Springer Protocols

Vladimir N. Uversky
A. Keith Dunker *Editors*

Intrinsically Disordered Protein Analysis

Volume 1
Methods and Experimental Tools

Humana Press

METHODS IN MOLECULAR BIOLOGY™

For further volumes:
http://www.springer.com/series/7651

Intrinsically Disordered Protein Analysis

Volume 1, Methods and Experimental Tools

Edited by

Vladimir N. Uversky

Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL, USA
Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia

A. Keith Dunker

Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University, Indianapolis, IN, USA

Editors
Vladimir N. Uversky, Ph.D.
Department of Molecular Medicine
College of Medicine
University of South Florida
Tampa, FL, USA

Institute for Biological Instrumentation
Russian Academy of Sciences
Pushchino, Moscow Region
Russia

A. Keith Dunker, Ph.D.
Department of Biochemistry
and Molecular Biology
Center for Computational Biology
and Bioinformatics
Indiana University
Indianapolis, IN, USA

Preface

Over the past decade, we have witnessed an explosive development of research dedicated to intrinsically disordered proteins (IDPs), which are also known as natively unfolded proteins among various other names. The existence of biologically active but extremely flexible proteins is challenging the century-old structure-to-function paradigm according to which a rigid well-folded 3D structure is required for protein function. Many structural biologists now recognize that the functional diversity provided by disordered regions complements the functional repertoire of ordered protein regions. The high abundance of IDPs in various organisms, their unique structural features, numerous functions, and crucial associations with different diseases show that there are enough grounds to conclude that these proteins should be considered as a unique entity, an unfoldome.

In comparison with "normal" globular proteins, IDPs possess increased amounts of disorder that can be detected by many physico-chemical methods that were originally developed to characterize protein self-organization. On the other hand, due to the highly dynamic nature of IDPs, new and existing experimental methods need to be developed and extended, respectively, for the structural and functional analysis of these IDPs. These methods represent an instrumental foundation for experimental unfoldomics.

Information based on modern protocols is provided herein on virtually every experimental method used both to identify IDPs and to analyze their structural and functional properties. Hence, this book will be of interest to all scientists and students studying IDPs, whether the focus is on an IDP's (lack of) structure or on its function.

The general audience for this book includes scientists working in the fields of biochemistry, biophysics, molecular medicine, biotechnology, pharmacology and drug discovery, molecular and cellular biology; Students of Medical Schools, departments of Biochemistry, Biophysics, Molecular Biology, Biotechnology, and Cell Biology, to name a few. We are aware that many scientists have encountered IDPs in their research, but have shied away from deeper studies due to the lack of knowledge of what to try next. By collecting the current methods for the analysis of IDPs in one place, our goal is to help such scientists further their investigations of these fascinating, dynamic molecules.

Tampa, FL, USA   *Vladimir N. Uversky, Ph.D.*
Indianapolis, IN, USA   *A. Keith Dunker, Ph.D.*

Contents

PART III   VIBRATIONAL SPECTROSCOPY

PART IV   OTHER SPECTROSCOPIC TECHNIQUES

Contributors

LEONARDO G. ALONSO • *Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, Buenos Aires, Argentina*

DILETTA AMI • *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy*

BEATA BEKEI • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

VALÉRIE BELLE • *Bioénergétique et Ingénierie des Protéines, UMR 7281 CNRS and Aix-Marseille Université, Marseille, France*

PAU BERNADÓ • *Centre de Biochimie Structurale, CNRS UMR 5048 - UM 1 - INSERM UMR 1054, Montpellier, France*

MARTIN BLACKLEDGE • *Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA; CNRS, Grenoble, France*

M. BOKOR • *Research Institute for Solid State Physics and Optics, Hungarian Academy of Sciences, Budapest, Hungary*

WADE BORCHERDS • *Department of Cell Biology, Microbiology, and Molecular Biology and the Center for Drug Discovery and Innovation, University of South Florida, Tampa, FL, USA*

GINKA S. BUCHNER • *Department of Chemistry, University of Wyoming, Laramie, WY, USA*

LUCÍA B. CHEMES • *Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, Buenos Aires, Argentina*

EEFEI CHEN • *Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA, USA*

STEFANO CIURLI • *Department of Agro-Environmental Science and Technology, University of Bologna, Rimini, Italy*

GARY W. DAUGHDRILL • *Department of Cell Biology, Microbiology, and Molecular Biology and the Center for Drug Discovery and Innovation, University of South Florida, Tampa, FL, USA*

MANLI DENG • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

APRATIM DHAR • *Department of Chemistry, University of Illinois, Urbana-Champaign, IL, USA*

SILVIA MARIA DOGLIA • *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy*

ALEXANDER DOSE • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

DAVID ELIEZER • *Department of Biochemistry and Program in Structural Biology, Weill Cornell Medical College, New York, NY, USA*

YVES ENGELBORGHS • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

HANNAH GELMAN • *Department of Chemistry, University of Illinois, Urbana-Champaign, IL, USA*

MARTIN GRUEBELE • *Department of Chemistry, Department of Physics, and Center for Biophysics and Computational Biology, University of Illinois, Urbana-Champaign, IL, USA*

ANTOINE GRUET • *Architecture et Fonction des Macromolécules Biologiques, UMR 7257 CNRS and Aix-Marseille Université, Marseille, France*

BRUNO GUIGLIARELLI • *Bioénergétique et Ingénierie des Protéines, UMR 7281 CNRS and Aix-Marseille Université, Marseille, France*

ELISHA HAAS • *The Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel*

JOHNNY HABCHI • *Architecture et Fonction des Macromolécules Biologiques, UMR 7257 CNRS and Aix-Marseille Université, Marseille, France*

MAARIT HELLMAN • *Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland*

MICHAELA HERZIG • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

YUTAKA ITO • *Department of Chemistry, Tokyo Metropolitan University, Tokyo, Japan*

MALENE RINGKJØBING JENSEN • *Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA; CNRS, Grenoble, France*

STEPAN KASHTANOV • *Department of Physics, University of Idaho, Moscow, ID, USA*

DAVID S. KLIGER • *Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA, USA*

MALGORZATA KLUBA • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

EBERHARD KRAUSE • *Department of Chemical Biology and Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

JAN KUBELKA • *Department of Chemistry, University of Wyoming, Laramie, WY, USA*

IRINA M. KUZNETSOVA • *Laboratory of Structural Dynamics, Stability and Folding of Proteins of the Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia*

VLADIMIR LADIZHANSKY • *Department of Physics and Biophysical Interdepartmental Group, University of Guelph, Guelph, ON, Canada*

JENNIFER C. LEE • *Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA*

CASEY H. LONDERGAN • *Department of Chemistry, Haverford College, Haverford, PA, USA*

SONIA LONGHI • *Architecture et Fonction des Macromolécules Biologiques, UMR 7257 CNRS and Aix-Marseille Université, Marseille, France*

MARLÈNE MARTINHO • *Bioénergétique et Ingénierie des Protéines, UMR 7281 CNRS and Aix-Marseille Université, Marseille, France*

JESSIKA MEUVIS • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

TSUTOMU MIKAWA • *Advanced Science Institute, RIKEN, Saitama, Japan*

ANTONINO NATALELLO • *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy*

SANGEETA NATH • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

PAOLO NEYROZ • *Department of Biochemistry "G. Moruzzi", University of Bologna, Rimini, Italy*

MARÍA G. NOVAL • *Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, Buenos Aires, Argentina*

PERTTU PERMI • *Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland*

EUGENE A. PERMYAKOV • *Institute for Biological Instrumentation of the Russian Academy of Sciences, Pushchino, Moscow Region, Russia*

CANDACE M. PFEFFERKORN • *Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA*

OLGA I. POVAROVA • *Laboratory of Structural Dynamics, Stability and Folding of Proteins of the Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia*

GONZALO DE PRAT-GAY • *Protein Structure, Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas, Buenos Aires, Argentina*

MAXIM PRIGOZHIN • *Department of Chemistry, University of Illinois, Urbana-Champaign, IL, USA*

HONOR MAY ROSE • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

LOÏC SALMON • *Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA; CNRS, Grenoble, France*

SARAH SCHREURS • *Laboratory of Biomolecular Dynamics, University of Leuven, Leuven, Belgium*

DIRK SCHWARZER • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

REINHARD SCHWEITZER-STENNER • *Department of Chemistry, Drexel University, Philadelphia, PA, USA*

PHILIPP SELENKO • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

YOSEF SHAUL • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*

LICHI SHI • *Department of Physics and Biophysical Interdepartmental Group, University of Guelph, Guelph, ON, Canada*

BRIAN O. SMITH • *School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK*

JONATHAN B. SOFFER • *Department of Chemistry, Drexel University, Philadelphia, PA, USA*

HEIKE STEPHANOWITZ • *Department of Chemical Biology and Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

ANNA I. SULATSKAYA • *Laboratory of Structural Dynamics, Stability and Folding of Proteins of the Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia*

ROSSUKON THONGWICHIAN • *Department of NMR-assisted Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Berlin, Germany*

SIOBHAN TOAL • *Department of Chemistry, Drexel University, Philadelphia, PA, USA*

K. TOMPA • *Research Institute for Solid State Physics and Optics, Hungarian Academy of Sciences, Budapest, Hungary*

P. Tompa • *Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary*

Peter Tsvetkov • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*

Konstantin K. Turoverov • *Laboratory of Structural Dynamics, Stability and Folding of Proteins of the Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia*

Vladimir N. Uversky • *Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL, USA; Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia*

Daniel Verbaro • *Department of Chemistry, Drexel University, Philadelphia, PA, USA*

Hongwei Wu • *Department of Cell Biology, Microbiology, and Molecular Biology, the Center for Drug Discovery and Innovation, University of South Florida, Tampa, FL, USA*

Hailiu Yang • *Department of Chemistry, Haverford College, Haverford, PA, USA*

F. Marty Ytreberg • *Department of Physics, University of Idaho, Moscow, ID, USA*

Part I

Assessing IDPS in the Living Cell

Chapter 1

Determination of IUP Based on Susceptibility for Degradation by Default

Peter Tsvetkov and Yosef Shaul

Abstract

Intrinsically unstructured proteins (IUPs) like the structured proteins are subjected to proteasomal degradation. However, unlike the structured ones, there is no crucial need of protein unfolding step to access the IUPs to the 20S catalytic subunit of the proteasome. This distinctive behavior set the stage for operational definition of the IUPs based on their susceptibility to the 20S degradation in a cell free system. Numerous studies revealed that this is the case in the cells as well, although no comprehensive analysis was performed to date. IUPs are degraded by the 20S proteasome subunit by default, without being poly-ubiquitinated or undergoing any other modifications. IUPs escape the process of degradation by default by a number of mechanisms, of which a more general one is interaction with a partner named nanny. Based on these attributes one can define IUP by conducting a set of cell free and cell culture experiments as outlined in this chapter.

Key words: Protein degradation, 20S proteasome, In vitro proteasomal degradation, Degradation by default, Nanny, Intrinsically unstructured proteins, Pulse-chase, Protein half-life

1. Introduction

Protein homeostasis dictates cell fate. It is therefore critical to understand the molecular principles determining the steady state levels of proteins. The topic of this chapter is identifying intrinsically unstructured proteins (IUPs) in cells by their mode of accumulation. We describe a series of experiments that are important in understanding whether and to what extent a protein is unstructured in cells.

There are few molecular principles that pave the way for our experimental settings. It is long been believed that proteins are inherently stable unless are actively destabilized. This is the case with the ubiquitin proteasome system whereby a protein substrate needs to be actively recognized by a specific E3 ligase that covalently attaches the ubiquitin polychain, the protein death tag.

The polyubiquitin chain is then recognized by the 26S proteasomes (composed of the 19S and 20S particles) where the polyubiquitin chain is first removed, the substrate is then unfolded to access the 20S proteolytic particle for degradation (1, 2). Lately a novel proteasomal degradation pathway has been described that involves the 20S proteasome but not the 19S particle. This degradation pathway needs no prior modification and was therefore termed degradation by default (3). IUPs are selectively susceptible to degradation by the 20S proteasome due to their unstructureness (4). This differential behavior set the possibility to operationally define IUPs in cells (4).

The second principle is that the labile IUPs can be stabilized by an interacting partner that we termed "nanny" (5). For example, Hdmx, a p53 N-terminal binding protein, protects p53 from degradation by default (6). There are additional examples in the literature whereby the co expression of IUPs with its interacting partner results in the stabilization of the IUP. In many cases these attributes were overlooked, as the underlying mechanisms remained obscure. In fact this inherent feature of the IUPs can be utilized to identify new IUP interacting candidates in cells.

Thus the two experimental approaches, namely, IUPs undergoing degradation by default and escaping this degradation by binding partners (nannies), set the stage for experimental strategies in characterizing or identifying IUPs in the cells. This sort of experimental setting is challenging and might encompass difficulties but is doable in a conventional cell biology laboratory.

The critical steps are as follows:

1. Examine whether your protein is a potential IUP by utilizing protein structure prediction software.

2. Validate unstructureness by susceptibility to in vitro degradation by the 20S proteasome (Subheading 3.1).

3. Confirm in vivo degradation by default by demonstrating rapid ubiquitin-independent proteasomal degradation in cells (Subheading 3.2).

4. Demonstrate stabilization of the IUP candidate by an interacting partner both in vitro and in the cells (Subheading 3.3).

2. Materials

2.1. In Vitro Degradation by the 20S Proteasomes

2.1.1. 20S Proteasome Preparation

1. Livers extracted from C57BL mice.

2. Ammonium sulfate.

3. Equipment: POLYTRON homogenizer, Teflon homogenizer, Superpose 6 prep grade gel-filtration column (GE health care), resource-Q ion-exchange column (GE health care), Amicon Ultra centrifugal filter.

4. Buffer A: 20 mM Tris–HCl pH 7.5, 1 mM DTT, 1 mM EDTA, 250 mM sucrose.

5. Buffer B: 20 mM Tris–HCl pH 7.5, 1 mM DTT, 20 % Glycerol.

6. Buffer C: 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM MgCl, 1 mM DTT.

7. Glycerol gradient is made with Buffer C containing either 10 or 40 % glycerol.

2.1.2. Preparation of [^{35}S] Methionine Labeled IUPs

1. TNT Quick coupled transcription/translation system (Promega).

2. Plasmid containing the cDNA of the protein of interest under T7 promoter.

3. [^{35}S] methionine (10 mCi/ml; Amersham).

2.1.3. In Vitro Degradation by the 20S Proteasome

1. Degradation (DEG) buffer: 100 mM Tris–HCl 7.5, 150 mM NaCl, 5 mM MgCl$_2$, 1 mM DTT.

2. Water bath set at 37°C.

3. Laemmli sample buffer X3 (SBX3): 180 mM Tris–HCl 6.8, 6 % SDS, 30 % glycerol, 2.19 M β-mercaptoethanol, 0.01 % bromophenol blue.

4. SDS PAGE was comprised of the lower gel: 10–15 % acrylamide (depending on size of protein detected) 375 mM Tris–HCl 8.8, 0.1 % SDS, 0.1 % APS, and 0.001 % TEMED. The upper stacking gel: 4.5 % acrylamide, 0.13 M Tris–HCl 6.8, 0.1 % SDS, 0.1 % APS, and 0.001 % TEMED.

2.2. Rapid Ubiquitin-Independent Proteasomal Degradation of the IUP Candidate in Cells (Degradation by Default)

2.2.1. Protein Half-Life Determination

1. Cell types—NIH3T3, A31N-ts20 mouse fibroblasts. Human embryonic kidney (HEK) 293T cells.

2. Transfection reagents: for NIH3T3 cells JetPEI (PolyPlus Transfection), A31N-ts20 were transfected with Lipofectamine 2000 (Invitrogen) and HEK293T cell by the method of calcium phosphate.

3. Cycloheximide (CHX) stock solution of 100 mg/ml (Sigma).

4. Complete growth medium: Dulbecco's modified Eagle's Medium (DMEM, Invitrogen, Carlsbad, CA) containing 8 % fetal bovine serum.

5. Cold Sterile phosphate-buffered saline (PBS).

6. Trypsin solution (Biological industries).

7. Labeling medium (M−): Methionine free DMEM with dialyzed serum (M−) (Biological industries) with the addition of [^{35}S] methionine (10 mCi/ml; Amersham) at the final concentration of 0.2 mCi/ml.

8. Chase medium (M+): Complete growth medium with the addition of 1 mg/ml of L-Methionine (Sigma).

9. Flag/Ha beads (Sigma) or protein A/G beads preconjugated to an antibody of interest (421 mouse anti p53 antibody).

10. RIPA buffer: 50 mM Tris–HCI 8, 150 mM NaCl, 1 % NP40, 0.5 % deoxicholate, 0.1 % SDS. Prior to lysis, 1 mM DTT, 1/100 cocktail of protease inhibitors (Sigma), and 25 μM MG132 are added.

2.2.2. 26S Proteasome Reduction by Knocking Down Psmd1

1. Cell types: HEK 293T for virion production, MCF10A.

2. Plasmids: pTRIPZ Psmd1 shRNA (Open Biosystems), pCMV-dR8.91, and VSV-G plasmids for lenti virion production.

3. Antibiotics: puromycin, doxycycline.

3. Methods

For the determination of an IUP the first and the easiest step is to test the sequence of the protein of interest in silico. Numerous IUP prediction software packages are available and simple to use in the Web. The prediction methodologies are not described here, but the different prediction tools have been previously reviewed and specified (7). Once a protein is predicted (preferably by several prediction methods) to be an IUP it is suggested to move on to the next step of validation of the protein unstructured features in cells by the steps described below.

3.1. In Vitro Degradation by the 20S Proteasomes   IUPs are susceptible to the 20S proteasome degradation in vitro, whereas globular proteins are resistant (4). The rationale is that in the absence of the 19S regulatory subunit the structured proteins would not be unfolded and therefore inaccessible to the 20S proteolytic activity. Thus, digestion of a protein by the 20S proteasome in a cell free system provides the first evidence for it being an IUP. Furthermore, to substantiate this evidence the nanny principle can be employed. To this end, incubation with any interacting protein (may also include a specific monoclonal antibody) should rescue IUP from the 20S proteasomal degradation (Fig. 1).

3.1.1. 20S Proteasome Purification from Mice Livers   Commercial 20S samples are available that can be used for the experiments outlined below. Nevertheless, we describe our protocol of 20S proteasome preparation from mouse livers.

1. Mouse livers were homogenized in buffer A using POLY-TRON homogenizer and then Teflon homogenizer.

2. The homogenate was centrifuged using a 45TI rotor at $125,000 \times g$ for 1 h at 4°C and the supernatant was then
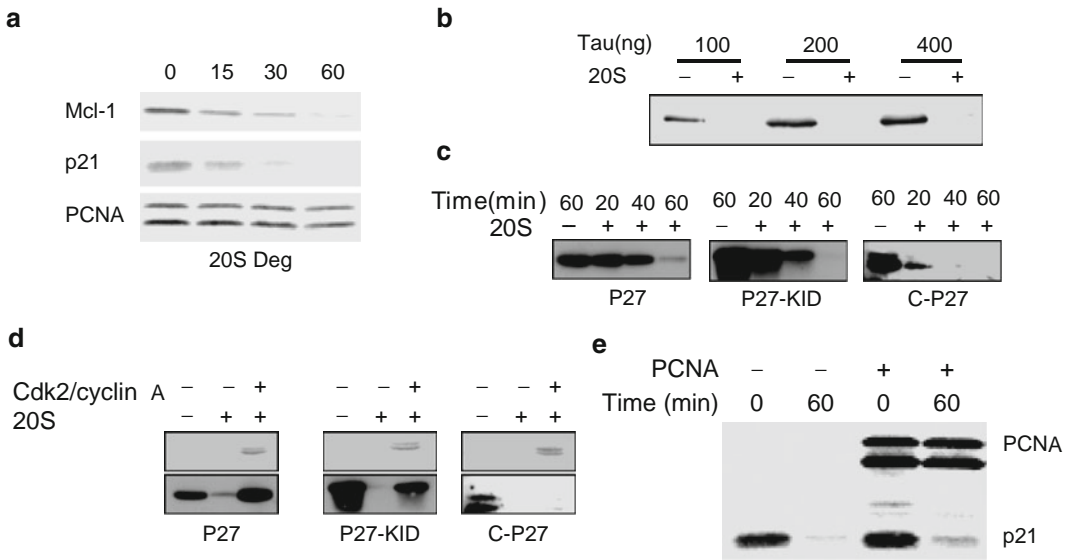
Fig. 1. In vitro degradation assay. (**a**) In vitro translated [^{35}S] methionine labeled proteins, Mcl-1 and p21 IUPs and PCNA a structured protein, were incubated in the degradation reaction with 1 μg of purified 20S proteasomes for the indicated time points. (**b**) the degradation of increasing concentrations of purified tau protein were examined following incubation with 1 μg of purified 20S proteasomes for 60 min at 37°C. (**c**) The full-length p27 (1–198), KID (1–98) and C terminus of p27 (98–198) were incubated in the presence of purified 20S proteasomes for the indicated times for up until an hour. (**d**) 1 mM of purified full-length p27 (1–198) KID-p27 (1–98) and C terminus p27 (98–198) were incubated in the presence of 1 μg purified 20S proteasomes for indicated time points. (**e**) p21 alone or in the presence of equal amount of PCNA were incubated in the presence of purified 20S proteasomes for 1 h at 37°C. In vitro translated proteins (p21, PCNA, Mcl-1) were detected by autography, tau by immunoblotting with mouse anti tau antibody and P27 truncations and wt were detected by immunoblotting with anti His antibody. Cdk2/Cyclin A complex was detected by Ponceau staining.

subjected to the first precipitation with 40 % (w/v) saturated ammonium sulfate followed by a second precipitation of the resulting supernatant with 80 % (w/v) saturated ammonium sulfate. The pellet that was obtained after centrifugation at 125,000 × g, using 45TI rotor for 1/2 an hour at 4°C was resuspended in a minimal volume of buffer B and dialyzed 5–16 h against buffer B (to get rid of ammonium sulfate).

3. The resuspended pellet was then loaded on to a Superpose 6 prep grade gel-filtration column. Collected fractions were analyzed for the presence of 20S proteasomes by Western blot analysis.

4. The fractions containing the proteasome were combined and loaded on a resource-Q ion-exchange column. Proteins were eluted with a gradient of NaCl and fractions were analyzed for the presence of 20S proteasomes by Western blot.

5. The fractions containing the proteasome were combined and dialyzed overnight against buffer C.

6. The sample was first concentrated with Amicon Ultra centrifugal filter and then loaded on an 11 ml linear 10–40 % glycerol gradient and centrifuged using SW41 rotor at $100,000 \times g$ for 16 h at 4°C. 500 μl fractions were collected and analyzed for the presence of 20S proteasomes by Western blot. The fractions containing the proteasome were combined and dialyzed overnight against buffer C.

7. The sample was concentrated with Amicon Ultra centrifugal filter, divided into aliquots, frozen, and stored at −80°C.

3.1.2. Preparation of IUPs for 20S Proteasomal Degradation Assay

The IUP candidate is prepared for the 20S proteasomal degradation assay by two main methods. (a) Recombinant protein is prepared from bacteria and purified by using conventional methods (which we shall not describe here). (b) The IUP candidate is in vitro translated and [^{35}S] methionine labeled. Due to the simplicity of the in vitro translation system it is particularly recommended when examining many different protein substrates. We utilize the TNT quick-coupled transcription/Translation system (Promega) that utilizes the T7 RNA polymerase to translate proteins of interest. For the in vitro reaction the plasmid of an IUP of interest needs to be inserted into a plasmid containing the T7 promoter and then translated according to the manufacturer's instructions. If examining the property of a protein it is best to utilize a purified protein to avoid adverse affects originated from the reticulocyte proteins.

3.1.3. In Vitro Degradation of the IUP Candidate by the 20S Proteasome

1. Add to each sample either 0.2–0.5 μl of the in vitro translated [^{35}S] methionine labeled protein mix or between 0.05 and 1 μg of bacterially expressed and purified protein (see Note 2).

2. Add 1 μg of purified 20S proteasomes and incubate the samples at 37°C. Different time points can be collected for the degradation experiment for up until 2 h (see Note 3 and Fig. 1).

3. The degradation reaction is performed in the final volume of 30 μl in DEG buffer. The final volume and the amount of protein can be proportionally increased if many time points are assayed.

4. Stop the reaction by the addition of 15 μl SBx3 and incubation at 95°C for 5 min. The samples are kept on ice or at −20°C until subjected to SDS-PAGE.

5. For detection the gels can be either dried under vacuum in gel dryer apparatus for 2 h at 65°C or the proteins can be transferred to nitrocellulose membrane that enables not only the identification of radiolabeled protein but also protein detection by immunoblot analysis.

6. Levels of [^{35}S] methionine labeled proteins are detected by autography and quantified, whereas levels of purified proteins

can be either detected by gel code staining or immunoblot analysis with the appropriate antibodies.

7. Positive control such as known IUPs as p21 and negative controls should be utilized in the experiments.

3.1.4. In Vitro Stabilization Assay of the IUP Candidate

1. The experiment in Subheading 3.1.3 is repeated only this time the known interacting partner (a nanny candidate) is added in the same quantities and incubated for 5 min in room temp before adding the 20S proteasome sample (see Note 4). In case the in vitro translated protein is utilized for this assay it is important to add an equal amount of reticulocyte extract and a nonbinding protein to the IUP as controls.

2. The reactions in the presence or absence of the interacting partner are then subjected to the 20S proteasomal degradation assay by the addition of 1 μg of purified 20S proteasomes.

3. Samples are incubated at 37°C for different time points and the reaction is stopped by the addition of SBx3 and incubation at 95°C for 5 min. Detection is done as described above.

3.2. Rapid Ubiquitin-Independent Proteasomal Degradation of the IUP Candidate in Cells (Degradation by Default)

The protein candidate that proved IUP positive based on in vitro (cell free system) analysis should be further validated in the cells. IUPs are predicted to be susceptible for degradation by the 20S proteasome in the cells as well, therefore shall exhibit *ubiquitin-independent* (UI) proteasomal degradation. There are several methodologies that can be utilized to demonstrate a UI process as previously described (8). The basic strategy is to demonstrate that IUPs undergo proteasomal degradation even when the ubiquitin system is inactive. In other words a protein can be suspected as an IUP in cells if inhibition of the ubiquitin-26S proteasome machinery does not compromise its instability. This is a crucial step for determining that it behaves indeed as an IUP in cells. The inactivation of the ubiquitin-26S proteasome pathway can be achieved as follows:

1. Preventing the ubiquitination by eliminating all the lysine residues in the protein of interest by site directed mutagenesis to substitute the amino acids lysines (K) to arginines (R) (K-less mutant). The polyubiquitin chain is attached to lysine residues. K-less mutant do not undergo polyubiquitination (except some cases where a protein can be N-terminally ubiquitinated). If the K-less mutant protein undergoes proteasomal degradation it is considered a ubiquitin independent process (9, 10) (see Subheading 3.2.1).

2. Inactivation of the E1 ubiquitin ligase. Inactive E1 ligase will result in the depletion of activated ubiquitin and therefore the process of protein ubiquitination is essentially eliminated. The inactivation of E1 can be achieved by utilization of specific cell lines such as A31N-ts20 or tsBN75 cells harboring an E1

temperature sensitive mutant. Under the restrictive temperature the ubiquitin system is inactivated, whereas the ubiquitin independent proteasomal degradation should remain active. For example under restrictive temperature (39°C) p53 undergoes proteasomal degradation (ref. 6 and Fig. 2c), suggesting that it undergoes UI degradation. This strategy was extensively utilized to show that numerous proteins undergo UI degradation such as Mcl-1 (10), c-Fos (11), and some other published proteins. Moreover, pharmacological inhibition (by compounds as PYR-41) or specific knockdown of the E1 ubiquitin ligase can be also utilized to achieve the results, although these strategies are not elaborated here (see Subheading 3.2.2).

3. Reducing the levels of the 26S proteasome by specific knockdown of crucial components in the 19S regulatory particle. As the ubiquitin-mediated degradation is executed by the 26S proteasome (and not the 20S), it is possible to block this process by selectively reducing the 26S proteasome levels in the cells. As there is no known specific inhibitor for the 26S proteasome the reduction of the 26S proteasome can be achieved by knockdown of one of the 19S complex subunits. We have found that depletion of Psmd1, a 19S subunit eliminates 19S complex formation (6) (see Subheading 3.2.3).

3.2.1. Stability of IUP Is Not Affected by Elimination of Lysines

The first strategy is construction and employment of a K-less mutant. If degradation is not much compromised under this condition the substrate is likely to undergo UI degradation, provided that the substrate does not undergo N-terminal ubiquitination (see Note 5). To measure substrate stability in the cells one has to determine its half-life. Therefore, we describe here two basic methodologies either by inhibiting protein translation using CHX or conducting pulse-chase experiments by radiolabeling the substrate using [^{35}S] methionine. These methodologies can be utilized for overexpression experiments to analyze a substrate and its K-less mutant half-life. Moreover these methodologies can be implemented in determining the rate of accumulation of endogenous and ectopically expressed proteins of interest in general (see also Fig. 2).

Determination of the Protein Half-Life by CHX

1. In the case of determination of endogenous protein half-life seed cells (NIH3T3 or cells of interest) at the density of 100,000 cells/cm². A plate for each time point is required.

2. 24 h after seeding treat cells with 100 μg/ml of CHX

3. Collect cells at different time points of both treated and untreated plates. Protein detection and analysis is performed as described below (Subheading 3.3.2)

4. For the determination of half-life of a transfected protein, also see Subheading 3.3.2
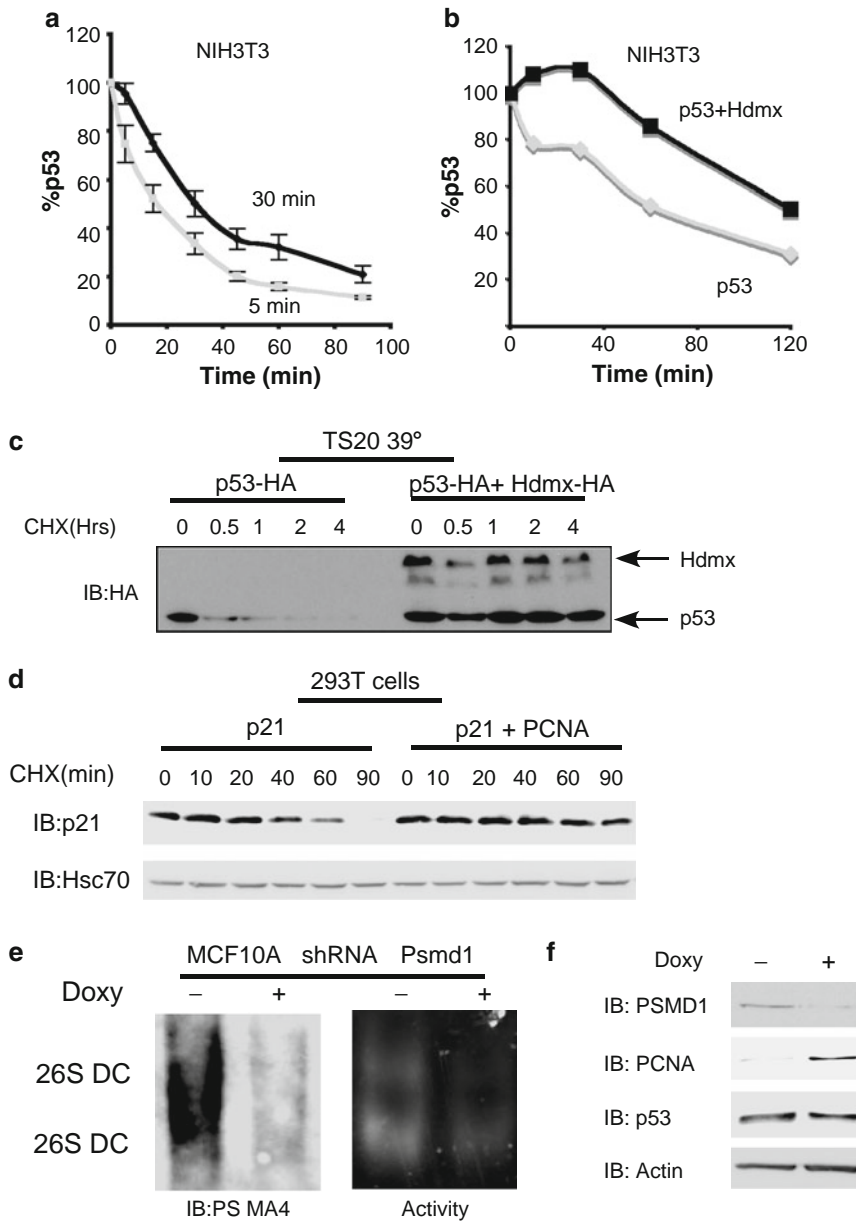
Fig. 2. (**a**) Pulse-chase experiment of endogenous p53 in NIH3T3 cells was performed at least three times for two different pulses (5, 30 min). Each time point was averaged and plotted with the standard deviation. (**b**) Pulse-chase experiment was conducted on wt p53 overexpressed in NIH3T3 cells in the presence or absence of Hdmx. (**c**) The half-life of p53 was analyzed by CHX treatment in TS20 cells following 16 h at restrictive temperature (39°C) in the presence or absence of Hdmx. (**d**) p21 was overexpressed in 293T cells in the presence or absence of PCNA. P21 protein levels were monitored following treatment with CHX for the indicated time points. (**e**, **f**) MCF10A cells were infected with TRIPZ Lentiviral vector with shRNAmir for the Psmd1 subunit of the 19S complex. 72 h after treatment with 1 μg/ml of doxycycline (to induce the shRNA expression) the reduction in the 26S proteasomal complex levels and activity was observed by the nondenaturing PAGE (**e**) structured protein such as PCNA accumulated whereas an IUP as p53 was not affected by the reduction in the 26S proteasomal complex (**f**).

Determination of
Endogenous Protein
Half-Life by Pulse-Chase
Experiments

Detection of endogenous proteins is not a trivial task while dealing with labile proteins. Therefore, as a first step, one has to select the appropriate cell line where the protein of interest is detectable. Here we describe the conditions for NIH3T3 cells metabolic labeling with [^{35}S] methionine. Most protocols recommend methionine starvation (its exclusion from the medium) before adding [^{35}S] methionine. This step is important to improve [^{35}S] methionine incorporation (up to tenfold). However, this step may introduce perturbations in protein homeostasis. Our recommendation is to avoid this step as long as the protein of interest is detectable without methionine starvation. Under such conditions it is expected that an IUP will exhibit rapid degradation and in some cases such as p53 might exhibit biphasic degradation with the initial phase being the 20S proteasomal degradation and the later the ubiquitin mediated 26S proteasomal degradation (6).

Given the fact that newly synthesized IUPs are expected to undergo very rapid degradation, it is important to detect the initial rapid degradation phase. To this end, one needs to make sure that the radiolabeling period (pulse time) of the cells is minimal and significantly shorter than the half-life of the protein. Our pulse-chase protocol has been modified to enable reproducible experiment of very short pulses (as short as 1 min). Instead of labeling the cells while attached to the plate, i.e., the conventional strategy, we advise that the cells are first collected and transferred to Eppendorf tubes where the pulse-chase experiment is conducted. The advantage of such a protocol is to enable very rapid manipulation of cells with minimal time errors as all the time points are labeled together in one tube (instead of a plate per each time point) and then further divided for the chase analysis. The only disadvantage of this method is that it can only be utilized for proteins with short half-life (less than an hour). Such a protocol has been successfully utilized by us (6) and with some modifications by others to detect DRiPs (should be discriminated from IUPs) (12, 13). This method can be further calibrated by performing time course pulse experiments to find the shortest pulse time where the IUP is detected.

1. Seed NIH3T3 cells at the concentration of 100,000 cells/cm^2 in 10 cm plates. Enable the cells to reach confluence. 48 h after seeding change growth medium and perform experiment 72 h after seeding.

2. Collect the cells with trypsin and transfer to Eppendorf tube. Prior to collecting the cells, starve them for methionine (see Note 6)

3. Resuspend the cell pellet with labeling medium for differential time points between (pulse) 1–30 min. Incubate at 37°C during [^{35}S] methionine labeling time (see Note 7).

4. Centrifuge at $1,000 \times g$ for 30 s, discard the labeling medium, and wash once with chase medium (1 ul). Spin again at $1,000 \times g$ for 30 s and discard the medium again.

5. Resuspend pellets in 900 μl chase buffer out of which take 100 μl and add to 500 μl chase buffer in Eppendorf prepared in advance and incubated at 37°C.

6. Incubate at 37°C for different time points and collect time points by pelleting the cells at $1,000 \times g$ for 1 min at 4°C. Discard the medium. Wash with 1 ml cold PBSx1 and pellet again by centrifugation at $1,000 \times g$ for 1 min at 4°C. Discard supernatant and freeze cells in liquid nitrogen.

7. After all the time points are collected, the cells are extracted with 400 μl RIPA extraction buffer. Protein extract is incubated on ice for 5 min and cleared by centrifugation at $16,200 \times g$ for 15 min at 4°C.

8. Protein extract is then subjected to immunoprecipitation (IP) with protein A/G beads preincubated with antibody of interest. The IP reaction is performed on a roller at 4°C between 1 and 12 h (based on antibody known qualities).

9. Following IP the beads are extensively washed ($\times 5$) with 400 ul RIPA buffer and then 40 μL SBx2 is added to beads and incubated 5 min at 95°C. Do spin down on the samples and load sup on SDS-PAGE.

10. For detection the gels can be either dried under vacuum in gel dryer apparatus for 2 h at 65°C or the proteins can be transferred to nitrocellulose membrane that can enable not only the identification of radiolabeled protein but can allow immunoblot analysis as well.

11. The dried gel or nitrocellulose membrane is then covered with plastic wrap and exposed on a phosphor imager screen for 1–24 h depending on the signal.

12. The screen is scanned by phosphor imager scanner and analyzed by appropriate software

13. Quantify the relative band intensity subtracting the general background (should be the same for all points) utilizing the phosphor imager scanner software.

3.2.2. Inhibiting Cellular Ubiquitin System Does Not Compromise IUP Instability

Examining the accumulation or the half-life of a protein while the ubiquitin system is inactive can teach us if a protein has the IUP attributes. One of the most extensively used strategies is inhibiting the ubiquitin-activating enzyme E1, a crucial enzyme in the ubiquitination pathway. The A31N-ts20 cells contain a thermo sensitive ubiquitin-activating enzyme E1 (14). At the permissive temperature (32°C) the ubiquitin system is functional whereas at the restrictive temperature (39°C) the E1 is unstable resulting in

the accumulation of proteins that are normally degraded by the ubiqutin-26S proteasomal pathway. Other strategies to inactive the E1 ubiquitin activating enzyme can be utilized such as specific knockdown or pharmacological inhibition by commercially available drugs (such as (PYR-41)).

1. Both the endogenous and the transfected protein of interest can be examined in the ts20 cells. Grow the cells at the 32°C permissive temperature. For transfection let the cells reach 80 % confluence before transfecting with Lipofectamine 2000 according to the manufacturer's protocol.

2. 12–24 h after transfection transfer the cells to grow under 39°C restrictive temperature for 5–16 h.

3. Examine the protein half-life by CHX or pulse-chase labeling methods described above.

4. An IUP is expected to be degraded although the ubiquitin system is inactive. It is highly important to use positive and negative controls. P21, Mcl-1 p53 each can be used as a positive control (as protein degraded in a ubiquitin impendent manner). As a negative control a structured protein that undergoes ubiquitin dependent degradation PCNA and GFP (for transfected experiments) can be utilized.

3.2.3. Measuring 20S Proteasomal Degradation in the Cells

A candidate IUP that so far proved positive next has to be challenged by 20S proteasomal degradation. Currently there are no specific inhibitors of the 26S proteasome that do not inhibit the 20S proteasome as well; therefore, the inhibition of the 26S proteasome can be achieved by knockdown strategy. The basic idea is to employ inducible knockdown of one of the subunits of the 19S complex unique to the 26S proteasome by reducing the subunit levels the whole complex is destabilized resulting in reduced levels of intact 26S proteasomes in the cell. In our experience knocking down the Psmd1 (Rpn2) subunit results in sharp reduction at the level of 26S proteasomes.

IUP that is susceptible for 20S proteasomal degradation should be refractory or poorly affected by this manipulation unlike the structured proteins. Described below is the protocol for the conditional knockdown of the Psmd1 subunit of the 19S complex that results in the reduction of cellular 26S proteasomes (Fig. 2e, f).

1. We found that knockdown of the PSMD1 subunit is very efficient to obtain efficient reduction at the level of the 26S proteasome. To this end, we utilize Lentiviral vectors with shRNAmir against 26S proteasome subunit Psmd1 and control shRNAmir (Open Biosystems). The elimination of the 26S proteasome by the Psmd1 knockdown can be lethal to the cells; therefore, we utilize pTRIPZ lentivirus shRNAmir, a tetracycline inducible knockdown vector. Transducing

lentivirus particles were produced according to the manufacturer's protocol.

2. Cells of interest are infected with the lentivirus particles (infections conditions vary between cell lines). Selection with puromycin is employed for a week (the concentration has to be determined based on the cell type).

3. To induce the knockdown the cells are treated with doxycycline between 0.5 and 1.5 μg/ml. A reduction of the 26S proteasomal complex should be observed after 48 h (see Fig. 2). The efficiency of the reduction in the 26S proteasomal complex can be analyzed by subjecting the cellular extract to nondenaturing PAGE as previously described (15). This enables to visualize the reduction in the 26S proteasomal complex and not only the subunit (as examined by the SDS-PAGE).

4. Endogenous protein levels and the half-life are examined as described above in the control (untreated) or knockdown cells (treated with doxycycline).

5. In cases where overexpressed proteins are analyzed it is recommended to transfect the protein before the induction of the shRNAmir expression. Once the protein is introduced, the protocol is the same as for endogenous proteins.

3.3. Stabilization of an IUP by an Interacting Protein (Nanny) in Cells

Another attribute of IUP in cells is the stabilization by an interacting partner, the nanny principle (5). Similar to the in vitro assay in the cells the nanny of the IUP should inhibit its ubiquitin-independent 20S proteasomal degradation. The experiment is relatively simple: overexpression of an IUP in the presence of its interacting partner and the half-life determined either by inhibiting protein translation by CHX or by pulse-chase methodology (Fig. 2). In certain cases one has to uncouple between the biological function of the interaction and the stabilizing effect. One such example is p53 and Mdm2. Mdm2 is the E3 ligase of p53 that can ubiquitinate p53 and target it for 26S proteasomal degradation. However, Mdm2 also binds the unstructured N-terminus of p53 preventing its degradation by the 20S proteasome in vitro. The stabilizing effect of Mdm2 on p53 is overlooked while overexpressing the two genes, since the E3 ligase activity of Mdm2 will be dominant over the nanny function. However, it is possible to uncouple these two opposing roles by either mutating the E3 ligase domain of Mdm2 or by conducting the experiments under conditions whereby the ubiquitination process is inactive (6).

3.3.1. IUP Stabilization by a Binding Partner; Analysis by Pulse-Chase Experiments

1. As described above we utilize NIH3T3 cells but other cells can be utilized for this purpose. Transfection is performed after cells reach ~80 % confluence with JetPI reagent as described in the manufacturer's protocol. 9-cm plates are each transfected

either with the IUP alone or in the presence of the interacting partner (see Note 8)

2. 24 h after transfection the medium is changed and 48 h after transfection perform pulse-chase experiment as described above (Subheading 3.2.1).

3.3.2. IUP Stabilization by a Binding Partner; Analysis by CHX Experiments

1. To this end many different cell lines can be utilized. We chose either HEK 293T cells or A31N-ts20 mouse fibroblasts.

2. HEK 293T cells are seeded at ~30 % confluence in 10-cm (55 cm^2) plates. 16 h after the seeding the transfection is performed by the method of calcium phosphate.

3. 8 h after transfecting the medium is changed.

4. 24 h after transfection the cells are split into 6-well plates (9.4 cm^2) so that each well contains 1/18 of the original 10-cm plate. The number of plates seeded should be according to number of time points to be analyzed (see Note 9).

5. 16–24 h after seeding 100 μg/ml of CHX is added to each well and cells are collected at different time points (see Note 10).

6. In the case of A31N-ts20 cells and the determination of ubiquitin-independent IUP half-life, the transfection is performed at the permissive temperature (32°C) utilizing Lipofectamine 2000 transfection reagent. 24 h after transfection the cells are split as described for 293 cells and after 8 h after seeding the cells are transferred to the restrictive temperature (39°C) for 16 h.

7. Discard the medium and wash the cells once with 1 ml of cold PBSx1.

8. Remove the cells by gentle scraping with rubber policeman. The cells are then transferred to 1.5 ml Eppendorf tubes and pelleted in a cooled table centrifuge (microfuge) at 3,500 × *g* for 1 min at 4°C. The sup is discarded and the cells are immediately frozen in liquid nitrogen.

9. Each cell pellet is lysed in 150 μl RIPA extraction buffer.

10. Protein extract is incubated on ice for 5 min and the cleared by centrifugation at 16,200 × *g* for 15 min at 4°C.

11. The sup protein extract is transferred to new Eppendorfs and protein concentration is detected by the Bradford methodology (16).

12. Equal amounts of protein (in the range of 10–30 μg) (see Note 11) are taken for each time point and mixed with SBx3 and incubated at 95°C for 5 min.

13. Samples are run on SDS-PAGE and then transferred to a nitrocellulose membrane

14. The membranes are blocked for 1 h at room temp with 7 % milk in PBS-T.

15. Membranes are incubated with primary and secondary antibodies according to the manufacturer's guidelines.

16. The signals are detected by exposing the membrane to Ez-ECL kit and then either exposure to X-ray films or detection by ImageQuant RT ECL imager (GE) and analyzed by Image-Quant TL (GE) or other suitable software.

