2 26 nm.[37] (E) Cowpea chlorotic mosaica virus 28 nm.[38] (F) Cowpea mosaic virus 31 nm.[39] (G) Murine polyomavirus 48.6 nm.[40] Colors indicate domains or subunits. All structures were rendered using Chimera[25] or Viper.[41]

or contrast agents on 'biocompatible platforms' can help resolve these issues.[42–44] Cowpea chlorotic mottle virus (CCMV), Fig. 2E, can be modified on its internal and exterior surfaces, in addition to its capacity to produce non-infectious particles using *in vitro* expression systems.[45] CCMV also has a porous capsid surface, through which added gadolinium ions (Gd^{3+}) can freely diffuse and bind to the 180 interior metal-binding domain of the capsid with a binding affinity of 31 μM.[46] Gadolinium-labeled CCMV forms a paramagnetic nanoparticle with high molecular relaxivity, whereby coordination of the gadolinium ions caused by attachment to the virus capsid slows their molecular rotation in solution. This Gd-modified VLP can then be used as a platform for magnetic resonance imaging (MRI).[46] A slightly different strategy using MS2 bacteriophage, Fig. 2D, and Gd-chelates attached to the MS2 capsid internally through tyrosines and externally through lysines has produced an attractive MRI contrast agent as well.[47] CPMV, Fig. 2F, has also been decorated with Gd and will be discussed further in Section 3.4.2.[48]

2.2. *Use of VLPs and VNPs in Vaccine Development*

Plant viruses and bacteriophages are becoming more frequently utilized for vaccine development because they are non-infectious, inexpensive to produce and provide multivalent antigen display. These multivalently displayed antigens increase the efficacy of the immune response, often requiring no adjuvant since they are composed of repeating units of a foreign protein. One can often evoke a strong immune response using only a single immunization.

Current research follows in the footsteps of several successful examples of VLP-based vaccines already on the market. Hepatitis B surface antigen (HBsAg) is found in two forms during a typical infection: the infectious 42 nm particle and a hollow particle 22 nm in diameter. In 1982, researchers were able to produce the 22 nm particles in the yeast *Saccharomyces cervisiae* and successfully used this nucleic acid-lacking VLP to induce a protective immune response to Hepatitis B.[49] Also using VLP technology, several groups building off of each others' advances in the 1990s were able to produce VLPs capable of eliciting a neutralizing response to the majority of the more than 100 known human papillomaviruses.[50,51] Derivatives of these advancements are used for the current vaccination arsenal under the

better-known commercial names: Engerix-B, Recombivax HB, Gardasil, and Cervarix.

Recent VNPs show equal promise. Through the genetic addition of a highly antigenic epitope from gp41 of human immunodeficiency virus 1 to the coat protein of potato virus X, researchers were able to induce a neutralizing antibody response in the murine model system after intranasal or peritoneal immunization.[52] Alfalfa mosaic virus, modified with antigenic determinants from rabies virus nucleoprotein and glycoprotein, was protective against an infectious challenge after parenteral immunization.[53] Mice peritoneally or subcutaneously immunized with tobacco mosaic virus (TMV), Fig. 2C, that had been genetically fused with a short epitope from the spike protein of murine hepatitis virus (MHV), were able to survive lethal challenge with MHV.[54] Intranasal injection of this modified TMV was also able to elicit a specific immune response; however, ten times more immunogen was needed than when subcutaneously introduced.[54] In a separate study, larger antigens attached via a linker to TMV were more immunogenic.[55] Mice immunized with VNPs of bacteriophage Qbeta (Qß) modified with IL-1ß were found to be strongly protected from arthritis as well.[56] Impressively, the insect virus flock house virus (FHV) that has been modified with protective antigen from *B. anthracis* was able to function as both an antitoxin and a vaccine in a single administration without adjuvant.[57]

2.3. *VNP Use in Targeted Drug Delivery*

Viruses offer great opportunities in the field of targeted drug delivery, perhaps most notably for cancer therapeutics. Cancer cells and cancer neovasculature are different from normal tissues, and display unique surface markers.[58] In addition the vasculature of tumors is leaky and provides a microenvironment that is physically different in accessibility and retention of macromolecules.[59,60] Studies using polymers of vinyl alcohol or liposomes have shown that maximum tumor accumulation and retention size of particles is 100 nm or smaller.[61,62] This is known as the enhanced permeability and retention effect (EPR). VNPs are this ideal size and can be engineered to take advantage of these specialized tumor markers, all while protecting a drug payload from non-specific internalization or toxicity. The ability to multivalently modify multiple sites on the interior and exterior surfaces of VNPs permits the addition of multiple functionalities to each particle. It is becoming relatively straightforward to attach a drug

to the interior of a VLP, thus protecting it from cellular internalization and metabolism, and then add one or more targeting moieties to the outer surface to direct the particle to a highly specific cell type. Direct VNP surface interactions with the vasculature and tumor can be controlled through the addition of polyethylene glycol (PEG) linkers between VNPs and targeting molecules, thus further highlighting the targeting moieties.[63–65]

Several VNPs have been successfully utilized to accomplish payload delivery and targeting. The interior and exterior surfaces of MS2 bacteriophage have been conjugated with mock payloads and target moieties.[66–68] Filamentous M13 phage, popular for use in phage display technologies, was modified to target alpha-integrins and deliver a quantum dot payload to tumors.[69] Payload-modified canine parvovirus is capable of human cell entry through its native receptor, the transferrin receptor, which is upregulated in many types of cancer.[70,71] A heat-shock protein (HSP) cage, Fig. 2B, was also shown to be capable of exterior targeting modifications with optional carrying of a therapeutic payload.[72]

All of these VNPs exhibit broad utility; however, one virus has been used for imaging, vaccine development and targeted drug delivery.

3. Cowpea Mosaic Virus, a Versatile Nanoparticle

Cowpea Mosaic Virus (CPMV) is the type member of the genus *Comovirus*, and is part of the picornavirus superfamily, which spans both the plant and animal kingdoms. CPMV has a positive sense, single-stranded bipartite RNA genome in which RNA1 encodes the replication and protease machinery, and RNA2 encodes both capsid protein subunits as well as a movement protein.[73,74] CPMV is inexpensive to cultivate, and one can isolate high quantities of virus (1 mg per gram of infected leaf mass) from its native plant host the black-eyed pea plant (*Vigna unguiculata*) using non-sterile techniques.[75] These and the physical properties of the CPMV virion, including its structure and the ability to chemically modify its capsid, have led to its use in biomedical nanotechnology.