4. Notes

1. A general note: Substrates to be used must be cautiously tagged if at all. Although tagging is of much help in detecting the substrate but at the same time it might affect the level of unstructureness. Degradation of the most of the protein substrates is directional and is initiated from either through the N- or C-terminus. If tagging is necessary try not to tag the initiating terminus. If no information is available on the initiating end, prepare and separately test each of the two ends possible tagged substrates.

2. The amount of protein taken should be the minimal needed for detection.

3. An IUP is expected to be degraded within an hour. This is the case with Mcl-1, p21 and tau shown in Fig. 1. Exceeding 2 h of incubation in the degradation reaction can cause nonspecific results.

4. The amount of interacting protein is largely dependent on the stoichiometry and the affinity. In our experience it is best to take in excess and start with at least ×2 of the IUP amount. Furthermore, the stabilizing effect depends on the binding capacity; therefore, it is advised to optimize the binding reaction prior to degradation assay.

5. It has been previously reported that some proteins can undergo N-terminal ubiquitination (17). Therefore, when eliminating all the lysines in a protein, it is still important to indeed validate that the K-less mutant does not undergo ubiquitination.

6. Methionine starvation can increase the signal by up to tenfolds but can also perturb normal cellular function. Therefore it is recommended to avoid this step if the labeled substrate is detectable without the methionine starvation.

7. The optimal pulse time is the shortest possible pulse that is reproducible and enables the convenient detection of the labeled protein. The length of pulse is advised to be significantly shorter than the half-life of the protein.

8. The amount of the transfected expression plasmid should be minimal but enough for detection of the IUP and the interacting protein (putative nanny) on the one hand, and on the other hand enough DNA to achieve the interaction between the IUP and the nanny is required (needs to be calibrated by coimmunoprecipitation experiments).

9. We find that the method by which the cells are first transfected and then split to small wells is preferred over performing the transfection on each separate well. This step reduces the variability between the experiments that can arise from the transfection.

10. CHX treatment inhibits the cellular translation machinery. It is therefore recommended to treat the cells for a shortest possible period up to 4 h. The more the protein is stable the longer treatment time is required. Overall if the desired half-life is not observed after 4 h of treatment with CHX, pulse-chase experiment is the preferred choice.

11. Minimal amount of protein is advised to be loaded on gel for detection in the half-life determination experiments.

References

1. Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–847

2. Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 82:373–428

3. Asher G, Reuven N, Shaul Y (2006) 20S proteasomes and protein degradation "by default". Bioessays 28:844–849

4. Tsvetkov P, Asher G, Paz A, Reuven N, Sussman JL, Silman I, Shaul Y (2008) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. Proteins 70:1357–1366

5. Tsvetkov P, Reuven N, Shaul Y (2009) The nanny model for IDPs. Nat Chem Biol 5:778–781

6. Tsvetkov P, Reuven N, Prives C, Shaul Y (2009) Susceptibility of p53 unstructured N terminus to 20S proteasomal degradation programs the stress response. J Biol Chem 284:26234–26242

7. Dosztanyi Z, Tompa P (2008) Prediction of protein disorder. Methods Mol Biol 426:103–115

8. Jariel-Encontre I, Bossis G, Piechaczyk M (2008) Ubiquitin-independent degradation of proteins by the proteasome. Biochim Biophys Acta 1786:153–177

9. Forsthoefel AM, Pena MM, Xing YY, Rafique Z, Berger FG (2004) Structural determinants for the intracellular degradation of human thymidylate synthase. Biochemistry 43:1972–1979

10. Stewart DP, Koss B, Bathina M, Perciavalle RM, Bisanz K, Opferman JT (2010) Ubiquitin-independent degradation of antiapoptotic MCL-1. Mol Cell Biol 30:3099–3110

11. Bossis G, Ferrara P, Acquaviva C, Jariel-Encontre I, Piechaczyk M (2003) c-Fos proto-oncoprotein is degraded by the proteasome independently of its own ubiquitinylation in vivo. Mol Cell Biol 23:7425–7436

12. Qian SB, Princiotta MF, Bennink JR, Yewdell JW (2006) Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control. J Biol Chem 281:392–400

13. Schubert U, Anton LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404:770–774

14. Chowdary DR, Dermody JJ, Jha KK, Ozer HL (1994) Accumulation of p53 in a mutant cell line defective in the ubiquitin pathway. Mol Cell Biol 14:1997–2003

15. Elsasser S, Schmidt M, Finley D (2005) Characterization of the proteasome using native gel electrophoresis. Methods Enzymol 398:353–363

16. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

17. Ciechanover A, Ben-Saadon R (2004) N-terminal ubiquitination: more protein substrates join in. Trends Cell Biol 14:103–106

Chapter 2

In-Cell NMR of Intrinsically Disordered Proteins in Prokaryotic Cells

Yutaka Ito, Tsutomu Mikawa, and Brian O. Smith

Abstract

In-cell NMR, i.e., the acquisition of heteronuclear multidimensional NMR of biomacromolecules inside living cells, is, to our knowledge, the only method for investigating the three-dimensional structure and dynamics of proteins at atomic detail in the intracellular environment. Since the inception of the method, intrinsically disordered proteins have been regarded as particular targets for in-cell NMR, due to their expected sensitivity to the molecular crowding in the intracellular environment. While both prokaryotic and eukaryotic cells can be used as host cells for in-cell NMR, prokaryotic in-cell NMR, particularly employing commonly used protein overexpression systems in *Escherichia coli* cells, is the most accessible approach. In this chapter we describe general procedures for obtaining in-cell NMR spectra in *E. coli* cells.

Key words: In-cell NMR, *Escherichia coli*, Heteronuclear multidimensional NMR, Rapid NMR measurement, Protein structure, Protein dynamics

1. Introduction

Recent improvements in the measurement sensitivity of NMR spectra as the result of developments in hardware, pulse sequences, and stable isotope labeling techniques allow the observation of high-resolution heteronuclear multidimensional NMR spectra of proteins inside living cells (in-cell NMR) (1–8). Due to the noninvasive character of NMR spectroscopy and its ability to provide data at atomic resolution, in-cell NMR is an excellent tool for investigating protein behavior in detail inside living cells under macromolecular crowding (9).

In order to observe in-cell NMR spectra, the proteins of interest must be abundant and labeled selectively with NMR active stable isotopes such as ^{13}C and ^{15}N. In-cell NMR experiments can be performed both in prokaryotic and eukaryotic cells, although different labeling procedures are required. For prokaryotic in-cell NMR

studies (10–12) the proteins of interest are labeled by utilizing protein expression systems in the host cells. The basic principle relies on the fact that recombinant protein production driven by strong promoters leads to the rapid intracellular accumulation of large amounts of the expression system-encoded proteins to the almost complete exclusion of new endogenous protein synthesis. Thus, switching from unlabeled to isotope-labeled growth medium shortly before recombinant protein induction results in the selective labeling of the protein of interest. For eukaryotic in-cell NMR, labeled proteins are first purified, and then incorporated into cells by microinjection (13–15), by combining with a cell-penetrating peptide (16), or through resealable pores (17).

Prokaryotic, particularly *Escherichia coli*, in-cell NMR is the easiest system, and thus has been used for the investigation of various protein systems and biological events, such as protein–protein interactions (4, 18), protein–DNA interactions (19), intracellular protein dynamics (20), protein stability (21), and the screening of small molecule interactor libraries (22). We have recently demonstrated three-dimensional protein structure determination exclusively on the basis of information obtained by in-cell NMR (23).

The behavior of intrinsically disordered proteins (IDPs) in cells is also an important target for in-cell NMR. Recently, in-cell NMR using *Xenopus* oocytes has been applied to the human amyloid protein Tau, one of the largest known IDPs (15). Prokaryotic in-cell NMR has also been used for the investigation of IDPs since the inception of this method, and it has been demonstrated that the bacterial protein FlgM is completely unfolded in vitro, but appears partially folded when analyzed in the context of live bacteria (24). With interest in the structure–function relationship of IDPs, particularly inside cells, growing and with the technique's relative experimental convenience, we anticipate the application of prokaryotic in-cell NMR to various IDPs in the near future.

Thus, we describe here a general procedure for *E. coli*-based in-cell NMR experiments, including sample preparation, rapid NMR measurement, data processing, and analysis. Being critical for the validation of in-cell NMR, procedures for checking the sample condition and the viability of the cells are also described.

2. Materials

In this section, we list the materials required in our typical procedure for the preparation of ^{15}N-labeled *E. coli* cell samples and in-cell NMR measurement. The modifications required for other stable isotope labeling patterns are described in Note 1. The details of apparatus required for standard molecular biological and biochemical experiments, e.g., devices for aseptic techniques, incubators, centrifuges, etc., are omitted.

2.1. Large Scale E. coli Culture and Stable Isotope Labeling

1. Expression plasmid encoding the protein of interest (see Note 2).

2. *E. coli* strains: e.g., JM109 (DE3) *E. coli*, BL21 (DE3) *E. coli* (see Note 3).

3. LB medium: 10 g/l Bacto tryptone, 5 g/l Bacto Yeast Extract, 10 g/l NaCl. Sterilize by autoclaving.

4. $MgSO_4$ stock solution: 1 M $MgSO_4$. Store at room temperature.

5. $CaCl_2$ stock solution: 50 mM $CaCl_2$. Store at room temperature.

6. $FeCl_3$ stock solution: 5 mM $FeCl_3 \cdot 6H_2O$. Store at 4°C.

7. Metal mixture stock solution: 4 mM $ZnSO_4 \cdot 7H_2O$, 1 mM $MnSO_4 \cdot 5H_2O$, 4.7 mM H_3BO_3, 0.7 mM $CuSO_4 \cdot 5H_2O$. Store at 4°C.

8. Thiamine stock solution: 0.3 M Thiamine hydrochloride. Store at −20°C.

9. M9 minimal medium (unlabeled, 100 ml): 1.2 g Na_2HPO_4, 0.6 g KH_2PO_4, 0.1 g NH_4Cl, 0.1 g NaCl, 0.2 g D-glucose, 200 µl $MgSO_4$ stock solution, 200 µl $CaCl_2$ stock solution, 40 µl $FeCl_3$ stock solution, 40 µl thiamine stock solution, 100 µl Metal mixture stock solution. Prepare just before use. Filter-sterilized.

10. M9 minimal medium (^{15}N-uniformly labeled, 100 ml): Same as the composition of unlabeled M9 minimal medium with the exception that unlabeled NH_4Cl was replaced by the same concentration of $^{15}NH_4Cl$. Prepare just before use. Filter-sterilized (see Note 1).

11. Reagent for induced expression of the target protein. For example, isopropyl thio-β-D-thiogalactoside.

12. Antibiotics stock solution. For example, ampicillin stock solution: 50 mg/ml Ampicillin sodium. Store at −20°C.

2.2. NMR Measurement

1. NMR spectrometer (for specifications and pulse sequences see Notes 4 and 5, respectively).

2. Linux computer system for data processing and analysis of spectra (for specifications and installed software see Notes 6 and 7, respectively).

3. Methods

In this section, we focus on the methods required for typical in-cell NMR experiments of proteins overexpressed in *E. coli* cells; thus, standard procedures and troubleshooting for the construction of protein expression systems in *E. coli* are omitted.

3.1. Preparation of Proteins Inside Living E. coli Cells

1. Transform *E. coli* cells with the overexpression plasmid.

2. Grow the *E. coli* cells in 2 ml LB media at 37 °C with shaking to a high OD_{600} of ~2.0 (see Note 2).

3. Subculture the *E. coli* cells (100 μl) in 100 ml unlabeled M9 media, and incubate the culture at 37 °C until the OD_{600} reaches 0.5–0.6.

4. Centrifuge the culture at ~800 × g for 20 min at room temperature.

5. Decant the supernatant. Centrifuge the pellet again at ~800 × g for 5 min at room temperature.

6. Remove the supernatant by pipetting and resuspend the cells in 100 ml stable isotope-labeled M9 media (see Note 8).

7. Incubate the cells at 37 °C without shaking for 1 h.

8. Induce the production of the target protein (for example, by adding isopropyl thio-β-D-thiogalactoside to a final concentration of 0.5 mM).

9. Continue protein expression with shaking at optimal temperature (for example, 3 h at 37 °C in our previous in-cell NMR study of *T. thermophilus* HB8 TTHA1718 (23)).

10. Harvest the cells by centrifugation at ~400 × g for 30 min at room temperature.

11. Remove the supernatant by aspiration and resuspend the cells by adding small amounts of unlabeled M9 media (140–160 μl) and carefully pipetting the solution up and down until the entire cell pellet has been suspended. The cells are harvested by gentle centrifugation and placed as a ~60 % slurry with M9 medium.

12. Add D_2O (10 % of the final sample volume) to the bacterial slurry, and transfer the sample into an NMR tube (see Note 9).

3.2. Checking the Sample Conditions for Sample Stability and Viability of the Cells During In-Cell NMR Experiments (see Note 10)

1. Insert the in-cell NMR sample into the magnet and tune the probehead (see Note 11).

2. Perform a series of short, diagnostic NMR experiments (e.g., 2D ^1H-^{15}N HSQC experiment with 20 min duration). The appearance of significant changes in the spectra indicates the limit of sample stability.

3. Between each NMR experiment, transfer 10 μl from the NMR sample to a microfuge tube and centrifuge at high speed to pellet the bacteria. Store the pellet and supernatant separately for SDS-PAGE (see Note 12) to check the stability of the expressed protein.

4. Check the viability of the cells during NMR experiments can be investigated by a plating colony test. Spread a small volume of the NMR sample (~10 μl) taken before and after the experiments on LB plates containing the appropriate antibiotic.

Incubate the plates overnight at 37 °C and count the colonies. In our laboratories the acceptable percentage of viability limits is set to be 80–85 %.

3.3. NMR Measurements, Data Processing, and Analysis

1. Insert the in-cell NMR sample into the magnet and tune the probehead.

2. Check the sample's condition by measuring 1D ^{1}H-NMR spectra and 1D (or 2D) ^{1}H-^{15}N HSQC spectra.

3. Collect 2D/3D NMR spectra (see Notes 13 and 14). Figure 1a shows the 2D ^{1}H-^{15}N HSQC spectrum of an IDP, the C-terminal region of *S. cerevisiae* Mre11, in living *E. coli* cells, in comparison with the spectrum from the purified sample (Fig. 1b). Due to the higher viscosity inside cells (25), which increases the rotational correlation time and apparent molecular mass of proteins, and the inhomogeneity of in-cell NMR samples, the line shape for both ^{1}H and ^{15}N dimensions of cross peaks are much broadened when compared to the in vitro sample. The duration of the NMR experiments for an in-cell
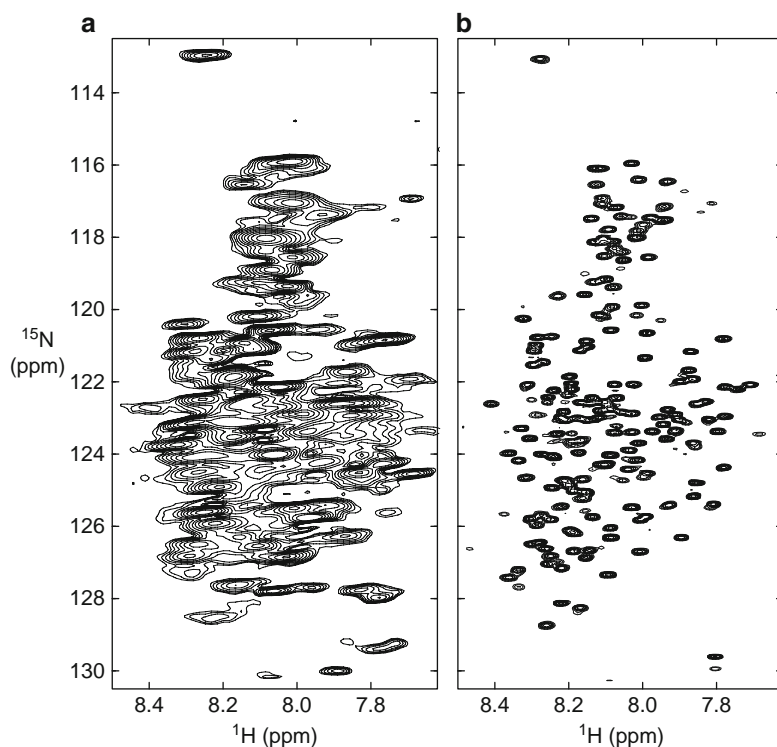
Fig. 1. 2D ^{1}H-^{15}N HSQC spectra of the C-terminal region of *S. cerevisiae* Mre11 in *E. coli* cells (**a**) and in vitro (**b**). The spectra were acquired with eight transients and a total of 1,024 (t_2, ^{1}H^N) × 64 (t_1, ^{15}N) complex points at a probe temperature of 37 °C. For in-cell NMR measurement, *E. coli* strain BL21 (DE3) was used. Resolution enhancement for the t_1 dimension was achieved by applying linear prediction prior to Fourier transform on Azara v2.8 software.

NMR sample must be set considering the life time of the *E. coli* cells under the measurement conditions (see Note 15).

4. Proteins may leak from the cells, invalidating the measured data, so to make sure that the measured NMR data are definitely from the protein molecules inside cells, remove the cells from the sample after measurements by centrifugation (e.g., ~16,000 × g for 5 min). Retain the supernatant of the measured sample, make it up with unlabeled M9 medium to the initial sample volume, and measure a ^{1}H-^{15}N HSQC experiment of it. Check that no or only very weak cross peaks due to the target protein is observed in the spectrum (see Note 16).

5. To make sure that the target protein is inside the cells, lyse the harvested cells, fill up with unlabeled M9 medium to the initial sample volume, and measure ^{1}H-^{15}N HSQC experiment of it. Check that much sharper cross peaks are observed in the spectrum (see Note 16).

6. Process NMR spectra with appropriate procedure (e.g., maximum entropy reconstruction for nonlinearly sampled data) (see Notes 5 and 7).

7. Analyze NMR spectra with interactive NMR spectrum analysis software (see Note 7).

4. Notes

1. The choice of stable isotope labeling pattern depends on the purpose of the experiments. For simple observation of or monitoring the behavior of proteins of interest inside *E. coli* cells by 2D ^{1}H-^{15}N correlation experiments, uniform ^{15}N-labeling is sufficient. On the other hand, when resonance assignments are required from spectra acquired from in-cell NMR samples, uniform ^{13}C/^{15}N-labeling is required and ^{13}C/^{15}N-labeled M9 minimal medium is used for sample preparation, in which unlabeled D-glucose in the ^{15}N-labeled M9 medium is replaced by the same concentration of uniformly ^{13}C-labeled glucose, e.g., U-^{13}C-glucose.

Since IDPs tend to provide ^{1}H-^{15}N correlation spectra with poorer peak separation than proteins with defined structures (see Fig. 1a), selective labeling procedures may be very useful. For selective ^{15}N-labeling of amino acids which are end products of biosynthetic pathways in *E. coli* (e.g., lysine), addition of a supply of the ^{15}N-labeled amino acid(s) to the unlabeled M9 minimal medium is sufficient (11, 26). Though we have not tested them in our groups, auxotrophic *E. coli* strains lacking some transaminase activity may be required for the cases where isotope scrambling is expected during biosynthesis (27).
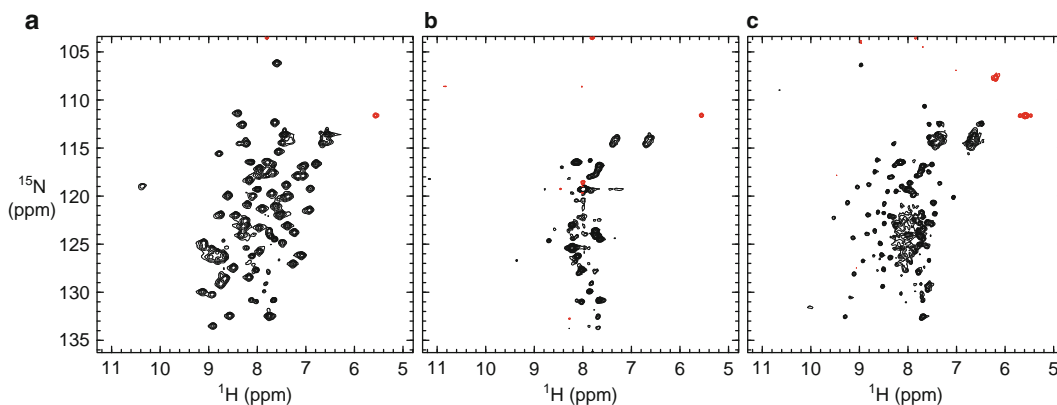
Fig. 2. 2D in-cell ^{1}H-^{15}N HSQC spectra of *T. thermophilus* HB8 TTHA1718 (**a**), TTHA1431 (**b**) and TTHA1912 (**c**) in *E. coli* strain JM109 (DE3). The expression plasmids encoding TTHA1431, TTHA1718 and TTHA1912 (51) are purchased from RIKEN BioResource Center (http://www.brc.riken.jp/lab/dna/en/thermus_en.html). The spectra were acquired with eight transients and a total of 1,024 (t_2, ^{1}H^N) × 64 (t_1, ^{15}N) complex points at a probe temperature of 37 °C.

Selective ^{13}C-labeling at side-chain methyl groups is another useful option. Serber et al. employed methyl ^{13}C-labeling at methionine residues as probes in in-cell NMR (28). In our previous in-cell NMR study (23), aimed at observing NOEs involving methyl groups, we employed selectively ^{1}H/^{13}C-labeling at methyl groups of aliphatic amino acid residues by using amino acid precursors (29). Typically, 10 mg of ^{13}C- or ^{2}H/^{13}C-labeled amino acids or precursors, e.g., [3-^{13}C] alanine (ISOTEC), [u-^{13}C, 3-^{2}H] α-ketoisovalerate (Cambridge Isotope Laboratories) for leucine and valine residues, and [u-13C, 3,3-^{2}H] α-ketobutyrate (Cambridge Isotope Laboratories) for isoleucine residues, are supplemented in 100 ml of M9 minimal medium prepared using 100 % D$_2$O.

In addition, ^{19}F-labeling utilizing fluoro amino acids (30), as well as site-specific incorporation of labeled nonnatural amino acids (31), has been reported recently.

2. Higher expression level does not always promise good in-cell NMR spectra. Despite high intracellular concentration, no, or very weak cross peaks are sometimes observed from the target protein in in-cell NMR spectra. Fig. 2a, b show 2D ^{1}H-^{15}N HSQC spectra of *Thermus thermophilus* HB8 TTHA1718 and TTHA1431 in *E. coli* cells, respectively. While both of these proteins express very well inside *E. coli* cells, theTTHA1431 sample shows very poor spectra in which only background cross peaks are observed, in contrast to TTHA1718. The reasons for such results will be case-specific with motional restrictions due to nonspecific interactions with other cellular components inside *E. coli* cells being a possible explanation. Usually drastic improvement cannot be guaranteed, but it is worth exploring alternative

protein expression conditions, varying, e.g., *E. coli* host strains, timing of induction, temperature, and incubation times.

3. We usually try at least two *E. coli* strains, e.g., JM109 (a K-12 strain) and BL21 (a B strain) when starting in-cell NMR experiments with a new protein. In many cases, we obtain similar results from both strains, but we have also experienced the case that good ^{1}H-^{15}N HSQC spectra were observed when using JM109, while broadened cross peaks were observed when using BL21.

4. NMR spectrometers and probeheads must be equipped for heteronuclear ^{1}H-detection experiments or triple-resonance (^{1}H/^{13}C/^{15}N) experiments. Hardware enabling pulsed field gradients is highly recommended, since efficient water suppression in in-cell NMR experiments is very difficult to achieve without them. As sensitivity is one of the limiting factors of in-cell NMR experiments, a cryogenic probehead and a magnetic field strength corresponding to a ^{1}H frequency of at least 500 MHz are recommended.

5. A major hurdle for in-cell NMR experiments is the limited lifetime of the cells inside the NMR sample tube. Standard 3D NMR experiments usually require 1–2 days of data collection, which is an unacceptably long time for live cells. NMR techniques for rapid measurement of multidimensional experiments are therefore very important. As was demonstrated in the in-cell NMR study using human cultured cells (16), application of SOFAST-HMQC and related techniques (32, 33) to 2D/3D experiments will be advantageous. In addition, nonlinear sampling (non uniform sampling) techniques for the indirectly acquired dimensions (34–36) are also very helpful. Spectrometers should be prepared so as to run these new techniques (they are starting to be usable as standard tools on very recent NMR spectrometers). For nonlinear sampling, the pulse sequences have to be prepared so as to control sampling points according to sampling schedule lists as opposed to conventional sampling of every point on a regularly spaced grid (26, 36).

6. Data processing and spectral analysis can in principle be run on a computer with a single processor. In case the project includes structure calculations, multicore PC systems or PC clusters are preferable.

7. Software for data processing must be installed. Particularly, when employing nonlinear sampling schemes for obtaining multidimensional NMR spectra, the software must be able to handle sparsely sampled data. In our laboratories, the two-dimensional maximum entropy method (2D MaxEnt) (37, 38) on AZARA 2.8 software (W. Boucher, http://www.bio.cam.ac.uk/azara) is used. The Rowland NMR Toolkit (http://webmac.rowland.

org/rnmrtk/) has also been used for this type of processing. Note that, in addition to MaxEnt, other processing procedures, such as multidimensional decomposition (39) and multidimensional Fourier transform (40, 41), can also be used.

An interactive NMR spectrum analysis software is required. In our laboratories, either an OpenGL version of ANSIG 3.3 (42, 43) or CcpNmr Analysis (44) is used. Alternatively, programs such as NMRView (One Moon Scientific, inc) (45, 46), Sparky (http://www.cgl.ucsf.edu/home/sparky/), XEASY (47) (http://www.mol.biol.ethz.ch/groups/wuthrich_group/software), and CARA (http://www.nmr.ch/) can be used.

When the project includes structural analysis, software for structure calculations has to be installed. In our laboratories, CYANA 3.0 (http://www.cyana.org) (48) and ARIA 2.3 (http://aria.pasteur.fr) (49) are used for automated NOESY assignment and structure calculation.

8. In order to produce a sample of *E. coli* cells in which only the target protein is labeled with ^{13}C and ^{15}N, the cells harboring the expression plasmid are first grown in unlabeled LB medium, and then transferred into M9 minimal medium containing stable isotopes where protein expression is induced.

9. The concentration of the expressed protein in *E. coli* cells can be estimated by comparing the density of the Coomassie-stained bands in SDS-PAGE gels with those of proteins with similar molecular size and known concentration.

10. To ensure that the observed in-cell NMR spectrum represents intracellular protein and that the signals are not caused by proteins released from the bacteria, the sample conditions and the viability of the cells during the experiments has to be checked before the "real" measurements.

11. Prior to the measurement of the in-cell NMR sample, shim the magnetic field with a separate NMR sample containing unlabeled M9 media (10 % D_2O), which is prepared with the same sample length as for the in-cell NMR sample.

12. Sometimes we experienced leakage of the target protein from cells during NMR experiments. Figure 2c shows a typical in-cell NMR spectrum when leakage has occurred. Sharp cross peaks emerge and increase their intensity during the timecourse of measurements. In another case, leakage was largely prevented by reducing the temperature for protein expression and NMR measurement.

Li et al. reported that encapsulation in alginate microcapsules stabilizes *E. coli* cells and prevents leakage (20).

13. It is highly recommended to estimate the allowable maximum experimental duration under the measurement conditions. In our previous in-cell NMR study of *T. thermophilus* HB8

TTHA1718, the virtual identity of ^{1}H-^{15}N HSQC spectra recorded immediately after sample preparation and after 6 h in an NMR tube at 37 °C shows that TTHA1718 in-cell NMR samples are stable for at least 6 h, which was also confirmed by plating colony test (the viability of the bacteria in the in-cell samples after 6 h of NMR measurements was 85 ± 11 %) [23].

14. The present protocol does not require a specific set of NMR spectra. If duration of experiments and sensitivity permit, any of the common 2D and 3D experiments can be applied. In our previous study of in-cell structure determination of *T. thermophilus* HB8 TTHA1718 [23, 26], we measured three 3D NOESY-type experiments in addition to 2D ^{1}H-^{15}N and ^{1}H-^{13}C HSQC experiments. For application to IDPs, another 3D experiments which are widely used for backbone resonance assignment of IDPs, such as HNN and HN(C)N [50], can be considered, though we have not tested them on our in-cell NMR samples yet.

15. We recommend repeatedly monitoring the stability of the *E. coli* samples by 2D ^{1}H-^{15}N HSQC spectra followed by plating colony tests, which allows for a comparison between the initial and current health of the cells.

    With the nonlinear sampling scheme the duration of each 3D experiment is reduced to 2–3 h. Repeat the measurement of each 3D experiment several times interleaved with a short 2D ^{1}H-^{15}N HSQC experiment used to monitor the condition of the sample. Combine these 3D data to generate a new data set with improved signal-to-noise ratio up to the point that the 2D spectra exhibit significant changes.

16. These results were corroborated by SDS-PAGE, demonstrating that the contribution of extracellular protein to the observed signals is negligible.

Acknowledgments

References

1. Serber Z, Dotsch V (2001) In-cell NMR spectroscopy. Biochemistry 40:14317–14323

2. Serber Z, Corsini L, Durst F, Dötsch V (2005) In-cell NMR spectroscopy. Method Enzymol 394:17–41

3. Serber Z, Selenko P, Hänsel R, Reckel S, Löhr F, Ferrell JE, Wagner G, Dötsch V (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. Nat Protoc 1:2701–2709

4. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). Nat Methods 3:91–93

5. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) In-cell NMR for protein-protein interactions (STINT-NMR). Nat Protoc 1:146–152

6. Reckel S, Hänsel R, Löhr F, Dötsch V (2007) In-cell NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 51:91–101

7. Selenko P, Wagner G (2007) Looking into live cells with in-cell NMR spectroscopy. J Struct Biol 158:244–253

8. Pielak GJ, Li C, Miklos AC, Schlesinger AP, Slade KM, Wang GF, Zigoneanu IG (2009) Protein nuclear magnetic resonance under physiological conditions. Biochemistry 48:226–234

9. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26:597–604

10. Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dötsch V (2001) High-resolution macromolecular NMR spectroscopy inside living cells. J Am Chem Soc 123:2446–2447

11. Serber Z, Ledwidge R, Miller SM, Dötsch V (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. J Am Chem Soc 123:8895–8901

12. Wieruszeski JM, Bohin A, Bohin JP, Lippens G (2001) *In vivo* detection of the cyclic osmoregulated periplasmic glucan of *Ralstonia solanacearum* by high-resolution magic angle spinning NMR. J Magn Reson 151:118–123

13. Selenko P, Serber Z, Gade B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. Proc Natl Acad Sci USA 103:11904–11909

14. Sakai T, Tochio H, Tenno T, Ito Y, Kokubo T, Hiroaki H, Shirakawa M (2006) In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. J Biomol NMR 36:179–188

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

16. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature 458:106–109

17. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I (2009) Observation of NMR signals from proteins introduced into living Mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J Am Chem Soc 131:10834–10835

18. Burz DS, Shekhtman A (2008) In-cell biochemistry using NMR spectroscopy. PLoS One 3:e2571

19. Augustus AM, Reardon PN, Spicer LD (2009) MetJ repressor interactions with DNA probed by in-cell NMR. Proc Natl Acad Sci USA 106:5065–5069

20. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, Young GB, Macdonald JM, Pielak GJ (2008) Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. J Am Chem Soc 130:6310–6311

21. Charlton LM, Barnes CO, Li C, Orans J, Young GB, Pielak GJ (2008) Residue-level interrogation of macromolecular crowding effects on protein stability. J Am Chem Soc 130:6826–6830

22. Xie J, Thapa R, Reverdatto S, Burz DS, Shekhtman A (2009) Screening of small molecule interactor library by using in-cell NMR spectroscopy (SMILI-NMR). J Med Chem 52:3516–3522

23. Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 458:102–105

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

25. Williams SP, Haggie PM, Brindle KM (1997) ^{19}F NMR measurements of the rotational mobility of proteins in vivo. Biophys J 72:490–498

26. Ikeya T, Sasaki A, Sakakibara D, Shigemitsu Y, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Nietlispach D, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y (2010) NMR protein structure determination in living *E. coli* cells using nonlinear sampling. Nat Protoc 5:1051–1060

27. McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q Rev Biophys 23:1–38

28. Serber Z, Straub W, Corsini L, Nomura AM, Shimba N, Craik CS, de Montellano PO, Dötsch V (2004) Methyl groups as probes for proteins and complexes in in-cell NMR experiments. J Am Chem Soc 126:7119–7125

29. Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol 263:627–636

30. Li C, Wang GF, Wang Y, Creager-Allen R, Lutz EA, Scronce H, Slade KM, Ruf RA, Mehl RA, Pielak GJ (2010) Protein ^{19}F NMR in *Escherichia coli*. J Am Chem Soc 132:321–327

31. Jones DH, Cellitti SE, Hao X, Zhang Q, Jahnz M, Summerer D, Schultz PG, Uno T, Geierstanger BH (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. J Biomol NMR 46:89–100

32. Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015

33. Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

34. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. J Magn Reson 73:69–77

35. Schmieder P, Stern AS, Wagner G, Hoch JC (1994) Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain. J Biomol NMR 4:483–490

36. Rovnyak D, Frueh DP, Sastry M, Sun ZYJ, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. J Magn Reson 170:15–21

37. Laue ED, Mayger MR, Skilling J, Staunton J (1986) Reconstruction of phase sensitive 2D NMR spectra by maximum entropy. J Magn Reson 68:14–29

38. Hoch JC, Stern AS (1996) NMR data processing. Wiley-Liss, New York

39. Jaravine V, Ibraghimov I, Orekhov VY (2006) Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. Nat Methods 3:605–607

40. Marion D (2005) Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. J Biomol NMR 32:141–150

41. Kazimierczuk K, Zawadzka A, Kozminski W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

42. Kraulis PJ (1989) ANSIG: a program for the assignment of protein ^{1}H 2D NMR spectra by interactive computer graphics. J Magn Reson 84:627–633

43. Kraulis PJ, Domaille PJ, Campbell-Burk SL, Van Aken T, Laue ED (1994) Solution structure and dynamics of Ras p21-GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry 33:3515–3531

44. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

45. Johnson BA, Blevins RA (1994) NMR View: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

46. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278:313–352

47. Bartels C, Xia T, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 6:1–10

48. Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

49. Linge JP, O'Donoghue SI, Nilges M (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol 339:71–90

50. Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. J Biomol NMR 20:135–147

51. Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S (2000) Structural genomics projects in Japan. Nat Struct Biol 7 Suppl: 943–945

Chapter 3

In-Cell NMR in *Xenopus laevis* Oocytes

Rossukon Thongwichian and Philipp Selenko

Abstract

For the purpose of studying IDPs inside cells of higher organisms, several eukaryotic in-cell NMR systems have been developed over the past years. In this chapter we will focus on high-resolution in-cell NMR applications in *Xenopus laevis* oocytes, the first eukaryotic cellular model system to be established. In contrast to prokaryotic in-cell NMR samples, eukaryotic in-cell NMR specimens are prepared by cytoplasmic delivery of an exogenously produced, isotope-labeled protein into the non-isotope-labeled environment of the respective "host" cell. In-cell NMR applications in *Xenopus* oocytes rely on intracellular sample deposition by direct microinjection into the oocyte cytoplasm. Here, we describe the preparation of oocyte in-cell NMR samples for IDP studies in this cellular model environment.

Key word: *Xenopus laevis* oocytes

1. Introduction

In-cell NMR applications in eukaryotic cells enable high-resolution investigations of IDPs in cellular environments that exhibit many of the biological activities typically encountered in higher organisms (1). While NMR experiments that are typically employed to study IDPs in vitro are essentially the same than for in-cell NMR studies, Chapters 3–6 outline different protocols for protein delivery into eukaryotic cells, rather than specific NMR methods. These can roughly be divided into two parts: Protein delivery protocols for large amphibian cells (this chapter) and protein transduction methods for very much smaller, mammalian cells (Chapters 4–6). Especially with regard to IDP studies addressing the structural in vivo effects of posttranslational protein modifications (PTMs), eukaryotic in-cell NMR methods offer unique advantages over prokaryotic in-cell NMR measurements. Because the proteins to be delivered into the eukaryotic cellular environment are
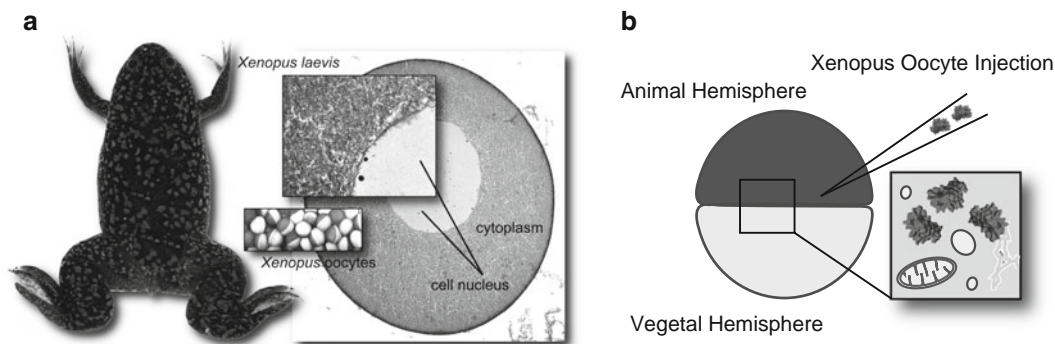
Fig. 1. (**a**) The African clawed frog *Xenopus laevis* and its oocyte cells. A thin-section, low-resolution microscopy image of a single oocyte (HE stained) depicts the large degree of macromolecular crowding in the oocyte cytoplasm. The cell nucleus occupies about 20 % of the total cell volume. The close-up view shows the nuclear envelope boundary that separates the cell nucleus from the cytosol. (**b**) Schematic representation of the spherical structure of the *Xenopus* oocyte. The darkly pigmented animal pole is clearly separated from the brightly colored vegetal hemisphere. Oocyte in-cell NMR samples are prepared by microinjection of isotope-labeled proteins into the cellular cytoplasm.

recombinantly produced in bacteria, they effectively lack the PTMs that are characteristic for many eukaryotic IDPs (2). Upon intracellular sample deposition, endogenous cellular enzymes "realize" this PTM deficiency and actively establish missing modifications, guided by the specific conditions of the respective cellular settings. In this way, IDPs "receive" physiologically relevant sets of PTMs the establishments of which-, and their structural consequences-, are directly observable by time-resolved in-cell NMR experiments (3).

Oocytes from the African clawed frog *Xenopus laevis* have long served as a standard cellular model system in many areas of biology (Fig. 1a) (4). Their large size (~1 μL cell volume) enables the direct delivery of biological compounds such as RNA, DNA or proteins by simple microinjection (Fig. 1b). Because isotope labeling occurs in a different cellular environment than the actual in-cell NMR measurements (i.e., *E. coli* versus *X. laevis* oocytes, respectively), no background labeling artifacts are encountered. While *Xenopus* oocytes may not represent a truly native cellular context for many eukaryotic IDPs, their general physical properties in terms of macromolecular crowding and cellular viscosity closely resemble most other eukaryotic cells. In that sense, crowding induced changes in IDP conformations for example, are likely to also occur in this cellular setting (5). In addition, *Xenopus* oocytes are naturally cell-cycle synchronized. In the ovary lobe of female frogs, oocytes mature into stage VI cells and arrest at the G2/M boundary of the first meiotic division. During oocyte to egg maturation, a hormonal trigger activates the synchronized progression through both meiotic cycles, until all cells arrest in metaphase of meiosis II. Concomitant changes in cellular PTM activities are uniformly displayed in all cells, which allows for "discrete" PTM studies even when multiple

cells are used. Furthermore, concerted cell-cycle progression can be activated in vitro by the addition of the hormone progesterone (PG), which renders this system an important laboratory tool for studying cell-cycle dependent signaling pathways and cellular kinase activities (6). In-cell NMR studies of biomolecules in *Xenopus* oocytes have been performed on folded (3, 7, 8) and natively unfolded proteins (9), as well as on RNA and DNA (10). Detailed descriptions of oocyte in-cell NMR sample preparations have been reported in these papers and also in (11). The goal of this chapter is to provide an updated manual that comprehensively integrates the experimental advances made over the recent years in the preparation of in-cell NMR samples employing *Xenopus laevis* oocytes.

2. Materials

1. Equipment: The protocol assumes that access to *Xenopus laevis* frogs is available and that proper animal care and safety protocols are obeyed at all times. Institutional licenses for housing and manipulating frogs must be in place. A standard laboratory setup for surgically removing oocytes and for manipulating oocytes by microinjection is needed. This includes a glass capillary puller to prepare injection needles (or access to a commercial vendor), a dissecting microscope with a built-in micrometer scale, and a standard pneumatic oocyte injection device (Harvard Instruments). Alternatively, large-scale oocyte manipulations can be performed with fully automated injection robots like the Roboinject™, or Robocyte™ (Multichannel Systems). In addition, standard laboratory equipment for recombinant protein production and purification, including a fast protein liquid chromatography (FPLC) system, is needed. Access to high-field (>500 MHz) solution-state NMR spectrometers must be available.

2. Pregnant mare serum gonadotropin (PMSG, Sigma Aldrich, USA).

3. Anesthetizing buffer, AB: 1 L of 0.25 mM Na_2CO_3 + 1 g of Tricaine mesylate, TMS.

4. OR2 buffer 82.5 mM NaCl, 2.5 mM KCl, 1 mM $MgCl_2$, 5 mM HEPES, pH 7.6.

5. ND96 buffer 96 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2 \times 2\ H_2O$, 1 mM $MgCl_2 \times 6H_2O$, 5 mM HEPES, pH 7.6.

6. Collagenase (Sigma Aldrich, USA).

3. Methods

3.1. Recombinant Protein Expression and Purification

1. Express the protein of interest recombinantly, incorporating the NMR-active isotope label or labels. Commonly, the gene encoding the protein of interest is inserted in a T7 inducible-expression vector transformed into BL21 (DE3) *E. coli* cells. Grow bacteria in labeled growth medium to an OD_{600} of 0.6 and induce expression with IPTG (see Note 1).

2. Purify the protein using standard protein chromatography steps (see Note 2).

3. Concentrate the pure protein to ~1 mM using a centrifugal filter unit, or any other appropriate device (see Note 3).

3.2. Oocyte Preparation

1. For microinjections, prime female frogs with 0.5 mL of 200 U/mL PMSG between 2 and 5 days prior to harvesting the oocytes (see Note 4).

2. Anesthetize frogs by placing them in 1 L of AB for 20 min. All subsequent steps of this section should be carried out at 18 °C.

3. Make a small incision through the abdominal tissue and remove the ovary lobes with forceps.

4. Suture the incision closed and allow the frogs to recover in isolation for 24 h before returning them to a communal tank (see Note 5).

5. Place the freshly harvested ovaries in a glass Petri dish filled with OR2 buffer.

6. To remove the vascular follicle layer that surrounds the oocytes prior to injections, incubate ovaries in OR2 supplemented with 5 % Collagenase, 1 % Trypsin inhibitor and 1 % BSA for 2 h on a rocking platform (see Note 6).

7. Decant collagenase solution and wash oocytes several times in OR2 until the final solution is no longer turbid.

8. Wash twice with ND96.

9. Manually sort out healthy looking stage VI oocytes. Allow cells to recover for 3–6 h in ND96 buffer prior to microinjection (see Note 7).

3.3. Oocyte Microinjection

1. Pull glass injection needles with apertures of 15 μm or greater (see Note 8).

2. Centrifuge the concentrated protein sample for 10 min at $20,000 \times g$ in a tabletop centrifuge at 4 °C to pellet any particular matter or precipitate that might otherwise clog the injection needle (see Note 9).

3. Mount the glass needle on the pneumatic injection device and calibrate by dispensing individual drops of protein sample in mineral oil. Measure the drop diameter under a microscope using the built-in micrometer scale. Choose injection-time and -pressure settings that yield injection volumes of 50 nL, or less (see Note 10).

4. For microinjections, seed prepared oocytes onto injection grids. Align cells with identical orientations in order to allow for rapid injections and comparable settings for needle penetration and sample deposition. These manipulations are carried out under a dissecting microscope.

5. Penetrate the immobilized oocytes at the equatorial plane separating the animal and the vegetal hemisphere of the cell with the injection needle. After sample deposition, withdraw the needle with equal precision and care.

6. Upon completion of oocyte injections, inspect all manipulated cells for the degree of incision and sort out any obviously damaged cells. Allow cells to recover for a minimum of 3 h (see Note 11).

7. Alternatively, oocyte manipulations can be carried with fully automated injection devices (see Note 12).

8. Following injections, transfer the oocytes to a large glass Petri dish. Wash cells thoroughly five times with excess volumes of ND96 at 45 min intervals and allow them to recover for at least 3 h.

3.4. Loading the NMR Tube

1. Transfer the injected oocytes to ND96 buffer containing 10 % D_2O.

2. Fill a Shigemi™ NMR tube with ND96/D_2O buffer (see Note 13). Carefully collect the oocytes with a pipette and add them individually to the tube, allowing them to sediment by gravity. Occasionally swirl the tube to ensure optimal settling and packing.

3. Count the number of oocytes per NMR sample and note the resulting volume of the specimen, as this will enable the accurate correlation of the effective NMR concentration (C_{NMR}) to the intracellular concentration (C_{Cell}) of the injected protein (see Note 14).

4. Proceed to the NMR spectrometer for in-cell NMR recordings (see Note 15).

5. Protein leakage from *Xenopus* oocytes should be assessed before-, and after in-cell NMR experiments. Carefully decant oocyte buffer and check for the presence of injected protein in the supernatant by Western blotting (12).

4. Notes

1. The concentration of IPTG, the induction temperature, and the duration of the induction can all be varied to find optimal conditions for maximal, soluble protein expression. The volume of cells required depends entirely on the protein yield; a single *Xenopus* oocyte in-cell NMR sample will typically require in excess of 2.5×10^{-8} mol of pure protein (~0.5–1 mM). Detailed protocols for recombinant protein expression and purification (13), as well as for stable-isotope labeling for NMR purposes (14), are available elsewhere.

2. Large affinity or fluorescent tags, such as GST, MBP, or GFP, are not recommended unless a specific protease cleavage site is introduced enabling the subsequent removal of the tag. Commonly used proteases include Factor X and TEV. Small peptide tags, especially hexahistidine, appear relatively innocuous in *Xenopus* oocytes, but will yield NMR resonances similar to those of IDPs and may obscure some chemical shifts of interest. Choose the final protein buffer considering the physiology of the oocytes and the need to keep the protein soluble. A good starting point is 50 mM sucrose, 150 mM NaCl, 25 mM HEPES, pH 7.5, 1 mM DTT.

3. Remember that subsequent injection steps will dilute the pure protein by a factor of ~20 upon oocyte delivery. This number is based on the notion that 50 nL of protein are injected into a cell volume of ~1 μL. Thus, if ~50 μM are envisaged as an arbitrary target for the effective concentration of the in-cell NMR sample (C_{NMR}), the oocyte experiment will require a starting protein concentration of at least 1 mM.

4. Alternatively, oocytes can be obtained without hormone priming. This is especially suitable during northern hemisphere winter months (*X. laevis* is an amphibian from South Africa and its intrinsic "seasonal" clock suggests a favorable summer climate when in fact winter prevails in the northern hemisphere). Oocyte quality may vary greatly from frog to frog. It is therefore suggested to hormone prime a couple of individuals in order to only select the best oocytes. On average, the number of healthy stage VI oocytes that can be obtained from a single frog, usually several hundred, suffice for multiple in-cell NMR experiments.

5. Oocytes are essentially prepared as described in ref. 15. This protocol represents the most up-to-date version of classical oocyte preparation routines (4). Frogs can also be sacrificed after oocyte removal, if laboratory regulations permit to do so.

6. Alternatively, the follicle layer can be removed by manually peeling it off under a dissecting microscope (16). This procedure

requires considerable practice and is quite time-consuming. The advantages over treating with collagenase are that follicle removal is complete, the oocytes are typically healthier, and components of the extracellular membrane have not undergone as harsh a treatment (which might not be critical for in-cell NMR measurements of intracellular proteins).

7. Stage VI oocytes are easily identified as the largest oocytes in the population, and they are typically healthy when the darkly pigmented hemisphere (the animal pole) appears rich and uniformly colored. About 200 oocytes are needed for a single in-cell NMR experiment. ND96 buffer can be substituted with MBS buffer, without calcium (88 mM NaCl, 1 mM KCl, 1 mM MgSO$_4$, 2.5 mM NaHCO$_3$, 5 mM HEPES, pH 7.55) at any stage of the oocyte preparation.

8. Concentrated protein solutions are typically more viscous than RNA that is most often injected into *Xenopus* oocytes. Therefore, the aperture of the needle needs to be wide enough to allow the protein solution to pass through while also small enough to minimize the trauma inflicted upon the oocytes. The aperture diameter should be determined empirically for each protein. Alternatively, pre-pulled needles can be obtained from commercial sources and have been found to consistently yield highly reproducible results, especially when used in combination with robotic injection devices (11, 17).

9. Some IDPs are aggregation prone and the highly concentrated injection solution may result in the formation of insoluble precipitates. Sedimentation of those species by an additional centrifugation step alleviates clogging problems. This step is particularly important when working with IDPs.

10. Most dissecting microscopes contain built in micrometer scales, visible through the ocular, which are used to determine the drop volume by employing the arithmetic diameter–volume relation assuming a perfect sphere shape (18).

11. Sakai et al. have reported the coinjection of an inert fluoroscopic dye in order to visually inspect for sample leakage (19). Possible dye binding to the protein of interest should first be ruled out by in vitro NMR measurements of a dye–protein mixture.

12. These systems can be employed to perform large-scale manipulation routines within a minimal amount of time (<2 min per 96-well plate). Mount pre-pulled injection needles (#38GC100TF-10, nozzle aperture ~15 μm, Multi Channel Systems) on a "standard" pneumatic oocyte injection device for sample loading through the needle tip and conventional drop-volume calibration in mineral oil. We found that this approach was superior to "back-loading" via the larger needle opening at

the mounting end, as the formation of trapped air bubbles is consistently avoided. Transfer the injection needle containing the labeled protein sample onto the injection arm of the robotic injection device and lock 96-well plates, containing one oocyte cell per well, onto the injection platform. Align the injection needle and 96-well plate manually according to the manufacturer's instructions. Automated injection procedures are typically carried out with settings of 0.1 bar holding pressure, 0.7 bar injection pressure, 200 ms injection time, and 500 μm injection depth, which corresponds to calibrated sample volumes of 50 nL per oocyte/injection (STD ±10 % or ±5 nL).

13. Bodart et al. have reported the usage of 20 % Ficoll in the final in-cell NMR buffer (9). The higher overall density of this solution significantly enhances the time span over which oocytes maintain their intact morphology. This appears to be due to reduced packing damage. Do not insert the Shigemi™ plunger, nor remove the excess buffer from the NMR tube for the in-cell experiment.

14. The effective NMR concentration (C_{NMR}) of the final in-cell NMR sample will be determined by the intracellular concentration of the "delivered" protein (C_{Cell}), the number of cells in the NMR sample (N_{Cell}, typically 200), their cell volume (V_{Cell}, 1 μL on average), the final NMR sample volume (V_{NMR}, typically 250 μL), and the corresponding volume dilution factor, VDF ($V_{NMR}/V_{Cell} \times N_{Cell}$). For *Xenopus* oocytes the VDF is usually 1.25. According to $C_{NMR} \times V_{NMR} = C_{Cell} \times (V_{Cell} \times N_{Cell})$ the effective NMR concentration can be calculated as $C_{NMR} = C_{Cell}/VDF$. For example, in order to obtain a spectrum with an effective NMR concentration (C_{NMR}) of 10 μM of isotope labeled protein in 250 μL of NMR sample volume (V_{NMR}), for 200 cells (N_{Cell}) with an average cell volume (V_{Cell}) of 1 μL a total intracellular protein concentration (C_{Cell}) of 12.5 μM of isotope-labeled IDP must be reached. Given the dilution factor upon oocyte injection (50 nL injection volume into 1 μL of cell volume, i.e., ~20) the protein concentration in the injection needle has to be 250 μM.

15. Oocyte in-cell NMR samples are inherently inhomogeneous. For this reason, any attempts to shim the sample will be unsuccessful and should be omitted altogether. NMR pulse calibrations are best set up by fast one-dimensional heteronuclear correlation experiments varying the individual pulse lengths in small increments. Due to the high viscosity of the intracellular environment, pulses are typically longer than for pure in vitro samples (although this increase in pulse lengths is typically less drastic for IDPs with high internal mobility).

References

1. Ito Y, Selenko P (2010) Cellular structural biology. Curr Opin Struct Biol 20:640–648

2. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 10 Suppl 1: S7

3. Selenko P, Frueh DP, Elsaesser SJ, Haas W, Gygi SP, Wagner G (2008) In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. Nat Struct Mol Biol 15:321–329

4. Murray AW (1991) *Xenopus laevis*: practical uses in cell and molecular biology. Methods Cell Biol 36:1–718

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

6. Crane R, Ruderman J (2006) Using *Xenopus* oocytes to study signal transduction. Methods Mol Biol 322:435–445

7. Sakai T, Tochio H, Tenno T, Ito Y, Kokubo T, Hiroaki H, Shirakawa M (2006) In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. J Biomol NMR 36:179–188

8. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. Proc Natl Acad Sci USA 103:11904–11909

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

10. Hansel R, Foldynova-Trantirkova S, Lohr F, Buck J, Bongartz E, Bamberg E, Schwalbe H, Dotsch V, Trantirek L (2009) Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. J Am Chem Soc 131:15761–15768

11. Serber Z, Selenko P, Hansel R, Reckel S, Lohr F, Ferrell JE Jr, Wagner G, Dotsch V (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. Nat Protoc 1:2701–2709

12. Lee C (2007) Western blotting. Methods Mol Biol 362:391–399

13. Hewitt L, McDonnell JM (2004) Screening and optimizing protein production in *E. coli*. Methods Mol Biol 278:1–16

14. Serber Z, Corsini L, Durst F, Dotsch V (2005) In-cell NMR spectroscopy. Methods Enzymol 394:17–41

15. Sive HL, Grainger RM, Harland RM (2010) Isolation of *Xenopus* oocytes. Cold Spring Harb Protoc. pdb prot5534

16. Sive HL, Grainger RM, Harland RM (2010) Defolliculation of *Xenopus* oocytes. Cold Spring Harb Protoc. pdb prot5535

17. Schnizler K, Kuster M, Methfessel C, Fejtl M (2003) The roboocyte: automated cDNA/mRNA injection and subsequent TEVC recording on *Xenopus* oocytes in 96-well microtiter plates. Receptors Channels 9:41–48

18. Sive HL, Grainger RM, Harland RM (2010) Calibration of the injection volume for micro-injection of *Xenopus* oocytes and embryos. Cold Spring Harb Protoc. pdb prot5537

19. Sakai T, Tochio H, Inomata K, Sasaki Y, Tenno T, Tanaka T, Kokubo T, Hiroaki H, Shirakawa M (2007) Fluoroscopic assessment of protein leakage during *Xenopus* oocytes in-cell NMR experiment by co-injected EGFP. Anal Biochem 371:247–249

In-Cell NMR in Mammalian Cells: Part 1

Beata Bekei, Honor May Rose, Michaela Herzig, Alexander Dose, Dirk Schwarzer, and Philipp Selenko

Abstract

Many mammalian IDPs exert important biological functions in key cellular processes and often in highly specialized subsets of cells. For these reasons, tools to characterize the structural and functional characteristics of IDPs inside mammalian cells are of particular interest. Moving from bacterial and amphibian in-cell NMR experiments to mammalian systems offers the unique opportunity to advance our knowledge about general IDP properties in native cellular environments. This is never more relevant than for IDPs that exhibit pathological structural rearrangements under certain cellular conditions, as is the case for human α-synuclein in dopaminergic neurons of the *substantia nigra* in the course of Parkinson's disease, for example. To efficiently deliver isotope-labeled IDPs into mammalian cells is one of the first challenges when preparing a mammalian in-cell NMR sample. The method presented here provides a detailed protocol for the transduction of isotope-labeled α-synuclein, as a model IDP, into cultured human HeLa cells. Cellular IDP delivery is afforded by action of a cell-penetrating peptide (CPP) tag. In the protocol outlined below, the CPP tag is "linked" to the IDP cargo moiety via an oxidative, disulfide-coupling reaction.

Key words: Cell-penetrating peptides, Disulfide coupling, Alpha-synuclein, HIV-Tat, HeLa cells

1. Introduction

Although isotope-labeled IDPs can in principle be microinjected into mammalian cells, such procedures are hardly feasible in practice because millions of cells would need to be individually manipulated in order to produce a single in-cell NMR sample (mammalian cells are typically several orders of magnitudes smaller than *Xenopus* oocytes). For this reason, all mammalian in-cell NMR applications today exploit protein delivery schemes that target many cells in parallel, and in a batch-like manner. In essence, two different procedures for intracellular protein transduction have been described for in-cell NMR applications in mammalian cells (1, 2). The first takes advantage of the unique properties of cell-penetrating peptide

Fig. 1. (**a**) Schematic representation of CPP-mediated protein delivery into mammalian cells. CPP–cargo transduction mostly occurs via endosomal uptake routes. Once released into the reducing environment of the cytoplasm, the CPP–cargo disulfide bond is broken. This leads to the accumulation of free cargo in the cytoplasm of the host cell. (**b**) Outline of CPP–cargo constructs employed in this study. Cargo protein (α-synuclein) was produced with N- or C-terminal cysteines that are used for disulfide coupling to the CPP. The inclusion of a fluorescence label enables assessment of CPP–cargo coupling efficiency and intracellular protein uptake. (**c**) Structure of the HIV-Tat (CPP) construct described in this protocol (MW 2.4 kDa). The peptide is modified to include a N-terminal *S*-3-nitro-2-pyridinesulphenyl (Npys) activated cysteine and C-terminal Fluorescein ($\lambda_{\text{Excitation}}$ 492 nm, $\lambda_{\text{Emission}}$ 517 nm).

(CPP) tags (1). The second exploits bacterial pore-forming toxins that will be discussed in the next chapter (2).

CPPs come in many different flavors (3). They usually comprise 10–15 amino acid residues, are rich in basic amino acids, sometimes occur in the form of an amphipathic alpha-helix (4), and have the ability to translocate through biological membranes (5, 6). For these reasons they are often employed as "carrier peptides" to deliver biomolecules into live cells (Fig. 1a). A CPP–cargo construct can be produced several ways. In the following sections, we will describe the coupling of a synthetic CPP to a recombinantly produced cargo IDP (specifically human α-synuclein). Coupling is achieved via engineered cysteines in the CPP-, and cargo protein moieties. As this method involves the separate production of the CPP and the cargo protein, it offers several advantages for in-cell NMR experiments. First, only the IDP portion can be prepared in an isotope-labeled form and thus constitutes the only detectable species in the in-cell NMR experiment (the CPP remains "invisible"). Second, because the CPP is coupled to the IDP in a separate reaction, a number of different CPPs can be prepared in advance and then coupled to the IDP cargo of interest to quickly assess the best CPP–cargo combination for optimal protein delivery (see below). Third, because the method described here involves CPP production by chemical solid-phase peptide synthesis (SPPS), additional chemical entities like a fluorescence label, a biotin tag, or a sulfhydryl activating group, can easily be incorporated and exploited to verify cellular uptake (see Chapter 6). Fourth, once

the CPP–cargo has been successfully delivered into mammalian cells, the reducing environment of the cytoplasm breaks the disulfide bond between the CPP and the cargo and thereby ensures efficient cargo release. Alternative CPP "coupling" methods such as the recombinant production of a single CPP–cargo fusion protein will not be discussed.

CPPs mediate cellular protein transduction by different biological mechanisms and by different uptake routes (7–11). Thus, for any given CPP–cargo–cell line combination, uptake efficiencies may vary greatly (12). In most instances, the choice of cargo protein and cell line to be targeted is given by the nature of the research project. Therefore, the choice and design of the CPP–cargo construct should be evaluated carefully. As the method presented here may be applied to different CPPs and IDPs we have chosen general terms throughout the protocol. In our case, "cargo" refers to isotope-labeled human alpha-synuclein that has been engineered to contain a N- or C-terminal cysteine residue to which the CPP can be coupled (Fig. 1b). "CPP" refers to SPPS synthesized HIV-Tat (aa 47–57) containing a fluorescein dye and a *S*-3-nitro-2-pyridinesulphenyl (Npys) activating group to enhance the coupling efficiency (Fig. 1c).

2. Materials

1. Equipment: The following protocol assumes that standard laboratory equipment for recombinant protein production and purification, including a fast protein liquid chromatography (FPLC) system, is available. In addition, CPP synthesis requires a SPPS setup including a reversed-phase high-pressure liquid chromatography (HPLC) system, or access to a commercial peptide production facility. Cell culture equipment for maintaining and manipulating mammalian cells, including a sterile workbench, a CO_2 incubator and a basic tissue culture microscope are needed. Access to high-field (>500 MHz) solution-state NMR spectrometers must be available.

2. Cargo: Cargo proteins should contain single cysteines at either the N- or the C-terminus. Cysteines that are part of the cargo protein sequence may also be used for coupling. The cargo protein should be available in an appropriately isotope-labeled form for the envisaged in-cell NMR experiment. Detailed protocols for recombinant protein expression and purification (13), as well as for stable-isotope labeling for NMR purposes (14) are described elsewhere.

3. CPP: The CPP to be coupled to the cargo protein must contain a cysteine residue that is preferably *S*-3-nitro-2-pyridinesulphenyl

(Npys) activated. An additional fluorescence dye can be incorporated for convenient in-cell detection by microscopy methods (see Chapter 6). Other chemical entities like a Biotin tag for example, can also be added.

4. CPP coupling buffer (CB): 20 mM Phosphate or 50 mM HEPES, 150 mM NaCl, pH 7.0.

5. Appropriate gel-filtration (size-exclusion) and ion-exchange columns (GE Healthcare, USA).

6. SDS-PAGE equipment.

7. Coomassie Blue staining solution.

8. Immobilized-TCEP reducing column (Thermo scientific, USA, Meridian Rd.).

9. Suitable cell culture medium for HeLa cells: Complete DMEM (low Glucose, 5 mM Glutamine, 10 % Fetal Bovine Serum, (FBS) (PAA Laboratories, Canada)).

10. Phosphate buffered saline (PBS), cell culture grade, without Calcium/Magnesium (PAA Laboratories, Canada).

11. 6-Well cell culture plates/T175 cell culture flasks.

12. 0.25× Trypsin-EDTA (PAA Laboratories, Canada).

13. Low melting agarose (USB Affymetrix, USA, California).

14. In-cell NMR buffer: DMEM (serum-free), 5 mM HEPES, pH 7.2, 90 mM D-glucose, 5 % D_2O, or DMEM (serum-free), 10 % D_2O.

15. RIPA buffer (denaturing): Tris 50 mM, NaCl 150 mM, 0.1 % SDS, 0.5 % Na-Deoxycholate, 1 % Triton X 100 or NP40, protease inhibitors.

16. Hemocytometer.

3. Methods

3.1. CPP–Cargo Coupling

1. Before starting the procedure, carefully assess the required number of mammalian cells to prepare the in-cell NMR sample (see Note 1), the target volume of the final in-cell NMR sample (see Note 2) and the target concentration of the isotope-labeled protein inside the cells (see Note 3).

2. Produce the CPP by SPPS. Extend the CPP sequence by a N- or C-terminal S-3-nitro-2-pyridinesulphenyl (Npys) activated cysteine. Reversed-phase HPLC purify the CPP. 10–40 mg of lyophilized CPP is needed as a starting material for a single in-cell NMR sample (see Note 4).

3. Produce NMR isotope-labeled, recombinant cargo protein. The cargo protein must contain at least one cysteine residue for CPP coupling. This can also be engineered by mutagenesis (preferable at the N- or C-terminus of the protein). The cysteine residue must be in a reduced state (i.e., as sulfhydryl) prior to coupling. The required concentration of cargo protein before coupling is ~2 mM in a total volume of 2 mL CB (see Note 5).

4. Initially test CPP–cargo coupling efficiencies in small-scale experiments. Non-isotope-labeled cargo proteins may be used at this point. For small-scale coupling trials, reduce all volumes indicated below by a factor of 10.

5. Dissolve CPP to a final concentration of 4 mM in 2 mL of CB, adjust pH to 7.0 (see Note 6).

6. Mix 2 mL CPP and 2 mL cargo stock solution (2 mM). The CPP will be in ~4-fold molar excess over the cargo protein (see Note 7).

7. Incubate the coupling reaction at room temperature (RT) for 10–60 min. Note that steps 5–7 will require optimization to ensure preferential formation of the desired CPP–cargo product (see Note 8).

8. Determine coupling efficiency by running a nonreducing SDS-PAGE (Fig. 2) (see Note 9).

9. Optimize coupling conditions accordingly and proceed to large scale coupling reaction (step 5).

10. After coupling, concentrate CPP–cargo mixture to a final volume of 2 mL using a centrifugal filter unit, or any other appropriate device.

11. Apply and separate by size-exclusion chromatography (see Note 10).

12. Collect fractions containing the desired CPP–cargo and further purify by ion-exchange chromatography (see Note 11).

13. Concentrate CPP–cargo to 1 mM (~2 mL) (see Note 12).

3.2. Cellular CPP–Cargo Delivery

1. To determine the efficiency of cellular CPP–cargo delivery, small-scale trials are carried out first. Seed $2.5–3 \times 10^5$ cells per well in a 6-well plate (surface area 9.6 cm^2/well) (see Note 13).

2. Incubate cells for 24 h in complete DMEM at 37 °C in a CO_2 incubator. Cells should be ~80 % confluent at the beginning of the experiment (~$5–6 \times 10^5$ cells) (see Note 14).

3. *Optional* Remove culture medium and wash twice with PBS (3 mL/well). Add 0.5 μL/well of 50–250 μM Pyrenebutyrate (PA) in PBS. Incubate for 3–10 min at 37 °C (see Note 15).

4. Add 0.5 mL/well of serial CPP–cargo dilutions (0.05–0.5 mM) in prewarmed PBS to test for optimal uptake concentrations.
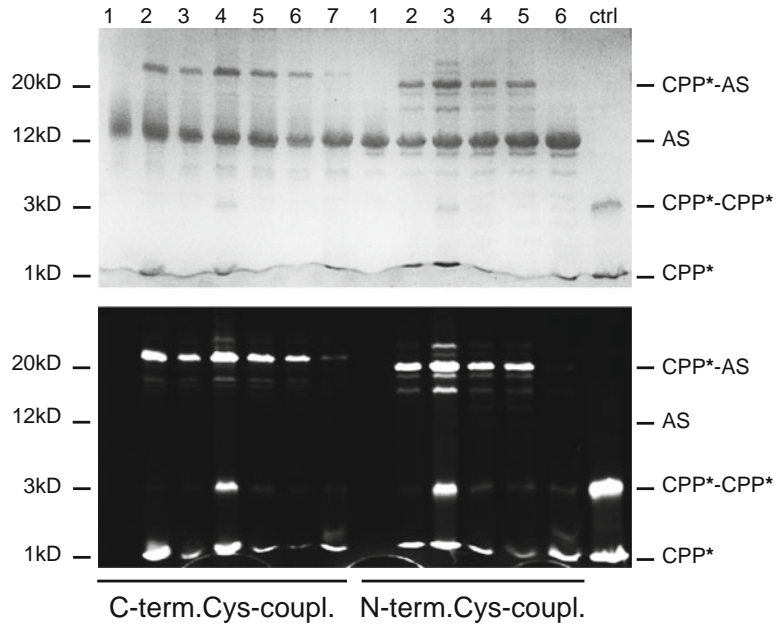
Fig. 2. Assessment of CPP–cargo coupling efficiencies by nonreducing SDS-PAGE. *Top panel*: Coomassie stained PAGE of C-, or N-terminal cysteine α-synuclein (AS)/CPP coupling reactions. Lanes 1 are AS input samples. Differences in AS/CPP coupling efficiencies and levels of CPP–CPP dimers are readily discerned for coupling reactions at 37 °C with incubation times of 10 min (*lane 2*), 30 min (*lane 3*) ,and 1 h (*lane 4*). Reactions at 4 °C, ranging from 10 min (*lanes 6 and 7*) to 1 h (*lane 5*) display similar differences in coupling efficiencies. The control lane at the far right (ctrl) indicates the extent of CPP–CPP dimer formation in the absence of AS (37 °C for 1 h). The bottom panel shows the same PAGE under UV illumination. The CPP moiety contains the fluorescein label (CPP*).

5. Incubate cells with CPP–cargo mixtures for 10–60 min at 37 °C in a CO_2 incubator (see Note 16).

6. *Optional* Cellular CPP–cargo incubations can be repeated multiple times for enhanced protein uptake. If this option is chosen, allow cells to recover for 1 h in complete DMEM at 37 °C in between the individual incubation steps (see Note 17).

7. Remove and collect CPP–cargo solution and wash cells twice with prewarmed PBS. Add fresh complete DMEM and allow cells to recover for 1 h at 37 °C in a CO_2 incubator (see Note 18).

8. Determine cellular CPP–cargo uptake by semiquantitative Western blotting of lysates prepared from manipulated cells, or by suitable alternative methods (see Note 19).

3.3. In-Cell NMR Sample Preparation

1. Once optimal delivery conditions have been found, the protein transduction procedure is scaled up for in-cell NMR sample preparation. To this end, seed 4–5×10^6 cells in one T175 cell culture flask. (It should be noted that a fluorescent tag is not necessarily desired on the CPP–cargo construct used for the final in-cell NMR sample preparation).

2. Incubate cells for 24 h in complete DMEM at 37 °C in a CO_2 incubator. Cells should be ~80 % confluent at the beginning of the experiment (~0.8–1 $\times$ 10^7 cells).

3. Replace medium with 4 mL/flask prewarmed PBS, containing CPP–cargo at the concentration that was determined to yield the best protein uptake results.

4. Incubate cells with CPP–cargo under optimized conditions (see Note 20).

5. Remove and collect CPP–cargo solution and wash cells twice with prewarmed PBS. Add fresh complete DMEM and allow cells to recover for 1 h at 37 °C in a CO_2 incubator.

6. In order to transfer the manipulated cells to the NMR tube, they first have to be detached from the cell culture flask. Wash once with prewarmed PBS to remove FBS.

7. Incubate with 4 mL (minimal surface volume) 0.25 % Trypsin/ EDTA in prewarmed PBS for 3 min (see Note 21).

8. Detach cells by tapping the side of the cell culture flask.

9. Add 20 mL (5$\times$ volume) prewarmed complete DMEM to inactivate Trypsin.

10. Transfer the cell suspension to a 50 mL centrifugation tube. Remove a 20 μL aliquot and sediment by low-speed centrifugation (~400 $\times$ g).

11. Resuspend the pellet in 20 μL PBS and add 20 μl 0.4 % Trypan Blue in PBS.

12. Mix cells carefully and remove 10 μL for a Trypan Blue cell viability test using a hemocytometer (see Note 22).

13. Sediment the remaining ~24 mL of the original cell suspension (step 10) by low-speed centrifugation (~400 $\times$ g).

14. Wash cell slurry twice in the final in-cell NMR buffer (see Note 23)

15. Resuspend cells in 500 μL (or 300 μL, depending on the size of the NMR tube) in-cell NMR buffer (see Note 24).

16. Transfer cells to the NMR tube and proceed to in-cell NMR experiments.

17. *Optional* Test for protein leakage and cell viability before and after in-cell NMR measurements as outlined in Chapter 6.

4. Notes

1. On average, 1–3 $\times$ 10^7 cells are needed for a single in-cell NMR sample. This number depends on the kind of cells that are used and on their individual cell volumes, i.e., ~2 pL for human HeLa cells (15).

2. The final volume of the in-cell NMR sample is determined by the kind of NMR tube that is to be used. On narrow-bore NMR probes, standard 5 mm (~500 μL), or Shigemi™ NMR tubes (~300 μL) may be employed. Their difference in volume will require different numbers of cells.

3. The effective NMR concentration (C_{NMR}) of the final in-cell NMR sample will be determined by the intracellular concentration of the "delivered" protein (C_{Cell}), the number of cells in the NMR sample (N_{Cell}), their individual cell volume (V_{Cell}), the final NMR sample volume (V_{NMR}) and the corresponding volume dilution factor, VDF ($V_{NMR}/V_{Cell} \times N_{Cell}$). According to $C_{NMR} \times V_{NMR} = C_{Cell} \times (V_{Cell} \times N_{Cell})$ the effective NMR concentration can be calculated as $C_{NMR} = C_{Cell}/VDF$. For example, in order to obtain a spectrum with an effective NMR concentration (C_{NMR}) of 10 μM of isotope labeled protein in 300 μL of NMR sample volume (V_{NMR}), for 1×10^7 cells (N_{Cell}) with an average cell volume (V_{Cell}) of 2 pL (i.e., 2×10^{-6} μL) a total intracellular protein concentration (C_{Cell}) of 150 μM of isotope-labeled IDP must be reached (VDF in this case is 15). It is therefore suggested to first determine the minimum protein concentration that is required for suitable in vitro NMR results (i.e., the lower limit of C_{NMR}). This will define the benchmark intracellular protein concentration (C_{Cell}) that has to be reached in a defined number of cells (N_{Cell}) of volume (V_{Cell}), for satisfactory in-cell NMR results.

4. The indicated amount of CPP starting material is based on average yields of CPP-coupled cargos that can be obtained with the outlined protocol. Individual coupling efficiencies, losses during CPP–cargo purification steps and overall yields may therefore vary for different CPP–cargo combinations. Additional chemical entities like a fluorescence label for in-cell detection by microscopy methods, or a biotin-tag for affinity pull-down experiments, may be incorporated at this point.

5. The required amount of isotope-labeled cargo protein is based on average coupling yields (>60 %). Efficient reduction of cysteine residues is achieved by running the cargo through a Tris(2-carboxyethyl)phosphine (TCEP) column prior to the coupling reaction. No other reducing agents must be present at this point.

6. Some CPPs may not be sufficiently soluble in aqueous solutions to achieve the suggested concentration of the stock solution. In such instances, test different pH ranges for solubilizing and coupling.

7. A molar excess of CPP generally ensures that the desired species, i.e., CPP–cargo is preferably formed. Ratios from 1:1 to 4:1 (CPP–cargo) are good starting points. One drawback of

using large excess of CPP is the generation of CPP–CPP disulfide dimers that constitute unwanted side-products and have to be removed in turn (see below).

8. The following parameters have been shown to greatly influence coupling efficiencies: incubation time, i.e., minutes to hours, incubation temperature, i.e., 4–37°C, and CPP to cargo molar ratio. These parameters should be optimized in parallel for best coupling results.

9. To determine overall coupling efficiencies, small aliquots of the coupling reaction are removed at different time-points and run on a SDS-PAGE that is then stained with Coomassie Blue. Successful coupling is assessed by the appearance of the desired CPP–cargo protein band. Whenever the CPP contains a fluorescence dye, coupling efficiency can also be determined by UV illumination. In most instances, two protein populations will be present: High molecular weight (MW) components, i.e., CPP–cargo, cargo–cargo, and uncoupled cargo and low MW species, i.e., free CPP and CPP–CPP dimers.

10. To initially separate high-, from low-MW components (see above) a gel filtration chromatography (GF) step is employed. Some FLPC systems allow for detection of the emission wavelength of the CPP fluorophore, which may assist in the identification of the CPP–cargo product. After the GF run, the high MW fractions containing the desired CPP–cargo, as well as free cargo molecules are pooled. Fractions containing the low MW species, i.e., CPP and CPP–CPP dimers can be discarded.

11. In order to purify the CPP–cargo from the pool of nonreacted cargo molecules an additional chromatography step is employed. The positive charge of the CPP usually alters the pI of the CPP–cargo product in a manner that is sufficient to separate it from free cargo molecules by ion exchange chromatography. The optimal loading pH and slope of the salt gradient to elute the desired species must be experimentally determined for best separation results. If the final eluate contains a salt concentration above 150 mM, dialysis of the CPP–cargo solution is suggested.

12. When measuring final CPP–cargo concentrations, beware that the presence of a fluorescence label on the CPP will necessitate a correction factor for accurately determining the precise amount of product by UV/VIS spectrophotometry.

13. Seeding density will vary depending on the cell line used. The number given here is based on the average cell size of human HeLa cells. It is generally recommended to employ cells with a low passage number.

14. HeLa cells have an average doubling time of 24 h. Therefore, this incubation time will vary depending on the cell line used.

15. Incubation of cells with pyrenbutyrate (PA) has been shown to enhance CPP–cargo uptake ([16](16)). It is recommended to perform initial PA toxicity tests to determine suitable PA concentrations and incubation times for a given cell line. The initial PBS wash is required to remove FBS, as it will diminish PA activity. If PA is applied, a cell recovery step (30–60 min) in DMEM (serum-free) is suggested ([1](1)). The CPP–cargo incubation solution that is then added must also contain PA.

16. Cellular CPP–cargo uptake will vary depending the CPP–cargo concentrations and chosen incubation times. It is therefore advisable to optimize these parameters accordingly. In our experience, incubation with 300 μM CPP–cargo results in 10–20 % uptake efficiency. Thus, a single incubation step yields an intracellular CPP–cargo concentration of 30–60 μM and a resulting effective NMR concentration (C_{NMR}) of 4–8 μM, in a 300 μL in-cell NMR sample volume (V_{NMR}) with 1×10^7 HeLa cells (N_{Cell}).

17. Inomata et al. have reported four rounds of CPP–cargo incubations for their mammalian in-cell NMR samples ([1](1)). They achieved an effective NMR concentration (C_{NMR}) of 30 μM in 200μL of NMR sample volume (V_{NMR}). With a total number of 1×10^7 cells (N_{Cell}) this equates to an intracellular protein concentration (C_{Cell}) of approximately 150 μM, according to the equation outlined above. Assuming equal amounts of intracellular protein deposition for every round of incubation, this relates to a protein uptake efficiency of 15 % in each incubation step.

18. CPP–cargo solutions can be recycled for additional incubation steps. As only 10–20 % of CPP–cargo is taken up per round of incubation, substantial amounts of noninternalized CPP–cargo remain present in the supernatant. This portion of CPP–cargo can easily be repurified by conventional chromatography methods and reused at later points in time.

19. To prepare cell lysates for Western-blotting, add RIPA buffer directly to the cells and detach them from the culture dish. Separate proteins by SDS-PAGE and perform Western-blotting according to standard protocols ([17](17)) with antibodies against the cargo protein. Determine cellular protein uptake by comparing cargo signal intensities to a series of known cargo concentrations run on the same gel. Quantification is afforded by commercial imaging readout software packages. Additional methods for the quantification of CPP–cargo uptake are outlined in Chapter [6](6).

20. At this point, optimal conditions for efficient cellular CPP–cargo uptake should be known. If PA treatment, and multiple rounds of CPP–cargo incubations, were advantageous for cellular protein uptake in small-scale experiments, they should also be used here.

21. Trypsin incubation times may vary for different cell lines. Cell morphology/viability should be assessed at this point by bright field microscopy. Another beneficial aspect of the Trypsin treatment is the proteolytic degradation of membrane-attached, but not internalized CPP–cargo molecules that might compromise the in-cell NMR readout.

22. This cell count assay is used to determine overall cell viability after CPP–cargo delivery (18). On average, more than 80 % of the cells should score as viable (i.e., nonblue) in the Trypan Blue assay.

23. Inomata et al. use 5 mM HEPES (pH 7.2) with 90 mM D-glucose (supplemented with 5 % D_2O) as the final in-cell NMR buffer (1). In our hands, serum-free DMEM (10 % D_2O) works equally well.

24. Alternatively, cells can be embedded in a solid support matrix to avoid sedimentation during the in-cell NMR experiment. This often yields better overall in-cell NMR results. We sometimes employ low-melting agarose as a biologically and spectroscopically inert embedding material. The low-speed centrifugation cell pellet obtained in step 13 is resuspended in half of the final in-cell NMR sample buffer (i.e., 250 μL or 150 μL, depending on the size of the NMR tube). An equal volume of a 0.4 % low-melting agarose in PBS (20 % D_2O) solution (37 °C) is added. After careful mixing of the two suspensions, the sample is transferred to the NMR tube and solidified by briefly placing it at 4 °C. The resulting 0.2 % agarose mixture is sufficient to prevent sedimentaion of 1×10^7 HeLa cells for up to 24 h.

References

1. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature 458:106–109

2. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I (2009) Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J Am Chem Soc 131:10834–10835

3. Dietz GP, Bahr M (2004) Delivery of bioactive molecules into the cell: the Trojan horse approach. Mol Cell Neurosci 27:85–131

4. Herbig ME, Weller K, Krauss U, Beck-Sickinger AG, Merkle HP, Zerbe O (2005) Membrane surface-associated helices promote lipid interactions and cellular uptake of human calcitonin-derived cell penetrating peptides. Biophys J 89:4056–4066

5. El-Andaloussi S, Holm T, Langel U (2005) Cell-penetrating peptides: mechanisms and applications. Curr Pharm Des 11:3597–3611

6. Foerg C, Merkle HP (2008) On the biomedical promise of cell penetrating peptides: limits versus prospects. J Pharm Sci 97:144–162

7. Ferrari ME, Nguyen CM, Zelphati O, Tsai Y, Felgner PL (1998) Analytical methods for the characterization of cationic lipid-nucleic acid complexes. Hum Gene Ther 9:341–351

8. Fittipaldi A, Ferrari A, Zoppe M, Arcangeli C, Pellegrini V, Beltram F, Giacca M (2003) Cell membrane lipid rafts mediate caveolar endocytosis of HIV-1 Tat fusion proteins. J Biol Chem 278:34141–34149

9. Wadia JS, Stan RV, Dowdy SF (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. Nat Med 10:310–315

10. Kaplan IM, Wadia JS, Dowdy SF (2005) Cationic TAT peptide transduction domain enters

cells by macropinocytosis. J Control Release 102:247–253

11. Richard JP, Melikov K, Brooks H, Prevot P, Lebleu B, Chernomordik LV (2005) Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors. J Biol Chem 280:15300–15306

12. Ruzza P, Biondi B, Marchiani A, Antolini N, Calderan A (2010) Cell-penetrating peptides a comparative study on lipid affinity and cargo delivery properties. Pharmaceuticals 3:1045–1062

13. Hewitt L, McDonnell JM (2004) Screening and optimizing protein production in *E. coli*. Methods Mol Biol 278:1–16

14. Serber Z, Corsini L, Durst F, Dotsch V (2005) In-cell NMR spectroscopy. Methods Enzymol 394:17–41

15. Zhao L, Sukstanskii AL, Kroenke CD, Song J, Piwnica-Worms D, Ackerman JJ, Neil JJ (2008) Intracellular water specific MR of microbead-adherent cells: HeLa cell intracellular water diffusion. Magn Reson Med 59:79–84

16. Takeuchi T, Kosuge M, Tadokoro A, Sugiura Y, Nishi M, Kawata M, Sakai N, Matile S, Futaki S (2006) Direct and rapid cytosolic delivery using cell-penetrating peptides mediated by pyrenebutyrate. ACS Chem Biol 1:299–303

17. Lee C (2007) Western blotting. Methods Mol Biol 362:391–399

18. Strober W (2001) Trypan blue exclusion test of cell viability. Curr Protoc Immunol Appendix 3, Appendix 3B

Chapter 5

In-Cell NMR in Mammalian Cells: Part 2

Beata Bekei, Honor May Rose, Michaela Herzig, and Philipp Selenko

Abstract

Delivery of isotope-labeled IDPs into mammalian cells for the purpose of generating suitable in-cell NMR samples can also be facilitated by action of pore-forming bacterial toxins. In the course of this procedure, mammalian cell membranes are permeated for short periods of time in order to enable the influx of exogenous proteins via a concentration gradient between the outside and the inside of the targeted "host" cells. In contrast to CPP-mediated IDP uptake, toxins offer the advantage that cellular protein transduction does not rely on active biological processes like endocytosis, but on simple passive diffusion. Therefore, proteins that are to be delivered into mammalian cells are not required to contain additional "targeting" sequences, and can be employed in their native contexts. The protocol outlined here employs isotope-labeled human α-synuclein, adherent human HeLa cells, and the *Streptococcus pyogenes* endotoxin Streptolysin O (SLO).

Key words: Pore-forming toxins, Streptolysin O, α-Synuclein, HeLa cells

1. Introduction

In the previous chapter we outlined a protocol for cellular IDP transduction via the HIV-derived cell penetrating peptide, Tat. Here, we present an alternative method for IDP delivery that exploits the cell permeating properties of the pore-forming toxin Streptolysin O (SLO). Cellular uptake of exogenous protein material by action pore-forming toxins offers the possibility to deliver isotope-labeled purified IDPs into the cytoplasm of live cells without the need to first couple the protein to a carrier molecule (i.e., a cell penetrating peptide).

Streptolysin O is a bacterial endotoxin produced by *Streptococcus pyogenes*. The biological function of this and other toxins is to perforate mammalian cell membranes in the course of the bacterial infection cycle and to thereby destabilize host cell integrity ([1]). Pore-forming toxins can be categorized according to the tertiary
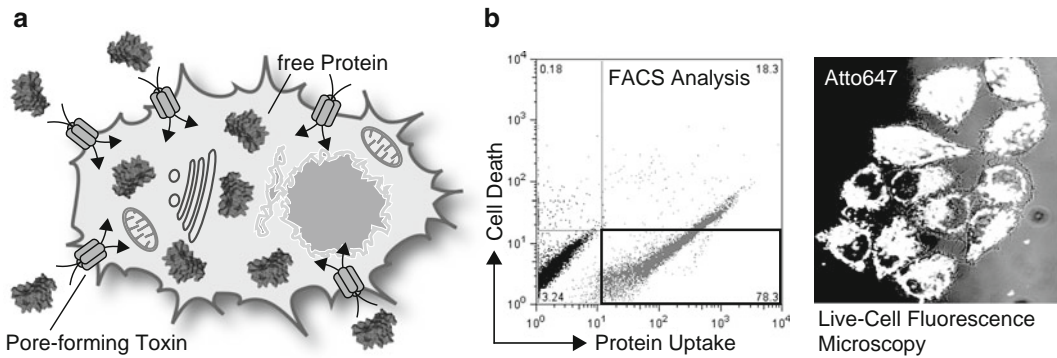
Fig. 1. (**a**) Schematic representation of Streptolysin O (SLO)-mediated protein delivery into mammalian cells. SLO forms homooligomeric, amphipathic pores in mammalian plasma cell membranes, through which proteins can enter into the cytoplasm along a concentration gradient. After protein delivery, cell membranes are resealed with Ca^{2+}-containing buffer solutions. (**b**) Assessment of cellular α-synuclein (AS) uptake and overall cell viability for the SLO-mediated protein delivery protocol by Flow Cytometry (FCM) (*left image*). Scatter plot analysis of untreated HeLa cells (*black*) and of 50 ng/mL SLO-, 200 μM AS- (Atto647-labeled) treated HeLa cells (*gray*). 78 % of the cells contain AS and are viable (*lower right quadrant, boxed*). 18 % of the cells contain AS but are not viable (*upper right quadrant*). 3 % of viable cells (*lower left quadrant*) and 0.2 % of dead cells (*upper left quadrant*) do not contain AS. The overlay of transmission and fluorescence live cell microscopy images of HeLa cells treated with 50 ng/mL SLO and 200 μM AS (Atto647-labeled, *bright white color*) is *shown on the right*.

protein structures into which they fold upon plasma membrane insertion, and based on their different membrane binding characteristics (2, 3). SLO is a member of the cholesterol-dependent cytolysin (CDC), β-pore forming toxin family (4). It is secreted in a monomeric form and assembles into barrel-like structures upon contact with cholesterol-containing membranes (5). The resulting homooligomeric, amphipathic pores extend well into the host membrane and lead to severe cellular leakage (6). In essence, these toxin pores produce an open passage between the intracellular and extracellular compartment, through which molecules may move in a concentration gradient-dependent manner (Fig. 1a) (7). At high concentrations, pore-forming toxins cause irreversible membrane damage and cell death by lysis. When applied at subcytolytic concentrations, cells are typically able to recover from such injuries through a series of homeostatic responses that quickly reseal the perforated membranes (8). Upon mild SLO exposure for example, cellular recovery is accomplished in less than 1 h and via mechanisms that also involve cell-surface patching with membranes from intracellular organelles such as lysosomes, endosomes, and in some instances the Golgi apparatus (9). It is therefore the sizes of the pores formed by SLO (diameters between 25 and 30 nm) and the ability to promote membrane recovery through the re-establishment of cellular calcium homeostasis that enables this toxin to be exploited for the delivery of different biological macromolecules (10–14). The treatment of mammalian cells with even small amounts of a pore-forming toxin like SLO does, however, bear some risks. For the purpose of

generating a suitable mammalian in-cell NMR specimen, a fine balance between efficient protein transduction and cellular toxicity has to be found (Fig. 1b). Additionally, aberrant activation events of intracellular signaling pathways, as well as changes in epigenetic modification states, have to be considered. Both types of unwanted "side-effects" have been reported for SLO treated cells (15–18). Therefore, the SLO delivery method described here may be less suitable for in-cell NMR analyses of biological events that require more unperturbed cellular settings. While this protocol focuses on SLO-mediated protein delivery, a range of other bacterial pore-forming toxins have been employed for controlled cell permeabilization (2, 19, 20). We detail the delivery of α-synuclein into adherent human HeLa cells, but as this method can be adapted to transduce different IDPs to adherent or suspension cells, we have first specified a set of useful experiments (suitable for either cell type) that help to optimize the overall efficiency of protein uptake.

2. Materials

1. Equipment: The following protocol assumes that standard laboratory equipment for recombinant protein production and purification, including a fast protein liquid chromatography (FPLC) system, is available. In addition, cell culture equipment for maintaining and manipulating mammalian cells, including a sterile workbench, a CO_2 incubator, and a basic tissue culture microscope are needed. For optimizing SLO-mediated uptake efficiencies and to determine cell viability and SLO toxicity, a Flow Cytometer (FCM) and a confocal fluorescence microscope are required. Access to high-field (>500 MHz) solution-state NMR spectrometers must be available.

2. The protein to be delivered should be available in an appropriately isotope-labeled form for the envisaged in-cell NMR experiments. Detailed protocols for recombinant protein expression and purification, as well as for stable-isotope labeling for NMR purposes, are described elsewhere (21, 22).

3. Streptolysin O from *S. pyogenes* (Sigma Aldrich, USA).

4. Suitable cell culture medium for HeLa cells: Complete DMEM (low Glucose, 5 mM Glutamine, 10 % Fetal Bovine Serum, (FBS) (PAA Laboratories, Canada)).

5. Phosphate buffered saline (PBS), cell culture grade, without Calcium/Magnesium (PAA Laboratories, Canada).

6. 6-Well cell culture plates/T175 cell culture flasks.

7. 0.25× Trypsin-EDTA (PAA Laboratories, Canada).

8. 20 mM Phosphate buffer pH 7.4.

9. HBSS buffer: 137 mM NaCl, 5.4 mM KCl, 0.25 mM Na$_2$HPO$_4$, 0.44 mM KH$_2$PO$_4$, 4.2 mM NaHCO$_3$, 0.1 % (w/v) D-glucose, 30 mM HEPES, pH 7.2.

10. Membrane resealing buffer: HBSS + 2 mM CaCl$_2$.

11. N-Hydroxysuccinimide fluorescence dye (i.e., Atto647 NHS succinimide ester, Sigma Aldrich, USA).

12. Sephadex G-25 column (GE Healthcare, USA).

13. Trypan Blue 0.4 % (Sigma Aldrich, USA).

14. Lactate dehydrogenase (LDH) Assay Kit (Abcam, USA).

15. *Optional.* Low melting agarose (USB Affymetrix, USA).