3.1. *CPMV Structure*

CPMV is a non-enveloped virus, which forms an icosahedral capsid made up of sixty identical asymmetric units.[73] Each asymmetric unit is composed of one small (S) and one large (L) protein subunit in a pseudo- $T = 3$

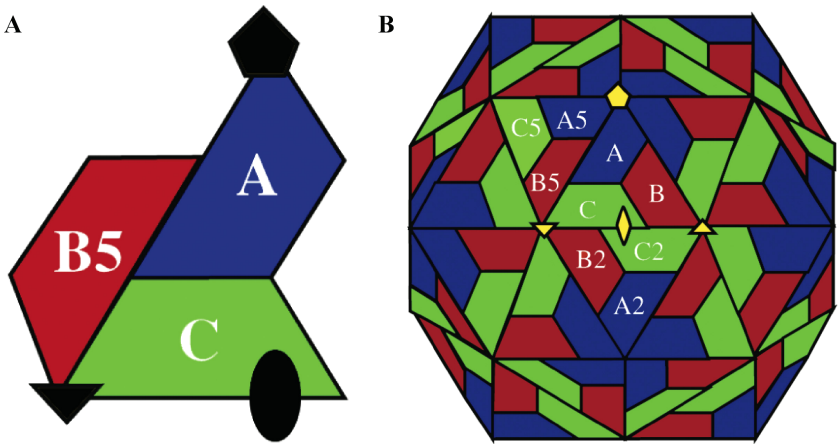


Fig. 3. The CPMV icosahedral viral capsid structure. Comprised of two proteins, the S subunit is in blue and the two domains of the L subunit in green and red. These three domains form a quasi-equivalent T = 3 surface. Icosahedral symmetry axis indicated where pentagon = 5-fold, triangle = 3-fold, and oval = 2-fold. (A) Asymmetric unit. (B) Schematic of capsid. Figure courtesy of Dr Jack Johnson.

configuration (Fig. 3). The X-ray crystallographic structure of the CPMV capsid was determined to 2.8 Å resolution by Lin, *et al.*[39] The small subunits occupy the five-fold axis, and the two β jelly-roll domains alternate around the three-fold axis. This arrangement is ideal for multivalent display because any modification to an individual subunit is repeated sixty times on the entire virion in a highly specific array.

3.2. *Genetic Manipulation of CPMV*

The generation of CPMV mutants was pioneered by George Lomonossoff's laboratory. Full-length CPMV complimentary DNA (cDNA) clones, wild-type (wt) or mutant, were created by placing RNA 1 and RNA 2 downstream of the cauliflower mosaic virus 35S promoter. When these plasmids are linearized and then inoculated on plants, replication is initiated and intact and infectious virions can be produced.[76] Harvest and purification of these virions can be accomplished in about one day using PEG precipitation and sucrose gradient centrifugation.[75]

This genetic manipulation method has been extremely productive in analyzing the function of the CPMV movement protein (MP) and both capsid subunits in the virus life cycle. The functional domains

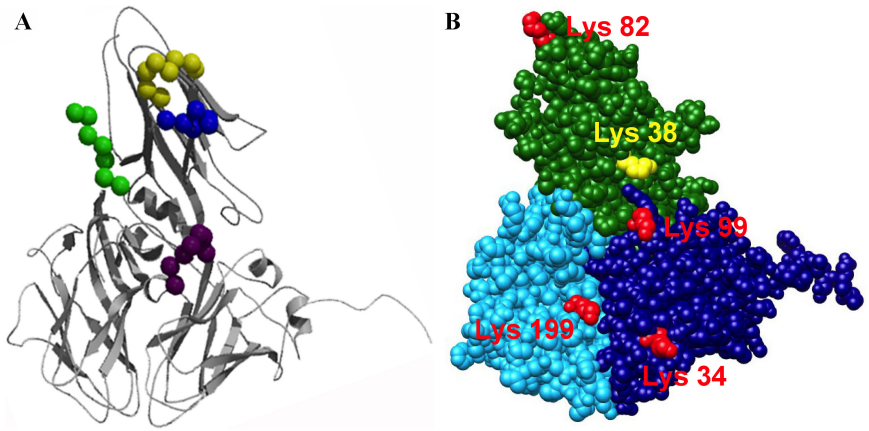


Fig. 4. Asymmetric unit of CPMV highlighting genetic insertion sites and reactive surface lysine residues on the capsid. (A) On the S subunit: the βB-βC loop is highlighted in yellow and the C-terminus of the small subunit is in green. On the L subunit: the ßE-αF loop is highlighted in purple and the C′-C″ loop in blue. (B) Depicted are the small subunit (green) and two β-jelly-roll domains of the large subunit (light and dark blue) with capsid surface reactive lysines indicated in red. Lysine 38 of the small subunit, the most reactive surface lysine, is indicated in yellow. Adapted from Chatterji, *et al.* (2004).[85]

of and transport tubules formed by MP have been elucidated using this mutagenesis.[77–83] There are also several solvent-exposed loops in each asymmetric unit in which short peptide sequences can be inserted. Common modification sites for the small subunit are the βB-βC loop and the C-terminus, and for the large subunit the ßE-αF loop and C′-C″ loops.[84–86] These insertion sites are illustrated in Fig. 4A. Display of large peptides, however, has had mixed success and polypeptides cloned into the βB-βC loop are susceptible to up to 80% proteolytic cleavage.[86]

3.3. *Chemical Modification of CPMV*

The stability of the CPMV capsid and the solvent exposure of several amino acids allows for numerous different chemical attachment schemes. The presence of reactive lysines on the capsid surface and the stability of their reactivity in a wide variety of conditions including temperature, pH, and organic solvents facilitates a wide variety of chemical conjugations.[87] Through mutational analysis the five surface-exposed lysines of each asymmetric unit were subsequently found to be uniquely reactive in

comparison to one another as depicted in Fig. 4B.[85] Small molecules, or large molecules connected to a linker, can be easily attached to these surface lysines through use of an N-hydroxysuccinimidyl ester (NHS).