16. Redigrad™ solution (GE Healthcare, USA).

17. In-cell NMR buffer: Suspension cell growth medium containing 20 % D$_2$O and 30 % Redigrad™ or serum free DMEM, 10 % D$_2$O.

3. Methods

3.1. IDP and SLO Stock Solutions

1. Before starting the procedure, carefully assess the required number of mammalian cells to prepare the in-cell NMR sample (see Note 1), the target volume of the final in-cell NMR sample (see Note 2) and the target concentration of the isotope-labeled protein inside the cells (see Note 3).

2. Produce isotope-labeled IDP stock solution in 20 mM Phosphate, 150 mM NaCl buffer (pH 7.4) at a concentration of ~1 mM in 1 mL (see Note 4).

3. *Optional* For optimization trials, the IDP to be delivered into cells may be labeled with a fluorescence dye for direct intracellular detection by fluorescence microscopy or Flow Cytometry (see Note 5). The protein of interest should be purified in the appropriate buffer for the dye-coupling reaction. Below, we describe labeling of human α-synuclein with a lysine-conjugated dye. Whenever fluorescence labeling is not desired proceed to step 9.

4. Prepare 100–200 μL of 2 mg/mL of the desired dye in DMSO (see Note 6).

5. Adjust pH with fresh NaHCO$_3$ (1 M) according to the manufacturers' instructions for optimal coupling efficiency (for most dyes this is pH 7.8).

6. Mix 2–5 mg of the protein in 500 μL of 20 mM Phosphate buffer, pH 7.8 with a two-fold molar excess of the dye.

7. Incubate at room temperature for 1–2 h on a shaker plate.

8. Separate the labeled product from nonreacted dye using a Sephadex G-25 column and measure the protein concentration of the fluorescent product by UV/VIS spectrophotometry (see Note 7).

9. Dissolve lyophilized SLO powder in a suitable protein buffer (0.2–1 mg/mL) (see Note 8). This stock is further diluted to 20–50 ng/μL in HBSS buffer. 50 μL aliquots are frozen and stored at −20 °C (avoid repeated freeze–thaw cycles). These aliquots constitute 1,000× stocks for the recommended concentrations to be directly added to mammalian cells (i.e., 20–50 ng/mL) (see Note 9).

3.2. Protein Delivery into Adherent Cells

The goal of this section is to determine the concentrations of SLO, and of SLO–protein mixtures that can safely be applied to mammalian cells for efficient protein delivery and without adverse effects on cell viability (see Note 10). In a first step, the maximum tolerable SLO concentration is determined.

1. Seed 2.5–3 × 10^5 cells/well in a 6-well plate to test multiple SLO concentrations in parallel.

2. Incubate the cells for 24 h in complete DMEM at 37 °C in a CO_2 incubator. The cells should be ~80 % confluent at the beginning of the experiment (~5–6 × 10^5 cells).

3. Wash the cells twice with 3 mL/well of prewarmed PBS to remove FBS (which contains Ca^{2+} and inhibits SLO pore formation).

4. Wash the cells twice with 3 mL/well of ice-cold HBSS (see Note 11).

5. Add SLO concentrations ranging from ~20 to 50 ng/mL in 500 μL/well of HBSS to the cells and incubate for 30 min at 4 °C (see Note 12).

6. Remove non-incorporated, excess SLO by washing the cells twice with 3 mL/well of ice-cold HBSS.

7. Add 500 μL/well of fresh HBSS and incubate the cells for 30 min at 37 °C in a CO_2 incubator. Once, the optimal SLO concentration has been determined (step 5) the protein of interest to be delivered into the cells is added to the HBSS solution at this point (see Note 13).

8. Check for successful SLO pore-formation. This is best done with a lactate dehydrogenase (LDH) assay, which, similar to the Trypan Blue staining procedure described in the previous chapter, is a simple test for cell membrane integrity. Specifically, this assay is used to "quantify" successful membrane perforation by measuring the amount of intracellular LDH that is released into the culture medium following SLO treatment (see Note 14).

This step can be omitted once a suitable SLO concentration has been found.

9. Wash the cells once with 3 mL/well of ice-cold HBSS.

10. Add ice-cold membrane resealing buffer (HBSS/Ca^{2+}, 500 μL/well), incubate for 30 min at 4 °C (see Note 15).

11. Remove HBSS/Ca^{2+}, wash the cells twice with 3 mL/well of prewarmed PBS.

12. At this point, check for overall cell viability. This can be done by a second LDH assay (see above), Trypan Blue staining, or by alternative means (see Note 16). Cell viability may also be tested after the following recovery step (see Note 17). Once the above routines are executed with the target IDP in place (added in step 7) it is suggested to additionally verify cellular protein uptake by fluorescence microscopy, Flow Cytometry, or other appropriate methods (see Note 18).

13. Add 2 mL/well of complete DMEM and incubate at 37 °C for 1 h in a CO$_2$ incubator.

14. Choose optimal SLO concentrations and repeat steps 7–13 including the protein to be delivered. Evaluate protein uptake efficiency and changes in cell viability parameters (see Note 19).

3.3. Protein Delivery into Suspension Cells

As the only other published SLO/in-cell NMR application describes the manipulation of non-adherent mammalian cells (human embryonal 293F kidney cells) (14), we also outline a tailored protocol for SLO-mediated protein delivery into suspension cells. Many of the required routines are similar to the ones outlined for adherent cells (see above).

1. To determine tolerable SLO concentrations, collect ~1.0 × 10^6 cells for every SLO concentration to be tested and sediment cells by centrifugation (~400 × g).

2. Discard supernatant and transfer cell slurry to 2 mL Eppendorf tubes.

3. Perform optimization steps as outlined in Subheading 3.2, reducing the indicated volumes to 100–200 μL for each incubation and wash step (see Note 20).

3.4. In-Cell NMR Sample Preparation Using Adherent Cells

1. For adherent (i.e., HeLa) cells, seed 4–5 × 10^6 cells in a T175 cell culture flask.

2. Incubate in complete DMEM for 24 h at 37 °C in a CO$_2$ incubator, until the cells are 80 % confluent (0.8–1 × 10^7 cells/T175 flask).

3. Wash the cells twice with 15 mL prewarmed PBS to remove FBS (see Note 21).

4. Wash the cells twice with 15 mL ice-cold HBSS.

5. Add determined concentration of SLO in 4 mL HBSS and incubate the cells for 30 min at 4 °C.

6. Remove non-incorporated, excess SLO by washing the cells twice with 15 mL ice-cold HBSS.

7. Add 4 mL of fresh HBSS containing the amount of IDP that has been shown to yield the best protein uptake results (300–600 μM on average (see Note 22)). (It should be noted that a fluorescent tag is not necessarily desired on the IDP used for the final in-cell NMR sample preparation).

8. Incubate the cells for 30 min at 37 °C in a CO_2 incubator.

9. Wash the cells once with 15 mL ice-cold HBSS.

10. Add ice-cold membrane resealing buffer (HBSS/Ca^{2+}, 4 mL) and incubate for 30 min at 4 °C.

11. Remove HBSS/Ca^{2+}, wash the cells twice with 15 mL prewarmed PBS.

12. *Optional* Add 30 mL fresh complete DMEM and allow the cells to recover for 1 h at 37 °C in a CO_2 incubator.

13. In order to transfer the manipulated adherent cells to the NMR sample tube, incubate the cells with 4 mL (minimal surface volume) 0.25 % Trypsin/EDTA in prewarmed PBS for 3 min. If a recovery step was performed, wash once with 15 mL prewarmed PBS to remove FBS.

14. Add 20 mL prewarmed complete DMEM to inactivate Trypsin and transfer cell suspension to a 50 mL centrifugation tube.

15. If testing for cell viability is desired, remove a 20 μL aliquot and proceed with Trypan Blue staining protocol (for details see Chapter 6).

16. Sediment cell suspension by centrifugation (~400 × g). Wash cell slurry twice in the final in-cell NMR buffer.

17. Resuspend in 500 μL (or 300 μL, depending on the size of the NMR tube) in-cell NMR buffer.

18. Transfer the cells to the NMR tube and proceed to perform the in-cell NMR experiment.

3.5. In-Cell NMR Sample Preparation Using Suspension Cells

1. Prepare 2 × 10^7 suspension cells in suitable suspension culture flasks.

2. Collect the cells in 50 mL centrifugation tube and sediment at ~400 × g.

3. Wash cell slurry twice with 2 mL prewarmed PBS to remove FBS (which contains Ca^{2+} and inhibits SLO pore formation).

4. Wash the cells once with 2 mL ice-cold HBSS.

5. Discard supernatant and resuspend cell slurry in 1–2 mL HBBS containing the optimized concentration of SLO (~20–40 ng/mL) (see Note 23).

6. Incubate the cells according to optimized conditions on a roller platform to avoid cell settling.

7. Remove nonincorporated, excess SLO by two centrifugation/wash steps with 2 mL ice-cold HBSS.

8. Add 2 mL HBSS containing the amount of IDP that has been shown to yield the best protein uptake results (300–600 μM on average) (see Note 24). (It should be noted that a fluorescent tag is not necessarily desired on the IDP used for the final in-cell NMR sample preparation).

9. Incubate the cells for 30 min at 37 °C in a CO_2 incubator on a roller platform.

10. Wash the cells twice with 2 mL ice-cold HBSS.

11. Add ice-cold membrane resealing buffer (HBSS/Ca^{2+}, 2 mL), incubate for 30 min at 4 °C on a roller platform.

12. Remove HBSS/Ca^{2+}, wash the cells twice with 2 mL pre-warmed PBS.

13. Resuspend cell suspension in appropriate cell culture medium and transfer to suitable cell culture flasks. Allow the cells to recover for 1 h at 37 °C in a CO_2 incubator.

14. Transfer cell suspension to a 50 mL centrifugation tube. Remove a 20 μL aliquot for Trypan Blue staining.

15. Sediment remaining suspension by centrifugation (~400 × g).

16. Resuspend in 500 μL (or 300 μL, depending on the size of the NMR tube) in-cell NMR buffer (see Note 25).

17. Transfer the cells to the NMR tube and proceed to in-cell NMR experiments.

4. Notes

1. On average, $1–3 \times 10^7$ cells are needed for a single in-cell NMR sample. This number depends on the kind of cells that are used and on their individual cell volumes, i.e., 2 pL for human HeLa cells (23).

2. The final volume of the in-cell NMR sample is determined by the kind of NMR tube that is to be used. On narrow-bore NMR probes, standard 5 mm (~500 μL), or Shigemi™ NMR tubes (~300 μL) may be employed. Their difference in volume will require different numbers of cells.

3. The effective NMR concentration (C_{NMR}) of the final in-cell NMR sample will be determined by the intracellular concentration of the "delivered" protein (C_{Cell}), the number of cells in the NMR sample (N_{Cell}), their individual cell volume (V_{Cell}), the final NMR sample volume (V_{NMR}), and the corresponding volume dilution factor, VDF ($V_{NMR}/V_{Cell} \times N_{Cell}$). According to $C_{NMR} \times V_{NMR} = C_{Cell} \times (V_{Cell} \times N_{Cell})$ the effective NMR concentration can be calculated as $C_{NMR} = C_{Cell}/$ VDF. For example, in order to obtain a spectrum with an effective NMR concentration (C_{NMR}) of 10 μM of isotope-labeled protein in 300 μL of NMR sample volume (V_{NMR}), for 1×10^7 cells (N_{Cell}) with an average cell volume (V_{Cell}) of 2 pL (i.e., 2×10^{-6} μL) a total intracellular protein concentration (C_{Cell}) of 150 μM of isotope-labeled IDP must be reached (VDF in this case is 15). It is therefore suggested to first determine the minimum protein concentration that is required for suitable in vitro NMR results (i.e., the lower limit of C_{NMR}). This will define the benchmark intracellular protein concentration (C_{Cell}) that has to be reached in a defined number of cells (N_{Cell}) of volume (V_{Cell}), for satisfactory in-cell NMR results.

4. Initial optimization trials may also be performed with non-isotope-labeled IDPs.

5. For direct live cell imaging or Flow Cytometry, fluorescence dyes must be coupled to the IDP. We routinely use Atto647 or Atto488, which are conjugated to recombinant proteins via lysine or cysteine residues. Dyes should be chosen based on desired detection wavelengths and available functional groups for coupling. For details see Chapter 6.

6. Many fluorescence dyes are light sensitive and must be stored protected from light.

7. A correction factor for the specific absorbance of the fluorophore must be applied when the concentration of the conjugated product is determined by UV/VIS spectrophotometry.

8. SLO is usually obtained as a lyophilized powder, prepared from 10 mM Tris, 3 mM Na-azide, 5 mM, EDTA, 1 mM PMSF solutions. In such instances, the SLO stock is prepared by directly resuspending the SLO powder in distilled H_2O, or in distilled H_2O containing 2–5 mM DTT. DTT has been reported to further "activate" the toxin by reducing its single cysteine (10, 24, 25). However, addition of reducing agents does not seem to be strictly required for SLO activity (11, 12, 14).

9. Many vendors refer to SLO's "activity" in terms of enzymatic units (U). Ogino et al. employ 40–80 U/mL SLO in 300 μL to perforate 293F cells (14). Concentrations employed in this protocol correspond to 30–70 U/mL SLO for adherent HeLa cells.

10. As different cell lines possess different tolerance levels for bacterial toxins, suitable SLO concentrations need to be determined first. The aim is to perforate 70–90 % of the cultured cells while maintaining maximum cell viability.

11. Ice-cold HBSS is used to adjust cells to the following SLO incubation at 4 °C (see step 5 and below).

12. Incubation at 4 °C promotes accumulation of SLO on cell membranes and subsequently pore-formation (26). Alternatively, cells can also be treated with SLO at 37 °C, as less robust cells might not tolerate low temperature incubation (12, 24, 25). Choosing this option may improve overall cell viability, but result in less efficient pore formation.

13. During this incubation step, "functional" SLO pores form in the plasma membrane and cellular protein uptake eventually occurs.

14. Several LDH-based toxicity kits are commercially available. They comprise a colorimetric assay for dead or plasma membrane-leaked cells. LDH present in the culture supernatant (due to SLO pore formation) participates in a coupled reaction that converts a yellow tetrazolium salt into a red formazan-class dye, which is measured by absorbance at 492 nm. The amount of formazan is directly proportional to the amount of LDH in the culture, which is in turn directly proportional to the number of SLO pores (and/or dead cells).

15. Alternatively, complete DMEM may be added at this point. FBS contains sufficient amounts of Ca^{2+} to ensure membrane resealing. For less robust mammalian cells, this option is suggested (10, 12, 25).

16. A range of other (commercial) cell viability assays may be employed at this point (see also Chapter 6).

17. For many cell lines, overall viability and resealing efficiency may increase with the following recovery step in complete DMEM. For less robust mammalian cells in particular, this option is suggested.

18. A detailed description of quality control methods to determine cellular protein uptake is provided in Chapter 6.

19. The intracellular accumulation of the delivered IDP is likely to affect overall cell viability (in a concentration-dependent manner). Therefore, it is advisable to carefully adjust SLO- and target protein concentrations for most satisfactory results. In many instances, a compromise between highest tolerable SLO- and highest-possible IDP concentrations will have to be found for the preparation of suitable in-cell NMR samples.

20. The smaller incubation volumes afforded by suspension cells and the concomitant requirements of reduced amounts of IDPs and SLO offer some advantages over adherent cell protocols.

21. Protein delivery into adherent cells can also be conducted under suspension conditions. While this bears several advantages (see previous note), it may not be suitable for all cell types. In addition, problems arise regarding the separation of nonviable from viable cells. When working with adherent cells, this is conveniently achieved in the course of the repeated washing steps. To manipulate adherent cells under suspension conditions, incubate cells from step 3 with 4 mL 0.25 % Trypsin/EDTA in prewarmed PBS for 3 min. Inactivate Trypsin with 20 mL complete DMEM. Continue with IDP delivery protocol **steps 2–16** of Subheading 3.5, which is customized for suspension cells.

22. Assuming an average protein uptake efficiency of 2–5 % (14), incubation with 300–600 µM IDP yields intracellular concentrations in the range of 6–30 µM.

23. This concentration resulted in a 50 % pore forming-, and 70–80 % membrane resealing efficiency in 293F suspension cells (14).

24. Ogino et al., incubated cells with 1 mM thymosin β4 to achieve a reported intracellular protein concentration (C_{Cell}) of 50 µM. Approximately a 5 % protein uptake efficiency (14).

25. Ogino et al. resuspended human embryonal 293F kidney cells in cell culture medium containing 20 % D_2O and 30 % RediGrad. RediGrad is a nontoxic medium, composed of colloidal silica particles, which is covalently coated with silan. This commercially available product is used for density gradient centrifugation. It is also suitable to reduce cell sedimentation in the NMR tube during the NMR experiment (14). An alternative method to avoid cell sedimentation is to use a 0.2 % low melting agarose NMR buffer solution, as described in Chapter 4.

References

1. Bischofberger M, Gonzalez MR, van der Goot FG (2009) Membrane injury by pore-forming proteins. Curr Opin Cell Biol 21:589–595

2. Parker MW, Feil SC (2005) Pore-forming protein toxins: from structure to function. Prog Biophys Mol Biol 88:91–142

3. Gonzalez MR, Bischofberger M, Pernot L, van der Goot FG, Freche B (2008) Bacterial pore-forming toxins: the (w)hole story? Cell Mol Life Sci 65:493–507

4. Iacovache I, Bischofberger M, van der Goot FG (2010) Structure and assembly of pore-forming proteins. Curr Opin Struct Biol 20:241–246

5. Palmer M (2001) The family of thiol-activated, cholesterol-binding cytolysins. Toxicon 39:1681–1689

6. Tweten RK (2005) Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. Infect Immun 73:6199–6209

7. Shatursky O, Heuck AP, Shepard LA, Rossjohn J, Parker MW, Johnson AE, Tweten RK (1999) The mechanism of membrane insertion for a cholesterol-dependent cytolysin: a novel paradigm for pore-forming toxins. Cell 99:293–299

8. Idone V, Tam C, Goss JW, Toomre D, Pypaert M, Andrews NW (2008) Repair of injured

plasma membrane by rapid Ca2+-dependent endocytosis. J Cell Biol 180:905–914

9. McNeil PL, Kirchhausen T (2005) An emergency response team for membrane repair. Nat Rev Mol Cell Biol 6:499–505

10. Fawcett JM, Harrison SM, Orchard CH (1998) A method for reversible permeabilization of isolated rat ventricular myocytes. Exp Physiol 83:293–303

11. Walev I, Bhakdi SC, Hofmann F, Djonder N, Valeva A, Aktories K, Bhakdi S (2001) Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O. Proc Natl Acad Sci USA 98:3185–3190

12. Fu H, Ding J, Flutter B, Gao B (2008) Investigation of endogenous antigen processing by delivery of an intact protein into cells. J Immunol Methods 335:90–97

13. Brito JL, Davies FE, Gonzalez D, Morgan GJ (2008) Streptolysin-O reversible permeabilisation is an effective method to transfect siRNAs into myeloma cells. J Immunol Methods 333:147–155

14. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I (2009) Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J Am Chem Soc 131:10834–10835

15. Hamon MA, Batsche E, Regnault B, Tham TN, Seveau S, Muchardt C, Cossart P (2007) Histone modifications induced by a family of bacterial toxins. Proc Natl Acad Sci USA 104:13467–13472

16. Aroian R, van der Goot FG (2007) Poreforming toxins and cellular non-immune defenses (CNIDs). Curr Opin Microbiol 10:57–61

17. Huffman DL, Abrami L, Sasik R, Corbeil J, van der Goot FG, Aroian RV (2004) Mitogenactivated protein kinase pathways defend against bacterial pore-forming toxins. Proc Natl Acad Sci USA 101:10995–11000

18. Stassen M, Muller C, Richter C, Neudorfl C, Hultner L, Bhakdi S, Walev I, Schmitt E (2003) The streptococcal exotoxin streptolysin O activates mast cells to produce tumor necrosis factor alpha by p38 mitogen-activated protein kinase- and protein kinase C-dependent pathways. Infect Immun 71:6171–6177

19. Bhakdi S, Weller U, Walev I, Martin E, Jonas D, Palmer M (1993) A guide to the use of pore-forming toxins for controlled permeabilization of cell membranes. Med Microbiol Immunol 182:167–175

20. Miyamoto K, Yamashita T, Tsukiyama T, Kitamura N, Minami N, Yamada M, Imai H (2008) Reversible membrane permeabilization of mammalian cells treated with digitonin and its use for inducing nuclear reprogramming by *Xenopus* egg extracts. Cloning Stem Cells 10:535–542

21. Hewitt L, McDonnell JM (2004) Screening and optimizing protein production in *E. coli*. Methods Mol Biol 278:1–16

22. Serber Z, Corsini L, Durst F, Dotsch V (2005) In-cell NMR spectroscopy. Methods Enzymol 394:17–41

23. Zhao L, Sukstanskii AL, Kroenke CD, Song J, Piwnica-Worms D, Ackerman JJ, Neil JJ (2008) Intracellular water specific MR of microbeadadherent cells: HeLa cell intracellular water diffusion. Magn Reson Med 59:79–84

24. Jun Y, Kim E, Jin M, Sung HC, Han H, Geraghty DE, Ahn K (2000) Human cytomegalovirus gene products US3 and US6 downregulate trophoblast class I MHC molecules. J Immunol 164:805–811

25. Brito JL, Brown N, Morgan GJ (2010) Transfection of siRNAs in multiple myeloma cell lines. Methods Mol Biol 623:299–309

26. Palmer M, Valeva A, Kehoe M, Bhakdi S (1995) Kinetics of streptolysin O self-assembly. Eur J Biochem 231:388–395

Chapter 6

In-Cell NMR in Mammalian Cells: Part 3

Beata Bekei, Honor May Rose, Michaela Herzig, Heike Stephanowitz, Eberhard Krause, and Philipp Selenko

Abstract

Irrespective of how isotope-labeled proteins are delivered into mammalian cells, laboratory routines are needed to assess the quality of the resulting in-cell NMR samples. These include methods to evaluate overall cell viability, protein transduction efficiency, intracellular protein concentration, localization, and stability. In addition, quality control experiments to assess protein leakage from manipulated cells are of particular importance for in-cell NMR experiments. The purpose of this chapter is to outline qualitative and quantitative methods to determine general biological properties of in-cell NMR samples in order to ensure the highest possible standards for in-cell NMR studies.

Key words: Trypan blue, Flow cytometry (FCM), Fluorescence live cell imaging, Immunofluorescence microscopy, Western blotting, Quantitative mass spectrometry, Quality control

1. Introduction

In the previous chapters we have outlined a collection of methods to deliver isotope-labeled IDPs into amphibian-, and cultured mammalian cells. While different protocols can be employed to prepare in-cell NMR samples, generic quality control experiments to verify the overall fitness of the obtained in-cell NMR specimens are of general importance. Questions concerning cell viability, protein delivery efficiency, as well as intracellular protein localization and stability need to be addressed in order to critically assess the biological relevance of in-cell NMR results. In the following, we will outline a compendium of stringent quality control experiments that should additionally be performed on every in-cell NMR sample. Most of the suggested experiments follow standard laboratory routines in cell biology. Therefore, we will only focus on those aspects of the protocols that are of particular importance for in-cell NMR experiments.

First and foremost, methods to determine overall cell viability at any stage of the delivery routines-, and without having to sacrifice the sample under investigation-, are especially useful. Here, conventional phase-contrast light microscopy offers many advantages. Cytopathological changes in cell morphology, such as swelling, blebbing, or fraying of cell membranes, are readily observed and provide clear indications of imminent cell death (1). Thus, counting the fraction of morphologically intact, or compromised cells under a light microscope provides a simple means to quickly assess overall cell viability. More than 80 % of cells should score as "viable" at any point of the individual manipulation schemes, as well as before and after the in-cell NMR experiment. More specialized assays such as the Trypan blue cell staining protocol (see below) and lactate dehydrogenase (LDH)-based cytotoxicity tests should be employed at specified stages of the delivery protocols (see Chapters 4 and 5). Flow Cytometry (FCM) offers the additional advantage to test cell viability and protein uptake in parallel. Because both mammalian cell delivery protocols outlined in this volume entailed fluorescence-labeling schemes for either the CPP moiety-, or the delivered target protein-, FCM control experiments are suitable means to assess these two important parameters. In addition, the availability of fluorescence labels enables straightforward intracellular protein localization studies by live cell fluorescence microscopy imaging, which is also described in this chapter. This technique can additionally provide invaluable insights into aberrant cellular distributions of exogenously delivered proteins, which are essential for properly interpreting in-cell NMR results. While fluorescence microscopy can verify cellular protein uptake and assess cellular protein localization, it is not suited to determine intracellular protein concentrations. However, quantitative knowledge about cellular levels of delivered proteins is indispensable for evaluating protein uptake efficiencies, for optimizing individual steps in protein delivery protocols, and for ultimately determining in-cell NMR sample concentrations. Therefore, reliable methods for quantitative and semi-quantitative measurements of "delivered" protein concentrations are essential. In this chapter, we describe how specifically tailored Western-blotting routines, quantitative mass spectrometry (MS) and in-cell NMR measurements themselves can be employed to assess of the quality of mammalian in-cell NMR samples.

2. Materials

1. Equipment: Standard gel electrophoresis and Western blotting equipment, Phase contrast microscope, hemocytometer, sterile collagen-coated round 22 mm glass coverslips, 6-well cell culture plate, fluoresence confocal microscope, microscope

chamber for live cell imaging, FCM equipment, sonicator, tandem electro-spray mass spectrometer (MS), NMR spectrometer.

2. Appropriate cell culture media: For HeLa cells, complete DMEM (low Glucose, 5 mM Glutamine, 10 % Fetal Bovine Serum, (FBS), PAA Laboratories, Canada).

3. Phosphate buffered saline (PBS), cell culture grade, without Calcium/Magnesium (PAA Laboratories, Canada).

4. 0.25% Trypsin-EDTA solution in PBS (PAA Laboratories, Canada).

5. Standard SDS-PAGE solutions and buffers, Western blot incubation buffers and loading controls.

6. Denaturing lysis buffer, i.e., RIPA: Tris 50 mM, NaCl 150 mM, SDS 0.1 %, Na-Deoxycholate 0.5 %, Triton X 100 or NP40 1 %, (Protease inhibitors should be freshly added.).

7. Native phosphate lysis buffer: 20 mM Potassium phosphate, 150 mM NaCl, pH 7.5.

8. Propidium Iodide (PI) staining solution: 10 mg/mL in PBS (Sigma Aldrich, USA).

9. 7-AAD staining solution: 1 mg/mL in PBS (Sigma Aldrich, USA).

10. 0.4 % Trypan blue staining solution (Sigma Aldrich, USA).

11. Primary antibody against IDP to be delivered.

12. Appropriate secondary HRP-conjugated antibody for enhanced chemoluminescence (ECL) detection.

13. Appropriate secondary antibody, fluorescence dye conjugated.

3. Methods

3.1. Trypan Blue Staining

This section outlines a mammalian cell viability assay that can be employed in combination with cell morphology investigations using a conventional bright field light microscope. The advantage of this method is that it can easily be performed at various stages of the in-cell NMR sample preparation routines, whenever manipulated cells are in suspension. Each Trypan blue assay requires $\sim 1 \times 10^4$ cells.

1. To suspend adherent cells: Wash cells once with pre-warmed PBS and incubate with minimum surface volume of 0.25 % Trypsin/EDTA in prewarmed PBS until all cells detach from the culture dish/flask (typically 2–4 min). Add 5× volume of prewarmed complete DMEM and collect cell suspension.

2. Remove a 20 μL aliquot, transfer to a 1.5 mL centrifugation tube and sediment cells at ~400 × g for 5 min.

3. Discard supernatant and resuspend cell pellet in 20 μL PBS.

4. Add equal volume (20 μL) 0.4 % Trypan blue staining solution and mix thoroughly.

5. Place a cover slip onto a hemocytometer and dispense 10–20 μL of the cell suspension into each counting chamber.

6. Count the number of viable (non-blue) and non viable (blue) cells, within 5 min of staining (over time even viable cells will begin to take up Trypan blue). The cell dilution should be adjusted to minimum cell number of ~200 on the hemocytometer grid.

7. Calculate the proportion of viable cells, adjusting for the 2× dilution factor of the Trypan blue solution in your sample.

3.2. Flow Cytometry   The Trypan blue staining procedure outlined above provides information about overall cell viability only. For in-cell NMR applications it is especially useful to correlate intracellular protein uptake to changes in cell viability. Certain protein delivery schemes may retain large numbers of healthy cells at the expense of intracellular protein uptake. In turn, high levels of protein delivery may be contrasted with deleterious effects on cell viability. In that sense, mammalian in-cell NMR sample preparation schemes need to strike a balance between levels of exogenous protein uptake-, and correspondingly intracellular protein concentration-, and overall cell viability. One way to simultaneously assess cell viability and protein uptake is flow cytometry (FCM). For FCM applications, the protein to be delivered into mammalian cells has to be labeled with a suitable fluorescent dye. Dual-channel cell sorting is achieved by the different fluorescence properties of the protein-, and cell viability dye.

1. Choose protein cargo labeling dye according to the type of FCM experiment that is to be performed (see Note 1).

2. Couple fluorescence dye to the protein to be delivered into mammalian cells (see Note 2).

3. Purify dye-coupled protein (i.e., remove nonincorporated dye) according to manufacturer's instructions and execute the cellular protein delivery protocol of choice.

4. Perform FCM measurements at desired stages of the delivery schemes and according to standard FCM protocols (2).

5. Interpret FCM results with respect to whether adherent or suspension cells were assayed (see Note 3).

6. Determine whether cells that score positive for exogenous protein uptake contain internalized protein, or whether unspecific membrane binding skews the FCM readout (see Note 4).

7. Check for changes in FCM scores upon delivery of different concentrations of exogenous protein (see Note 5)

8. Optimize individual steps of the chosen delivery protocol accordingly.

3.3. Live-Cell Fluorescence Imaging

The fluorescently labeled versions of target proteins produced for FCM measurements can also be detected inside live cells by microscopy techniques. These methods afford the possibility to qualitatively assess cellular protein uptake, as well as to determine intracellular protein localization. For the outlined protein delivery schemes, both properties are of fundamental importance. Many of the CPP-mediated protein delivery processes for example, involve endocytotic uptake routes (3). Therefore, CPP-cargo proteins are often trapped inside endosomal vesicles and inefficiently "released" into the cytoplasm (4). Similarly, toxin-mediated protein transduction procedures often result in lysosomal protein deposition, as a result of cell toxicity. Both scenarios are detrimental for in-cell NMR analyses and to know about their occurrence is important for optimizing cell delivery protocols and for assessing overall in-cell NMR sample quality.

1. Perform live cell fluorescence microscopy imaging experiments according to standard protocols (5).

2. Correlate experimentally determined intracellular localization properties to known cellular distributions of the delivered protein (see Note 6)

3. Inspect microscopy images for unusual localization properties, in particular punctuated, or speckled intracellular distributions that might indicate organelle co-localization, or membrane attachment (Fig. 1) (see Note 7).

4. Choose alternative fluorescence dyes when in doubt about the possible effects of dye properties on intracellular protein localization behaviors (see Note 8).

5. Consider co-staining with known organelle markers such as 6-diamidino-2-phenylindole (DAPI), Hoechst 43332, or Mito-Tracker™, LysoTracker™ to determine the possible biological nature of non-physiological localization behaviors (see Note 9).

6. Perform time-course experiments to determine cellular protein stability and turnover rates, as well as time-dependent changes in intracellular localization (see Note 10).

7. Interpret changes in cellular localization properties with regard to expected durations of in-cell NMR experiments (see Note 11).
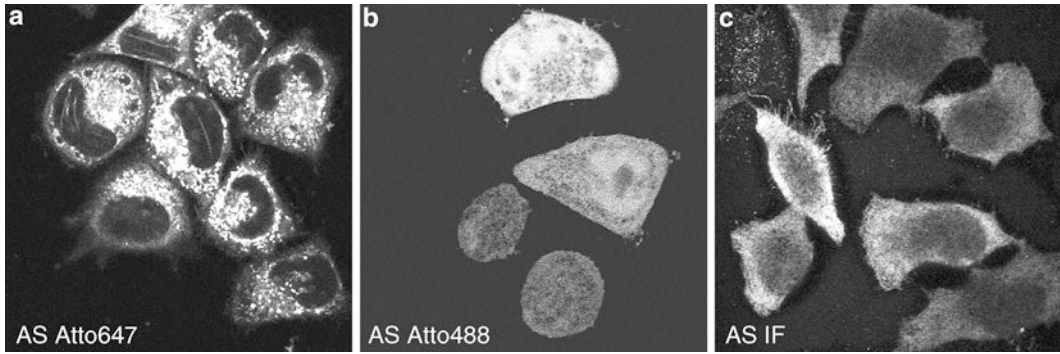
Fig. 1. (**a**) Live-cell fluorescence microscopy image of SLO transduced human α-synuclein (AS) in cultured HeLa cells. The Atto647 fluorescence dye was conjugated to recombinant AS via a lysine side-chain coupling reaction. 200 μM of exogenous AS was employed in the protein delivery protocol. A strongly punctuated, cytoplasmic staining is observed. (**b**) Live-cell fluorescence microscopy image of SLO transduced Atto488 conjugated human AS in cultured HeLa cells. The same lysine side-chain coupling procedure and protein concentration in the "delivery" solution was employed. This time, a uniform cytoplasmic distribution, as well as nuclear staining is detected. (**c**) Fixed-cell immunofluorescence (IF) microscopy analysis of SLO transduced, unmodified AS (200 μM) in HeLa cells. Primary antibodies against human AS and fluorescence labeled secondary antibodies were employed. IF-microscopy indicates a uniform cytoplasmic distribution and no nuclear localization of the protein. All microscopy images were recorded with same apertures and gain settings.

8. Determine intracellular protein localization and overall changes in cell viability upon different sample embedding procedures (see Note 12).

3.4. Fixed-Cell Fluorescence Imaging

Whenever fluorescence labeled proteins are not available, or when conjugating proteins to fluorescence dyes results in non-physiological intracellular protein distributions (see above), cellular protein uptake and localization can also be verified by fixed-cell, immunofluorescence (IF) microscopy using antibodies against the delivered protein. This procedure requires cells to be "fixed," which denotes a process that involves immobilizing the antigen (i.e., the protein of interest) inside the cell and to permeabilize the plasma membrane so that the antibody can enter the cytoplasm and bind to the antigen. Therefore, fixed-cell microscopy employs specimens that are no longer viable. In addition, it has been demonstrated that different fixation protocols can lead to drastic differences in observed intracellular localization properties of CPP-delivered proteins, for example (6, 7).

1. Carefully choose a suitable fixation protocol for the envisaged IF microscopy experiment (see Note 13).

2. Perform fixed-cell IF experiment according to standard protocols (5).

3. Correlate experimentally determined intracellular localization properties to known cellular distribution characteristics of

the delivered protein and to the live-cell imaging results (see Note 14).

4. Determine the effect of different fixation protocols on intracellular protein localization (see Note 15).

5. Check for changes in intracellular protein localization at different intracellular concentrations of delivered protein and during time-course experiments (see Note 16).

3.5. Semi-quantitative Western Blotting

Cellular protein detection by Western blotting provides a simple routine to qualitatively and semi-quantitatively assess intracellular protein concentrations, as well as address potential protein leakage problems (8, 9). Antibodies against targeted proteins need to be available in order to perform such experiments. The main purpose of this section, is to provide a general protocol to accurately determine the amount of successfully delivered protein in mammalian cells. In the following section, we demonstrate how Western-blot signals, obtained with lysates of protein-transduced mammalian cells, can be employed to semi-quantitatively determine intracellular concentrations of successfully delivered proteins.

1. Choose an appropriate lysis buffer for cell extract preparation (see Note 17).

2. Prepare ~2×10^6 protein-transduced cells. If adherent cells are to be analyzed, this number corresponds to two 9.6 cm^2 cell culture dishes, 80 % cell density, cell volume 1–4 pL.

3. Detach cells by Trypsin treatment, sediment by centrifugation (~$400 \times g$) and wash cell pellet twice with PBS.

4. Resuspend cells in 500 μL PBS (1×10^6 cells/mL), and transfer 20 μL to a hemocytometer to count cell numbers (see Note 18).

5. Collect 1×10^6 cells accordingly, sediment, and discard supernatant.

6. Resuspend cell pellet in 100 μL lysis buffer (see Note 19).

7. Lyse cells by brief sonication or repeated freeze–thaw cycles (see Note 20).

8. Centrifuge extract at $16,000 \times g$ for 10–20 min at 4 °C and collect supernatant (i.e., the final cell lysate).

9. Determine the total protein concentration by a Bradford assay, or other appropriate techniques (see Note 21).

10. Load ~4 μL (per slot) of cell lysate (with appropriate sample buffer) onto a SDS-PAGE (~20 μg total protein/lane). Add a reference concentration series of known amounts of input protein (10–300 ng) on the same gel.

11. Run gel electrophoresis, transfer proteins onto an appropriate membrane (i.e., Nitrocellulose, or PVDF). Perform Western-

blotting routines by standard protocols (10) using primary antibodies against the protein of interest and relevant HRP-conjugated secondary antibodies.

12. Quantify Western signal intensities by densitometry, or by commercial software packages when using digital chemoluminescence detection setups.

13. The intracellular protein concentration is determined based on the calibration curve given by the input reference samples. For a detailed outline of the individual calculation steps, refer to the indicated note (see Note 22).

14. Optimize delivery protocols accordingly.

3.6. Quantitative Mass Spectrometry

Given that ^{13}C and ^{15}N labeled proteins possess similar physical and chemical properties as their unlabeled counterparts, stable isotope labeling in combination with mass spectrometry (MS) has been widely used for relative quantifications of proteins (11). Absolute quantification methodologies based on isotope dilution strategies allow the determination of protein concentrations in biological samples (12, 13). Intracellular concentrations of proteins can also be accurately determined by MS using labeled peptides as internal standards (14, 15). In fact, MS turns out to be especially useful for assessing intracellular protein concentrations of in-cell NMR samples, because levels of isotope-labeled proteins can be correlated to known quantities of non-isotope-labeled internal standard proteins. To this end, cell lysates prepared from in-cell NMR samples are "spiked" with defined amounts of unlabeled protein and jointly analyzed by MS after proteolytic digestion with Trypsin (or an alternative enzyme such as chymotrypsin, Lys-C, Asp-N, or others). Within a certain dynamic range, the mass spectrometer can recognize both the labeled and unlabeled forms of tryptic peptides and relative quantification can be achieved by comparing their respective signal intensities. If so, the mean ratio of all quantified peptides accurately reflects the heavy/light ratio of a protein species, which can be used to calculate in-cell NMR sample concentrations.

1. For a given protein to be delivered into mammalian cells, defined mixtures of unlabeled (UL) and stable-isotope labeled (SIL) forms of the protein are prepared first. Concentration ratios may range from 0.001 to 100 (UL/SIL). Absolute concentrations are less critical since MS is highly sensitive and detection limits are not usually a problem. 25–100 ng is a good starting point (see Note 23).

2. Separate mixtures by SDS-PAGE, Coomassie stain and excise the corresponding protein band (see Note 24).

3. Digest proteins in-gel with appropriate enzyme(s) (see Note 25).
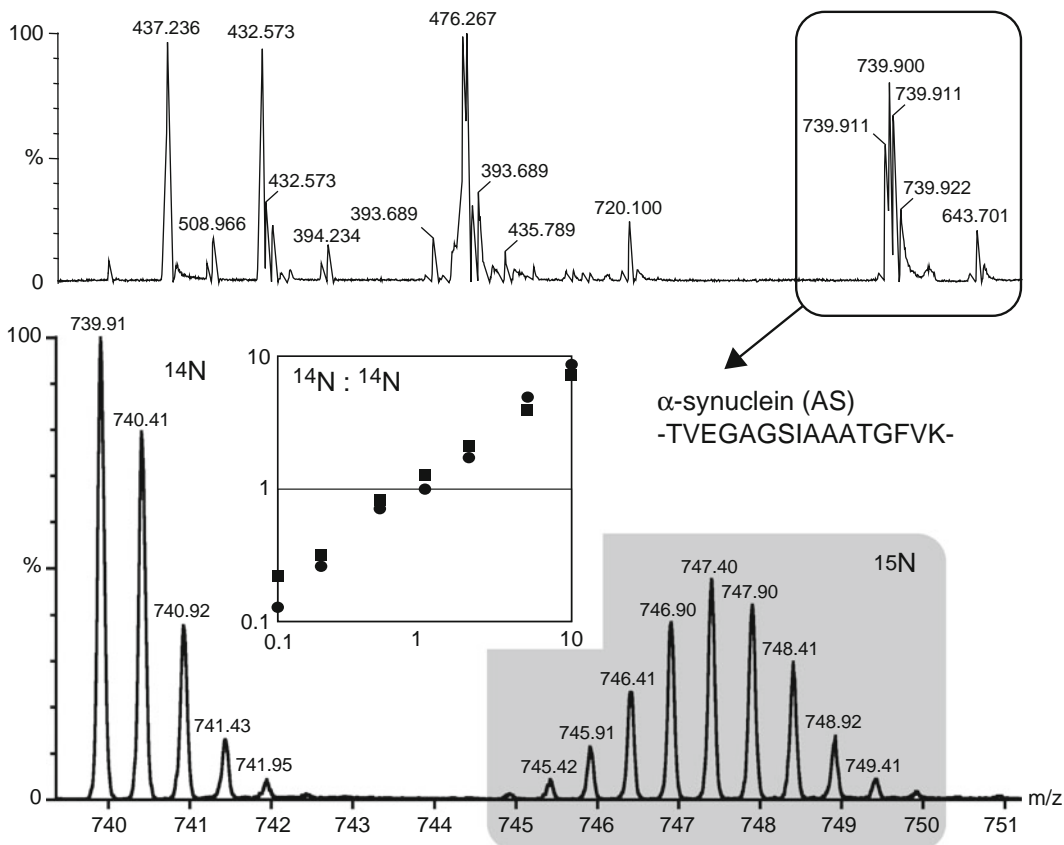
Fig. 2. Quantification of stable-isotope labeled protein concentrations by quantitative mass spectrometry (MS). NanoLC tandem MS run of an in-gel Trypsin digested mixture (equimolar) of ^{15}N-labeled and non-labeled (^{14}N) α-synuclein (AS) (*upper panel*). Several tryptic AS peptides and their corresponding masses are detected. The *lower panel* depicts the MS/MS fragmentation run and isotope-resolved distribution pattern of the 739.9 Da peptide species (corresponding to aa 81–96 of human AS). Clear separation of stable isotope-labeled (^{15}N) and non-isotope-labeled (^{14}N) signals enables accurate concentration measurements over a dynamic range spanning two orders of magnitude (*inset*).

4. Analyze peptide mixtures by nanoLC tandem MS (preferably nanoLC coupled with an ESI-QTOF or ESI-LTQ-Orbitrap instrument) and select suitable proteolytic (tryptic) peptides to confirm peptide sequences by MS/MS fragmentation(s) (Fig. 2) (see Note 26).

5. Analyze MS spectra of selected peptides (see Note 27).

6. Integrate MS signals from unlabeled and the corresponding labeled species (see Note 28).

7. Inspect the quality of the fit of experimentally obtained-, versus calculated concentration ratios. Choose a suitable dynamic range for all subsequent steps. In most instances, ratios between 0.1 and 10 (UL/SIL) yield good correlations and produce the most accurate results (Fig. 2).

8. For in-cell NMR sample analysis, prepare a cell lysate from a defined number of stable-isotope labeled, protein transduced cells.

9. Measure the total protein concentration of the cell lysate and add defined amounts (~25–100 ng) of unlabeled target protein. An estimate for the intracellular protein concentration of the labeled component may already be available at this point (from Western-blot analyses, for example). This will help to determine the most suitable concentration of unlabeled protein to be added for optimal correlation results (however, this is not strictly required) (see Note 29).

10. Separate protein lysate by SDS-PAGE, Coomassie stain and excise gel band at the expected molecular weight of the delivered protein (see Note 30).

11. Digest proteins in-gel with appropriate enzyme(s), extract isotope-labeled and non-isotope-labeled peptide fragments, and analyze peptide mixture by nanoLC tandem MS as indicated above (steps 3–5).

12. Integrate MS signals and calculate heavy/light ratios as indicated above. Determine lysate concentration of isotope-labeled protein using the mean value of all quantified peptides. Refer to Subheading 3.5 for calculating the effective intracellular protein concentration of the delivered substrate based on the number of cells initially used to prepare the lysate.

3.7. In-Cell NMR Spectroscopy

In the previous Subheadings 3.5 and 3.6 we outlined methods to experimentally determine solid estimates for intracellular protein concentrations (C_{Cell}) of successfully delivered substrates in mammalian in-cell NMR samples. By using the equation provided in Chapters 4 and 5, Subheading 3.1 Note 3, we can now calculate the effective NMR concentrations (C_{NMR}) that the different in-cell NMR samples should have. In order to assess whether expected-, and experimentally observed in-cell NMR signal intensities match, simple 1D in-cell NMR experiments provide a good starting point (see below). Why are those correlations important? Above all, because expected and experimentally determined in-cell NMR signal intensities will only match when most of the delivered protein is "tumbling freely" in the cytoplasm of the targeted cells. Protein interactions with cellular components like membranes, or other large macromolecular biomolecules (such as cytoskeletal proteins or DNA, for example) will result in severe NMR line broadening (16) so that NMR signal intensities will be greatly diminished. In-cell NMR measurements themselves can thereby identify possible biological interactions of intracellular proteins and provide important clues towards the qualitative nature of the environment that the delivered protein experiences.

1. Based on the methods outlined above, determine the intracellular protein concentration (C_{Cell}) of the in-cell NMR sample and correspondingly the expected, effective NMR concentration (C_{NMR}).

2. Prepare a "mock" in vitro in-cell NMR sample to which the isotope-labeled protein sample is directly added at the effective NMR concentration (C_{NMR} see Note 31).

3. Run 1D hetero-nuclear NMR experiments with settings that will also be applicable and used for the true in-cell NMR specimen (see Note 32).

4. Determine experimentally obtained NMR signal intensities as reference for the in-cell NMR sample.

5. Run mammalian in-cell NMR sample with identical settings as previously employed for the reference sample.

6. Compare NMR signal intensities and interpret accordingly (see Note 33).

7. Proceed to multi-dimensional NMR experiments when results are satisfying.

8. Consider re-running intermittent 1D NMR quality tests to evaluate changes in NMR signal intensities over time and interpret accordingly (see Note 34).

4. Notes

1. The choice of fluorescence dye is determined by the desired cell viability assay that the FCM experiment is supposed to report. Because dual-channel FCM measurements will be employed, the spectroscopic properties of the two different fluorescence dyes (i.e., absorbance and emission wavelengths) must be sufficiently set apart in order to enable their simultaneous readouts. Standard cell viability dyes for FCM analyses include 7-AAD, or propidium iodide (PI) (17).

2. Different fluorescence dyes require different reactive groups for coupling (i.e., lysine or cysteine side-chains, for example). Therefore, check for the availability of these residues in the protein of interest. Note that certain residues might be important for protein function and/or localization. Nuclear localization sequences (NLS), for example, are rich in lysines and arginines. To modify their chemical properties will impair nuclear import of the target protein and should therefore be avoided. Note that certain buffers are not compatible with these coupling reactions, for example Tris (TBS), glycine and ammonium salt buffers contain free amide groups, that greatly affect coupling efficiencies.

3. For protein delivery into adherent mammalian cells, keep in mind that non-viable cells typically detach from cell culture dishes/flasks and are conveniently removed by the individual wash steps of the respective protocols. For suspension cells, this is not the case. Therefore, adherent cell viability counts are usually larger than for suspension cells, although this does not necessarily reflect the true situation of the manipulated system.

4. An important point to consider is that fluorescence protein signals detected by FCM measurements may originate from molecules inside the manipulated cells, as well as from molecules stuck to the extracellular cell membrane, thereby scoring as protein uptake positive cells without containing properly internalized protein. This problem may become particularly evident for cellular protein transduction schemes that involve membrane interactions (i.e., CPP-mediated protein delivery). A way to test whether FCM results are skewed by unspecific protein/membrane binding events is to employ a Trypsin treatment step. Trypsin proteolytically degrades cell-surface-bound proteins, but does not affect successfully delivered, intracellular proteins (18). A direct comparison of Trypsin-treated versus non-reacted cells provides an accurate measure for intracellular protein uptake, and the degree of unspecific membrane binding. (This step is more important for suspension cell lines, as adherent cells will be treated with Trypsin to detach them from the culture dishes prior to FCM measurements.)

5. Increasing the intracellular concentration of a delivered protein eventually leads to toxicity problems. As the range of intracellular protein concentrations for in-cell NMR samples is typically high (up to hundreds of μM) different proteins may lead to cytotoxicity problems at different intracellular protein concentrations. It is therefore necessary to carefully assess maximally tolerable protein levels for every protein that is delivered and for every cell line that is targeted.

6. Most proteins to be studied by in-cell NMR spectroscopy have been characterized by other biological methods, including cell microscopy. To assess the physiological relevance of cellular distribution patterns of exogenously delivered proteins, it is important to know beforehand where the protein of interest localizes to in its native setting (also in respect to the cell type to be targeted). Note that for CPP-mediated protein delivery protocols for example, the fluorescence dye is part of the CPP moiety and not of the cargo protein. Intracellular release of the CPP from the cargo, by nature of the reducing environment of the mammalian cytoplasm and reduction of the disulfide bond in between the CPP and the protein, means that the observed fluorescence signal most likely originates from free cytosolic CPP. The distribution of the fluorescence signal and its cellular

lifetime will therefore reflect properties of the CPP, rather than those of the delivered protein. In order to overcome this problem, the cargo protein may be directly detected by immunofluorescence microscopy, see Subheading 3.4.

7. As stated before, whenever a delivered protein displays intracellular localization characteristics that are not in line with its known physiological properties, it should be a matter of concern. Unusually high intracellular protein concentrations-, as required for in-cell NMR measurements-, may lead to non-physiological protein distributions. A notion that is also well known in transient protein overexpression studies in cell biology. In that sense, irregular protein localization may result as a consequence of intracellular protein levels. In a few instances, aberrant intracellular protein targeting may also be artificially introduced by the chemical nature of the fluorescence dye that is used to visualize the delivered protein. In one such example, we found that α-synuclein (AS), lysine-coupled to Atto488, or to Atto647 displayed vastly different localization properties in human HeLa cells. Depending on which fluorescent dye was used, AS strongly colocalized with lysosomal vesicles, i.e., Atto647 (Fig. 1a), or displayed a uniform distribution in the cytosol and cell nucleus, i.e., Atto488 (Fig. 1b). By contrast, immunofluorescence microscopy detection of the non-modified protein using AS-specific-antibodies and fixed HeLa cells revealed a homogenous cytoplasmic staining with no nuclear localization (Fig. 1c).

8. Whenever experimental evidence suggests that the chemical properties of a fluorescence dye might influence the intracellular localization behavior of a delivered protein, alternative dyes may be chosen for microscopy analyses. It is important to emphasize, however, that intracellular localization properties may be different for non-fluorescence dye coupled proteins. As the actual in-cell NMR experiment will not be performed with dye-coupled-, but rather with non-modified protein, localization studies need to be considered as indicative of possible intracellular localizations only.

9. Several dyes are commercially available to quickly identify cellular structures/organelles by co-staining procedures. This is especially helpful for identifying potential causes for non-physiological protein localizations. In many instances, such analyses will reveal protein co-localization with lysosomes, vesicular organelles that function as cellular "trash-bins" processing "superfluous" biomolecular materials. This may indicate rapid cellular clearance of the delivered protein, which is detrimental for in-cell NMR analyses.

10. Time-dependent evolution of changes in intracellular localization behaviors are of particular importance for in-cell NMR studies as samples are typically analyzed over extended periods of time (up to several hours). Therefore, it is advisable to perform live cell imaging analyses at time intervals that correspond to expected experimental NMR settings. In this regard, it is also suggested to analyze samples under "mock" in-cell NMR conditions, i.e., after incubating the manipulated cells in NMR sample tubes to mimic the effects of limited aeration and nutrient supply.

11. Most cytotoxic effects accumulate over time and consequently lead to increased problems in intracellular protein stability. It is important to assess beforehand the level of in-cell NMR sample deterioration that is expected for an envisaged in-cell NMR experiment.

12. As both mammalian in-cell NMR sample preparation protocols contain optional steps for embedding manipulated cells in high-density support matrices (i.e., low-melting agarose, or Redigrade™) check how these procedures affect intracellular protein localization and cell viability. Manipulated cells can be mock treated with embedding material on microscopy cover slips and similarly analyzed by fluorescence imaging. For this, apply the staining procedure with organelle-specific dye (e.g., DAPI, Hoechst 43332, LysoTracker™), then harvest cells and prepare solid support matrix.

13. Fixation methods commonly utilize a combination of organic solvents and cross-linking reagents. Many different combinations are available which can be applied to best suit the target antigen.

14. Similar to live cell fluorescence imaging experiments outlined above, intracellular protein distribution must be assessed against the known localization properties of the delivered protein. Comparing live-cell and fixed-cell imaging results may help to identify problems in fixation procedures, as well as possible sources for imaging artifacts.

15. As different fixation procedure can lead to different localization properties, it is advisable to test several routines in parallel.

16. Similar to live-cell imaging experiments, localization analyses by IF microscopy should also be performed with samples containing different intracellular concentrations of delivered protein and at different time intervals (see above).

17. There are essentially two options: Manipulated cells can be lysed with either a denaturing (i.e., RIPA) or a non-denaturing buffer. Denaturing buffers will solubilize all proteins, while non-denaturing buffers will only yield the soluble fraction of cytoplasmic proteins (i.e., no membrane-, or

membrane-attached, or vesicular proteins). For determination of the total intracellular protein concentration of a delivered substrate, a denaturing buffer should be employed. If an estimate for the fraction of soluble delivered protein is to be obtained, a non-denaturing buffer may be chosen instead, or in addition. Note that proteins observable by in-cell NMR spectroscopy should largely be contained in the soluble protein fraction, but may not necessarily do so. It is therefore advised to comparatively analyze manipulated cells by preparing extracts with both types of buffer.

18. It is important to precisely determine the number of cells used to prepare the extract. This number will be used to finally calculate the molar concentration of delivered protein substrate per cell. Cell counting is performed in a Neubauer hemocytometer (a cell counting chamber) under a light microscope according to the manufacturer's instructions.

19. Protease/phosphatase inhibitors should be added freshly. Note that detergents such as TX-100, NP40, Tween20 and SDS may not be suitable for lysate protein concentration measurements by Bradford-type assays (see below).

20. When non-denaturing buffer conditions are used, complete cell lysis should be confirmed by light microscopy.

21. The total protein concentration of the cell lysate will be ~5 μg/μL at this point. The linear range for Bradford assays is typically 0.1–1 μg/μL. Dilute an aliquot of the cell lysate accordingly and measure total protein concentration of the resulting extract.

22. If, for example, 40 ng of delivered protein in 4 μL of cell lysate has been determined in this way, the effective lysate concentration is 0.01 μg/μL. Since 1×10^6 cells (~2 pL individual cell volume, 2 μL total cell volume) were initially employed to prepare the extract (in 100 μL) a ~50-fold dilution factor going from cell volume to extract volume has to be taken into account. Doing so, we obtain a total protein concentration of 0.5 μg in 1×10^6 cells (or 0.5 μg/μL). On a single cell basis, this equates to 1 pg of protein/cell (2 pL cell volume). Assuming a protein molecular weight (MW) of 14,500 Da for example, we reach an intracellular protein concentration (C_{Cell}) of 35 μM.

23. These mixtures should contain different ratios of concentrations in order to initially determine the dynamic range over which MS produces accurate quantification results.

24. For a mixture of pure proteins, there should only be one protein band on the gel.

25. Use Trypsin wherever applicable and follow standard laboratory routines for MS-based protein identification and quantification used in proteomic studies ([19–21]).

26. Multiple proteolytic peptide species should be sequenced and selected for most accurate quantification results.

27. MS spectra display characteristic isotopic patterns of MS signals. These reflect natural abundance isotope distributions (mainly C^{12}/C^{13}), as well as different ionization species. Because recombinantly produced proteins are never 100 % isotope-labeled, this results in an additional mass distribution of the "heavy" peptide species (Fig. 2).

28. For accurate results integrate and sum up all signals of the isotopic pattern of the individual species and calculate heavy/light ratios for each of the selected peptide using at least three proteolytic peptides.

29. This corresponds to the expected range of concentrations of successfully delivered proteins in in-cell NMR sample lysates (with ~20 μg of total protein). Note that the targeted cell line may already contain endogenous amounts of non-isotope-labeled protein. In such instances, first determine the concentration of the endogenous protein by spiking untreated cell extracts with defined amounts of isotope-labeled protein (i.e., the reverse reaction). Analyze endogenous protein levels accordingly and subtract from in-cell NMR sample calculations.

30. Whenever target proteins are posttranslationally modified by cellular enzymes, their SDS-PAGE migration behaviors-, as well as their individual molecular masses-, will change in unpredictable manners and complicate MS analyses. Check for PTM events by appropriate alternative means.

31. It is important to mimic in-cell NMR sample conditions as closely as possible. If, for example, cells are embedded in low-melting agarose or resuspended in Redigrade™, then the corresponding "mock" sample should also contain these components. This will also enable to determine the NMR behavior of leaked protein molecules in the true in-cell NMR sample.

32. Employ settings (especially temperature and pH conditions) that will be used for and experienced by the final in-cell NMR sample. It is essential to avoid differences in chemical exchange behavior in these analyses, as these will affect NMR signal intensities and compromise quantitative comparison.

33. Note that some differences in signal intensities are to be expected. These result from sample and magnetic field inhomogeneities, as well as from differences in intracellular viscosity and macromolecular crowding that are unavoidable in in-cell NMR samples. These differences should, however, be within an expected range of effects and not orders of magnitude off the anticipated results.

34. As sample aging is clearly observed in many in-cell NMR studies, it is advisable to repeat the outlined 1D experiments at different time points of in-cell NMR measurements. An overall decrease in experimentally obtained signal intensities with time will be apparent and conversely reflect changes in overall cell viability parameters. Based on these observations, suitable time frames for in-cell NMR measurements in mammalian cells should be chosen.

References

1. Orrenius S, Zhivotovsky B, Nicotera P (2003) Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol 4:552–565

2. Recktenwald DJ (1991) Multicolor immuno-fluorescence and flow cytometry. Methods 2:192–199

3. Duchardt F, Fotin-Mleczek M, Schwarz H, Fischer R, Brock R (2007) A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. Traffic 8:848–866

4. Richard JP, Melikov K, Vives E, Ramos C, Verbeure B, Gait MJ, Chernomordik LV, Lebleu B (2003) Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. J Biol Chem 278:585–590

5. Miller DM, Shakes DC (1995) Immunofluo-rescence microscopy. Methods Cell Biol 48:365–394

6. Moulton HM, Hase MC, Smith KM, Iversen PL (2003) HIV Tat peptide enhances cellular delivery of antisense morpholino oligomers. Antisense Nucleic Acid Drug Dev 13:31–43

7. Melan MA, Sluder G (1992) Redistribution and differential extraction of soluble proteins in permeabilized cultured cells. Implications for immunofluorescence microscopy. J Cell Sci 101(Pt 4):731–743

8. Barnes CO, Pielak GJ (2011) In-cell protein NMR and protein leakage. Proteins 79: 347–351

9. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, Young GB, Macdonald JM, Pielak GJ (2008) Differential dynamical effects of macro-molecular crowding on an intrinsically disor-dered protein and a globular protein: implications for in-cell NMR spectroscopy. J Am Chem Soc 130:6310–6311

10. Lee C (2007) Western blotting. Methods Mol Biol 362:391–399

11. Ong SE, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 1:252–262

12. Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, Vandenesch F, Garin J (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. Mol Cell Proteomics 6:2139–2149

13. Kippen AD, Cerini F, Vadas L, Stocklin R, Vu L, Offord RE, Rose K (1997) Development of an isotope dilution assay for precise determina-tion of insulin, C-peptide, and proinsulin levels in non-diabetic and type II diabetic individuals with comparison to immunoassay. J Biol Chem 272:12513–12522

14. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA 100:6940–6945

15. Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protoc 1:1029–1043

16. Selenko P, Wagner G (2007) Looking into live cells with in-cell NMR spectroscopy. J Struct Biol 158:244–253

17. Telford WG, Komoriya A, Packard BZ, Bagwell CB (2011) Multiparametric analysis of apopto-sis by flow cytometry. Methods Mol Biol 699:203–227

18. Richard JP, Melikov K, Brooks H, Prevot P, Lebleu B, Chernomordik LV (2005) Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors. J Biol Chem 280: 15300–15306

19. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207

20. Lange S, Sylvester M, Schumann M, Freund C, Krause E (2010) Identification of phosphorylation-dependent interaction part-ners of the adapter protein ADAP using quantitative mass spectrometry: SILAC vs (18)O-labeling. J Proteome Res 9:4113–4122

21. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1:2856–2860

Chapter 7

Fourier Transform Infrared Microspectroscopy of Complex Biological Systems: From Intact Cells to Whole Organisms

Diletta Ami, Antonino Natalello, and Silvia Maria Doglia

Abstract

Fourier transform infrared (FTIR) microspectroscopy is a powerful tool for the study of complex biological systems. Indeed, it is employed to characterize intact cells, tissues, and whole model organisms such as nematodes, since it allows to obtain a chemical fingerprint of the sample under investigation, giving information on the molecular composition and structures. The successful application of this technique for the in situ study of biological processes requires specific sample preparations, in order to obtain reliable and reproducible results. In the present work, we illustrate the optimized procedures to prepare biological samples for IR measurements and the method to collect and analyze their FTIR spectra. In particular, we describe here the investigations on bacterial cells, intact eukaryotic cells, and whole intact nematode specimens.

Key words: Amyloid aggregation, *Caenorhabditis elegans*, Embryonic stem cell differentiation, FTIR microspectroscopy, Inclusion bodies, Protein aggregation, Recombinant protein expression

1. Introduction

In the last few years, FTIR microspectroscopy (1) has been found to be a powerful tool for the characterization of complex biological systems (2, 3) such as intact cells, tissues, and whole model organisms (4–8). In particular, this noninvasive and label-free technique allows to obtain, within a single experiment, a unique molecular fingerprint that provides information on the composition and structure of the main cellular biomolecules—such as proteins, nucleic acids, lipids, and carbohydrates (see Fig. 1). In this way, FTIR microspectroscopy provides an invaluable tool to tackle the study—in situ—of biological processes that take place simultaneously within biological systems. For instance, the potential of this spectroscopic approach has been widely exploited to monitor in
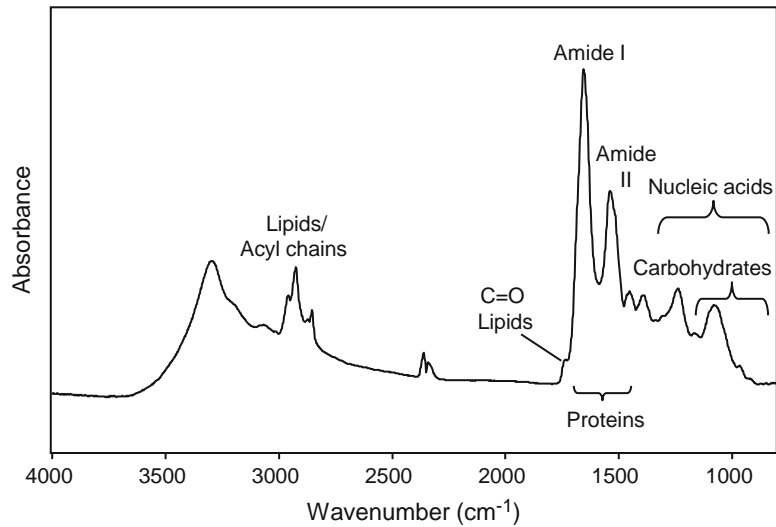
Fig. 1. FTIR absorption spectrum of intact murine embryonic stem cells. The measured absorption spectrum taken from a cell pellet (less than 10^2 cells) is reported without any corrections. Cells—deposited on a BaF_2 window—were measured in transmission by the IR microscope UMA 500 coupled to the FTIR spectrometer FTS-40A (both from Digilab), at a resolution of 2 cm^{-1}. The main absorption regions of cell biomolecules are indicated.

situ protein aggregation (9–12), cell differentiation (5, 13), and cancerogenesis (14, 15).

In particular, protein aggregation is a central issue in several fields, with important biomedical and biotechnological implications. This process is involved, indeed, in the so-called "amyloid diseases," such as Alzheimer, Parkinson, and prion diseases (16). With regard to this point, we should underline that in the last decade transgenic invertebrate models, such as the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*, have been developed for the in vivo study of amyloid diseases (17–20).

In addition, protein aggregation in form of inclusion bodies (IBs) is involved in the setup of microbial factories for the production of recombinant proteins (21). Moreover, IBs have been also found to share common features with amyloid aggregates, offering an interesting model system for the in vivo study of protein aggregation-based diseases (22). Interestingly, IBs enriched in the Aβ peptides—the well-known disordered peptides responsible of the Alzheimer's disease—have been recently characterized by Dasari and colleagues (23), highlighting the possibility to study amyloids in a cellular environment. Indeed, FTIR (micro)spectroscopy has been successfully applied in this field, since it provides not only information on the protein secondary structures but also on protein aggregates, thanks to a specific marker band falling between 1,630 and 1,620 cm^{-1}, due to the formation of intermolecular

β-sheet structures (24–26). Interestingly, the possibility to follow protein aggregation has been proven not only in vitro but also in situ, enabling the study of the aggregation process in intact cells, tissues, and also in nematodes (9–12, 27, 28).

Another promising application of FTIR (micro)spectroscopy in cell biology is the study of cell differentiation (5, 6, 13, 29). In particular, this technique allows to study, in a noninvasive way, the main physiological processes that take place—simultaneously—within intact cells. Indeed, changes in the protein secondary structures, lipid and carbohydrate composition, and nucleic acid conformations can be monitored within a single measurement from a limited number of cells, even at the single cell level.

Considering the increasing interest in the application of stem cells in regenerative medicine, FTIR (micro)spectroscopy appears as a powerful new tool to provide unique stemness markers, enabling first to discriminate between undifferentiated and differentiated cells (5, 6, 13, 29). Moreover, the possibility to obtain information on the main cell biomolecule content could make it possible to disclose the cell fate, identifying the spectral features typical of the new incoming phenotype (5, 30).

In the present work, first, we illustrate the use of FTIR microspectroscopy to monitor the formation of IBs within intact *Escherichia coli* cells. We, then, report in details the in situ study of the Aβ peptide expression within single *C. elegans* specimens.

Finally, to describe the procedure that we employed for the IR study of intact mammalian cells, we report and discuss the protocol optimized to characterize the IR response of murine embryonic stem cells, at different stages of their spontaneous differentiation process.

2. Materials

2.1. FTIR Instrumentation and Accessories

The FTIR absorption spectra of the different biological systems illustrated in this work were collected—from 4,000 to 800 cm^{-1}—using the UMA 500 infrared microscope equipped with a nitrogen cooled MCT detector (narrow band, 250 μm), coupled to a FTS-40A spectrometer (both from Bio-Rad, Digilab Division, MA, USA), under accurate dry-air purging (1).

For IR microscope analyses, samples were deposited onto IR transparent ZnSe or BaF$_2$ windows—depending on the samples—of 13 mm diameter and 2 mm thickness (Korth Kristalle GmbH, Germany).

2.2. Software

The Win-IR software (Bio-Rad, MA, USA) was employed for spectral collection. Spectral analysis, including water vapor subtraction and second derivative calculation, was performed by the GRAMS/32 software (Galactic industries corporation, Salem, NH, USA).

2.3. E. coli Recombinant Strains	We studied the infrared absorption of the following *E. coli* strains:

NP 366: recombinant strain expressing the human interferon alpha-2b (hIFN-alpha-2b); NP 406: recombinant strain expressing the hIFN-alpha-2b at three- to fivefold higher level than NP 366; NP 260: control strain not expressing the hIFN-alpha-2b (27, 31).

2.4. C. elegans Strains	To explore the IR response of intact nematode specimens, we measured four well-characterized taxa, among which a wild type *C. elegans* strain (8).

For the IR study of the Aβ peptide expression, we measured the temperature-inducible *C. elegans* transgenic strain (CL4176) expressing the muscle-specific $A\beta_{1-42}$ and the control strain (CL802), not expressing the peptide (see, for details, ref. 9).

2.5. Murine Embryonic Stem Cells	Murine embryonic stem cells were isolated from the inner cell mass of the blastocyst and spontaneously differentiated up to 14 days (for details, see ref. 5).

3. Methods

3.1. FTIR Microspectroscopy of Recombinant E. coli Cells	To illustrate the FTIR method to monitor the expression of heterologous proteins in form of inclusion bodies (IBs) within intact bacterial cells, we describe here the procedure that we followed to characterize the production of hIFN-alpha2b in *E. coli* cells. In particular, we examined the IR absorption of two strains producing respectively a low (NP366) and a high (NP406) level of IFN-alpha 2b and of the control strain NP 260, not expressing the recombinant protein (27, 31, 32).

3.1.1. Protocol

1. Recover bacterial cells by centrifugation from stirred-flask fermentations and normalize to the same cell density.

2. Resuspend each cell pellet in distilled water and deposit about 15 μl of the resulting suspension (containing approximately 10^6 cells) onto an infrared transparent window (ZnSe, BaF_2, etc.), in a spot of about 1 mm. In order to remove the excess water, let the samples dry at room temperature for at least 30 min, in a protected environment, such as a laminar flow hood (see Note 1).

3. For FTIR microscope measurements in transmission mode, set the microscope diaphragm aperture to the size of the sample that you would like to measure. Choose the instrument parameters: 25 kHz scan speed for MCT detector, a resolution of 2 or 4 cm^{-1} is recommended, and—to obtain spectra with

excellent signal to noise ratio—at least 512 scan coadditions. Triangular apodization function is preferred to truncate the interferogram.

4. Place the sample window on the microscope stage. Pull down the plastic cylinder around the objective to protect the sample from external water vapor, and—after having focused the beam on the sample—center an empty area of the window, keeping the fixed diaphragm aperture. Optimize, then, the IR beam focus by adjusting the Cassegrain condenser and the spectrometer sensitivity, and, at this point, collect the background spectrum (the number of scan coadditions should be the same or higher than the sample spectrum). To obtain a good baseline, collect the background just before each sample measurement.

5. Select an area (of the same size of the background spectrum) within the sample and collect the sample absorption spectrum (see Note 2).

6. To collect the water vapor spectrum on the same sample window, first choose an empty area, collect a background spectrum, then pull up the plastic cylinder to obtain a high vapor signal, and—on the same background area—collect an absorption spectrum. The same instrument parameters used for the sample should be used for the vapor spectrum. A vapor spectrum should be collected after each set of FTIR experiments.

7. Analyze the measured spectra, first correcting them—when necessary—for water vapor absorption (see Note 3).

8. To perform the second derivative analysis, in order to enhance the apparent spectral resolution, first the raw measured spectra should be smoothed (see Note 4). Then, calculate the second derivative spectrum, where the band components of the measured spectrum appear as negative peaks at the same wavenumbers (see Note 5). To monitor the IB formation in intact cells, concentrate your attention on the Amide I region, between 1,700 and 1,600 cm^{-1}, and in particular on the appearance of a spectral component around 1,630–1,620 cm^{-1}, due to intermolecular β-sheet formation (24–26).

9. In the case of low expression levels, it could be useful to subtract from the producer strain absorption spectrum that of the control strain (not expressing the recombinant protein), to better evaluate the response due to IB formation (see Note 6). Then, the resulting spectrum could be analyzed as explained at steps 7 and 8.

In Fig. 2 we reported the FTIR characterization of the recombinant hIFN-alpha-2b overexpressed in *E. coli* (27, 31, 32). The second derivative spectrum of the intact control cells (not producer
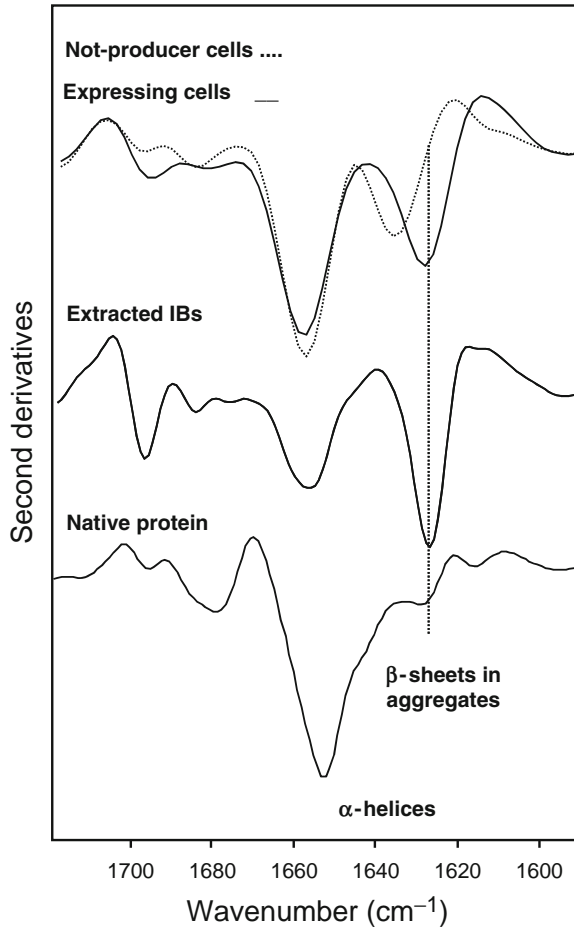
Fig. 2. FTIR analysis of human rIFN-alpha-2b overexpression in *Escherichia coli*. The second derivative of the FTIR absorption spectrum of intact *E. coli* cells overexpressing rIFN-alpha-2b is compared to that of the nonexpressing control strain. The rising of the band around 1,628 cm^{-1}, due to intermolecular β-sheets in aggregates, allows to study the kinetics of IB formation within intact cells. The second derivative spectrum of extracted IBs reveals the presence of intermolecular β-sheet structures (1,628 cm^{-1}) and of residual native-like structure (1,657 cm^{-1}). The second derivative spectrum of native IFN-alpha-2b in solution is also reported for comparison. Intact cells and extracted IBs (deposited on ZnSe windows) were measured in transmission by the IR microscope UMA 500 coupled to the FTIR spectrometer FTS-40A (both from Digilab), at a resolution of 2 cm^{-1}. The native protein was measured at a concentration of 10 mg/ml (resolution 2 cm^{-1}) using a temperature-controlled transmission cell, at room temperature.

strain) is characterized by two main components at ~1,658 cm^{-1} and at ~1,638 cm^{-1}, respectively due to the α-helix and intramolecular β-sheet structures (24–26) of the whole protein content of the cells. The derivative spectrum of the producer strain (NP 406) displays again the α-helix component at ~1,658 cm^{-1}, even if at lower intensity than the control strain, while the intramolecular

β-sheet component at ~1,638 cm^{-1} disappeared and a new peak at ~1,628 cm^{-1}, due to intermolecular β-sheets of protein aggregates, arises indicating the formation of hIFN-alpha-2b IBs. To obtain detailed structural information, IBs were then extracted and measured, following the same procedure used for the intact cell measurements and analysis, as described in steps 2–6. Unexpectedly, their second derivative spectrum (Fig. 2) is also characterized by the α-helix component at ~1,657 cm^{-1} and by a peak at ~1,627 cm^{-1}, assigned to protein aggregates, accordingly to what found in intact cells. Interestingly, the α-helix component represents the residual native-like structures of the polypeptides embedded within IBs, as confirmed by the comparison with the native recombinant hIFN-alpha-2b spectrum (see Fig. 2) (31, 32).

3.2. FTIR Microspectroscopy of Single C. elegans Specimens

To illustrate the FTIR microspectroscopy approach for the analysis of intact nematode specimens, we report the procedure that we first set up to characterize the IR absorption of different nematode species (8) and that we later applied to the study of the temperature-induced expression of Aβ peptide (Aβ$_{1-42}$) in *C. elegans* (9).

3.2.1. Protocol

1. Pick single nematode specimens from the agar plate and place them into a water drop. To remove any agar trace, wash extensively single specimens by transferring them to several water drops.

2. Transfer single specimens on BaF$_2$ windows within a water drop of a few microliters and, as discussed for bacterial cells, let the samples dry at room temperature for at least 30 min (see step 2, Subheading 3.1 and Note 1).

3. For FTIR microscope measurements, set the microscope diaphragm aperture in order to select the area of the nematode body that you want to measure (this will depend on the body length) (see Fig. 3a). Set the instrument parameters: 25 kHz scan speed for MCT detector, a resolution of 4 cm^{-1} may be preferable in this case (see Note 7), and at least 512 scan coadditions are suggested.

4. Optimize the infrared signal and collect the background spectrum (see step 4, Subheading 3.1 for details).

5. Center the microscope diaphragm aperture on the selected body region (pharynx, intestine, or tail; see Fig. 3a) and collect the absorption spectrum.

6. Collect the water vapor spectrum on the same sample window (see step 6, Subheading 3.1).

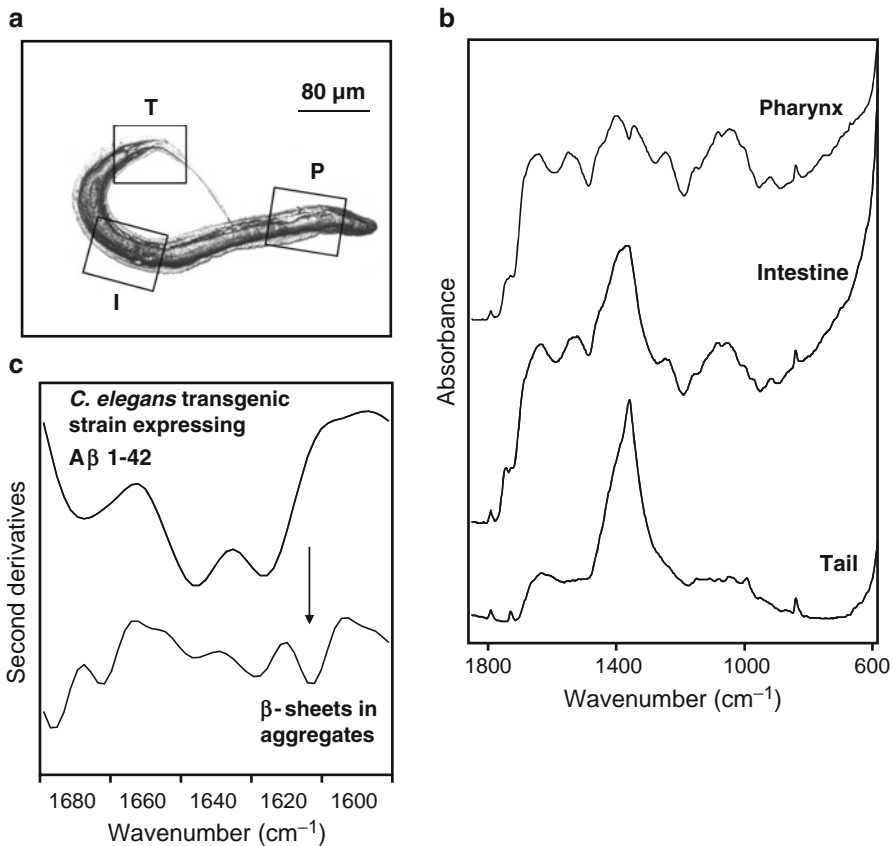7. Correct the absorption spectra for water vapor absorption, when necessary (see Note 3).

Fig. 3. FTIR microspectroscopy of *Caenorhabditis elegans*. (**a**) Optical image of an intact *C. elegans* specimen. The regions selected by the infrared microscope diaphragm aperture are indicated by the rectangular frames. (**b**) Raw FTIR absorption spectra of pharynx, intestine, and tail body regions, measured within an intact specimen, deposited on ZnSe windows. (**c**) Second derivatives spectra of a temperature-induced *C. elegans* transgenic strain expressing the Aβ peptide$_{1-42}$, measured before (time 0) and 48 h after the temperature induction. All the spectra were collected from the pharynx region within intact specimens (deposited on BaF$_2$ windows). Spectra were normalized to the tyrosine band around 1,515 cm^{-1} to compensate for the total protein content. The arrow points to increasing aggregate component (at about 1,623 cm^{-1}). FTIR measurements were performed in transmission using the UMA 500 IR microscope coupled to the FTIR spectrometer FTS-40A (both from Digilab), at resolution of 4 cm^{-1}.

8. To perform the second derivative analysis, after the proper smoothing of the raw spectra, follow the procedure illustrated in step 8, Subheading 3.1.

In Fig. 3b we report the absorption spectra collected from the pharynx, intestine and tail within an intact *C. elegans* specimen. As shown in the figure, the absorption spectra of the three body parts are different, as expected. In particular, a higher reproducibility was found for the pharynx region, whose spectral response reflected the main molecular components of the cuticle and muscles. Interestingly, the IR absorption of the pharynx region enabled us to discriminate among different nematode species (8).

The same procedure was then applied to study the aggregation kinetics of a *C. elegans* strain expressing the $A\beta_{1-42}$, involved in the Alzheimer's disease (9). In particular, we collected the IR spectrum from the pharynx region of single intact specimens—where the highest peptide expression was expected to occur—at different times after the peptide thermal induction. We measured also, as a control, the nonexpressing strain.

The second derivative analysis of the expressing strain in the Amide I region enabled us to follow the aggregation process through the monitoring of the intermolecular β-sheet spectral component around $1,623$ cm^{-1}—absent in the control strain—whose intensity was observed to increase at longer time after induction (see Fig. 3c).

Interestingly, we studied also the in situ effects of tetracyclines—compounds that inhibit protein aggregations—on the secondary structures of the Aβ peptide. To this aim, we collected the IR absorption spectrum from the pharynx of intact nematode specimens expressing the $A\beta_{1-42}$, fed with tetracycline. Noteworthy, in this case the aggregate spectral component was significantly reduced with respect to that of the untreated expressing strain (9). We believe that the results obtained in this work highlight the potential of FTIR microspectroscopy to better understand the aggregation mechanisms of the amyloid proteins in vivo, in the presence of the physiological factors able to modulate their aggregation, such as the intracellular crowding and the cellular systems that assist folding and/or protein degradation. Furthermore, the possibility to study the effects of antiamyloidogenic compounds in vivo would allow a better comprehension not only of their mechanisms of actions but also of their possible toxic effects.

3.3. FTIR Microspectroscopy of Murine Embryonic Stem Cells

To illustrate the procedure followed to measure intact eukaryotic cells, we describe the FTIR study of murine embryonic stem cells at different times of spontaneous differentiation (5, 33).

3.3.1. Protocol

1. Recover the cells from flasks by centrifugation ($1,000 \times g$, for 10 min at 4 °C) and wash them extensively, at least three times, with NaCl 0.9 % by centrifugation (at $1,000 \times g$ for 5 min at 4 °C). Resuspend the cell pellet in NaCl 0.9 %, in order to have a number of cells of about 10^3–10^4.

2. Deposit onto a BaF$_2$ window about 10–15 μl of cell suspension in drops of 2–3 μl. Let the samples dry at room temperature for at least 30 min (see Note 1).

3. For FTIR microscope measurements, follow the procedure illustrated at steps 3–5 of Subheading 3.1 (see Note 8).

4. Analyze the second derivative spectra (see steps 7 and 8 of Subheading 3.1. and Notes 5, 9, and 10).
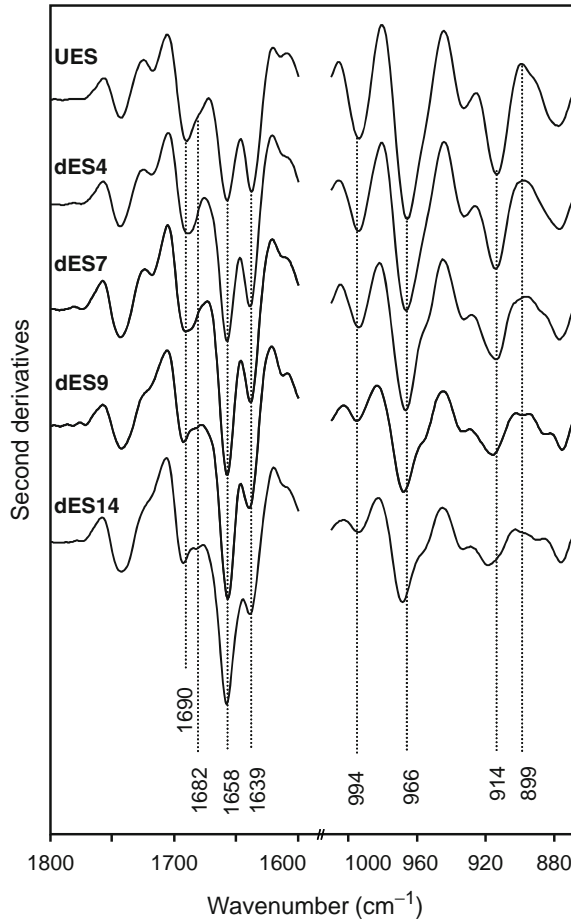
Fig. 4. FTIR microspectroscopy of undifferentiated and differentiating murine ES cells. Second derivatives of the FTIR absorption spectra of murine ES cells undifferentiated and at different stages of spontaneous differentiation are reported in the Amide I region and in the nucleic absorption between 1,000 and 860 cm^{-1}. Spectra were normalized at the tyrosine band around 1,515 cm^{-1}. The intensity of the spectra in the nucleic acid absorption regions was magnified by a factor three. Absorption spectra were collected by the IR microscope UMA 500 coupled to the FTIR spectrometer FTS-40A (both from Digilab), at resolution of 2 cm^{-1} (BaF$_2$ windows were employed).

Following the approach illustrated above, we studied the spontaneous differentiation of murine embryonic stem cells. Important changes in the main cell biomolecules were observed during the differentiation process. In particular, by analyzing the nucleic acid spectral response between 1,000 and 800 cm^{-1}, reported in Fig. 4, we monitored the formation of the DNA–RNA hybrid, detected by the simultaneous presence of the ribose ring at about 914 cm^{-1} and of the deoxyribose ring at 899 cm^{-1}, occurring after 4–7 days of differentiation. In this way, we were able to identify the differentiation time in which the switch of the new phenotype took place.

Moreover, the rising of the new phenotype was characterized by important changes in the whole lipid and protein content (5, 33).

Concerning proteins, between 4 and 7 days of differentiation, the analysis of the Amide I band clearly showed an increase of the α helix component around 1,658 cm^{-1} and the appearance of β turn structures, around 1,682 cm^{-1} (see Fig. 4). These results, together with the increased intensity—upon differentiation—of four glycogen marker bands between 1,160 and 1,020 cm^{-1}, allowed to identify the new phenotype, namely, cardiomyocyte precursors, as then confirmed by the PAS (periodic acid-Schiff) reaction analysis (5, 33). Concerning the nucleic acid response, we should add that, in agreement with previous spectroscopic studies (34, 35), a strong reduction of the RNA content, monitored by the intensity of the bands at 994 cm^{-1} (ribose phosphate main chain) and at 914 cm^{-1} (ribose ring mode), was found to occur during the differentiation process, reflecting its involvement in protein synthesis. Indeed, the RNA decrease—likely messenger RNA—could be taken as an important marker of ES cell differentiation process (34, 35).

3.4. Multivariate Statistical Analysis

A final remark should be added about the multivariate statistical analysis (36), an essential tool that allows to tackle the FTIR analysis of complex systems, usually dependent by more than one statistical variable. In particular, the multivariate method is crucial to validate and better comprehend the spectroscopic data.

Different approaches can be employed, depending on the aim of the analysis (37). For instance, to verify the reproducibility of the spectral data, the cluster analysis—a statistical method used to segregate the data into groups (clusters) on the basis of their similarity—could be a valid choice (38).

However, for the study of the temporal evolution of complex cellular processes, such as cell differentiation, it could be useful to disclose the spectral features that contain the most significant information on the molecular events simultaneously occurring within the system. To this aim, we suggest to employ the combined principal component–linear discriminant analysis (PCA-LDA) (5, 37, 39), which allows to discriminate among the different sample classes (i.e., differentiation times), identifying the most relevant wavenumbers that contribute to the inter-spectral variance (5, 33, 37).

4. Notes

1. It is necessary to evaluate the time required to dry the samples in a reproducible way. To this aim, it is possible to measure in a control experiment the FTIR absorption spectrum of the sample at different times of dehydration, until no more changes in the Amide I region (1,700–1,600 cm^{-1}) are observed.

Concerning this point, we should add that, recently, it has been demonstrated that a rapid desiccation of cells doesn't affect their spectroscopic response (40), confirming that the dry-fixing provides a suitable method for the study of intact cells by FTIR microspectroscopy.

We should also underline that to obtain a good spectrum, it would be useful to measure a homogeneous "film" of cells—a homogeneous monolayer for superior eukaryotes—in order to reduce light scattering artifacts.

2. In order to explore the possible heterogeneity of the sample and to verify the reproducibility of the IR response, it is recommended to repeat several measurements on different areas within the same drop and on the different sample drops. Moreover, at least three independent experiments (i.e., three independent sample preparations) should be performed.

3. When performing the water vapor subtraction, check its absorption in the Amide I region, enlarging the spectrum between $1,700$ and $1,600$ cm^{-1}. Concentrate your attention in particular on the intensity of the vapor peak around $1,653$ cm^{-1}: subtract vapor spectrum until you obtain a smooth spectrum around this peak. As a control, check for the absence of the vapor signal in a spectral range free from sample absorption (for example, between $1,850$ and $1,800$ cm^{-1}) (41).

4. To smooth the raw measured spectra, we usually employ the binomial function, applying 9–13 smoothing points, or more, depending on the quality of the measured spectrum. We suggest to choose the opportune number of smoothing points by checking the noise level in a spectral range free from absorption (i.e., between $1,850$ and $1,800$ cm^{-1}).

5. To calculate second derivative spectra, several functions can be employed; usually we apply the Savitzky–Golay method (42). To choose the number of smoothing points, you should check the noise level in an area free from absorption (see Note 4). In the interpretation of second derivative spectra, it is important to keep in mind that the spectrum resolution is increased, since the band half width is reduced of about a factor 2.7. Furthermore—being the peak height proportional to the original height, but inversely proportional to the square of the original half width—sharp peaks, as those due to water vapor, noise, and interference fringes, are particularly enhanced with respect to the spectral components. For this reason, an accurate purging of the instrument and high signal-to-noise ratio are crucial to obtain spectra that can be then processed by second derivative analysis. It is recommended to apply the same parameters for raw spectra smoothing and second derivative calculation when comparing different spectral data. To compare second derivative spectra (for instance, from different samples),

we usually normalize them at the tyrosine peak around 1,515 cm^{-1}, usually well resolved, to compensate for the protein content ([31]). It is also possible to normalize the measured spectra at the Amide I band area.

6. When performing the spectral subtraction, it is necessary to take care of the subtraction factor of the control strain spectrum, in order to compensate for the difference in optical path of the absorption measurements. This can be expected considering that the thickness of the cell layer cannot be accurately controlled. In particular, you should choose the subtraction factor checking that no negative absorption results in the subtracted spectrum.

7. We first employed a resolution of 2 cm^{-1}, but—likely due to the high complexity of the sample and to light scattering—it was then very difficult to analyze their second derivative spectra. For this reason, we decided to use a resolution of 4 cm^{-1}, however very satisfactory ([9]). Recently, the development of an algorithm for Mie scattering correction could allow to obtain high quality spectra from highly scattering biological samples ([43]), therefore enabling to increase the spectral resolution.