There are several other capsid attachment schemes in addition to the use of surface lysines. Mutational addition of cysteine residues on the capsid surface, or use of the 30 accessible natural cysteines on the interior of the capsid, allow for another layer of chemical attachment through maleimide chemistry.[88] Nickel binding chemistries can be used via histidine mutation.[89] Decoration of carboxylate groups exposed on the surface of the virion is also possible.[90] Further attachment strategies using copper(I)-catalyzed azide-alkyne cycloaddition reactions ('click chemistry') allow for attachment of complex moieties without the need for protecting groups.[91,92] Researchers armed with an array of possible alteration strategies for CPMV have put these options to use for a wide variety of applications. CPMV has been used in material science, molecular electronics, and the creation of multi-layer assemblies.[93–95] Numerous moieties have been conjugated to the viral surface for material or biological use including display of peptides,[84] polysaccharides,[91,96,97] nucleic acid,[98] biotin,[95] folic acid,[63] gadolinium,[99] and other synthetic structures.[100]

3.4. *In vivo Applications and Mammalian Bioavailability of CPMV*

CPMV VNPs have made many contributions in vaccine development, imaging and targeted drug delivery. Importantly, it was found that CPMV virus particles are non-toxic to the mammalian system, even in high doses as large as 100 mg per kilogram body weight.[99] It is also possible to create virions devoid of RNA that are stable for several weeks, can be chemically modified and are structurally resolved.[101]

3.4.1. *Studies using genetically modified CPMV VNPs*

Modified CPMV particles have been used in many vaccine studies. As with other VNPs, the foreign and highly repetitive capsid of CPMV is often capable of inducing an immune response with minimal use of adjuvants. The first study demonstrating that CPMV could be used as a platform for antigen display genetically inserted an epitope from the VP1 protein of foot-and-mouth disease virus (FMDV) into the small subunit

capsid protein. The resulting chimeras were capable of productive virion formation and systemic spread in the plant host, with the additional ability to react with FMDV-specific antiserum.[102] These VNPs rapidly reverted to wild-type following serial passage, however.

Through the work of Porta, *et al.* in 1994 it was possible to produce stable CPMV chimeras displaying epitopes from human immunodeficiency virus or human rhinovirus 14,[103] of which the latter was also immunogenic in rabbits. Through further refinement, a CPMV chimera bearing a genetic insertion of a mink enteritis virus epitope provided protection from disease and prevention of viral shedding after a single subcutaneous immunization.[104] Several other short peptide epitopes have been successfully displayed on CPMV and were shown to be capable of inducing a specific antibody response, including the *Staphylococcus aureus* fibronectin-binding protein[105] and gp41 from human immunodeficiency virus type 1.[106,107] Most recently CPMV has been used for heterologous protein production of vaccine antigens in plants, without virus purification, to serve as edible vaccines.[108–115]

These applications of the CPMV genetic chimera technology have been successful, but not without limitations. Expression of a peptide on the CPMV capsid surface is limited by the length and isoelectric point of the inserted sequence. Peptides that are long in length or that have high isoelectric points hinder systemic plant infection.[116] In addition, wild-type CPMV (wt-CPMV) and CPMV mutants undergo proteolytic cleavage depending on the position of the genetic insertion.[117,118] It is hypothesized that these different proteolytic cleavages happen in the plant host prior to virus isolation, and loss of genetically altered sequences or reversion is common.[86]

3.4.2. *Studies using chemically modified CPMV VNPs*

Chemical modification of purified virions allows addition of desired moieties without exposing them to proteolytic cleavage in the plant host. There have been several landmark studies using chemically modified CPMV particles *in vivo*, each addressing a key goal of biomedical nanotechnology.

The groups of Huang and Finn created a CPMV VNP used to elicit an immune response to a carbohydrate moiety upregulated on the surface of cancer cells to create an antitumor vaccine. The Tn glycan,

a tumor associated carbohydrate antigen linked with tumor progression, was conjugated selectively to a cysteine mutant of CPMV using either maleimide or bromoacetamide chemistry. When this chemically modified VNP was injected into mice, it was able to induce an immune response yielding high titers of IgG capable of recognizing Tn on the surface of cancer cell lines. This work is a significant advancement because carbohydrates alone typically have low immunogenicity. Attachment to the CPMV capsid allowed for the creation of a strong humoral response *in vivo* and this method is currently being applied to other carbohydrates as well.[119]

Gd has been attached to the CPMV capsid to form a MRI contrast agent using two separate chemical schemes. Gd^{3-} ions have been coordinated to the viral nucleoprotein and Gd(DOTA) analogues have been covalently attached to the capsid surface through azide-alkyne cycloaddition ('click chemistry'). Both methods yielded two- to three-fold enhancement in Gd relaxivity and additional layering of function through specific targeting of these VNPs is been pursued.[48]

Lewis, *et al.* conjugated an estimated 70 fluorescent dyes (AlexaFluor 555) to the CPMV capsid-surface lysines through the use of NHS esters to create a superior vasculature imaging agent. The resultant fluorescent particles (CPMV-A555) were capable of intravital imaging *in vivo* at a depth of 500 microns for at least 72 hours in both normal and tumor vasculature. This improvement in imaging depth over past platforms was attributed to the high-density labeling of fluorophores and the stability of the capsid. The labeled CPMV particles were bound and internalized by the vascular endothelium, yet differentially between venous and arterial vessels. When the CPMV particles were then labeled with polyethylene glycol (PEG) to make (CPMV-PEG-FITC) to inhibit capsid interactions with the vascular endothelium, the mammalian cells were unable to internalize the modified particles. This can be clearly seen in Fig. 5 where endothelial cell internalization of CPMV-A555 can be seen as bright dots within the cells (left panel), and no such internalization dot decoration is seen with CPMV-PEG-FITC (right panel) in the same vasculature when both VNPs were coperfused. The green color observed in the right panel is CPMV-PEG-FITC continuing to circulate in the bloodstream.[44]

The technique of PEG-coating CPMV was further utilized to redirect the targeted delivery of CPMV particles to cancer cells through the folic acid receptor. Folic acid with a PEG spacer was conjugated, via click

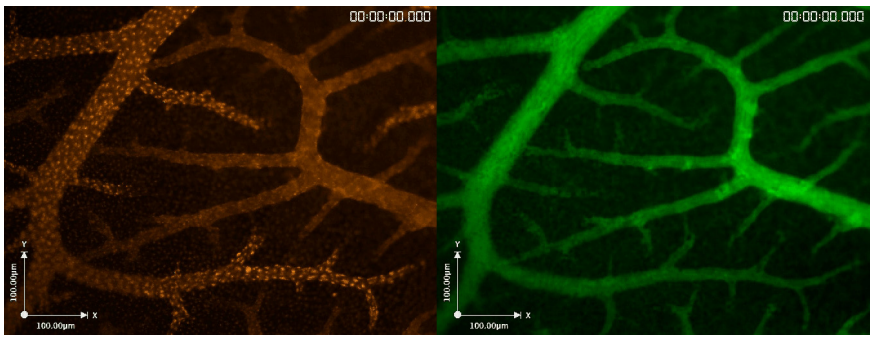


Fig. 5. Intravital imaging illustrates CPMV capsid mediated mammalian uptake. Fluorescent imaging of chick embryo CAM vasculature at 10X magnification, 30 minutes after CPMV-A555 (left) and CPMV-PEG-FITC (right) coperfusion. Figure courtesy of Dr John Lewis.