8. We should highlight that, in the case of cell differentiation studies, it is better to employ a spectral resolution of at least 2 cm^{-1}, since this allows to better resolve the overlapping spectral components due to the different biomolecules.

9. In the case of cellular studies involving several complex processes—such as cell differentiation—it is strongly suggested to analyze the spectral response of all the main cell biomolecules (proteins, nucleic acids, lipids, etc.). In this way, it could be possible to better characterize and comprehend the different processes that simultaneously take place within differentiating cells. To do this, it could be useful to explore the cell IR response over a wide range of frequencies, where changes in the biomolecule composition and conformations could be detected. In this way, it would be hopefully possible to disclose the marker bands of the process under investigation, which will depend on the cell system you are investigating.

10. We reported, in the following, a brief description of the most studied spectral ranges useful to analyze complex biological systems.

In particular, the lipid absorption falls mainly in the 3,050–2,800 cm^{-1} and 1,480–1,330 cm^{-1} spectral ranges, dominated by the absorption of the acyl ([4], [44]). Moreover, the spectral features between 1,750 and 1,700 cm^{-1}, mainly due to the C=O stretching of esters ([4], [44]), also give information on the sample lipid content.

As already discussed, protein analysis is mainly performed in the Amide I region, between 1,700 and 1,600 cm^{-1}, due to the

C=O stretching vibration of the peptide bond. Since this band is particularly sensitive to the protein secondary structures, it is very useful to get information on possible changes in the protein conformation and/or aggregation, even within complex systems.

The IR response of carbohydrates occurs mainly between 1,200 and 800 cm^{-1}, with absorption bands whose positions and intensities are specific for every polysaccharide (45). We should underline that the infrared response of glycogen, of interest for tissue and cell studies, is characterized by specific spectral features, with marker bands around 1,154 cm^{-1} (C–O stretching), 1,081 cm^{-1} (C–C stretching), 1,032 and 1,022 cm^{-1} (C–O–H bending) (5, 29, 33, 46, 47).

Concerning nucleic acids, we should consider that their complex IR absorption covers a wide range of frequencies (48–50). Among the most studied spectral region, of particular interest for cell biology is that between 1,250 and 1,000 cm^{-1}, characterized by absorption bands sensitive to nucleic acid backbone conformations. These spectral components arise mainly by vibrations along the sugar–phosphate chain, making it possible to find marker bands of the different DNA conformations (A, B, and Z). We should, however, note that this spectral range is particularly crowded, since different biomolecules absorb. For this reason, to characterize the nucleic acid response it could be useful to explore also the 1,000–800 cm^{-1} spectral region, which gives information mainly on nucleic acid sugar conformations. In particular, bands due to the different sugar puckering modes (S-N types)—varying depending on DNA cytosine methylation—could give information on this epigenetic modification, relevant in cell biology (51). Furthermore, spectral features typical of the different DNA conformations are also found in this region, in particular between 970 and 920 cm^{-1}, mainly due to the CC stretching of the backbone (5, 33, 48).

Acknowledgments

References

1. Orsini F, Ami D, Villa AM, Sala G, Bellotti MG, Doglia SM (2000) FT-IR microspectroscopy for microbiological studies. J Microbiol Methods 42:17–27

2. Shaw RA, Mantsch HH (1999) Vibrational biospectroscopy: from plants to animals to humans. A historical perspective. J Mol Struct 480–481:1–13

3. Heraud P, Tobin MJ (2009) The emergence of biospectroscopy in stem cell research. Stem Cell Res 3:12–14

4. Wood BR, Chernenko T, Matthäus C, Diem M, Chong C, Bernhard U et al (2008) Shedding new light on the molecular architecture of oocytes using a combination of synchrotron Fourier transform-infrared and Raman spectroscopic mapping. Anal Chem 80:9065–9072

5. Ami D, Neri T, Natalello A, Mereghetti P, Doglia SM, Zanoni M et al (2008) Embryonic stem cell differentiation studied by FTIR spectroscopy. Biochim Biophys Acta 1783:98–106

6. Walsh MJ, Hammiche A, Fellous TG, Nicholson JM, Cotte M, Susini J et al (2009) Tracking the cell hierarchy in the human intestine using biochemical signatures derived by mid-infrared microspectroscopy. Stem Cell Res 3:15–27

7. Kretlow A, Wang Q, Kneipp J, Lasch P, Beekes M, Miller L et al (2006) FTIR-microspectroscopy of prion-infected nervous tissue. Biochim Biophys Acta 1758:948–959

8. Ami D, Natalello A, Zullini A, Doglia SM (2004) Fourier transform infrared microspectroscopy as a new tool for nematode studies. FEBS Lett 576:297–300

9. Diomede L, Cassata G, Fiordaliso F, Salio M, Ami D, Natalello A et al (2010) Tetracycline and its analogues protect *Caenorhabditis elegans* from β amyloid-induced toxicity by targeting oligomers. Neurobiol Dis 40:424–431

10. Doglia SM, Ami D, Natalello A, Gatti-Lafranconi P, Lotti M (2008) Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. Biotechnol J 3:193–201

11. Kneipp J, Miller LM, Joncic M, Kittel M, Lasch P, Beekes M et al (2003) In situ identification of protein structural changes in prion-infected tissue. Biochim Biophys Acta 1639:152–158

12. Choo LP, Wetzel DL, Halliday WC, Jackson M, LeVine SM, Mantsch HH (1996) In situ characterization of beta-amyloid in Alzheimer's diseased tissue by synchrotron Fourier transform infrared microspectroscopy. Biophys J 71:1672–1679

13. Tanthanuch W, Thumanu K, Lorthongpanich C, Parnpai R, Heraud P (2010) Neural differentiation of mouse embryonic stem cells studied by FTIR spectroscopy. J Mol Struct 967:189–195

14. Kelly JG, Singh MN, Stringfellow HF, Walsh MJ, Nicholson JM, Bahrami F et al (2009) Derivation of a subtype-specific biochemical signature of endometrial carcinoma using synchrotron-based Fourier-transform infrared microspectroscopy. Cancer Lett 274:208–217

15. Schultz CP, Liu KZ, Johnston JB, Mantsch HH (1997) Prognosis of chronic lymphocytic leukemia from infrared spectra of lymphocytes. J Mol Struct 408:253–256

16. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366

17. Alberts B (2010) Model organisms and human health. Science 330:1724

18. Markaki M, Tavernarakis N (2010) Modelling human diseases in *Caenorhabditis elegans*. Biotechnol J 5:1261–1276

19. Moloney A, Sattelle DB, Lomas DA, Crowther DC (2010) Alzheimer's disease: insights from *Drosophila melanogaster* models. Trends Biochem Sci 35:228–235

20. Link CD (1995) Expression of human beta-amyloid peptide in transgenic *Caenorhabditis elegans*. Proc Natl Acad Sci USA 92:9368–9372

21. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99:303–310

22. de Groot NS, Sabate R, Ventura S (2009) Amyloids in bacterial inclusion bodies. Trends Biochem Sci 34:408–416

23. Dasari M, Espargaro A, Sabate R, Lopez Del Amo JM, Fink U, Grelle G et al (2011) Bacterial inclusion bodies of Alzheimer's disease β-amyloid peptides can be employed to study native-like aggregation intermediate states. Chembiochem 12:407–423

24. Arrondo JLR, Goni FM (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. Prog Biophys Mol Biol 72:367–405

25. Haris PI, Severcan F (1999) FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media. J Mol Catal B: Enzym 7:207–221

26. Barth A, Zscherp C (2002) What vibrations tell us about proteins. Q Rev Biophys 35:369–430

27. Ami D, Bonecchi L, Calì S, Orsini G, Tonon G, Doglia SM (2003) FTIR study of heterologous

protein expression in recombinant *Escherichia coli* strains. Biochim Biophys Acta 1624:6–10

28. Ami D, Natalello A, Gatti-Lafranconi P, Lotti M, Doglia SM (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett 579:3433–3436

29. Heraud P, Ng E, Sally Caine S, Yu Q, Hirst C, Mayberry R et al (2010) Fourier transform infrared microspectroscopy identifies early lineage commitment in differentiating human embryonic stem cells. Stem Cell Res 4:140–147

30. Chan JW, Lieu DK (2009) Label-free biochemical characterization of stem cells using vibrational spectroscopy. J Biophotonics 2:656–668

31. Ami D, Natalello A, Taylor G, Tonon G, Doglia SM (2006) Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. Biochim Biophys Acta 1764:793–799

32. Natalello A, Ami D, Doglia SM (2007) Protein aggregation studied in intact cells by Fourier transform infrared spectroscopy. In: Uversky VN, Permyakov EA (eds) Methods in protein structure and stability analysis: vibrational spectroscopy. Nova Science Publ Inc, Hauppage, NY

33. Ami D, Natalello A, Mereghetti P, Neri T, Zanoni M, Monti M et al (2010) FT-IR spectroscopy supported by PCA-LDA analysis for the study of embryonic stem cell differentiation. Spectroscopy 24:89–97

34. Notingher I, Bisson I, Polak JM, Hench LL (2004) In situ spectroscopic study of nucleic acids in differentiating embryonic stem cells. Vibr Sprectros 35:199–203

35. Notingher I, Bisson I, Bishop AE, Randle WL, Polak JM, Hench LL (2004) In situ spectral monitoring of mRNA translation in embryonic stem cells during differentiation in vitro. Anal Chem 76:3185–3193

36. Rencher AC (2002) Methods of multivariate analysis. Wiley, Hoboken, NJ

37. Ami D, Mereghetti P, Natalello A, Doglia SM (2011) Fourier transform infrared microspectroscopy as a tool for embryonic stem cell studies. In: Atwood C (ed) Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications. InTech, Rijeka, Croatia

38. Manly BFJ (2004) Multivariate statistical methods. Chapman & Hall/CRC, London, UK

39. Fearn T (2002) Discriminant analysis. In: Chalmers JM, Griffiths PR (eds) Handbook of vibrational spectroscopy. Wiley, New York

40. Konorov SO, Schulze HG, Caron NJ, Piret JM, Blades MW, Turner RFB (2011) Raman microspectroscopic evidence that dry-fixing preserves the temporal pattern of non-specific differentiation in live human embryonic stem cells. J Raman spectros 42:576–579

41. Dong A, Huang P, Caughey WS (1990) Protein secondary structures in water from second-derivative amide I infrared spectra. Biochemistry 29:3303–3308

42. Susi H, Byler DM (1986) Resolution-enhanced fourier-transform infrared-spectroscopy of enzymes. Methods Enzymol 130:290–311

43. Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P et al (2010) Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst 135:268–277

44. Casal HL, Mantsch HH (1984) Polymorphic phase behaviour of phospholipid membranes studied by infrared spectroscopy. Biochim Biophys Acta 779:381–401

45. Kacurakova M, Wilson RH (2001) Developments in mid-infrared FT-IR spectroscopy of selected carbohydrates. Carbohydr Polym 44:291–303

46. Wang TD, Triadafilopoulos G, Crawford JM, Dixon LR, Bhandari T, Sahbaie P et al (2007) Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. Proc Natl Acad Sci USA 104:15864–15869

47. Steller W, Einenkel J, Horn LC, Braumann UD, Binder H, Salzer R et al (2006) Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. Anal Bioanal Chem 384:145–154

48. Banyay M, Sarkar M, Graslund A (2003) A library of IR bands of nucleic acids in solution. Biophys Chem 104:477–488

49. Zhizhina GP, Oleinik EF (1972) Infrared spectroscopy of nucleic acids. Russ Chem Rev 41:258–280

50. Tsuboi M (1961) Infrared spectra and secondary structure of deoxyribonucleic acid. Progr Theor Phys Suppl 17:99–107

51. Banyay M, Graslund A (2002) Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR. J Mol Biol 324:667–676

Chapter 8

Studying IDP Stability and Dynamics by Fast Relaxation Imaging in Living Cells

Apratim Dhar, Maxim Prigozhin, Hannah Gelman, and Martin Gruebele

Abstract

Fast relaxation imaging (FReI) temperature-tunes living cells and applies small temperature jumps to them, to monitor biomolecular stability and kinetics in vivo. The folding or aggregation state of a target protein is monitored by Förster resonance energy transfer (FRET). Intrinsically disordered proteins near the structured–unstructured boundary are particularly sensitive to their environment. We describe, using the IDP α-synuclein as an example, how FReI can be used to measure IDP stability and folding inside the cell.

Key words: Temperature jump, Fluorescence, In vivo, Intrinsically disordered protein, Thermal denaturation, Folding kinetics, Folding thermodynamics

1. Introduction

The free energy differences between folded, unfolded, and transition states on a protein's folding landscape are very small compared to those of covalent bond-forming reactions. Hence, proteins are particularly sensitive to their environment. There is growing evidence that the natural folding environment of a protein inside the living cell shifts and modulates protein kinetics and thermodynamics compared to in vitro measurements. Experimental (1–10) and theoretical (3, 9, 11–17) investigations suggest that the discrepancy arises from a combination of chemical environment and macromolecular crowding. Crowding in particular reduces the entropy of extended, denatured, or disordered protein states.

Intrinsically disordered proteins (IDPs) present a particularly interesting subject for live-cell experiments because the crowded cellular environment can increase their propensity for folding (18) by lowering the entropy of the disordered state. IDPs that lie close to the borderline on the charge–hydropathy plot (19, 20) are an
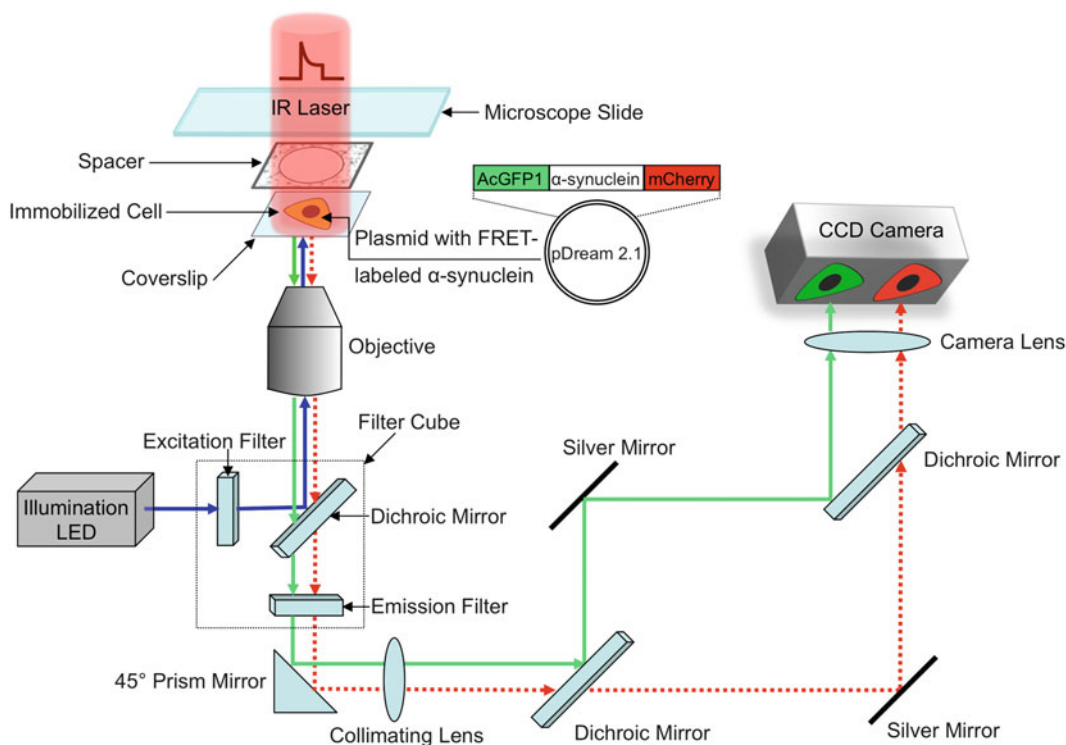
Fig. 1. A schematic diagram of the FReI microscope, showing various components and how two-color fluorescence images of live cells are obtained. Cells are transfected with a plasmid carrying the gene for α-synuclein FRET-labeled with AcGFP1 (FRET donor) and mCherry (FRET acceptor). A shaped pulse from an IR laser causes a rapid increase in temperature of the cells, causing the protein to unfold, and thereby triggering a change in the amounts of green and red fluorescence. This change is captured as a movie on the CCD camera and analyzed to extract thermodynamic and kinetics information about protein folding in live cells.

especially promising subgroup for such experiments. Even a slight perturbation of their energy landscape by macromolecular crowding could tip them over the folded–unfolded boundary, forcing such proteins to assume a biologically functional conformation that is difficult to access in vitro. For such proteins, test tube results could be misleading, accentuating the need for developing experimental methods suitable for the interrogation of protein folding, misfolding, and aggregation in its native environment—the living cell.

Fast Relaxation Imaging (FReI) is a new method that makes movies of fast protein dynamics inside living cells (21). FReI is a synergy of two techniques: Förster Resonance Energy Transfer (FRET) imaging of biomolecules, and temperature jump relaxation-induced biomolecular kinetics. Figure 1 shows the layout of our FReI microscope. For folding studies, a FReI sample consists of live cells expressing a recombinant protein of interest, sandwiched between two fluorescent proteins that comprise a FRET pair. For aggregation studies, the FReI sample consists of cells transfected with two proteins, each singly labeled (one by a donor and the other by an acceptor). The technique operates on the assumption that the average distance

between FRET pairs will change upon (un)folding or aggregation. Protein dynamics are measured by suddenly jumping the temperature of the cell by a few degrees with an infrared laser, and then monitoring FRET by imaging fluorescence in the donor (green) and acceptor (red) detection channels on a CCD camera.

Here, we describe experimental protocols specifically for using FReI to study the in-cell dynamics of α-synuclein, an IDP of tremendous clinical significance due to its role in the formation of insoluble fibrils that cause Parkinson's disease (22, 23). These protocols are also applicable to other IDPs. Details on how to set up FReI on an existing fluorescence microscope are given in refs. (21) and (24), which should be consulted before the protocols summarized below are implemented. Figure 1 illustrates the important additions to a conventional inverted fluorescence microscope: a temperature-controlled stage (custom-built or commercial), an infrared diode laser so the sample can be temperature jumped from above, and an external CCD camera with beam splitters, to record the donor and acceptor images of the cell separately for transfer to a computer and subsequent image processing. The filter box and dichroic mirrors that separate excitation light from fluorescence and the blue light source that excites the donor fluorescence are available as standard components for fluorescence microscopes.

2. Materials

The subsections have the same number as "Methods" subsections where the materials will be used.

2.1. Plasmid Assembly

1. pDream 2.1 vector (GenScript).
2. pAcGFP1 vector (Clontech).
3. pmCherry vector (Clontech).
4. Gene for α-synuclein (can be synthesized by several companies, such as GenScript).

2.2. Protein Expression and Purification

1. BL21 CodonPlus (DE3)-RIPL cells (Stratagene).
2. pDream 2.1 vector (GenScript) with FRET-labeled alpha-synuclein.
3. Lysogeny broth (LB) medium (Fisher Scientific).
4. Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Inalco).
5. Ni-NTA column (Qiagen).
6. Phenylmethylsulfonyl fluoride (PMSF, Sigma) may be stored for 1 week at 100 mM dissolved in ethanol or isopropyl alcohol.
7. DNase I (RNase-free), 2,000 units/mL (New England BioLabs).

8. Buffer A: 50 mM Na_2HPO_4, 300 mM NaCl, and 10 mM imidazole, pH 8.0.

9. Buffer B: 50 mM Na_2HPO_4, 300 mM NaCl, and 500 mM imidazole, pH 8.0.

2.3. Cell Culture and Transfection

1. Human osteosarcoma (U2OS) cells (ATCC). Other cell lines may be substituted.

2. DMEM media (HyClone).

3. OptiMEM media (Gibco).

4. Fetal Bovine Serum (FBS, Gibco).

5. Penicillin–streptomycin (PS, Invitrogen).

6. Lipofectamine 2000 (Invitrogen).

7. 35 mm Petri dishes (Corning).

8. Coverslips (Corning).

2.4. FReI Experiments

1. Coverslip (with cells expressing FRET-labeled alpha-synuclein adhered, see Subheading 3.3, step 7).

2. Microscope slide (Fisher).

3. Liebovitz L-15 media (Gibco).

4. FBS.

5. 100 μm spacer (Grace Bio Labs).

6. FReI microscope (based on an inverted fluorescence microscope, see refs. 21 and 24) with FCS2 chamber (Bioptechs).

2.5. Image and Data Analysis

7. LabVIEW (National Instruments).

8. MATLAB (MathWorks).

3. Methods

Unless otherwise specified, the procedures described below should be carried out at room temperature. The conversion of an inverted fluorescence microscope is not described in detail here; consult the references listed in Subheading 2.4, item 6 for details.

3.1. Design of α-Synuclein Protein Vectors

1. Assemble the gene for FRET-labeled α-synuclein by ligating the genes for AcGFP1 and mCherry to the 5′ and 3′ ends of the α-synuclein gene. A 6×His-tag should be added to the 5′ end of AcGFP1 to enable purification using a Ni-NTA column. This allows for comparison of in vivo results with in vivo results.

2. Clone this gene into the pDream 2.1 vector. This is a shuttle vector, meaning that the construct can be expressed in both mammalian cells (via the CMV promoter) as well as in *Escherichia coli* (via the T7 promoter) for in vitro studies. Companies such as GenScript offer services for the procedures listed in steps 1 and 2.

3.2. Protein Expression and Purification for In Vitro Comparison

1. Transform BL21 CodonPlus (DE3)-RIPL cells with the plasmid, as per the manufacturer's instructions. Plate the cells onto an agar plate containing LB and ampicillin and place the plate in an incubator at 37 °C for >12 h.

2. Pick a single colony from the agar plate using a sterile pipette tip and grow it in 20 mL of LB at 37 °C until the culture becomes cloudy. This generally takes 5–6 h.

3. Transfer the small culture to large flasks with LB media and grow the cells at 37 °C to an optical density at 600 nm of 0.8.

4. Add IPTG to the cells (1 mM final concentration) to induce protein expression and continue growing at 23 °C for 12 h.

5. Collect the cells by centrifugation 8,600 $\times$ g for 10 min at 4 °C and discard the supernatant.

6. Resuspend the cells in Buffer A and add PMSF (1 mM final concentration) and DNase (20 μL/liter culture expressed). Let sit for 20 min in ice and then lyse the cells using a sonicator.

7. Centrifuge the cell lysate at 20,000 $\times$ g for 25 min at 4 °C and collect the supernatant.

8. Filter the supernatant through a .22 micron syringe filter. Pass the supernatant over an Ni-NTA column charged according to the manufacturer's protocol and equilibrated with Buffer A.

9. Wash the column with 3 column volumes each of 20, 50, 100, 150, 200, 300, and 500 mM imidazole. Mix Buffer A and B in the appropriate ratios to achieve the desired imidazole concentrations.

10. Collect eluted buffer in fraction volumes equal to the column volume (or smaller).

11. Analyze all fractions by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

12. Combine the fractions containing FRET-labeled α-synuclein and dialyze overnight at 4 °C to remove remaining impurities.

13. Check final protein purity by electrospray ionization mass spectrometry (ESI-MS) and SDS-PAGE.

3.3. Transfection of Mammalian Cells for FReI Experiments

1. Grow U2OS cells (or another desired cell line) in 35 mm Petri dishes in DMEM containing 1 % PS and 10 % FBS to >70 % confluency.

2. 30 min before transfection, change the media in the Petri dishes to DMEM media supplemented only with 10 % FBS (i.e., containing no PS).

3. Prepare the following solutions (it is critical to maintain a ratio of 1:2.5 for the DNA mass to Lipofectamine volume concentration):

   (a) Tube A: 250 μL of Optimem I media + 2 μg plasmid DNA

   (b) Tube B: 250 μL of Optimem I media + 5 μL Lipofectamine

4. Tap gently to mix the contents of the two tubes and incubate for 5–20 min at room temperature.

5. Pipette the contents of tube B into A, mix by tapping gently, and incubate for 20–30 min at room temperature.

6. Pipette the contents drop-wise onto the cells.

7. After 5–6 h, aspirate the media in the petri dish and replace with fresh DMEM media (supplemented with 1 % PS and 10 % FBS). At this point, the cells can be split and plated onto coverslips for subsequent imaging experiments. Experiments can be performed 24–72 h posttransfection.

3.4. FRet Measurements in Cells and in In Vitro

1. Assemble the FCS2 chamber with the coverslip carrying the cells expressing FRET-labeled alpha-synuclein and fill the chamber with Liebovitz L-15 media supplemented with 30% FBS. Alternatively, a custom-built temperature-controlled stage may be used (see Notes 1 and 2).

2. Mount the FCS2 chamber (or custom stage) on the microscope stage.

3. Turn on the blue light source (e.g., LED in Fig. 1) for excitation of the donor FRET label.

4. Focus on a cell through the eyepiece and then direct the green and red fluorescence to the CCD by pulling the slider on the microscope. Visualize the cell on the computer monitor via camera software. CCD optics assembly is described in ref. (24).

5. Set the base temperature for the T-jump experiment by using the FCS2 temperature controller or a custom heating plate (i.e., to do a T-jump from 35 to 39 °C, set the heating plate to 35 °C).

6. Acquire a video of protein fluorescence from the cell imaged onto the CCD camera and transfer to a computer using software provided by the CCD manufacturer or custom-written imaging software (e.g., written in LabVIEW).

7. During video acquisition, measure a constant temperature signal for a few seconds, and then activate the shaped IR pulse to initiate a T-jump (Fig. 1 and ref. 24). The constant temperature measurement monitors thermal equilibrium, while the change in the green donor intensity D and red acceptor intensity A after the jump monitors the α-synuclein conformational dynamics.

8. Repeat steps 5–7 to obtain data at 10–12 different temperatures so that a full set of thermodynamic and kinetic parameters can be reliably determined during data analysis.

9. For in vitro measurements, prepare a slide by adding ~500 μL of the purified and 0.2 μm filtered protein (concentration of ~5 μM) to the FCS2 imaging chamber. Alternatively, use an imaging chamber (described in Note 2, below), filled with ~150 μL of the purified protein. The protein concentration
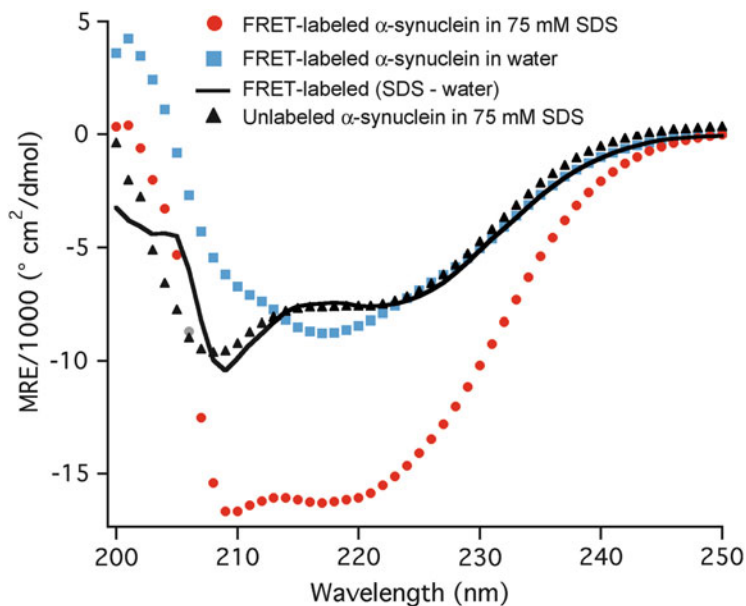
Fig. 2. Circular dichroism spectra (mean residue ellipticity = MRE) of FRET-labeled α-synuclein. In water (*blue squares*), the construct only shows the β-sheet signature contributed by the FRET labels. In 75 mM SDS (*red circles*), the construct has a much larger CD spectrum. The difference spectrum (*black line*) reveals that FRET-labeled α-synuclein contains coil and α-helical regions when bound to SDS, similar to unlabeled α-synuclein in SDS (*black triangles*).

should be chosen such that the in vitro fluorescence intensity is similar to that collected from cells.

10. Follow steps 2–8 above to obtain in vitro thermodynamics and kinetics analogous to the in cell data.

11. For in vitro analysis, it may also be useful to collect biophysical data as a function of temperature, such as CD spectra of FRET-labeled α-synuclein (see Fig. 2). Such in vitro data can be correlated with both in vivo and in vitro FRET data.

3.5. Image and Data Analysis

1. Open the acquired video using LabVIEW or other image processing software. The output from the CCD consists of two images of the cell in every frame, corresponding to the green and red fluorescence images (Fig. 1). The images may differ slightly in orientation and magnification, and require correction.

2. Select the green and red channel images in the two-color fluorescence image.

3. Align the two images by computing the cross-correlation function between the two selections, and rotate, scale, and translate one selection to maximize the cross-correlation (see Note 3).

4. Save the position of every pixel, its intensity D in the green donor channel, and its intensity A in the red acceptor channel,
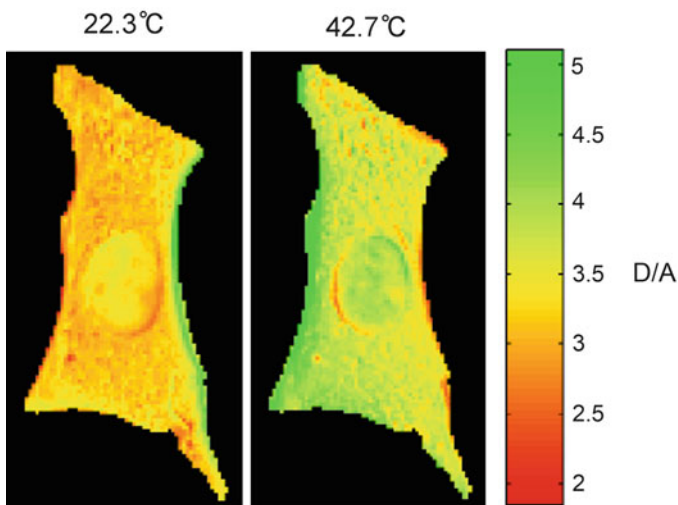
Fig. 3. Typical analysis output images of a cell with FRET-labeled α-synuclein at two different temperatures. At 22.3 °C (*redder image*), *D/A* is on average lower than at 49 °C (*greener image*), which suggests that the conformation of α-synuclein within the cell changes and FRET labels spatially separate as the temperature increases. By averaging the *D/A* (*green/red*) fluorescence intensity ratio over the cell, a cell-averaged FRET ratio can be obtained and plotted as a function of temperature. The *red and green rims* at the edge of the cell are due to temperature-dependent fluorescence interference because the slide, cell, and media differ in refractive index (25).

    for every time frame in the acquired video in a text file (see Note 4).

5. For in vitro data analysis, save an area similar to that of a cell in the green and red channels and proceed in the same manner as for in-cell data.

6. Cell-averaged analysis: Sum up the intensities of the green channel to get an overall D value for every frame. Sum up the intensities for the same pixels in the red channel to get an overall A value. This can be done in MATLAB or similar software.

7. Cell-averaged thermodynamic analysis

    (a) Compute the D/A ratio from the overall D and A values for all frames collected before each T-jump (Fig. 3). Average the values before each T-jump together to improve signal-to-noise ratio. If n T-jumps were measured starting at different temperatures, this yields n values of D/A at different temperatures. Normalize all D/A values by the value of D/A obtained at the lowest temperature. The result is a plot of D/A as a function of temperature, which starts with $D/A = 1$ at the lowest temperature (see Note 5).

(b) Determine the melting temperature T_m and the cooperativity of the unfolding transition $\Delta G^{(1)}$ by fitting the normalized D/A ratio to a sigmoid with the following functional form:

$$D/A = \frac{(1 - S_F) + (S_F - S_U)F_U}{S_F - (S_F - S_U)F_U}$$

$$\Delta G(T) = \Delta G^{(1)}(T - T_m), \quad F_U = \frac{1}{1 + e^{-\Delta G(T)/RT}}$$

Linear baselines $S_F = 0.5 + b(T - T_{min})$, $S_U = a + b(T - T_{min})$ for the folded and denatured states, commonly assumed in fitting protein denaturation data, can also be used here as a good approximation. T_{min} is the lowest temperature in the titration. F_U is the fraction of unfolded protein, and is given by the free energy difference ΔG between the folded and denatured states.

8. Cell-averaged kinetic analysis

(a) Determine $x = D/A$, where D and A are the unnormalized average donor and acceptor intensities before the T jump, as determined at the beginning of step 7(a).

(b) Use the cell averaged D and A values from step 6 above. Plot the difference $D - xA$ for each time point after the T-jump versus time to obtain a trace of α-synuclein kinetics averaged over the entire cell (see Note 6).

(c) Fit the pre-jump $D - xA$ values to a linear function to correct for any differential photobleaching effects in the two channels. Subtract this linear function from the $D - xA$ values to obtain the final, processed data to be fitted (see Note 6).

(d) Fit the data to the desired kinetic model. For example, fit to a stretched exponential:

$$y = y_0 + A(1 - e^{-(t/\tau)^\beta}),$$

where τ is the time constant and β is the stretching factor.

4. Notes

1. If using the commercial FCS2 chamber, add a drop of 1 M HCl to the center of the microaqueduct slide to dissolve the thermally conducting ITO (indium tin oxide) coating, so that the IR T-jump beam can be transmitted from above through the slide to the cells on the coverslip.

2. If using a custom-built stage (24), appropriate imaging chambers can be assembled by putting a 100 μm adhesive spacer (Grace Bio Labs) on a microscope slide, filling it with 150 μL of

L-15 media (supplemented with 30 % FBS), and then sealing the chamber by sticking the coverslip with cells onto the spacer.

3. Such alignment is provided by many image processing programs. The goal is to make sure that each pixel in the green image can be matched to a corresponding pixel in the red image for further analysis.

4. The text file can be read into MATLAB or similar software for future thermodynamic and kinetic analysis with pixel resolution. Here, we describe only an average analysis for the whole cell, where D and A are obtained by summing green or red channel intensities over all pixels of the cell.

5. This curve is the equivalent of thermal protein denaturation curves obtained in vitro. For example, if the protein is unfolding at higher temperature, one expects D/A to vary slowly at low T, suddenly increase as the protein unfolds and the green–red label distance increases, and then level off again.

6. For the most accurate thermodynamic and kinetic analysis, $A(T)$ (or $A(t)$ before and after the T-jump) should be scaled by the temperature-dependent quantum efficiency of the acceptor dye. This can be measured by exciting the acceptor directly with a light source at the peak donor emission wavelength, and measuring its fluorescence intensity $A_0(T)$ as a function of temperature. In practice, we find that $A_0(T)$ varies sufficiently slowly with temperature for mCherry that omitting this step does not introduce a significant error ($< 0.5\ ^\circ C$) in T_m or in the kinetic analysis.

References

1. Roque A, Ponte I, Suau P (2007) Macromolecular crowding induces a molten globule state in the C-terminal domain of histone H1. Biophys J 93:2170–2177

2. Perham M, Stagg L, Wittung-Stafshede P (2007) Macromolecular crowding increases structural content of folded proteins. FEBS Lett 581:5065–5069

3. Samiotakis A, Wittung-Stafshede P, Cheung MS (2009) Folding, stability and shape of proteins in crowded environments: experimental and computational approaches. Int J Mol Sci 10:572–588

4. Miklos AC, Li CG, Sharaf NG, Pielak GJ (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions. Biochemistry 49:6984–6991

5. Sakon JJ, Weninger KR (2010) Detecting the conformation of individual proteins in live cells. Nat Methods 7:203–205

6. Johnson CP, Tang HY, Carag C, Speicher DW, Discher DE (2007) Forced unfolding of proteins within cells. Science 317:663–666

7. Weiss M, Elsner M, Kartberg F, Nilsson T (2004) Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. Biophys J 87:3518–3524

8. Eggers DK, Valentine JS (2001) Crowding and hydration effects on protein conformation: a study with sol-gel encapsulated proteins. J Mol Biol 314:911–922

9. Minton AP (2006) How can biochemical reactions within cells differ from those in test tubes? J Cell Sci 119:2863–2869

10. Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, Gruebele M et al (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. Proc Natl Acad Sci USA 107:17586–17591

11. Cheung MS, Thirumalai D (2007) Effects of crowding and confinement on the structures of the transition state ensemble in proteins. J Phys Chem B 111:8250–8257

12. Kim YC, Best RB, Mittal J (2010) Macromolecular crowding effects on protein-protein binding affinity and specificity. J Chem Phys 133:2051011–2051017

13. Tsao D, Minton AP, Dokholyan NV (2010) A didactic model of macromolecular crowding effects on protein folding. PLoS One 5

14. Kim JS, Yethiraj A (2009) Effect of macromolecular crowding on reaction rates: a computational and theoretical study. Biophys J 96:1333–1340

15. Zhou HX, Rivas GN, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys 37:375–397

16. Zimmerman SB, Minton AP (1993) Macromolecular crowding – biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct 22:27–65

17. Kinjo AR, Takada S (2002) Effects of macromolecular crowding on protein folding and aggregation studied by density functional theory: statics. Phys Rev E 66:0319111–0319119

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

19. Uversky VN, Dunker AK (2010) Understanding protein non-folding. Biochim Biophys Acta 1804:1231–1264

20. Dunker AK, Oldfield CJ, Meng JW, Romero P, Yang JY, Chen JW et al (2007) The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9:1–26

21. Ebbinghaus S, Dhar A, McDonald D, Gruebele M (2010) Protein folding stability and dynamics imaged in a living cell. Nat Methods 7:319–323

22. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A et al (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047

23. Auluck PK, Caraveo G, Lindquist S (2010) a-Synuclein: membrane interactions and toxicity in Parkinson's disease. Annu Rev Cell Dev Biol 26:211–233

24. Dhar A, Gruebele M (2011) Fast relaxation imaging in living cells. Curr Protoc Protein Sci 65:28.1.1–28.1.19

25. Groves JT, Parthasarathy R, Forstner M (2008) Fluorescence imaging of membrane dynamics. Annu Rev Biomed Eng 10:311–338

Part II

NMR-Based Techniques

Chapter 9

Measurement and Analysis of NMR Residual Dipolar Couplings for the Study of Intrinsically Disordered Proteins

Loïc Salmon, Malene Ringkjøbing Jensen, Pau Bernadó, and Martin Blackledge

Abstract

Intrinsically disordered proteins (IDPs) are predicted to represent a significant fraction of all functional proteins. Their inherent plasticity allows them to sample more efficiently their surroundings and thereby increase the probability of interaction with one or several different biological partners. Due to their high flexibility, IDPs cannot be represented by a single, three-dimensional structure; rather, an ensemble description can be invoked, where the protein is assumed to interconvert between different conformations. This chapter focuses on the use of NMR spectroscopy to characterize the dynamic behavior of IDPs, in particular residual dipolar couplings, that provide highly sensitive tools for the study of intrinsic structural propensity and conformational transitions accompanying protein function.

Key words: Nuclear magnetic resonance, Intrinsically disordered protein, Residual dipolar coupling, Statistical coil, Unfolded protein

1. Introduction

Intrinsically disordered proteins (IDPs) are now recognized to represent a significant fraction of all functional proteins. Even though this fraction is limited to a few percent in prokaryotic organisms, it is estimated to reach more than 30 % for eukaryotic systems and even 40 % of the human proteome (1–3). The biological functions exerted by IDPs are very broad including cell cycle regulation, signaling, translation, and transcription and play key roles in neurodegenerative diseases and cancer (4).

The inherent plasticity of IDPs allows them to sample more efficiently their surroundings and thereby increase the probability of interaction with one or several different biological partners.

They often fold upon binding, although high flexibility can remain in the formed complexes (5). IDPs have large interaction surfaces, without requiring excessively high molecular weights, and can be efficiently regulated via high turnover rates. Due to their highly flexible nature, these proteins have until now escaped detailed biophysical characterization and it has become clear that the very existence of this class of proteins necessitates a reassessment of the classical structure–function paradigm. IDPs cannot be represented by a single, three-dimensional structure. Therefore, their biophysical properties have to be described using tools allowing for such flexible behavior. An ensemble description can be invoked, where the protein is assumed to interconvert between more or less equiprobable conformations. In general, descriptions based on probabilistic formalisms are expected to be more suitable for describing the properties of IDPs compared to standard structural approaches (6, 7).

Considering IDPs as completely unfolded systems (i.e., as random-coil states) is an oversimplification as many IDPs contain significant amounts of residual structure. This residual structure has been shown to be essential for controlling early molecular recognition events in IDPs undergoing disorder-to-order transitions upon binding to physiological partners (8). Therefore, many recent IDP developments have focused on characterizing such residual structure at atomic resolution.

The present chapter focuses on the use of nuclear magnetic resonance (NMR) spectroscopy to characterize the dynamical behavior of IDPs and other unfolded states (9). In particular we will discuss residual dipolar couplings, and emphasis will be put on their interpretation using a recently developed ensemble description of the disordered state, the so-called *Flexible-Meccano*.

Residual dipolar couplings (RDCs) are sensitive probes of biomolecular structure and dynamics. In principle this information exists both in folded and unfolded systems as long as RDCs can be measured (10–12).

The dipolar or dipole–dipole interaction results from the coupling of a given magnetic moment with any other magnetic moment in its surroundings, and this coupling is given by:

$$D_{IS} = -\frac{\gamma_I \gamma_S \mu_0 h}{16\pi^3 r_{IS}^3} \langle P_2 \cos(\theta_{IS}) \rangle \tag{1}$$

where γ is the gyromagnetic ratio for the two spins I and S, r_{IS} is the distance between the spins, μ_0 is the permeability of free space, and h is Planck's constant. Two cases can be envisaged: the isotropic case, such as that found in the liquid state, where interactions sample all orientations in free solution with equal probability. The measured value is then averaged to zero. The opposite occurs for the solid state case, where the absence of macroscopic motion leaves the dipolar interaction undiminished. The idea of using a mesomorphic or

liquid crystalline phase is to find a regime between these two extremes, where one can retain the essential geometric information content in a *residual* dipolar coupling that can be measured using the simplicity of liquid state NMR (13).

Residual dipolar couplings are therefore measured by dissolving the protein in a dilute liquid crystalline mixture. The diversity of dilute liquid crystals proposed to align biological samples is nowadays quite vast, but the proposed alignment systems behave principally as nematic or smectic phases (14). Schematically, nematic phases, named from the Greek word for thread, are rod shaped and usually exhibit ordered behavior in a single direction, and smectic phase, from the Greek word soap, are more ordered than the nematic as they present layered structures. Most of the alignment media make lyotrophic phases diagram is mainly dependent on two parameters, the temperature (as for thermotrophic ones) and the concentration of the chemicals that will build the structure into the solvent.

1.1. Alignment Media The first dilute liquid crystal used for aligning proteins was made of a mixture of two phospholipids: one long the DMPC (dimyristoylphosphatidylcholine) and one short the DHPC (dihexanoyl-phosphatidylcholine) which form disks with diameters around 400 Å (15). They are called bicelles, and their surface is constituted mainly from the long alkyl chain phospholipid and the border mainly from the short chains. Charged groups point into the water, whereas the aliphatic chains constitute the bulk of the bicelle. The alignment, which occurs at a temperature around 35 °C or above, is induced by the diamagnetic susceptibility of the bicelles, and their director is orthogonal to the magnetic field (13, 16). They mainly interact with the solute through steric interactions (17). Different compositions have now been proposed to extend the use of bicelles over a wider range of pH, temperature or to increase their time stability.

Bicelles can also be electrostatically charged using CTAB (hexadecyltrimethylammonium bromide) or SDS (sodium dodecyl sulfate) in order to switch to positive or negative electrostatic interaction with the solute (18, 19) or doped with lanthanides that will flip the director parallel to the magnetic field (20).

Another common alignment medium is filamentous bacteriophage such as fd and tobacco mosaic virus (TMV) (21) or more often the Pf1 (22) that presents an elongated rod shape of 20,000 Å long with a diameter of 60 Å. They are made of DNA encapsulated by negatively charged α-helical proteins and align through their intrinsic magnetic susceptibility apparently due to a repetitive distribution of planar peptide bonds in the shell protein. The interaction with a solute will be mainly due to negatively charged electrostatic interaction. The electrostatic interaction can be reduced by increasing the ionic strength of the solution, thereby avoiding attractive interaction, and binding, with positively charged solutes (23).

The mixture of {n}-alkyl-poly(ethylene glycol) (PEG) and {n}-hexanol provides another source of alignment forming lyotropic lamellar phases (24). Layers are organized with a hydrophobic interior, and PEG polar head-group in contact with water. They align parallel to the magnetic field and the spacing between layers can be tuned from several hundred to a few nanometers by increasing the concentration (16). Variations using different alkyl chains or by substituting the PEG with cetylpyridinium chloride or bromide have been proposed (25, 26). PEG–hexanol mixtures are cheap, while the large accessible range of pH and temperature makes them particularly attractive for studying a broad range of proteins.

Other liquid crystal based alignment media have been proposed, such as negatively charged purple membrane fragments containing bacteriorhodopsin (27, 28), cellulose microcrystallites that form rods of approximately 2000 Å, long with a diameter of 100 Å, that align orthogonally to the magnetic field (29) or stacked columns of d(GpG) (2′-deoxyguanylyl-(3′,5′)-2′-deoxyguanosine) which are compatible with detergents (30), and collagen (31).

A common form of alignment medium that is used for biomolecular systems results from mechanically straining polyacrylamide gels. They are obtained by compressing or stretching polyacrylamide gels and they can be obtained for example by compressing a gel in a Shigemi tube by pressure of the plunger or by reswelling dried gel in an NMR tube whose diameter is smaller than the one where the gel was polymerized (32, 33). This mechanical strain creates anisotropic cavities in the gel. The alignment is transferred to the solute through orientational restriction: if the molecule has for example an ellipsoidal shape, the main axis will have a higher probability to lie along the longest direction of the cavity than the shortest one.