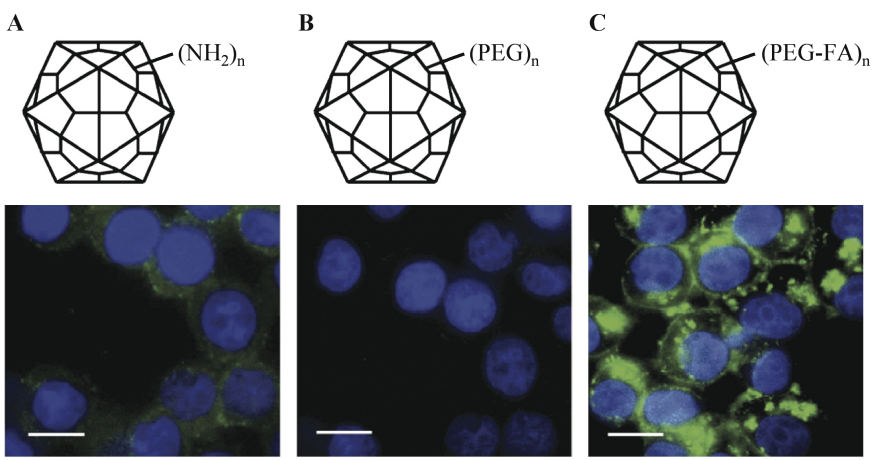


Fig. 6. Folic acid modified CPMV VNPs efficiently target tumor cells. (A) Wildtype CPMV, (B) PEG-modified CPMV or (C) PEG-FA-modified CPMV particles were incubated with KB human cancer cells. Binding and uptake of CPMV VNPs was visualized through fluorescent confocal microscopy where green is CPMV VNPs and blue is nucleic acid. Figure used and adapted with permission from *Chemistry and Biology*.[63]

chemistry, to the viral capsid lysines (CPMV-PEG-FA). Using confocal microscopy, KB human cancer cells incubated with wild-type CPMV (Fig. 6A) illustrated limited binding, PEG-modified CPMV (Fig. 6B) showed no binding, and CPMV-PEG-FA (Fig. 6C) demonstrated targeted

delivery of the CPMV VNPs to cancer cells *in vitro*. The linked folic acids of these modified particles were 10 times more efficient at binding to their receptor then equal molar concentrations of free folic acid *in vitro*.[63]

3.4.3. *Mammalian bioavailability of CPMV*

Although CPMV's replication is confined to plants, the CPMV capsid is bioavailable to the mammalian system, as shown in the vascular imaging and folic acid targeting studies. In addition, thin layers of CPMV were shown to be able to support mammalian cell proliferation and adhesion in tissue culture.[120] Surprisingly, when wt-CPMV was introduced into mice via intravenous injection or oral gavage, CPMV was detected through RT-PCR in wide variety of tissues including the liver, lung, kidney, spleen, stomach, small intestine, lymph nodes, bone marrow and brain.[121] It was also noted that CPMV particles remained intact when exposed to simulated gastric conditions *in vitro*.[121] In biodistribution studies CPMV was found to be cleared from the blood stream in about 20 minutes, and after 30 minutes the majority of the CPMV was found in the liver and spleen, and CPMV was non-toxic in high doses.[99]

Closer examination of the interaction between CPMV and the cell surface has uncovered two clues to mammalian binding and internalization. The Zeta potential of the virion and the Zeta potential of the cell surface were shown to have a role in binding and internalization.[122] Also a 54-kilodalton (kDa), unglycosylated mammalian cell surface protein was identified that specifically interacts with CPMV, Fig. 7, in a dose- and time-dependent manner.[123]

The ability of CPMV to enter cells derived from multiple kingdoms of life is surprising. Only a few viruses, such as flock house virus and bluetongue virus, have made phylogenetic leaps between kingdoms.[124–126] Even though there is currently no evidence that viruses that infect humans have yet come from plants,[127] these studies show that a plant virus does enter human cells *in vitro* and mammalian cells *in vivo*. It is of interest to note that CPMV is part of the picornavirus superfamily, whose members include polioviruses, rhinoviruses, and coxsackieviruses.[123] The further intricacies of how mammalian cells interact with this plant virus are poorly understood and are under further investigation in our laboratory.

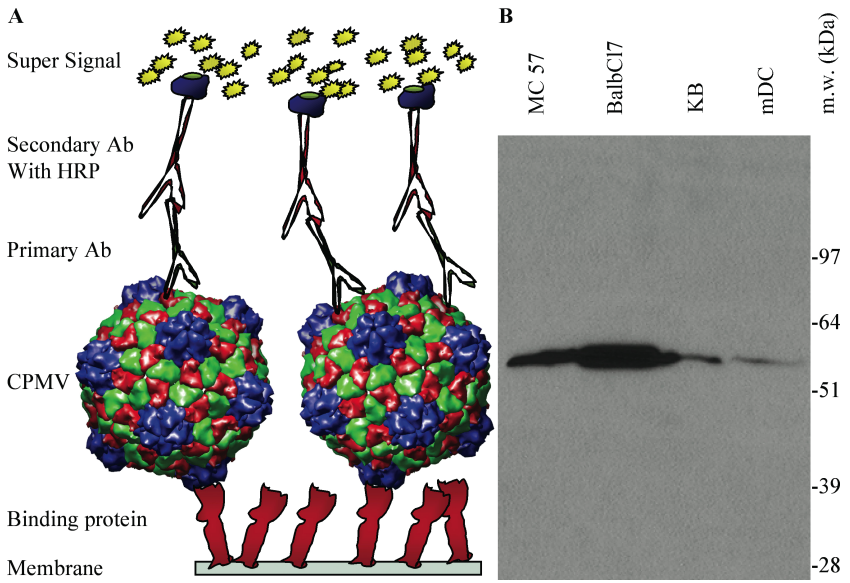


Fig. 7. A conserved 54-kDa mammalian cell membrane protein specifically interacts with CPMV in the Virus Overlay Protein Blot Assay (VOPBA). (A) Schematic of VOPBA method where cell membrane proteins are separated and immobilized on a PVDF membrane, incubated with CPMV particles, and CPMV binding identified by antibody detection much like a Western. (B) Through VOPBA a CPMV specific membrane protein, 54-kDa in size, is found in mouse fibroblast cells MC57 and BalbCl7, human cancer cells KB, and murine derived dendritic cells. Figure used and adapted with permission from *Journal of Virology*.[123]

4. Summary

Viruses have been thriving in a wide variety of biomedical nanotechnology applications including vaccine development, imaging and targeted drug delivery. Viruses are especially suited for these uses because of their unique physical properties: (i) their nanoscale size, (ii) their structure and (iii) the ease with which one may genetically and/or chemically modify their capsids. In the quest to create medicinal tools from these viruses, much has been learned about the viruses themselves.

Cowpea mosaic virus (CPMV)) is an excellent example of this. The CPMV virion's natural attributes allowed for utilization in many medical applications. Through these biomedical uses, CPMV was observed to specifically interact with mammalian cells and facilitate

internalization. Further studies investigating these properties revealed systemic bioavailability of CPMV and specific mammalian cell surface interactions. A better understanding of these qualities will in turn allow for more precise control over CPMV as a therapeutic agent. A constant beneficial cycle of discovery can be observed where physical properties fuel applications and applications uncover additional physical properties.

References

1. K. Petrak. Essential properties of drug-targeting delivery systems. *Drug Discov. Today* **10**:1667–73 (2005).
2. N.G. Portney and M. Ozkan. Nano-oncology: drug delivery, imaging, and sensing. *Anal. Bioanal. Chem.* **384**:620–30 (2006).
3. R.A. Freitas Jr. What is nanomedicine? *Nanomedicine* **1**:2–9 (2005).
4. M. Takeda *et al. In vivo* single molecular imaging and sentinel node navigation by nanotechnology for molecular targeting drug-delivery systems and tailor-made medicine. *Breast Cancer* **15**:145–52 (2008).
5. L. Pardanaud, C. Altmann, P. Kitos, F. Dieterlen-Lievre and C.A. Buck. Vasculogenesis in the early quail blastodisc as studied with a monoclonal antibody recognizing endothelial cells. *Development* **100**:339–49 (1987).
6. N.K. Jain and A. Asthana. Dendritic systems in drug delivery applications. *Expert Opin. Drug Deliv.* **4**:495–512 (2007).
7. A. Quintana, *et al.* Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor. *Pharm. Res.* **19**:1310–6 (2002).
8. X. Shi, *et al.* Dendrimer-entrapped gold nanoparticles as a platform for cancer-cell targeting and imaging. *Small* **3**:1245–52 (2007).
9. R.J. Lee and P.S. Low. Delivery of liposomes into cultured KB cells via folate receptor-mediated endocytosis. *J. Biol. Chem.* **269**:3198–204 (1994).
10. J.J. Gottlieb, K. Washenik, A. Chachoua and A. Friedman-Kien. Treatment of classic Kaposi's sarcoma with liposomal encapsulated doxorubicin. *Lancet* **350**:1363–4 (1997).
11. P.J. Lee and G.A. Peyman. Visualization of the retinal and choroidal microvasculature by fluorescent liposomes. *Methods Enzymol.* **373**:214–33 (2003).
12. D.P. O'Neal, L.R. Hirsch, N.J. Halas, J.D. Payne and J.L. West. Photo-thermal tumor ablation in mice using near infrared-absorbing nanoparticles. *Cancer Lett.* **209**:171–6 (2004).
13. C.S. Yang, *et al.* Nanoparticle-based *in vivo* investigation on blood-brain barrier permeability following ischemia and reperfusion. *Anal. Chem.* **76**:4465–71 (2004).
14. L.R. Hirsch, *et al.* Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. *Proc. Natl. Acad. Sci. USA* **100**:13549–54 (2003).

15. M.G. Harisinghani, *et al.* Noninvasive detection of clinically occult lymph-node metastases in prostate cancer. *N Engl. J. Med.* **348**:2491–9 (2003).
16. J.R. McCarthy, K.A. Kelly, E.Y. Sun and R. Weissleder. Targeted delivery of multifunctional magnetic nanoparticles. *Nanomed* **2**:153–67 (2007).
17. L. Josephson, M.F. Kircher, U. Mahmood, Y. Tang and R. Weissleder. Near-infrared fluorescent nanoparticles as combined MR/optical imaging probes. *Bioconjug. Chem.* **13**:554–60 (2002).
18. X. Wu, *et al.* Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nat Biotechnol.* **21**:41–6 (2003).
19. D.R. Larson, *et al.* Water-soluble quantum dots for multiphoton fluorescence imaging. *in vivo. Science* **300**:1434–6 (2003).
20. M. Cavazzana-Calvo, *et al.* Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**:669–72 (2000).
21. S. Hacein-Bey-Abina, *et al.* A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl. J. Med.* **348**:255–6 (2003).
22. M.G. Ott, *et al.* Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* **12**:401–9 (2006).
23. S.A. Nicklin, E. Wu, G.R. Nemerow and A.H. Baker, The influence of adenovirus fiber structure and function on vector development for gene therapy. *Mol. Ther.* **12**:384–93 (2005).
24. E.O. Saphire, *et al.* Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science* **293**:1155–9 (2001).
25. E.F. Pettersen, *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**:1605–12 (2004).
26. J. Nemunaitis, *et al.* Phase II trial of intratumoral administration of ONYX-015, a replication-selective adenovirus, in patients with refractory head and neck cancer. *J. Clin. Oncol.* **19**:289–98 (2001).
27. A.A. Rivera, *et al.* Combining high selectivity of replication with fiber chimerism for effective adenoviral oncolysis of CAR-negative melanoma cells. *Gene Ther.* **11**:1694–702 (2004).
28. T. Todo, R.L. Martuza, S.D. Rabkin and P.A. Johnson. Oncolytic herpes simplex virus vector with enhanced MHC class I presentation and tumor cell killing. *Proc. Natl. Acad. Sci. USA* **98**:6396–401 (2001).
29. S. Gleiter and H. Lilie, Cell-type specific targeting and gene expression using a variant of polyoma VP1 virus-like particles. *Biol. Chem.* **384**:247–55 (2003).
30. T. Yamada, *et al.* Novel tissue and cell type-specific gene/drug delivery system using surface engineered hepatitis B virus nano-particles. *Curr. Drug Targets Infect Disord.* **4**:163–7 (2004).
31. B.D. Anderson, T. Nakamura, S.J. Russell and K.W. Peng. High CD46 receptor density determines preferential killing of tumor cells by oncolytic measles virus. *Cancer Res.* **64**:4919–26 (2004).
32. K.W. Peng, *et al.* Oncolytic measles viruses displaying a single-chain antibody against CD38, a myeloma cell marker. *Blood* **101**:2557–62 (2003).