Gels can be charged during polymerization by including positively or negatively charged or zwitterionic monomer instead of neutral ones (34). This will modify the alignment properties of the medium by introducing electrostatic interactions. Gels can also be polymerized in the presence of phages or purple membranes, for example in order to give composite alignment media (35, 36).

For completely unfolded proteins, the existence of nonzero RDCs was not immediately intuitive, as the flexibility of the peptide chain was thought to potentially average the angular dependence of the RDCs to zero. However, it has become clear that RDCs in unfolded systems adopt bell-shaped curves with vanishing values at the extremities and nonzero mean values in the center of the chain (37). This can be rationalized for a steric alignment medium by considering an internuclear vector, e.g., an NH bond, in the center of the chain and one at the extremity. The alignment can be estimated by finding all possible bond vector orientations in the presence of the alignment medium and compare it to those in the absence of the medium (this is the principle underlying the

PALES approach). In the absence of the medium the internal and external bond vectors have the same accessible space. Therefore, the vector with the most restricted available space when the medium is introduced will be more aligned and will exhibit larger RDCs values. The probability to find a conformation where the external amide NH vector is close to the alignment medium elements (e.g., a bicelle) is much higher than for the internal vector and thus the central part of the chain should exhibit larger RDCs.

This qualitative reasoning has been more precisely studied by using a random walk model where the peptide chain is modeled as rigid segments (37). This approach has provided an analytical expression of the bell-shaped curve that gives reasonable agreement with experimental data (38). A major drawback of this approach is that each amino acid is treated in the same way and therefore does not allow different conformational propensies for each amino acid. In order to obtain a quantitative description of the unfolded state, it is essential to take into account amino-acid specific conformational propensies.

The deviation from the completely unfolded state can be monitored as modulations of this bell-shaped curve. For example, studies of the denatured states of apomyoglobin and acyl-CoA binding protein reveal deviations of the NH couplings from their negative bell-shaped profile (39). In some cases the NH values even become positive showing that the proteins contain significant amounts of residual structure in the denatured state. The observed change in sign can be explained because the mathematical function shown in Eq. 1 changes sign when the average orientation changes by 90°, as one would expect when comparing NH bond vectors that are perpendicular (extended) or parallel (α-helical) to the direction of the chain (Fig. 1).

It is worth noting that RDCs for unfolded proteins are often measured in steric alignment media because tensor estimation is relatively simple in these cases, relying on the overall shape of the molecule. Tensor estimation is more complicated in electrostatic alignment media as the distribution of charges in the protein has to be taken into account. Furthermore, as IDPs often exhibit a high proportion of charged residues, strong interactions with electrostatic alignment media may be more difficult to avoid.

2. Materials

In this section the experimental protocols for preparing two different alignment media are presented: PEG–hexanol mixture and bicelles. The volumes of the samples in both protocols are 250 μL that corresponds to the standard volume for
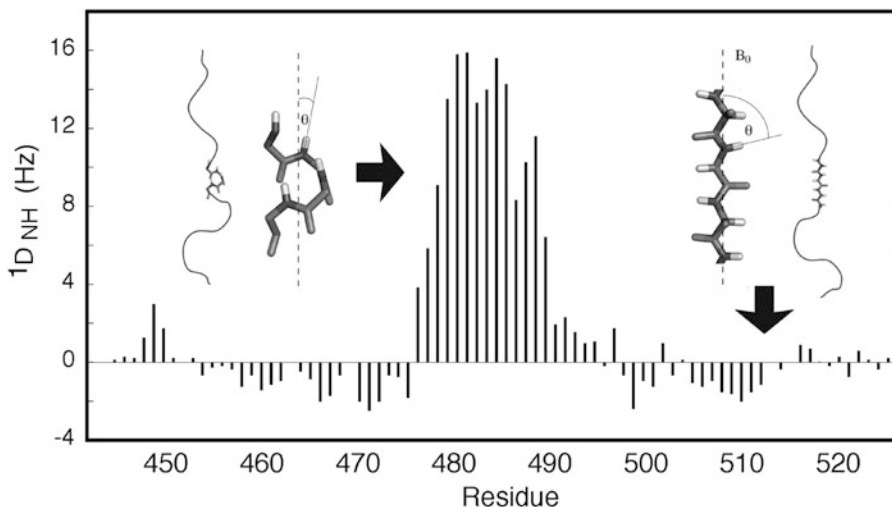
Fig. 1. NH RDCs become positive when the average orientation of the bond vector is parallel to the direction of the chain (when the residual structure is preferentially α-helical) and negative when the NH bond vectors are preferentially perpendicular to the direction of the chain (when there is no residual structure).

measurements in Shigemi tubes. The amount of reactants can be increased or decreased to modulate the level of alignment. The quantities given here are standard for aligning small to medium sized proteins.

2.1. PEG–Hexanol

1. Solution of the protein and corresponding buffer.
2. D_2O.
3. PEG: *n*-alkyl-poly(ethylene glycol) (pentaethyleneglycol monododecyl ether).
4. Hexanol.

2.2. Bicelles

1. Solution of the protein and corresponding buffer.
2. 10mM Sodium phosphate buffer.
3. D_2O.
4. Detergent: dihexyl-PC (1,2-di-*O*-hexyl-*sn*-glycero-3-phosphocholine).
5. Lipid: ditetradecyl-PC (1,2-di-*O*-tetradecyl-*sn*-glycero-3-phosphocholine).
6. CTAB (hexadecyltrimethylammonium bromide) or SDS (sodium dodecyl-sulfate) for doping.

2.3. Flexible-Meccano

The program can be downloaded at the following site: http://www.ibs.fr/science-213/scientific-output/software/flexible-meccano/.

3. Methods

3.1. PEG–Hexanol

In a microcentrifuge tube, prepare a sample of 236 μL containing the protein in the appropriate concentration (typically 0.2–1 mM) and 20 μL of D_2O.

1. Add 14 μL of PEG (see Note 1).

2. Add step by step the hexanol. As this step is the most critical for obtaining a good alignment, it is recommended to add the hexanol in steps of 0.5 μL. After each addition, vortex vigorously the sample (see Note 2).

3. Introduce the sample in the NMR tube (see Notes 3–5).

4. The level of alignment can be monitored by measuring the deuterium quadrupolar splitting. For a PEG–hexanol mixture the expected splitting is between 15 and 25 Hz (see Note 6).

3.2. Bicelles

1. Prepare a buffer solution of 5 mM sodium phosphate (NaH_2PO_4/Na_2HPO_4) at pH 6.0.

2. Prepare a stock solution of 15 % (mass/volume) of bicelles. The bicelles are made of lipid and detergent in a molar ratio of 3 to 1. As both lipid and detergent are solid, this solution can be prepared by dissolving 500 mg of lipid and 109.1 mg of detergent in 4.06 mL of buffer (see Notes 7–10).

3. Prepare in a microcentrifuge tube a solution of 166 μL containing the protein in the proper concentration and 20 μL of D_2O.

4. Add 84 μL of the bicelle solution to reach a final 5 % (mass/volume) of bicelles. Due to the viscosity of the bicelle stock solution, vortex the sample to properly mix the two fluids.

5. Introduce the sample in the NMR tube as described for the PEG–hexanol aligned sample.

6. The sample needs some time to equilibrate and the expected D_2O splitting is around 20 Hz.

3.3. Flexible-Meccano

Flexible-Meccano is an ensemble-based description of the unfolded state that consists of generating a large ensemble of structures whose average properties represent the unfolded state (40).

1. The backbone ϕ/ψ angles are selected in a database that is assumed to describe the unfolded state. This database is amino-acid specific, with some additional special cases such as amino-acids preceding prolines that are treated as specific amino-acids due to their restricted sampling. This database is an important feature of the model, and the first database was constructed by including all ϕ/ψ combinations adopted by residues in loop regions of 500 high resolution (<1.8 Å) X-ray structures (41).

2. A conformation, i.e., a ϕ/ψ combination, can be selected only if it does not generate steric clashes between spheres centered on the Cβ atoms.

3. For each conformer the desired quantities, e.g., RDCs, are calculated and averaged over all the members of the ensemble. RDCs are estimated using a static description, with the conformational disorder of the unfolded protein represented by the ensemble of states, and the tensor properties are determined using PALES ([42]). When dipolar couplings are averaged over a sufficient number of conformers (dependent on the nature/sequence of the peptide chain under consideration, this can range from many hundreds to many tens of thousands), they are assumed to fully sample the available conformational space, and the average is calculated. It has been demonstrated that convergence of RDCs towards experimental data improves when the protein is divided into small, uncoupled segments, but this does not allow the chain to be considered as an entire molecule ([43]).

A random coil description of intrinsically disordered or unfolded proteins as presented above provides a tool for calculating expected sterically induced RDC. This assumes that the protein is accurately described as a random coil, obeying local structural propensities and devoid of persistent local or long-range structure. Establishment of these approaches is essential for the identification of specific local or long-range conformational behavior. Significant efforts have been made in this direction, resulting in a quantitative description of conformational detail in regions of IDPs that exhibit local structure ([44], [45]). The *Flexible-Meccano* approach has indeed been successfully applied to various systems. The first application was the nucleocapsid-binding domain of Sendai virus phosphoprotein (PX) ([40]). Its biological partner Ntail, the C-terminal domain of the Sendai virus nucleoprotein, has also been investigated using *Flexible-Meccano*-based approaches ([46]–[48]).

The protein α-synuclein, was also investigated, where long-range contacts were necessary to reproduce experimental RDCs ([49], [50]), and during the study of the Tau protein, the combination of *Flexible-Meccano* and molecular dynamics simulations allowed the identification of β-turns in the vicinity of the functionally important microtubule binding, and aggregation nucleation sites ([51]). It was also applied to urea-denatured ubiquitin, where multiple RDCs measured for each peptide unit clearly indicated that the statistical coil model of the unfolded state needed to be refined to take into account the more extended nature of backbone conformational sampling in the presence of urea ([52], [53]).

4. Notes

1. As PEG tend to solidify at low temperature, it may be necessary to warm it up, for example in a water bath, to obtain a clear, transparent solution that will be easier to manipulate. After adding PEG the sample must be limpid.

2. When adding hexanol to PEG: During the addition the visual aspect of the sample will clearly change. After a few additions the sample becomes turbid and adding further hexanol will give a crystal clear solution. This solution corresponds to the desired lyotropic lamellar phase.

3. As the PEG–hexanol mixtures are relatively viscous, a centrifugation of the tube might be needed to get the entire sample to the bottom of the tube.

4. If you are using a Shigemi NMR tube for PEG–hexanol mixtures, introduce the plunger and try to avoid the formation of bubbles that will decrease the homogeneity of the sample and fix the plunger with parafilm.

5. If you are using a standard 5 mm NMR tube for PEG–hexanol mixtures, it is recommended to prepare a sample of around 450 μL.

6. The PEG–hexanol mixtures needs some time to get stable alignment in the magnetic field (typically 1 h).

7. To improve homogeneity of bicelle preparations: the solution has to be heated, vortexed, and cooled down several times.

8. The stock solution can be stored for several months at 4 °C.

9. In order to provide electrostatic alignment, the bicelles can be doped using CTAB to obtain positively charged bicelles. The protocol is identical, but the stock solution has to be prepared in a lipid-detergent-CTAB ratio of 30:10:1.

10. In order to provide negatively charged electrostatic alignment, the bicelles can be doped using SDS. The protocol is identical, but the stock solution has to be prepared in lipid–detergent–SDS ratio of 30:10:1.

References

1. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. Biochemistry 41:6572–6582

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

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

4. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human disease: introducing the D2 concept. Annu Rev Biophys 37:215–246

5. Wright PE, Dyson HJ (2009) Linking holding and binding. Curr Opin Struct Biol 19:31–38

6. Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 19:23–30

7. Bernadó P, Blackledge M (2010) Structural biology: proteins in dynamic equilibrium. Nature 468:1046–1048

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

9. Dyson HJ, Wright PE (2004) Unfolded proteins and protein holding studied by NMR. Chem Rev 104:3607–3622

10. Prestegard JH, Al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33:371–424

11. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

12. Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual bipolar couplings. Prog Nucl Magn Reson Spectrosc 46:23–61

13. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

14. Prestegard JH, Bougault CM, Kishore AI (2004) Residual bipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540

15. Sanders CR II, Schwonek JP (1992) Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine. Biochemistry 31:8898–8905

16. Gaemers S, Bax A (2001) Morphology of three lyotropic liquid crystalline biological NMR media studied by translational difussion anisotropy. J Am Chem Soc 123:12342–12352

17. Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 122:3791–3792

18. Losonczi JA, Prestegard JH (1998) Improved dilute solutions for high-resolution NMR of biological macromolecules. J Biomol NMR 12:447–451

19. Ramírez BE, Bax A (1998) Modulation of the alignment tensor of macromolecules dissolved in a diluted crystalline medium. J Am Chem Soc 120:9106–9107

20. Prosser RS, Hwang JS, Vold RR (1998) Magnetically aligned phospholipid bilayers with positive ordering: a new model membrane system. Biophys J 74:2405–2418

21. Clore GM, Starich MR, Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. J Am Chem Soc 120:10571–10572

22. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields bipolar coupling interactions. Nat Struct Biol 5:1065–1074

23. Zweckstetter M, Bax A (2001) Characterization of molecular alignment in aqueous suspension of Pf1 bacteriophage. J Biomol NMR 20:365–377

24. Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc 122:7793–7797

25. Prosser RS, Losonczi JA, Shiyanovskaya IV (1998) Use of a novel aqueous liquid crystalline médium for high resolution NMR of macromolecules in solution. J Am Chem Soc 120:11010–11011

26. Barrientos LG, Dolan C, Gronenborn AM (2000) Characterization of surfactant liquid crystalline phases suitable for molecular alignment and measurement of dipolar couplings. J Biomol NMR 16:329–337

27. Sass J, Cordier F, Hoffman A, Rogowski M, Cousin A, Omichinski JG, Löwen H, Grzesiek S (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. J Am Chem Soc 121:2047–2055

28. Koenig BW, Hu JS, Ottiger M, Hendler RW, Bax A (1999) NMR measurement of bipolar couplings to purple membrane fragments. J Am Chem Soc 121:1385–1386

29. Fleming K, Gray D, Prasannan S, Mathews S (2000) Cellulose crystallites: a new robust liquid crystalline médium for the measurement of residual bipolar couplings. J Am Chem Soc 122:5224–5225

30. Lorieau J, Yao L, Bax A (2008) Liquid crystalline phase of G-tetrad DNA for NMR study of detergent solubilized proteins. J Am Chem Soc 130:7536–7537

31. Ma J, Goldberg GI, Tjandra N (2008) Weak alignment of biomacromolecules in collagen gels: an alternative way to yield residual bipolar couplings by NMR. J Am Chem Soc 130:16148–16149

32. Tycko R, Blanco FJ, Ishii Y (2000) Alignment of bioploymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. J Am Chem Soc 122:9340–9341

33. Chou JJ, Gaemers S, Howder B, Louis JM, Bax A (2001) A simple apparatus for generating stretched polyacrilamide gels, yielding uniform alignment of proteins and detergent micelles. J Biomol NMR 21:377–382

34. Meier S, Häussinger D, Grzesiek S (2002) Charged acrylamide copolymer gels as media for weak alignment. J Biomol NMR 24:351–356

35. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J Biomol NMR 18:303–309

36. Riley SA, Giuliani JR, Augustine MP (2002) Capture and manipulation of magnetically aligned Pf1 with an aqueous polymer gel. J Magn Reson 159:82–86

37. Louhivuori M, Pääkkönen K, Fredriksson K, Permi P, Lounila J, Annila A (2003) On the origin of residual dipolar couplings from denatured proteins. J Am Chem Soc 125:15647–15650

38. Obolensky OI, Schlepckow K, Schwalbe H, Solov'yov AV (2007) Theoretical framework for NMR residual dipolar couplings in unfolded proteins. J Biomol NMR 39:1–16

39. Mohana-Borges R, Goto NK, Kroon GJ, Dyson HJ, Wright PE (2004) Structural characterization of unfolded status of apomyoglobin using residual dipolar couplings. J Mol Biol 340:1131.1142

40. Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual bipolar couplings and small-angle X-ray scattering. Proc Natl Acad Sci USA 102:17002–17007

41. Novell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450

42. Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Protocol 3:679–690

43. Marsh JA, Baker JM, Tollinger M, Forman-Kay J (2008) Calculation of residual bipolar couplings from disordered state ensembles using local alignment. J Am Chem Soc 130:7804–7805

44. Meier S, Blackledge M, Grzesiek S (2008) Conformational distributions of unfolded polypeptides from novel NMR techniques. J Chem Phys 128:052204

45. Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P, Blackledge M (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 17:1669–1685

46. Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, Blackledge M (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar: application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc 130:8055–8061

47. Jensen MR, Blackledge M (2008) On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. J Am Chem Soc 130:11266–11267

48. Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining conformational ensambles of intrinsically disordered and partially folded proteins directly from chemical shifts. J Am Chem Soc 132:1070–1072

49. Bernadó P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M (2005) Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. J Am Chem Soc 127:17968–17969

50. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M (2010) NMR characterization of long-range order in intrinsically disordered proteins. J Am Chem Soc 132:8407–8418

51. Mukrash MD, Markwick P, Biernat J, von Bergen M, Bernadó P, Griesinger C, Mandelkow E, Zweckstetter M, Blackledge M (2007) Highly populated conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J Am Chem Soc 129:5235–5243

52. Meier S, Grzesiek S, Blackledge M (2007) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. J Am Chem Soc 129:9799–9807

53. Bernadó P, Blackledge M (2009) A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J 97:2839–2845

Chapter 10

Distance Information for Disordered Proteins from NMR and ESR Measurements Using Paramagnetic Spin Labels

David Eliezer

Abstract

The growing recognition of the many roles that disordered protein states play in biology places an increasing importance on developing approaches to characterize the structural properties of this class of proteins and to clarify the links between these properties and the associated biological functions. Disordered proteins, when isolated in solution, do not adopt a fixed structure, but can and often do contain detectable and significant residual or transient structure, including both secondary and long-range structure. Such residual structure can play a role in nucleating local structural transitions as well as modulating intramolecular or intermolecular tertiary interactions, including those involved in ordered protein aggregation. An increasing array of tools has been recruited to help characterize the structural properties of disordered proteins. While a number of methods can report on residual secondary structure, detecting and quantifying transient long-range structure has proven to be more difficult. This chapter describes the use of paramagnetic spin labeling in combination with paramagnetic relaxation enhancement (PRE) in NMR spectroscopy and pulsed dipolar ESR spectroscopy (PDS) for this purpose.

Key words: Intrinsically unstructured, Natively unfolded, Unfolded state, Denatured state, Random coil, NMR, ESR, Paramagnetic relaxation enhancement, Pulsed dipolar spectroscopy

1. Introduction

The use of paramagnetic spin labels for the study of protein structure has experienced a dramatic resurgence over the past decade. PRE experiments are now performed routinely to assist with high-resolution structure determination for both isolated proteins and proteins in larger complexes. ESR distance measurements are also playing an increasing role in structural studies, especially of larger complexes where long-range constraints are difficult to obtain. In addition to the application of these methods to well-ordered systems, they have found increasing use in studies of more dynamic systems, including studies focused on protein–protein interactions that occur prior to the formation of stable complexes (1) and, of

particular interest for the subject of this monograph, studies focused on disordered protein states. Early applications included studies of unfolded forms of proteins that can fold into native structures (2–4), but more recent efforts have focused on intrinsically disordered proteins.

Both PRE and PDS measurements rely on the interaction between magnetic dipoles, with PRE detecting the interaction of nuclear magnetic moments with those of unpaired electrons, and PDS detecting interactions between pairs of unpaired electron spins. PDS experiments measure the magnitude of the dipolar coupling, allowing for the dipole–dipole interaction to be considered as a perturbation to the energy levels engendered by the Zeeman interactions of the individual spins with the external magnetic field and for nonsecular terms to be ignored. PRE experiments measure the effect of the dipolar interaction on the relaxation properties of the transverse component of the proton magnetic moment, necessitating the consideration of nonsecular "spin flip" terms. In both cases, a point dipole approximation is used for each spin. The relevant equations governing the interpretation of the observed data in terms of interdipole distances are not reproduced here but are easily found in the literature (5, 6).

The experiments required for the application of these methods to disordered proteins are not particularly different from experiments used in the study of more ordered systems and are fairly standard for practitioners of NMR or ESR spectroscopy. The purpose of this work, therefore, is not to describe their implementation. Rather, this chapter is intended to illustrate, to those without previous experience in this area, the basic requirements for preparation of samples suitable for measurements, and for the analysis and interpretation of the data obtained, which on a qualitative level is typically straightforward, but for which quantitative analysis is more complex and remains an area of active research and development.

2. Materials

Protocols for spin labeling of protons involve standard buffers and reagents, with the exception of nitroxide spin labels such as MTSL (S-(2,2,5,5-tetramethyl-2,5-dihydro-1 H-pyrrol-3-yl)methyl methanesulfonothioate) or metal chelating groups such as N-[S-(2-pyridylthio)cysteaminyl]EDTA. Concentrated stock solutions of MTSL can be prepared in organic solvents such as dimethyl sulfoxide (DMSO) while N-[S-(2-pyridylthio)cysteaminyl]EDTA stock solutions can be prepared in either organic solvents or aqueous solutions.

3. Methods

Both PRE and PDS methods require the introduction of unpaired electrons into the proteins of interest. This is most commonly achieved through the conjugation of either a radical group spin label, or a transition-metal containing paramagnetic tag to the protein. While many possibilities exist for conjugation methods, in practice the most popular and easily achievable approach, often referred to as site-directed spin labeling (7) involves the introduction of cysteine residues at the desired labeling sites (and removal of endogenous cysteines, if any, at other sites) using site-directed mutagenesis, followed by the conjugation of the spin label or paramagnetic tag using one of several cysteine-modifying chemistries, the most popular being alkylthiosulfonate-mediated disulfide bond formation. The most common spin labels contain nitroxide radicals, with the most popular of these being MTSL, although similar reagents with a relatively reduced degree of side chain mobility are also being used (8, 9). Paramagnetic metal tags are typically introduced by conjugating a chelating group to the desired cysteine residue followed by loading with the paramagnetic ion of choice. The most commonly used reagent is N-[S-(2-pyridylthio)cysteaminyl]EDTA.

Both PRE and PDS experiments require the use of control samples. In the case of PRE experiments, the control sample provides a measure of the proton transverse relaxation rate (R2) in the absence of the paramagnetic reagent. In the case of PDS, the control samples provide a means of testing for intermolecular contributions to the measured distances.

3.1. Preparation of Paramagnetic Samples

1. Mix an excess (5- to 30-fold is typical) of spin label or metal chelator with the protein (see Note 1).

2. Incubate (as short as 30 min at room temperature can suffice, although many protocols call for overnight incubation at 4 °C).

3. Remove excess label subsequent to the labeling reaction through buffer exchange (see Notes 2 and 3).

4. If a chelating group is used, load the desired paramagnetic ion by adding at slightly superstoichiometric concentrations.

5. Remove excess ions, which can bind nonspecifically to protein sites, through thorough buffer exchange.

6. Samples prepared for PDS measurements must be frozen prior to data acquisition (see Note 4).

3.2. Preparation of Control (Diamagnetic) Samples

An ideal PRE control experiment should involve a sample that is identical to the spin-labeled sample, but in which the paramagnetic effect is eliminated.

1. PRE control for nitroxide spin labels: The control is typically produced by reducing the nitroxide radical to its hydroxyl-amine using a reducing agent such as ascorbic acid. This produces a sample that is chemically nearly identical, but diamagnetic. Such a control can be produced by directly reducing the actual sample used for the PRE measurement. Alternately, an originally prepared sample can be split in two, one of the samples reduced via addition of ascorbic acid, and an equivalent volume added to the other to maintain matched protein concentrations. This allows for preservation of the paramagnetic sample for future experiments. An alternative to reduction by ascorbic acid is to separate the control prior to spin labeling and conjugate to it a diamagnetic analogue of the spin label to be used (such as N-acylated MTSL). However, this may result in an imperfect match in protein concentration, since manipulations to remove excess spin label or diamagnetic analogue may result in irreproducible losses. If single time point PRE measurements are used (see below), closely matched protein concentrations are desirable (see Note 5).

2. PRE control for chelating reagents: The control sample is prepared identically to the paramagnetic sample except that the chelating group is loaded with a diamagnetic metal ion. Samples precisely matched in protein-concentration may be difficult to obtain, and two time point PRE measurements (see below) are recommended.

3. PDS controls: Controls are used primarily to separate intermolecular from intramolecular distances. For this purpose, two primary strategies are employed. First, the effects of magnetic dilution of the sample with unlabeled protein on the measured distances can be determined. Contributions from intermolecular distances should decrease with magnetic dilution and can be thereby identified. A second method that can be used is to prepare controls samples using singly labeled proteins. In this case, any detectable dipolar coupling must necessarily reflect intermolecular distances.

3.3. PRE Experiments

PRE results from the interaction of the unpaired electron with nuclear spins, leading to an increase in the transverse relaxation rate (R2) of the nuclear spins, an effect that can be measurable at distances up to ~30 Å. The PRE contribution to the nuclear R2, typically referred to as $\Gamma 2$, is simply the difference between the intrinsic R2, measured using a control sample, and the R2 measured for the paramagnetically labeled sample.

1. Single time point measurements: Since NMR line widths, and therefore signal intensities, are directly related to R2, a popular approach is to determine the amid proton $\Gamma 2$ using the ratio of

the NMR signal intensities in matched 2D proton–nitrogen correlation (HSQC) spectra collected from the paramagnetic and diamagnetic samples (5). This approach, however, suffers from the unequal longitudinal relaxation rates of nuclear spins in the paramagnetic and controls samples, which can lead to significant errors, especially when metal ions are used (10).

2. Two time point measurements: An improved approach involves direct measurement of the amide proton R2 values (11). Traditionally, such measurements involve multiple experiments at different relaxation times in order to determine the exponential transverse relaxation rate constant, requiring a greatly increased data acquisition time compared to a single intensity measurement. However, in principle two measurements spaced suitable apart can suffice to accurately determine the rate constant, and this two time point approach provides greater accuracy without a dramatic increase in the required data acquisition time (10).

3.4. PDS Experiments   PDS distance measurements rely on the dipolar interaction between two unpaired electrons, which results in a splitting of the spectrum of each individual electron spin and can be detected at distances of up to 90 Å (12). Notably, the contributions of dipolar coupling to line broadening in continuous wave ESR spectra can also be measured and used to extract interspin distances, but typically the presence of substantial inhomogeneous broadening in CW-ESR spectra limits the utility of CW measurements to the characterization of shorter distances of up to ~20 Å. In PDS measurements, the generation of spin echoes removes inhomogeneous broadening, allowing more precise determination of contributions from dipolar coupling (6).

1. DEER: The most common experiment used to extract interspin distances is the double electron–electron resonance (DEER) pulse sequence, also known as PELDOR (pulsed electron double resonance). In this experiment, the intensity of a standard spin echo signal is modulated by recoupling the dipolar interaction (using frequency selective pulses) for variable times, resulting in a frequency modulation of the observed signal intensity. DEER pulse sequences are usually arranged in a constant-time fashion (the overall length of the pulse sequence is constant) in order to minimize the contributions from relaxation. The most commonly employed sequence is the so-called four-pulse DEER experiment, which improves upon the original three-pulse experiment by using a primary echo to generate the initial signal that is subsequently refocused in a second echo, thereby avoiding the need for simultaneous pulses during short evolution times.

2. DQC: A more recently developed class of PDS experiments employ hard (frequency nonselective) pulses and double quantum coherence (DQC) filters to remove signals that are not modulated by the interspin coupling, resulting in a decreased background (13). DQC experiments provide an advantage over DEER for short (<20 Å) distances and at lower protein concentrations (6). At present, PDS experiments are performed most commonly at Ku-band (17.3 GHz).

3.5. PRE Data Analysis   Analysis of PRE data begins with calculation of either the ratio of peak intensities in the control and diamagnetic spectra (single time point data) or with the fitting of the intensities or peak volumes from spectra at each time point to an exponential to extract the apparent R2, followed by the calculation of Γ2 as the difference between R2s from the control and paramagnetic spectra (two time point data). At this point, a qualitative interpretation of the data can be easily made by plotting either the intensity ratio or Γ2 versus residue number and establishing whether any PRE effect is observable at sites outside the window of residues that are covalently restricted to be in the proximity of the paramagnetic label, and therefore always experience PRE. For disordered proteins, the size and shape of this window can be estimated by employing one of several ideal polypeptide random coil models to calculate the average distance for a given residue from the labeling site, and using the appropriate form of the Solomon-Bloembergen equation (5) to calculate a predicted value for Γ2, and in the case of single point measurements, to then calculate the predicted peak intensity ratio (4, 14).

3.6. PRE Data Interpretation   Qualitative interpretation of PRE data can provide general insights into long-range interactions in disordered proteins, and how these may be affected by sequence variations or environmental conditions, with perhaps the best example being that of the Parkinson's disease associated protein alpha-synuclein (Fig. 1). Quantitative interpretation of PRE data is complicated for disordered proteins by the fact that the protein and the paramagnetic label are constantly in motion with respect to one another. Several groups have employed a strategy of converting measured PREs to distances, which are then used to restrain simulations, from which protein conformational ensembles are derived (15–17). An alternative approach involves the generation of an unrestricted ensemble for a given disordered protein, calculation of the PRE effect for each member of the ensemble based on the distance between each nucleus and the spin label, and selection of a subensemble for which the appropriately averaged PRE effect is consistent with experiment. Here it is important to note that the Solomon Bloembergen equation, which links Γ2 to the dipolar coupling and thence to the interspin distance, is derived assuming a fixed distance between the spins. A model free approach to including the effects
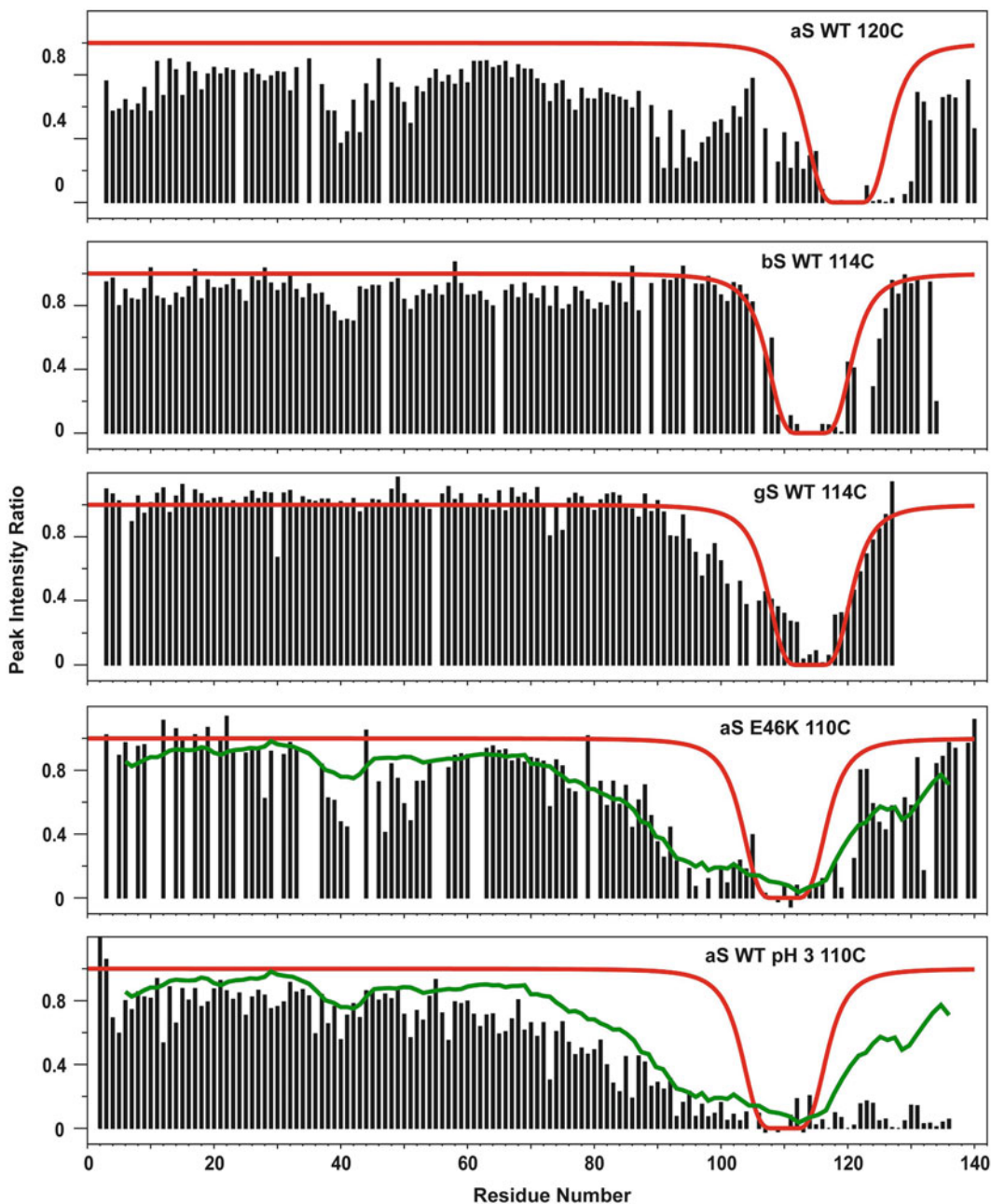
Fig. 1. PRE data from wild type (WT) alpha-synuclein (aS), family variants beta-synuclein (bS) and gamma-synuclein (gS), the Parkinson's disease associated E46K mutation, and the low pH form of the protein, illustrating that the effects of sequence variations and environment on long-range structure can be evaluated qualitatively (14, 26, 27). *Solid red lines* represent the expected PRE effect based on a Gaussian chain model of the protein. *Solid green lines* in the E46K and pH3 plots are smoothed data from the wild type protein labeled at the same position (110) at neutral pH and are shown for ease of comparison. In this case, there is a correlation with a decrease (bS and gS) or increase (E46K and low pH) in long-range interactions and a concordant decrease (bS and gS) or increase (E46K and low pH) in aggregation propensity.

of local motions of the spin label side chain on the measured $\Gamma 2$ values has been developed in the context of an otherwise well defined protein structure (18). The motions present in disordered proteins are far more complex than side chain motions and cannot easily be accounted for using this approach alone, but this method can be used to account for spin label mobility when performing distance calculations in the ensemble selection method (19).

3.7. PDS Data Analysis   To obtain structural information using PDS, time domain data need to be transformed into an average distance or a distance distribution. Weak intermolecular background contributions are typically removed by subtracting a linear or polynomial fit to the latter part of the time domain signal. Subsequently, the most straightforward approach is simple Fourier transformation into frequency space, followed by conversion of the observed frequency, which is a measure of the dipolar coupling, to an average distance. In well-ordered systems, this approach can yield distances that are accurate to within the variability introduced by the flexibility of the spin label side chain. In disordered systems, however, the distance is typically not well defined, and recovering a distance distribution is necessary. The observed time domain signal can be modeled as an integral of the probability-weighted contribution from each individual interspin distance, and inversion of the observed signal can be performed to recover the $P(r)$, or distance distribution, function. The problem, however, as is often the case for inverse problems, is not well posed and requires the application of a regularization method. Tikhanov regularization and/or maximum entropy methods are typically employed and require the careful choice of a regularization parameter, for which further algorithms exist (6).

3.8. PDS Data Interpretation   Interpretation of PDS data is in some ways simpler than for PRE data in that the ensemble of conformations being observed is invariant in time. Thus, each individual interspin distance in the ensemble is relatively well defined and does not experience motional averaging due to polypeptide disorder, although some degree of side chain motions can be retained even in frozen samples. Consequently, each individual interspin distance appears in the final distance distribution that is derived from the data. Not surprisingly, the large conformational ensemble sampled by disordered proteins invariably leads to very broad distance distributions. Such distributions are in many ways similar to those obtained from FRET or SAXS measurements (20, 21), and computational approaches have been combined with such measurements in order to determine or evaluate representative ensembles (22, 23). In contrast to PRE methods, PDS measurements only reflect distances between the two specifically labeled sites. Thus, information regarding potential long-range interactions
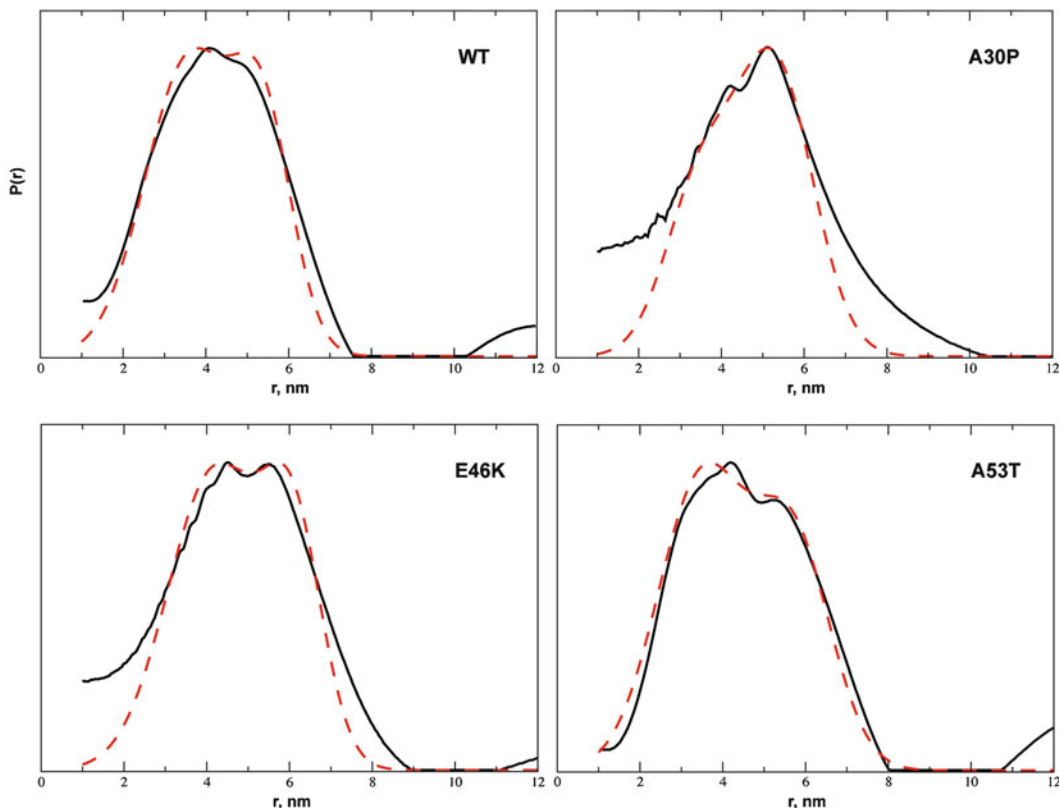
Fig. 2. PDS-derived distance distribution for WT alpha-synuclein and three Parkinson's linked mutants, A30P, E46K, and A53T, doubly spin-labeled at positions 24 and 72, illustrating the presence of two subensembles, one more compact with an average distance around 3.7 nm and one more extended with an average distance of around 5.7 nm (12). The distributions could be reasonably fit as a pair of Gaussians (*dashed red lines*), allowing for an estimate of their relative populations. While compact conformations can be directly inferred from PRE data such as those in Fig. 1, extended conformations are difficult to detect using PRE approaches.

involving other sites is not readily extractable. However, information on the presence of subpopulations with distinct distance distributions can be revealed (Fig. 2), whereas such subpopulations may be invisible to, or difficult to detect by, PRE methods.

4. Notes

1. Depending on the conditions used to purify and store the protein prior to spin labeling, reduction of the cysteine groups prior to labeling may be advisable and can be achieved using standard reducing agents such as dithiothreitol or TCEP (tris (2-carboxyethyl)phosphine), which should be removed through a rapid buffer exchange prior to addition of the label.

2. When introducing a single label into the protein, as is typically done for PRE experiments, a very low level of intermolecular disulfide-bond formation can be expected. Dimeric material can be removed using chromatography if desired.

3. When producing doubly labeled proteins, required for intramolecular PDS distance measurements, the potential for intramolecular disulfide bond formation is greater due to the increased effective local concentration of sulfhydryl groups and a greater excess of the conjugating reagent is recommended to maximize labeling efficiency. While any cross-linked multimeric species can be separated chromatographically, intramolecularly cross-linked molecules may be more difficult to remove. In the case of PDS, such molecules will be spectroscopically silent, but may affect the apparent protein concentration, and may also influence the behavior of the properly labeled protein molecules (for instance, by causing aggregation).

4. To minimize damage to or aggregation in protein samples, some form of cryoprotectant typically needs to be added to the samples. Glycerol is often used at around 30 % w/v, but sucrose can also be employed. Potential effects of the presence of cryoprotectants on the behavior of the protein sample are typically minimal, but should be investigated. Freezing is typically accomplished by immersing samples in liquid nitrogen, which leads to freezing times on the order of seconds for typical sample volumes (~50 μl). Faster freezing times can be obtained using cryogens with higher heat capacities and heat transfer rates. The influence of freezing rate on the resulting conformational ensemble remains to be more fully investigated.

5. A shortcoming of the methods described above for production of PRE control samples can result from a significant affinity of typical nitroxide spin label reagents to aromatic groups in proteins, which can lead to nonspecific binding of unconjugated spin label (5, 24, 25). In cases where such nonspecifically bound spin-label is not efficiently removed during the final removal of excess spin label from the sample, nonspecific PRE effects will be present in the paramagnetic sample, but not in the corresponding control sample. An alternative approach to eliminate this problem involves deconjugation of the spin label from the protein in the control sample through reduction of the disulfide bond linkage. In this case, nonspecifically bound spin labels will be present in both the paramagnetic and the control samples, and their PRE effects will be normalized out. However, this approach results in chemically distinct species in the paramagnetic and control samples, which can lead to chemical shift differences. These are usually confined to sites near the location of the spin label, which typically experience strong PRE effects and are therefore not of particular interest.

Acknowledgments

References

1. Clore GM, Tang C, Iwahara J (2007) Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. Curr Opin Struct Biol 17:603–616

2. Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 268:158–169

3. Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 268:170–184

4. Lietzow MA, Jamin M, Jane Dyson HJ, Wright PE (2002) Mapping long-range contacts in a highly unfolded protein. J Mol Biol 322:655–662

5. Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39:5355–5365

6. Borbat PP, Freed JH (2007) Measuring distances by pulsed dipolar ESR spectroscopy: spin-labeled histidine kinases. Methods Enzymol 423:52–116

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

8. McHaourab HS, Lietzow MA, Hideg K, Hubbell WL (1996) Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. Biochemistry 35:7692–7704

9. Zhang Z, Fleissner MR, Tipikin DS, Liang Z, Moscicki JK, Earle KA, Hubbell WL, Freed JH (2010) Multifrequency electron spin resonance study of the dynamics of spin labeled T4 lysozyme. J Phys Chem B 114:5503–5521

10. Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 184:185–195

11. Donaldson LW, Skrynnikov NR, Choy WY, Muhandiram DR, Sarkar B, Forman-Kay JD, Kay LE (2001) Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. J Am Chem Soc 123:9843–9847

12. Georgieva ER, Ramlall TF, Borbat PP, Freed JH, Eliezer D (2010) The lipid-binding domain of wild type and mutant alpha-synuclein: compactness and interconversion between the broken and extended helix forms. J Biol Chem 285:28261–28274

13. Borbat PP, Freed JH (1999) Multiple-quantum ESR and distance measurements. Chem Phys Lett 313:145–154

14. Sung YH, Eliezer D (2007) Residual structure, backbone dynamics, and interactions within the synuclein family. J Mol Biol 372:689–707

15. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127:476–477

16. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW (2008) Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. Proteins 71:587–598

17. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure 18:1094–1103

18. Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc 126:5879–5896

19. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M (2010) NMR characterization of long-range order in intrinsically disordered proteins. J Am Chem Soc 132:8407–8418

20. Ferreon AC, Moran CR, Gambin Y, Deniz AA (2010) Single-molecule fluorescence studies of

intrinsically disordered proteins. Methods Enzymol 472:179–204

21. Mylonas E, Hascher A, Bernado P, Blackledge M, Mandelkow E, Svergun DI (2008) Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry 47:10345–10353

22. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

23. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. Proc Natl Acad Sci USA 104:1528–1533

24. Bussell R Jr, Ramlall TF, Eliezer D (2005) Helix periodicity, topology, and dynamics of membrane-associated alpha-synuclein. Protein Sci 14:862–872

25. Card PB, Erbel PJ, Gardner KH (2005) Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. J Mol Biol 353:664–677

26. Rospigliosi CC, McClendon S, Schmid AW, Ramlall TF, Barre P, Lashuel HA, Eliezer D (2009) E46K Parkinson's-linked mutation enhances C-terminal-to-N-terminal contacts in alpha-synuclein. J Mol Biol 388:1022–1032

27. McClendon S, Rospigliosi CC, Eliezer D (2009) Charge neutralization and collapse of the C-terminal tail of alpha-synuclein at low pH. Protein Sci 18:1531–1540

Chapter 11

Using Chemical Shifts to Assess Transient Secondary Structure and Generate Ensemble Structures of Intrinsically Disordered Proteins

Stepan Kashtanov, Wade Borcherds, Hongwei Wu, Gary W. Daughdrill, and F. Marty Ytreberg

Abstract

The chemical shifts of backbone atoms in polypeptides are sensitive to the dihedral angles phi and psi and can be used to estimate transient secondary structure and to generate structural ensembles of intrinsically disordered proteins (IDPs). In this chapter, several of the random coil reference databases used to estimate transient secondary structure are described, and the procedure is outlined for using these databases to estimate transient secondary structure. A new protocol is also presented for generating a diverse ensemble of structures for an IDP and reweighting these structures to optimize the fit between simulated and experimental chemical shift values.