33. R.M. Myers, *et al.* Preclinical pharmacology and toxicology of intravenous MV-NIS, an oncolytic measles virus administered with or without cyclophosphamide. *Clin. Pharmacol. Ther.* **82**:700–10 (2007).
34. T.J. Stillman, *et al.* The high-resolution X-ray crystallographic structure of the ferritin (EcFtnA) of *Escherichia coli*; comparison with human H ferritin (HuHF) and the structures of the Fe(3+) and Zn(2+) derivatives. *J. Mol. Biol.* **307**:587–603 (2001).
35. K.K. Kim, R. Kim and S.H. Kim. Crystal structure of a small heat-shock protein. *Nature* **394**:595–9 (1998).
36. K. Namba, R. Pattanayek and G. Stubbs. Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 A resolution by X-ray fiber diffraction. *J. Mol. Biol.* **208**:307–25 (1989).
37. E. Grahn, *et al.* Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *Rna* **7**:1616–27 (2001).
38. J.A. Speir, S. Munshi, G. Wang, T.S. Baker and J.E. Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**:63–78 (1995).
39. T. Lin, *et al.* The refined crystal structure of cowpea mosaic virus at 2.8 A resolution. *Virology* **265**:20–34 (1999).
40. T. Stehle and S.C. Harrison. Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* **4**:183–94 (1996).
41. C.M. Shepherd, *et al.* VIPERdb: a relational database for structural virology. *Nucleic Acids Res.* **34**:D386-9 (2006).
42. C. Kirchner, *et al.* Cytotoxicity of colloidal CdSe and CdSe/ZnS nanoparticles. *Nano Lett.* **5**:331–8 (2005).
43. M.N. Rhyner, *et al.* Quantum dots and multifunctional nanoparticles: new contrast agents for tumor imaging. *Nanomed* **1**:209–17 (2006).
44. J.D. Lewis, *et al.* Viral nanoparticles as tools for intravital vascular imaging. *Nat. Med.* **12**:354–60 (2006).
45. S. Brumfield, *et al.* Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *J. Gen. Virol.* **85**:1049–53 (2004).
46. M. Allen, *et al.* Paramagnetic viral nanoparticles as potential high-relaxivity magnetic resonance contrast agents. *Magn. Reson. Med.* **54**:807–12 (2005).
47. A. Datta, *et al.* High relaxivity gadolinium hydroxypyridonate-viral capsid conjugates: nanosized MRI contrast agents. *J. Am. Chem. Soc.* **130**:2546–52 (2008).
48. D.E. Prasuhn Jr., R.M. Yeh, A. Obenaus, M. Manchester and M.G. Finn. Viral MRI contrast agents: coordination of Gd by native virions and attachment of Gd complexes by azide-alkyne cycloaddition. *Chem. Commun. (Camb)* **28**:1269–71 (2007).
49. P. Valenzuela, A. Medina, W.J. Rutter, G. Ammerer and B.D. Hall. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**:347–50 (1982).
50. R. Kirnbauer, *et al.* Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J. Virol.* **67**:6929–36 (1993).

51. J. Zhou, X.Y. Sun, D.J. Stenzel and I.H. Frazer. Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. *Virology* **185**:251–7 (1991).
52. C. Marusic, *et al.* Chimeric plant virus particles as immunogens for inducing murine and human immune responses against human immunodeficiency virus type 1. *J. Virol.* **75**:8434–9 (2001).
53. V. Yusibov, *et al.* Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine. *Vaccine* **20**:3155–64 (2002).
54. M. Koo, *et al.* Protective immunity against murine hepatitis virus (MHV) induced by intranasal or subcutaneous administration of hybrids of tobacco mosaic virus that carries an MHV epitope. *Proc. Natl. Acad. Sci. USA* **96**:7774–9 (1999).
55. M.L. Smith, *et al.* Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348**:475–88 (2006).
56. G. Spohn, *et al.* Active immunization with IL-1 displayed on virus-like particles protects from autoimmune arthritis. *Eur. J. Immunol.* **38**:877–87 (2008).
57. D.J. Manayani, *et al.* A viral nanoparticle with dual function as an anthrax antitoxin and vaccine. *PLoS Pathog.* **3**:1422–31 (2007).
58. E. Ruoslahti. Specialization of tumour vasculature. *Nat. Rev. Cancer* **2**:83–90 (2002).
59. P. Baluk, H. Hashizume and D.M. McDonald, Cellular abnormalities of blood vessels as targets in cancer. *Curr. Opin. Genet. Dev.* **15**:102–11 (2005).
60. H. Hashizume, *et al.* Openings between defective endothelial cells explain tumor vessel leakiness. *Am. J. Pathol.* **156**:1363–80 (2000).
61. Y. Tabata, Y. Murakami and Y. Ikada. Tumor accumulation of poly(vinyl alcohol) of different sizes after intravenous injection. *J. Control Release* **50**:123–33 (1998).
62. A. Nagayasu, K. Uchiyama and H. Kiwada. The size of liposomes: a factor which affects their targeting efficiency to tumors and therapeutic activity of liposomal antitumor drugs. *Adv. Drug Deliv. Rev.* **40**:75–87 (1999).
63. G. Destito, R. Yeh, C.S. Rae, M.G. Finn and M. Manchester. Folic Acid-Mediated Targeting of Cowpea Mosaic Virus Particles to Tumor Cells. *Chem. Biol.* **14**:1152–1162 (2007).
64. Y. Bae, W.D. Jang, N. Nishiyama, S. Fukushima and K. Kataoka. Multifunctional polymeric micelles with folate-mediated cancer cell targeting and pH-triggered drug releasing properties for active intracellular drug delivery. *Mol. Biosyst.* **1**:242–50 (2005).
65. S.M. Stephenson, P.S. Low and R.J. Lee. Folate receptor-mediated targeting of liposomal drugs to cancer cells. *Methods Enzymol.* **387**:33–50 (2004).
66. M. Wu, W.L. Brown and P.G. Stockley. Cell-specific delivery of bacteriophage-encapsidated ricin A chain. *Bioconjug. Chem.* **6**:587–95 (1995).
67. M. Wu, T. Sherwin, W.L. Brown and P.G. Stockley. Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanomedicine* **1**:67–76 (2005).
68. E.W. Kovacs, *et al.* Dual-surface-modified bacteriophage MS2 as an ideal scaffold for a viral capsid-based drug delivery system. *Bioconjug. Chem.* **18**:1140–7 (2007).

69. L. Chen, *et al.* Design and validation of a bifunctional ligand display system for receptor targeting. *Chem. Biol.* **11**:1081–91 (2004).
70. J.S. Parker, W.J. Murphy, D. Wang, S.J. O'Brien and C.R. Parrish. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. *J. Virol.* **75**:3896–902 (2001).
71. P. Singh, G. Destito, A. Schneemann and M. Manchester. Canine parvovirus-like particles, a novel nanomaterial for tumor targeting. *J. Nanobiotechnology* **4**:2 (2006).
72. M.L. Flenniken, *et al.* Melanoma and lymphocyte cell-specific targeting incorporated into a heat shock protein cage architecture. *Chem. Biol.* **13**:161–70 (2006).
73. G.P. Lomonossoff and J.E. Johnson. The synthesis and structure of comovirus capsids. *Prog. Biophys. Mol. Biol.* **55**:107–37 (1991).
74. G.P. Lomonossoff and M. Shanks. The nucleotide sequence of cowpea mosaic virus B RNA. *Embo. J.* **2**:2253–2258 (1983).
75. J. Wellink. Comovirus isolation and RNA extraction. *Methods Mol. Biol.* **81**:205–9 (1998).
76. J.T. Dessens and G.P. Lomonossoff. Cauliflower mosaic virus 35S promoter-controlled DNA copies of cowpea mosaic virus RNAs are infectious on plants. *J. Gen. Virol.* **74**:889–92 (1993).
77. J. Pouwels, *et al.* Identification of distinct steps during tubule formation by the movement protein of Cowpea mosaic virus. *J. Gen. Virol.* **84**:3485–94 (2003).
78. J. Pouwels, *et al.* Studies on the origin and structure of tubules made by the movement protein of Cowpea mosaic virus. *J. Gen. Virol.* **85**:3787–96 (2004).
79. P. Bertens, J. Wellink, R. Goldbach and A. van Kammen. Mutational analysis of the cowpea mosaic virus movement protein. *Virology* **267**:199–208 (2000).
80. C.M. Carvalho, J. Wellink, S.G. Ribeiro, R.W. Goldbach and J.W. Van Lent, The C-terminal region of the movement protein of Cowpea mosaic virus is involved in binding to the large but not to the small coat protein. *J. Gen. Virol.* **84**:2271–7 (2003).
81. K. Gopinath, *et al.* Intracellular distribution of cowpea mosaic virus movement protein as visualised by green fluorescent protein fusions. *Arch. Virol.* **148**:2099–114 (2003).
82. A. Lekkerkerker, *et al.* Distinct functional domains in the cowpea mosaic virus movement protein. *J. Virol.* **70**:5658–61 (1996).
83. J. Verver, J. Wellink, J. Van Lent, K. Gopinath and A. Van Kammen. Studies on the movement of cowpea mosaic virus using the jellyfish green fluorescent protein. *Virology* **242**:22–7 (1998).
84. A. Chatterji, *et al.* Cowpea mosaic virus: from the presentation of antigenic peptides to the display of active biomaterials. *Intervirology* **45**:362–70 (2002).
85. A. Chatterji, *et al.* New addresses on an addressable virus nanoblock; uniquely reactive Lys residues on cowpea mosaic virus. *Chem. Biol.* **11**:855–63 (2004).
86. T. Lin, C. Porta, G. Lomonossoff and J.E. Johnson. Structure-based design of peptide presentation on a viral surface: the crystal structure of a plant/animal virus chimera at 2.8 A resolution. *Fold Des.* **1**:179–87 (1996).
87. Q. Wang, E. Kaltgrad, T. Lin, J. Johnson and M. Finn. Natural supramolecular building blocks: wild-type cowpea mosaic virus. *Chem. Biol.* **9**:805–11 (2002).

88. Q. Wang, T. Lin, J.E. Johnson and M.G. Finn. Natural supramolecular building blocks. Cysteine-added mutants of cowpea mosaic virus. *Chem. Biol.* **9**:813–19 (2002).
89. A. Chatterji, W.F. Ochoa, T. Ueno, T. Lin and J.E. Johnson. A virus-based nanoblock with tunable electrostatic properties. *Nano Lett.* **5**:597–602 (2005).
90. N.F. Steinmetz, G.P. Lomonossoff and D.J. Evans. Cowpea mosaic virus for material fabrication: addressable carboxylate groups on a programmable nanoscaffold. *Langmuir* **22**:3488–90 (2006).
91. S.S. Gupta, K.S. Raja, E. Kaltgrad, E. Strable and M.G. Finn. Virus-glycopolymer conjugates by copper(i) catalysis of atom transfer radical polymerization and azide-alkyne cycloaddition. *Chem. Commun. (Camb)* **14**:4315–7 (2005).
92. Q. Wang, *et al.* Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**:3192–3 (2003).
93. P. Singh, M.J. Gonzalez and M. Manchester. Viruses and Their Use in Nanotechnology. *Drug Development Research* **67**:23–41 (2006).
94. N.F. Steinmetz and D.J. Evans. Utilisation of plant viruses in bionanotechnology. *Org. Biomol. Chem.* **5**:2891–902 (2007).
95. N.F. Steinmetz, *et al.* Assembly of multilayer arrays of viral nanoparticles via biospecific recognition: a quartz crystal microbalance with dissipation monitoring study. *Biomacromolecules* **9**:456–62 (2008).
96. K.S.W. Raja, Q., and M.G. Finn, Icosahedral virus particles as polyvalent carbohydrate display platforms. *Chembiochem.* **4**:1348–51 (2003).
97. E. Kaltgrad, *et al.* Anti-carbohydrate antibodies elicited by polyvalent display on a viral scaffold. *Chembiochem.* **8**:1455–62 (2007).
98. E. Strable, J.E. Johnson and M.G. Finn. Natural nanochemical building blocks: icosahedral virus particles organized by attached oligonucleotides. *Nano Lett.* **4**:1385–9 (2004).
99. P. Singh, *et al.* Bio-distribution, toxicity and pathology of cowpea mosaic virus nanoparticles *in vivo*. *J. Control Release* **120**:41–50 (2007).
100. N.G. Portney, *et al.* Organic and inorganic nanoparticle hybrids. *Langmuir* **21**:2098–103 (2005).
101. W.F. Ochoa, A. Chatterji, T. Lin and J.E. Johnson. Generation and structural analysis of reactive empty particles derived from an icosahedral virus. *Chem. Biol.* **13**:771–8 (2006).
102. R. Usha, *et al.* Expression of an animal virus antigenic site on the surface of a plant virus particle. *Virology* **197**:366–74 (1993).
103. C. Porta, *et al.* Development of cowpea mosaic virus as a high-yielding system for the presentation of foreign peptides. *Virology* **202**:949–55 (1994).
104. K. Dalsgaard, *et al.* Plant-derived vaccine protects target animals against a viral disease. *Nat. Biotechnol.* **15**:248–52 (1997).
105. A. Rennermalm, *et al.* Antibodies against a truncated Staphylococcus aureus fibronectin-binding protein protect against dissemination of infection in the rat. *Vaccine* **19**:3376–83 (2001).
106. Z. Durrani, *et al.* Intranasal immunization with a plant virus expressing a peptide from HIV-1 gp41 stimulates better mucosal and systemic HIV-1-specific IgA and IgG than oral immunization. *J. Immunol. Methods* **220**:93–103 (1998).

107. L. McLain, *et al.* Stimulation of neutralizing antibodies to human immunodeficiency virus type 1 in three strains of mice immunized with a 22 amino acid peptide of gp41 expressed on the surface of a plant virus. *Vaccine* **14**:799–810 (1996).
108. F. Sainsbury, P.O. Lavoie, M.A. D'Aoust, L.P. Vezina and G.P. Lomonossoff. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. *Plant Biotechnol J.* **6**:82–92 (2008).
109. M.C. Canizares, G.P. Lomonossoff and L. Nicholson, Development of cowpea mosaic virus-based vectors for the production of vaccines in plants. *Expert Rev. Vaccines* **4**:687–97 (2005).
110. M.C. Canizares, L. Nicholson and G.P. Lomonossoff, Use of viral vectors for vaccine production in plants. *Immunol Cell Biol.* **83**:263–70 (2005).
111. L. Liu, *et al.* Cowpea mosaic virus-based systems for the production of antigens and antibodies in plants. *Vaccine* **23**:1788–92 (2005).
112. I.A. Mechtcheriakova, *et al.* The use of viral vectors to produce hepatitis B virus core particles in plants. *J. Virol. Methods* **131**:10–5 (2006).
113. S.G. Yasawardene, G.P. Lomonossoff and R. Ramasamy, Expression & immunogenicity of malaria merozoite peptides displayed on the small coat protein of chimaeric cowpea mosaic virus. *Indian J. Med. Res.* **118**:115–24 (2003).
114. W. Monger, *et al.* An antibody derivative expressed from viral vectors passively immunizes pigs against transmissible gastroenteritis virus infection when supplied orally in crude plant extracts. *Plant Biotechnol. J.* **4**:623–31 (2006).
115. J.P. Phelps, N. Dang and L. Rasochova. Inactivation and purification of cowpea mosaic virus-like particles displaying peptide antigens from Bacillus anthracis. *J. Virol. Methods* **141**:146–53 (2007).
116. C. Porta, *et al.* Cowpea mosaic virus-based chimaeras. Effects of inserted peptides on the phenotype, host range, and transmissibility of the modified viruses. *Virology* **310**:50–63 (2003).
117. K.M. Taylor, *et al.* Influence of three-dimensional structure on the immunogenicity of a peptide expressed on the surface of a plant virus. *J. Mol. Recognit.* **13**:71–82 (2000).
118. K.M. Taylor, *et al.* Position-dependent processing of peptides presented on the surface of cowpea mosaic virus. *Biol. Chem.* **380**:387–92 (1999).
119. A. Miermont, *et al.* Cowpea mosaic virus capsid: a promising carrier for the development of carbohydrate based antitumor vaccines. *Chemistry* **14**:4939–47 (2008).
120. Y. Lin, *et al.* Layer-by-layer assembly of viral capsid for cell adhesion. *Acta Biomater.* **4**:838–43 (2008).
121. C.S. Rae, *et al.* Systemic trafficking of plant virus nanoparticles in mice via the oral route. *Virology* **343**:224–35 (2005).
122. Y. Zhang, *et al.* Zeta potential: a surface electrical characteristic to probe the interaction of nanoparticles with normal and cancer human breast epithelial cells. *Biomed. Microdevices* **10**:321–8 (2008).
123. K.J. Koudelka, C.S. Rae, M.J. Gonzalez and M. Manchester. Interaction between a 54-kilodalton mammalian cell surface protein and cowpea mosaic virus. *J. Virol.* **81**:1632–40 (2007).

124. D.M. White, C.D. Blair and B.J. Beaty. Molecular epidemiology of Bluetongue virus in northern Colorado. *Virus Res.* **6**:6 (2005).
125. L.E. McHolland and J.O. Mecham. Characterization of cell lines developed from field populations of Culicoides sonorensis (Diptera: Ceratopogonidae). *J. Med. Entomol.* **40**:348–51 (2003).
126. J.A. Chao, *et al.* Dual modes of RNA-silencing suppression by Flock House virus protein B2. *Nat. Struct. Mol. Biol.* **12**:952–7 (2005).
127. E.C. Holmes and A. Rambaut. Viral evolution and the emergence of SARS coronavirus. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **359**:1059–65 (2004).

Index