Key words: Ensemble structure, Random coil, Secondary chemical shift, Intrinsically disordered proteins, Nuclear magnetic resonance spectroscopy

1. Introduction

1.1. Chemical Shifts

Chemical shifts, determined using nuclear magnetic resonance (NMR) spectroscopy, provide valuable information about local atomic structure. In particular, the chemical shifts for the backbone atoms in polypeptides are sensitive to the dihedral angles, phi and psi. These dihedral angles can be reliably estimated from the backbone chemical shifts of ordered proteins. Using chemical shifts to estimate the backbone dihedral angles for IDPs is more challenging because the chemical shift represents a population weighted average of the local conformational preferences. For ordered proteins it is expected that there are no large fluctuations in the local conformational preferences for backbone atoms but the case is much different for IDPs. IDPs possess little or no tertiary or secondary structure

and they experience large-scale conformational fluctuations on multiple timescales (1–9). Because of the possible range of unique conformations that IDPs can sample, rigorously identifying the underlying distribution of structures that are responsible for the population weighted average chemical shifts represents a major challenge in structural biology. The approach that we present below depends on a broad sampling of conformational space to generate a large, diverse ensemble of structures. These structures are then subjected to a reweighting scheme that selects a subset of structures that are best able to reproduce the experimental measurement, in this case chemical shifts. While the approach does generate conformational ensembles that provide superior fits to the experimental data it is important to consider these ensembles as one of many possible representations of reality.

1.2. Using Chemical Shifts to Estimate Transient Secondary Structure

Secondary chemical shifts $(\Delta\delta)$ are a convenient and accurate way to determine the secondary structure of a polypeptide (10–13). They are a measure of how far a specific residue's chemical shifts have deviated from some random coil reference state. The random coil state can be defined as the state where the phi and psi dihedral angles populate a Boltzmann distribution, sampling three-dimensional conformational space freely (14). So the secondary chemical shift of the alpha carbon for a specific residue would be equal to its chemical shift minus the random coil reference for that amino acid. While structured polypeptides may show shifts as high as +3.1 ppm for the alpha carbon indicating a stable alpha helix, the secondary chemical shifts for IDPs are generally smaller. This is due to their ability to sample three-dimensional space more freely, and we are measuring the weighted average of a diverse population. These small deviations are indications of the residues overall bias toward any secondary structure, be it helical or extended (15). Though these shifts are smaller, they can still be used to calculate the percentage of the ensemble population that is displaying secondary structural characteristics (14). As an example, if a residue has an alpha carbon shift of +0.7 ppm, it could indicate 0.7/3.1 or 22.6 % helical structure, though the reader should note that it is the trends that should be focused on, not individual residues.

The ensemble-averaged $\Delta\delta$ values for IDPs can be below or very near the digital resolution of the experiment. It is therefore paramount that the random coil values be as accurate as possible. One method of increasing the applicability of a reference library is to account for any inherent effects that neighboring residues might possess. These sequence corrections are often small, and would have little impact on the $\Delta\delta$ values of a structured protein, except perhaps those of proline, but these small corrections can have a large impact on the $\Delta\delta$ values of an IDP.

The estimate transient secondary structure for IDPs three random coil chemical shift libraries developed by Mulder, Wishart, and

Schwarzinger, as well as a protein specific random coil library developed by assigning our test protein in 8 M urea, are used (15–17). As shown below, different random coil chemical shift libraries result in different $\Delta\delta$ values for IDPs, further complicating the reliable assignment of secondary structure based on chemical shift data.

1.3. Random Coil Reference Libraries

The three random coil chemical shift libraries described below were compiled using different strategies (see Table 1) (15–17). The Wishart library was determined using a set of 40 protected linear

Table 1

Random coil chemical shifts for backbone atoms

Residue	Wishart				Schwarzinger				Mulder			
	CA	CO	NH	N	CA	CO	NH	N	CA	CO	NH	N
A	52.5	177.8	8.24	123.8	52.8	178.5	8.35	125.0	52.6	178.4	8.16	123.9
B	55.4	174.6	8.43	118.6	55.6	175.5	8.54	118.7	N/A	N/A	N/A	N/A
C	58.2	174.6	8.32	118.8	58.6	175.3	8.44	118.8	58.3	174.9	8.41	119.1
D	54.2	176.3	8.34	120.4	53.0	175.9	8.56	119.1	54.3	177.0	8.22	120.2
E	56.6	176.6	8.42	120.2	56.1	176.8	8.40	120.2	56.7	177.1	8.30	120.8
F	57.7	175.8	8.30	120.3	58.1	176.6	8.31	120.7	57.9	176.4	8.11	120.1
G	45.1	175.9	8.33	108.8	45.4	174.9	8.41	107.5	45.2	174.6	8.31	108.8
H	55.0	174.1	8.42	118.2	55.4	175.1	8.56	118.1	56.0	175.3	8.31	118.9
I	61.1	176.4	8.00	119.9	61.6	177.1	8.17	120.4	61.2	176.9	7.96	120.5
K	56.2	176.6	8.29	120.4	56.7	177.4	8.36	121.6	56.4	177.2	8.22	121.4
L	55.1	177.6	8.16	121.8	55.5	178.2	8.28	122.4	55.3	178.0	8.09	121.9
M	55.4	176.3	8.28	119.6	55.8	177.1	8.42	120.3	55.6	177.0	8.21	120.0
N	53.1	175.2	8.40	118.7	53.3	176.1	8.51	119.0	53.2	175.8	8.37	118.7
P	63.3	177.3	N/A	N/A	63.7	177.8	N/A	N/A	63.2	177.5	N/A	N/A
Q	55.7	176.0	8.32	119.8	56.2	176.8	8.44	120.5	55.8	176.5	8.26	120.2
R	56.0	176.3	8.23	120.5	56.5	177.1	8.39	121.2	56.1	176.8	8.23	121.3
S	58.3	174.6	8.31	115.7	58.7	175.4	8.43	115.5	58.4	175.2	8.22	115.9
T	61.8	174.7	8.15	113.6	62.0	175.6	8.25	112.0	61.9	175.1	8.05	114.0
V	62.2	176.3	8.03	119.2	62.6	177.0	8.16	119.3	62.3	176.8	8.04	120.4
W	57.5	176.1	8.25	121.3	57.6	177.1	8.22	122.1	57.5	174.5	7.73	120.7
Y	57.9	175.9	8.12	120.3	58.3	176.7	8.26	120.9	57.8	176.3	8.03	120.2

hexapeptides with the general sequences of Gly-Gly-X-Ala-Gly-Gly, or Gly-Gly-X-Pro-Gly-Gly, where X refers to each of the 20 common amino acids. The measurements were made in the presence of 1 M urea in an effort to eliminate any possible structural propensities of the short peptides. Gly was chosen to be the flanking residues due to its inherent flexibility. Alanine was placed in the $i + 1$ position for its low steric hindrance and its inherent similarity to most of the amino acids. They also measured the values of each of the common amino acids with Proline in the $i + 1$ position because it has a large, systematic effect on the chemical shift residue X. This is the only other near-neighbor effect accounted for in the Wishart library (16).

The Schwarzinger library was developed accounting for the near-neighbor effects of each of the 20 common amino acids within two of residue i. To accomplish this they created 20 protected linear pentapeptides with the sequence Gly-Gly-X-Gly-Gly, where X once again represents each of the 20 common amino acids in turn. They measured the chemical shifts in 8 M urea at a pH of 2.3 to ensure that the peptides would be as close to their random coil shifts as possible. In addition to analyzing the chemical shift of amino acid X, Schwarzinger et al. determined the effects residue X had on the chemical shifts of each of the glycines, which permitted the generation of local sequence-corrected random coil shifts (17). When using this library, care needs to be taken when considering the residues with acidic side chains because at pH 2.3 these residues are protonated, which could influence their random coil chemical shifts.

The most recent random coil chemical shift library was developed by Mulder and is called the neighbor-corrected Intrinsically Disordered Protein database (ncIDP) (18). This database was generated using single value decomposition (SVD) to analyze the chemical shifts and the nearest neighbor effects of 14 confirmed IDPs. This analysis comprised 6,903 measured chemical shifts and results in lower root-mean-square-deviations from experimentally determined values than previous libraries, possibly due to conformational propensities of Alanine in the Wishart set, and due to the low pH and high urea used in the Schwarzinger set (15). However, due to the nature of IDPs there is a low sample size for residues that occur infrequently in IDPs (i.e., Tryptophan).

Finally analysis of $\Delta\delta$ based on a protein specific random coil library is presented. To develop this library, backbone resonance assignments are performed on an IDP dissolved in a high concentration of urea (8 M). This should eliminate all secondary structure, forcing the IDP into a "random coil" state.

1.4. Model System   The model system chosen to illustrate the differences and effectiveness of the various libraries is the N-terminal transactivation domain of the tumor suppressor protein p53 (p53TAD). It is a useful model to test these random coil libraries as it has been shown

```
Human MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMP-EAAPRV------
      ***.**: .::**********:**:*********.   . *:*:*:* *:.: :*: **  .*:***** :** .
Dog   MEESQSELNIDPPLSQETFSELWNLLPENNVLSSELCPAVDELLL-PESVVNWLDED--SDDAPRMPATSAPTAPGPAPS
      1            20              40              60
                        Amino acid position
```

Fig. 1. Alignment of human p53TAD (residues 1–73) and dog p53TAD (residue 1–77), showing 55.4 % identity for the entire sequence, and 90.9 % identity in the MDM2 binding region.

experimentally to be disordered, and has also been demonstrated to have transient helical characteristics in a binding domain, which forms a stable alpha helix when bound by its ligand MDM2 (19). We utilize a human construct with residues 1–73 and the dog orthologue with residues 1–77. These two domains share 55.4 % identity for the entire sequence but increases to 90.9 % identity in the MDM2-binding site (residues 17–26 for human) and the flanking regions (Fig. 1).

2. Methodology for Using Chemical Shifts to Assess Transient Secondary Structure in IDPs

The protocol described below assumes the user has access to the spreadsheet program Excel and that the backbone chemical shifts have already been assigned.

1. Use the following procedure to prepare a spreadsheet that associates the residue-specific random coil chemical shift with each amino acid in the protein sequence:

   (a) Create two columns containing the amino acid sequence and the random coil chemical shifts of choice (i.e., Wishart, Schwarzinger, or Mulder), associating the single-letter code for each amino acid with its corresponding random coil value. Create another column with the protein sequence using the single-letter code.

   (b) In a new column, use the VLOOKUP (in Microsoft Excel) function to associate the residue-specific random coil chemical shift with each amino acid in the protein sequence and transcribe it into a new cell. This function allows one to reference a table and transcribe values that have been associated with certain input criteria. The format for the VLOOKUP function is as follows: [=VLOOKUP(B2,$D$2:$E$22,2,FALSE)], where B2 refers to the first cell containing the amino acid sequence of the protein in question, the expression $D$2:$E$22 defines the cells that contain the residue-specific random coil chemical shift values, the 2 corresponds to the column in this range where the value is located. This is the value that will

ultimately appear in the new cell, and FALSE is a logical function that forces VLOOKUP to select only very close or exact matches. For this example if cell B2 contains an M for methionine, then the VLOOKUP will look for only an M in the range of cells that contain the amino acids, which should be the first column, then it will retrieve the random coil chemical shift value for M from the second column and transcribe it into the new cell. After this formula is entered for the first residue in the protein, the fill down command can be used to apply the VLOOKUP function to the remaining residues.

2. If you are using the Schwarzinger library use the following procedure to prepare an additional spreadsheet that corrects the residue-specific random coil chemical shifts to account for effects due to the residue type(s) of the nearest neighbors.

(a) This spreadsheet will contain two tables. In column 1 of Table 1 insert the protein sequence. In column 2, insert the protein sequence so that amino acid 1 in column 1 is aligned in the same row as amino acid three in column 2. This column represents the $i + 2$ sequence correction. In column 3, insert the protein sequence so that amino acid 1 in column 1 is aligned in the same row as amino acid 2 in column 3. This column represents the $i + 1$ sequence correction. Use the same logic to create two additional columns that represent the $i - 1$ and $i - 2$ sequence correction. The second table should contain five columns. The first column should contain the single-letter code for the 20 common amino acids. Columns 2–5 should contain the corrections to the random coil chemical shifts due to nearest neighbor effects. These columns should be organized in the same order as columns 2–5 of the first table but should not have any offset.

(b) To reference these values it is necessary to use the VLOOKUP function again. Make sure that the formula accounts for the fact that you have five columns.

(c) The near-neighbor effects must then be summed up and added to the uncorrected random coil chemical shift value retrieved in Subheading 2 step 1(b). This will yield the sequence-corrected random coil values for the protein. The list of sequence-corrected random coil values is then subtracted from experimentally determined chemical shifts, to yield $\Delta\delta$.

(d) These values can be plotted against the sequence or residue number for visual inspection of any patterns. Positive $\Delta\delta$ values are observed for C_α and CO nuclei in helical structures and negative values are observed for beta sheet and
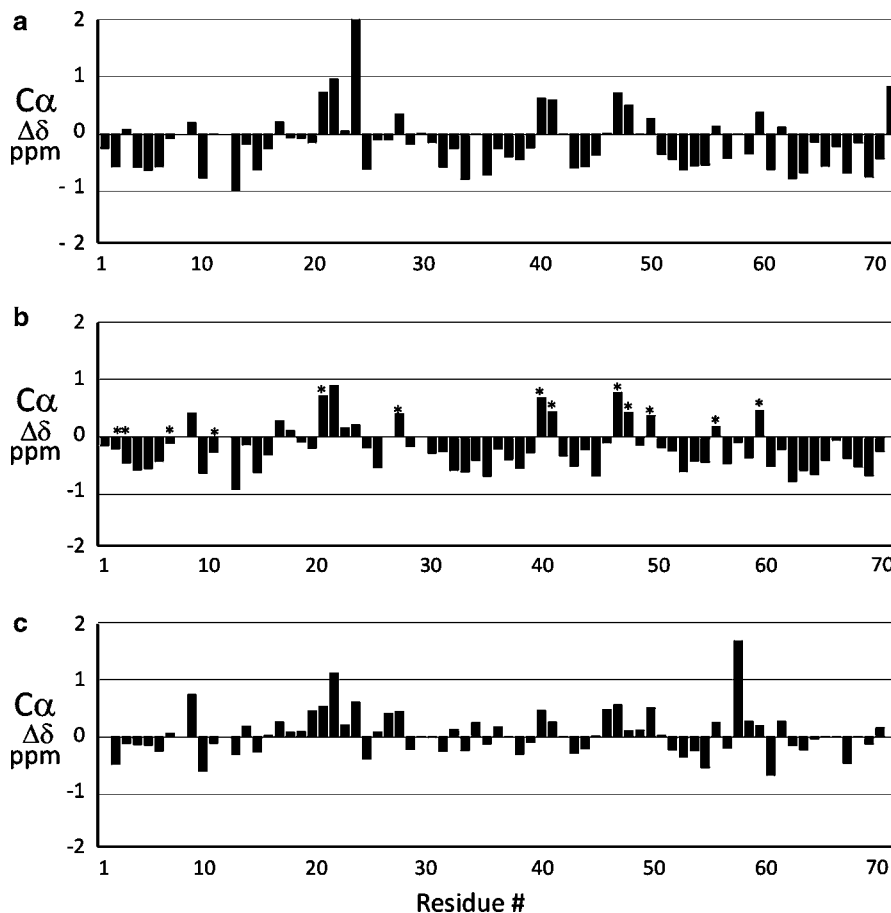
Fig. 2. The C$_\alpha$ $\Delta\delta$ values for human p53TAD determined using the random coil libraries developed by (**a**) Wishart, (**b**) Schwarzinger (Glu and Asp residues marked with *asterisk*), and (**c**) Mulder.

other extended structures. A detailed description of the relationship between $\Delta\delta$ and the type of secondary structure can be found in Case and Wishart ([20]).

3. It is recommended to make $\Delta\delta$ plots using all three random coil libraries. The different random coil libraries will give you different results as seen in Figs. [2] and [3]. When interpreting the $\Delta\delta$ plots it is important to look for general agreement between the different random coil libraries and for trends of positive or negative $\Delta\delta$ values that are observed for both C$_\alpha$ and CO nuclei and that extend over several residues. One will often observe one or two residues with positive or negative $\Delta\delta$ values. While it is formally possible for a single residue to have a preference for helical phi and psi dihedral angles, this is probably not an indication of a transiently forming helix.
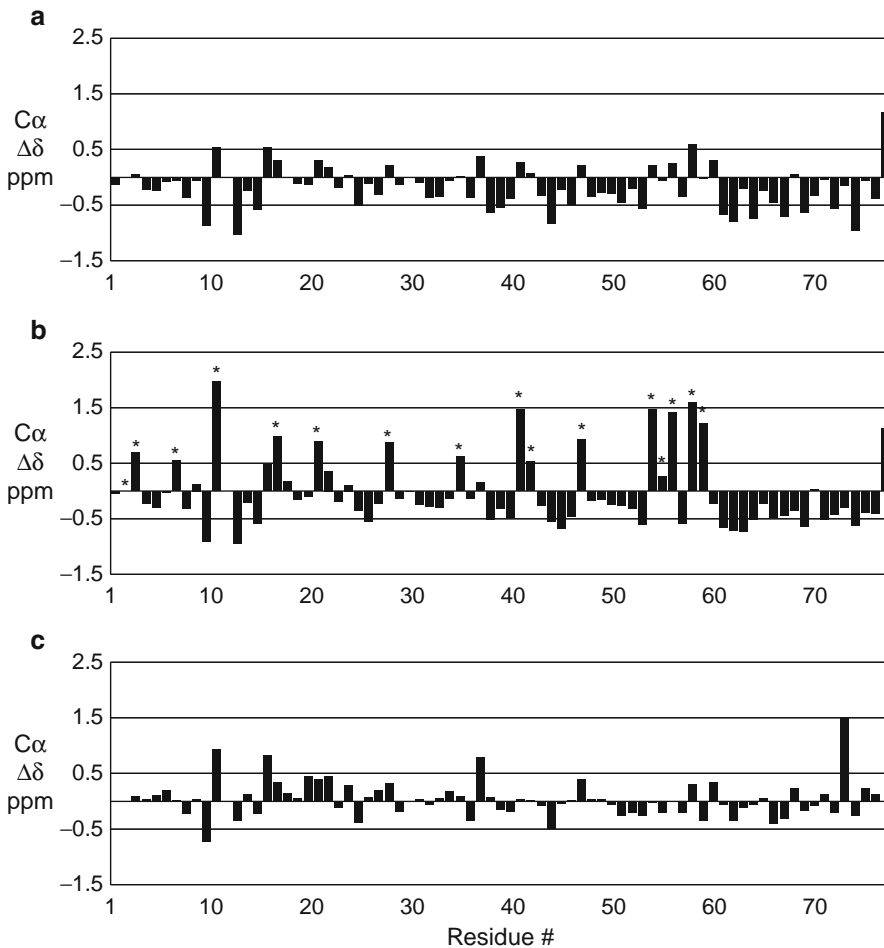
Fig. 3. The C_α $\Delta\delta$ values for dog p53TAD determined using the random coil libraries developed by (**a**) Wishart, (**b**) Schwarzinger (Glu and Asp residues marked with *asterisk*), and (**c**) Mulder.

4. There are instances where it is justifiable to mix the residue-specific random coil chemical shifts from multiple libraries. For instance, the random coil values from Schwarzinger were measured at pH 2.3. At this pH, Asp and Glu residues are protonated. This induces a systematic negative chemical shift for the C_α and CO nuclei which will result in a more positive $\Delta\delta$ value (see $\Delta\delta$ values with *asterisk* in Figs. 2b and 3b). In this case one could use the random coil chemical shifts for Asp and Glu from the Wishart or Mulder library.

5. One final recommendation is to develop a protein-specific random coil chemical shift standard. This involves making the resonance assignment for backbone nuclei in a denaturing solution. This is not that difficult for IDPs that are less than 100 residues. As shown in Fig. 4, this internally consistent standard appears to provide the most uniform patterns of $\Delta\delta$ values for an IDP.
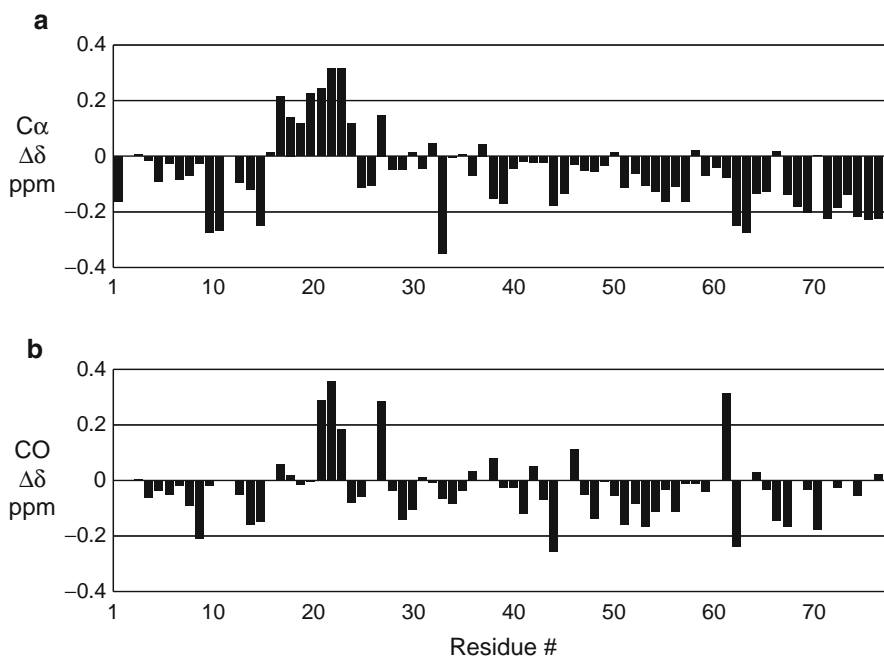
Fig. 4. The (**a**) C_{α} and (**b**) CO $\Delta\delta$ values for Dog p53TAD determined using the urea-denatured resonance assignments as the random coil reference.

3. Methodology for Determining Structural Ensembles of IDPs from Chemical Shifts

The broad ensemble generator with re-weighting (BEGR) approach combines the use of experimental and computational techniques to generate realistic structural ensembles for IDPs (25). Below we describe the BEGR method on human and dog p53TAD using chemical shift data. Note that other experimental data such as residual dipolar coupling or small-angle x-ray scattering can be used in a similar fashion.

1. Generate diverse pool of candidate structures for IDP.

    (a) Generate side chain coordinates (i.e., bond distances, bond angles, and torsion angles) for all amino acid types in p53TAD. For the results shown here we generated and used optimized geometries obtained via quantum mechanical calculation (26). Alternatively one could use the side chain coordinates provided with the Ribosome software package developed by Raj Srinivasen and George Rose (http://roselab.jhu.edu/~raj/Manuals/ribosome.html).

    (b) Starting from the N-terminus of the protein (in this case human and dog p53TAD) join the first two amino acids together using randomly generated Φ,Ψ backbone torsion angles. Check for a steric clash, i.e., determine if there are overlapping atoms. If there is a steric clash then join these

two amino acids together with a different set of random backbone torsions. Continue this process until a two amino acid structure is obtained with no steric clashes.

(c) Repeat Subheading 3 step 1(b) for each amino acid in the sequence until the C-terminal residue is reached but with the following modification: For a given amino acid (n) generate random values for the Φ,Ψ torsion angles ten times. If a steric-free structure is not found in ten attempts, generate new coordinates for amino acid $n - 1$ and then attempt to generate coordinates for amino acid n. Repeat this step ten times to try and identify a steric-free structure for amino acid n. If a steric-free structure is not identified then go back to amino acid $n - 2$, and as necessary amino acid $n - 3$ and $n - 4$, and so on.

(d) Repeat Subheading 3 step 1(b,c) to generate a pool of N candidate structures for the IDP. For the examples shown here 100,000 structures were generated for human and dog p53TAD.

(e) Relax each structure generated in Subheading 3 step 1 (d) to eliminate unphysical interactions using molecular dynamics simulation. Implicit solvent and low simulation temperatures should be used since the purpose of these simulations is to relax the structure, not to sample conformational space. For the examples shown here each structure was first minimized for 1000 steps and then equilibrated for 1.0 ps at 100 K. Generalized Born surface area implicit solvent was used for both minimization and equilibration, and cutoffs for electrostatics and van der Waals interactions were set to 1.0 nm. A 1.0 fs timestep and Langevin dynamics was used for the equilibration simulations.

(f) Convert the coordinates of the candidate structures into standard protein data bank (PDB) format (21). Note that because the backbone torsions are random and only steric interactions are considered, the conformations generated by this procedure do not utilize any information from folded proteins. This distinguishes the approach described above from other established structure-generating methods such as ROSETTA (22) and RANCH (23).

2. Calculate the backbone chemical shifts for each candidate structure.

(a) Download and install the shiftx software package (24). This software was developed to run on the Linux operating system.

(b) Process each structure generated in Subheading 3 step 1 above using shiftx with the default options, e.g., enter "/path/shiftx in.pdb out.file" where in.pdb is the name of the input structure in PDB format and out.file is the desired name of the output file. The output file will contain

chemical shift values for H-alpha (HA), C-alpha (CA), C-beta (CB), carbonyl or C′ (CO), N-amide (N), H-amide (HN), and side chain hydrogen atoms (H). For the results shown here we used the alpha-carbon chemical shift values, but the method can be used for any backbone chemical shift.

3. Reweight the candidate structures using metropolis Monte Carlo.

   (a) Assign an equal weight (w_i) to each structure so that $\sum_{i=1}^{N} w_i = 1.0$.

   (b) Calculate the fit between the simulated and experimental chemical shift

   $$\chi = \sum_{j=1}^{K} \left(I_j^{\text{sim}} - I_j^{\text{exp}} \right)^2,$$

   $$I_j^{\text{sim}} = \sum_{i=1}^{N} w_i \times I_{i,j}^{\text{sim}},$$

   where K is the number of data points (i.e, the number of residues in the IDP), I_j^{exp} are the experimental chemical shifts, and I_j^{sim} are the weighted, average simulated chemical shifts. For the example presented here $K = 73$ and $N = 100,000$.

   (c) Choose a single structure, k, at random from the pool and change its weight to a trial value $w_k^{\text{trial}} = w_k + \Delta w$ where $\Delta w = \pm 0.01/N$, and the positive and negative values are chosen at random in equal amounts.

   (d) Renormalize the set of trial weights via $w_i^{\text{trial}} = w_i^{\text{trial}} / \sum_{i=1}^{N} w_i^{\text{trial}}$.

   (e) Compute the fit between the simulated and experimental chemical shifts, χ^2_{trial} using the same relationship as in Subheading 3 step 3(b), but where $I_j^{\text{sim}} = \sum_{i=1}^{N} w_i^{\text{trial}} \times I_{i,j}^{\text{sim}}$.

   (f) If $\chi^2_{\text{trial}} < \chi^2$ then the trial weights are accepted and the old weights are overwritten with the new trial weights, i.e., $\chi^2 = \chi^2_{\text{trial}}$ and $w_i = w_i^{\text{trial}}$ for all structures i.

   (g) If $\chi^2_{\text{trial}} \geq \chi^2$ then pick a random number, R, between 0 and 1. If $R > \exp\left[(\chi^2 - \chi^2_{\text{trial}})/\tau\right]$ then the trial weights are accepted, i.e., $\chi^2 = \chi^2_{\text{trial}}$ and $w_i = w_i^{\text{trial}}$ for all structures i. If $R \leq \exp\left[(\chi^2 - \chi^2_{\text{trial}})/\tau\right]$ then the trial weights are rejected. The value of τ is discussed in Subheading 3 step 3(h).

   (h) Repeat the Metropolis Monte Carlo (MMC) in Subheading 3 step 3(c–g) until the fit to the experimental data stops improving. The value of τ is chosen to be large for the first MMC step and is reduced during the simulation setting $\tau = \alpha\tau$ for each subsequent MMC step and choosing $0 < \alpha < 1$. Reducing τ in this way is analogous to simulated annealing where the temperature is initially high to permit good sampling, and then is gradually reduced to allow the system to find the minimum. When the value of τ is large there is a higher probability that a trial weight will be

accepted when $\chi^2_{trial} \geq \chi^2$. For the results shown here the initial value of τ was 10^5 and the final value was 10^{-5}. This required 7×10^7 MMC steps, which means that $\alpha = \exp\left[\ln\left(\frac{\tau_{end}}{\tau_{begin}}\right)/N_{MMC}\right] \approx 0.99999967$, where N_{MMC} is equal to the number of MMC steps.

4. After the MMC procedure has converged the BEGR procedure is complete and the resulting weights can be thought of as the relative importance of each structure to fitting the experimental chemical shift data, i.e., the highest-weight structures are the most important to fitting the experimental chemical shift data. The "BEGR ensemble" is defined as the collection of structures with nonzero weights. Figure 5 shows the correlations between the experimental and simulated $\Delta\delta$ values for human and dog p53TAD. Figure 5a, c show the correlation between the experimental and simulated $\Delta\delta$ values calculated using all 100,000 structures in the initial pool. Figure 5b, d show the correlation
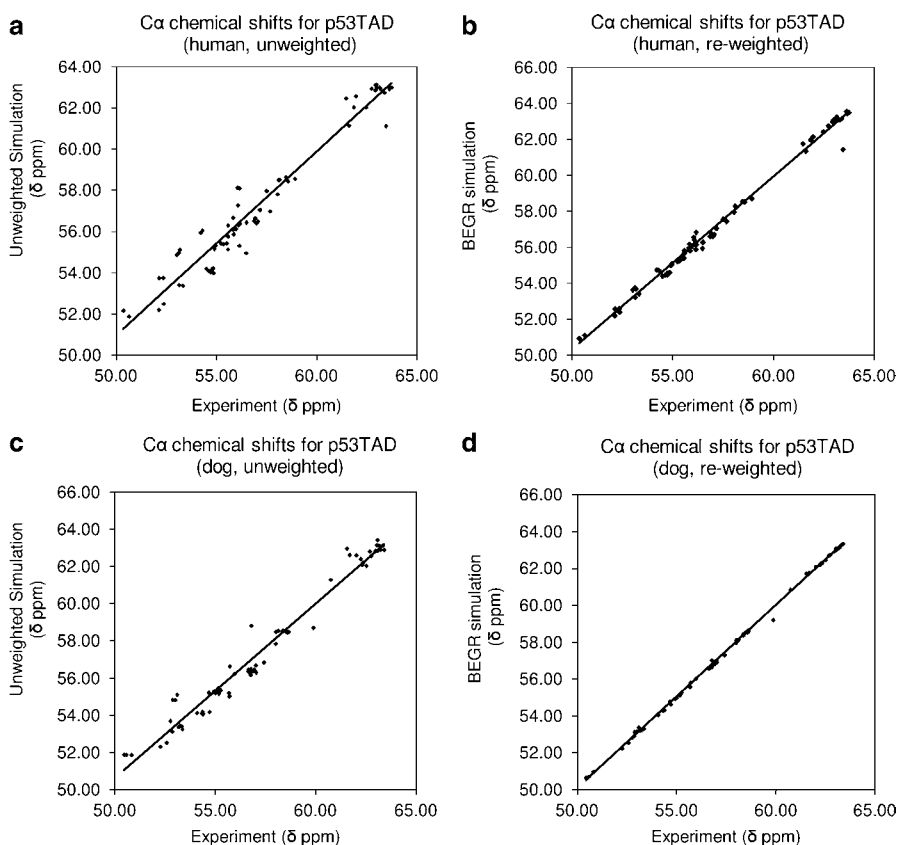
Fig. 5. Correlation plots for the experimental and calculated C_α $\Delta\delta$ values for human (**a**, **b**) and Dog (**c**, **d**) p53TAD, showing that the re-weighting used in the BEGR method significantly improves the fit between the simulation results and the experimental data. C_α $\Delta\delta$ values were calculated using the full ensemble for (**a**) and (**c**) and the BEGR ensemble for (**b**) and (**d**). For clarity, the chemical shift values for the G59 at 45 ppm is not shown.
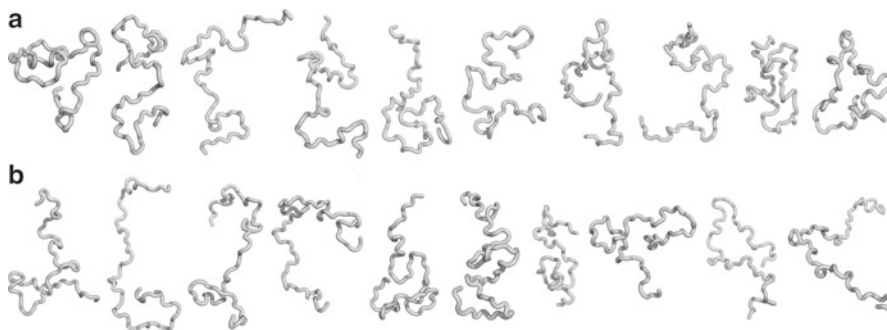
Fig. 6. Backbone worm models of the ten highest weight structures in the BEGR ensemble for (**a**) human and (**b**) dog p53TAD.

between the experimental and simulated $\Delta\delta$ values calculated using the structures in the BEGR ensemble. There are a total of 73 structures in the BEGR ensemble for human p53TAD and 77 structures for dog p53TAD. Figure 6 shows the ten highest weight structures in the BEGR ensemble for human (a) and dog (b) p53TAD.

5. It is important to recognize that the BEGR ensemble is not unique and repeating the procedure will result in a new collection of structures that will equally fit the data. We are currently determining how robust the approach is at identifying unique features in the ensemble. These findings will be presented in future manuscripts. Please stay tuned.

6. All of the steps above are implemented by the BEGR software. Free, open-source code can be found at http://begr.uidaho.edu.

Acknowledgments

References

1. Daughdrill GW (2010) Determining structural ensembles for intrinsically disordered proteins. In: Uversky VN, Longhi S (eds) Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation. John Wiley and Sons, Inc., Hoboken

2. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK (2005) Natively disordered proteins. In: Buchner J, Kiefhaber T (eds) Protein folding handbook, vol 3. WILEY-VCH, Darmstadt, pp 275–357

3. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN (2008) The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9(Suppl 2):S1. doi:1471-2164-9-S2-S1[pii]10.1186/1471-2164-9-S2-S1

4. Uversky VN (2002) What does it mean to be natively unfolded? Eur J Biochem 269 (1):2–12

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

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

7. Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 17(1):15–20

8. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. Biochemistry 41 (21):6573–6582

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

10. Wishart DS, Nip AM (1998) Protein chemical shift analysis: a practical guide. Biochem Cell Biol 76(2–3):153–163

11. Wishart DS, Sykes BD (1994) Chemical shifts as a tool for structure determination. Methods Enzymol 239:363–392

12. Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Methods Enzymol 338:3–34

13. Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 62:311–340

14. Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. In: Unfolded proteins, vol. 62. Adv Protein Chem. Academic Press Inc, San Diego, pp 311–340

15. Tamiola K, Acar Bi, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc 132(51):18000–18003. doi:10.1021/ja105656t

16. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) H-1, C-13 and N-15 random coil NMR chemical-shifts of the common amino-acids. 1. Investigations of nearest-neighbor effects. J Biomol NMR 5 (1):67–81

17. Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123 (13):2970–2978. doi:10.1021/ja003760i

18. Tamiola K (2009) Neighbor corrected Intrinsically disordered protein library. http://ww.protein-nmr.org/. Accessed Feb 2011

19. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW (2008) Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. Protein Struct Funct Bioinformatics 71(2):587–598. doi:10.1002/prot.21721

20. Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. In: Nuclear magnetic resonance of biologica macromolecules, Pt A, vol. 338. Methods Enzymol. Academic Press Inc, San Diego, pp 3–34

21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242

22. Simons KT, Bonneau R, Ruczinski I, Baker D (1999) Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins Suppl 3:171–176

23. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129 (17):5656–5664

24. Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. J Biomol NMR 26(3):215–240. doi:5123899 [pii]

25. Daughdrill GW, Kashtanov S, Stancik A, Hill S, Helms G, Muschol M, Receveur-Brechot V, Ytreberg FM (2012) Understanding the Structural Ensembles of a Highly Extended Disordered Protein. Mol BioSyst 8:308–319 doi:10.1039/C1MB05243H

26. Matta CF, Bader RFW (2002) Atoms-in-molecules study of the genetically encoded amino acids. II. Computational study of molecular geometries. Proteins 48:519–538

Chapter 12

Magic Angle Spinning Solid-State NMR Experiments for Structural Characterization of Proteins

Lichi Shi and Vladimir Ladizhansky

Abstract

Solid-state nuclear magnetic resonance (SSNMR) has become a prominent method in biology and is suitable for the characterization of insoluble proteins and protein aggregates such as amyloid fibrils, membrane–lipid complexes, and precipitated proteins. Often, the initial and the most critical step is to obtain spectroscopic assignments, that is, to determine chemical shifts of individual atoms. The procedures for SSNMR spectroscopic assignments are now well established for small microcrystalline proteins, where high signal-to-noise can be obtained. The sensitivity of the experiments and spectral resolution decrease with the increasing molecular weight, which makes setting SSNMR experiments in large proteins a much more challenging and demanding procedure. Here, we describe the protocol for the most common set of 3D magic angle spinning (MAS) SSNMR experiments. While the procedures described in the text are well known to SSNMR practitioners, we hope they will be of interest to scientists interested in extending their repertoire of biophysical techniques.

Key words: Solid-state NMR, Magic angle spinning, Protein structure, Spectroscopic assignments, Membrane proteins, Disordered proteins

1. Introduction

Magic angle spinning (MAS) solid-state nuclear magnetic resonance (SSNMR) is making a large impact in the field of biology. Its main areas of application are studies of insoluble proteins, protein aggregates, and large protein complexes, such as membrane-associated systems, amyloid fibrils, etc. (1–3). While initial applications of MAS SSNMR were on biomolecules with site-specifically introduced ^{13}C and/or ^{15}N labels, the focus recently started to shift towards proteins with uniform incorporation of isotopic labels (4, 5). The key advantage of such an approach is that a protein can be studied in a single sample, providing site-specific information on the whole molecule. On the other hand, the
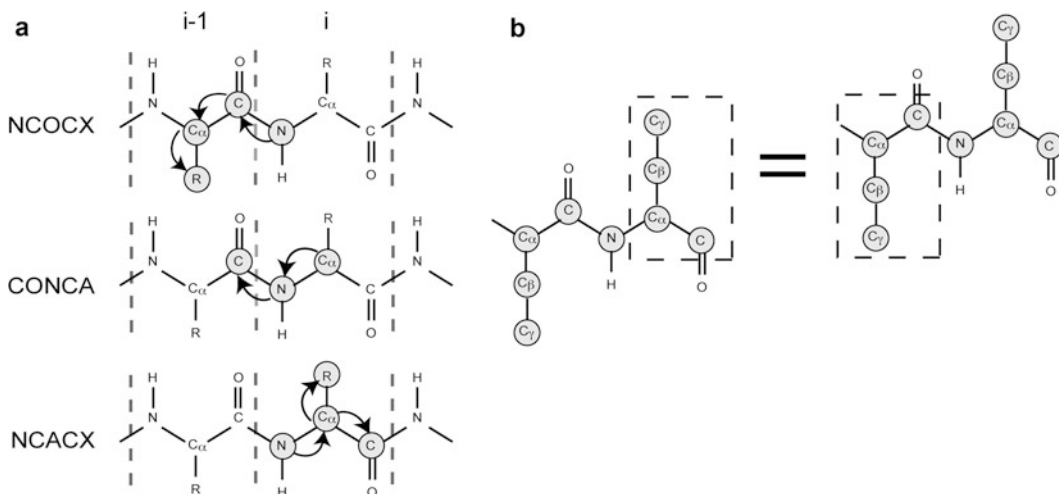
Fig. 1. (**a**) Solid-state NMR assignment strategies are based on three types of experiments that correlate chemical shifts of nitrogen and carbon atoms. The top NCOCX experiment correlates chemical shifts of nitrogen N[i] and backbone and side chain carbons of the preceding residue, CO[i–1], CX[i-1]. The bottom NCACX experiment correlates N[i], with CA[i] and CX[i]. CANCO experiment shown in the middle panel correlates backbone CA[i], N[i] and CO[i–1] shifts. (**b**) Two extended spin systems can be combined into a contiguous fragment by matching shifts of backbone and side chain atoms shown in *dashed boxes*.

complexity of experimental approaches and their analysis increases manyfold, while the sensitivity of already intrinsically insensitive technique decreases.

The presence of many magnetically active nuclei in a single sample results in heavily overlapped NMR spectra. For any site-specific studies of a molecule, it is imperative to obtain spectroscopic assignments, i.e., to map resonance positions onto a primary sequence of a protein, which is achieved using multidimensional NMR. To date, many of the available assignments were obtained using 2D NMR spectroscopy (6–8), but generally low resolution of such spectra limits the size of the proteins that can be studied. In this respect, 3D spectroscopy is much more powerful and can resolve most resonances (9–13).

The assignment strategies shown in Fig. 1 are based on a minimal set of three 3D experiments, which establish connectivities between chemical shifts of three or more nuclei. CANCO (or CONCA) correlates shifts of three backbone atoms: CA[i], N[i], and CO[i–1]. The NCACX experiment establishes an intraresidue backbone-side chain correlation between N[i], CA[i], and CX[i] atoms (CX = CO, CA, CB, CG, etc.), whereas NCOCX relates the backbone shift of N[i] to the carbon shifts of the preceding residue, CO[i–1], CX[i-1]. The triplets of shifts recorded in each experiment can be related to each other by matching at least two values: CANCO shares N and CO values with the NCOCX experiment, and N and CA values with the NCACX correlation. If a unique match is established, then one can build an extended spin system

containing the chemical shifts of many atoms, CX[i-1]- N[i]- CX[i], link the extended spin systems into contiguous fragments by matching CX shifts (Fig. 1b), and map the fragments onto the protein amino acid sequence. The procedure is more complicated in the case of spectral overlap, and a number of software tools are being developed toward automated procedures for SSNMR spectral assignments (14, 15).

The main purpose of this manuscript is to provide step-by-step protocols for recording high-quality magic angle spinning SSNMR data. These protocols are intended primarily for nonexperts, although some familiarity with basic SSNMR is highly desirable. The experimental procedures outlined below have been tested on two 7-helical integral membrane proteins proteorhodopsin (PR) (16, 17) and *Anabaena* Sensory Rhodopsin (ASR) (13). PR and ASR have similar molecular weight of ~27 kDa. Each of the experiments below could be recorded in about 3–4 days and required about ~8–11 mg of protein at 800 MHz magnetic field strength. Although the protocol described below was developed for the 800 MHz spectrometer, it can be extended to other higher/lower field strengths. It is essential, however, that studies of high-molecular-weight proteins be conducted at high fields (18).

2. Materials and Equipment

1. High-field NMR spectrometer equipped with at least three channels and high power amplifiers (recommended configuration: 1 kW for proton channel, and 500 W for ^{13}C and ^{15}N channels). The spectrometer should be capable of MAS experiments and equipped with standard accessories for SSNMR experiments, such as spinning frequency controller, high power preamplifiers, etc.

2. Triple channel $^{1}H/^{13}C/^{15}N$ magic angle spinning probe, rotor size 3.2 mm or 4 mm. In our experience, 3.2 mm E-free probe is one of the best available for routine protein NMR experiments.

3. Glycine powder (natural abundance) packed in a rotor.

4. Adamantane powder packed in a rotor.

5. Uniformly ^{13}C, ^{15}N-labeled microcrystalline protein (e.g., α-spectrin SH3 domain, ubiquitin, GB1, etc. see Note 1).

6. Protein sample of interest packed in a rotor (see Note 2). Any excess of buffer should be removed prior to sample packing. Sample should be constrained to the space contained within the NMR coil during the NMR experiment (~5 mm long middle portion of a standard 3.2 mm Bruker rotor). To prevent the sample from expanding during spinning, spacers should be used.

3. Methods

Prior to setting up a complicated multidimensional MAS SSNMR experiment, a certain amount of preparatory work should be done. We use the following notations common for Bruker spectrometers: d1, d2, ... denote delays; p1, p2, ... denote pulse lengths; pl1, pl2, ... denote power levels measured in dB; and sp0, sp1, ... are power levels of shaped pulses.

Setting up an NMR experiment requires calibration of power levels used in various experimental stages. We note that in modern spectrometers, the amplifier response is linear. Practically, this means that only a single power level value for each of the nuclei needs to be calibrated. Other power levels can be calculated according to the following relation:

$$pl_{cal} - pl = -20 \log \frac{v_{1cal}}{v_1} \tag{1}$$

where pl_{cal}, v_{1cal} are calibrated power values in dB and kHz, respectively, v_1 is the desired RF strength in kHz, and pl is its corresponding value in dB. An analogous equation holds for the parameters used for the power of shaped pulses, sp0, sp1, etc.

3.1. General Setup We assume that an approximate cross-polarization (CP) (19) has been established for standard samples such as adamantane and glycine during the initial stage of setting up the spectrometer. These samples should be used initially for chemical shift referencing and for setting the magic angle. It is assumed that after each sample or temperature change, each channel of the probe has to be tuned and matched.

1. *Referencing ppm scale using adamantane.* Use slow spinning frequency of 5–6 kHz and collect CPMAS spectrum (Fig. 2a), with p1 of 5 μs and with pl11 set to a value corresponding to zero power (120 dB for Bruker spectrometers), on adamantane. Adjust B_0 magnetic field strength so that the downfield shifted ^{13}C resonance is 40.48 ppm (20).

2. *Magic angle adjustment using glycine.* Setup the CPMAS experiment of Fig. 2a, with pl11 set to a value corresponding to zero power, p1 of 5 μs, and an acquisition length of ~30 ms. Set the spinning frequency and temperature that will later be used in the 3D experiments on the protein of interest. At a high field of 800 MHz, we recommend ~14–14.5 kHz for the routine applications discussed below. Collect the spectrum and observe the width of the carbonyl line. At 800 MHz, it should typically be ~50 Hz. If a larger line width is observed, change the magic angle slightly by turning the magic angle adjustment knob. Collect the spectrum again and observe the CO line width.
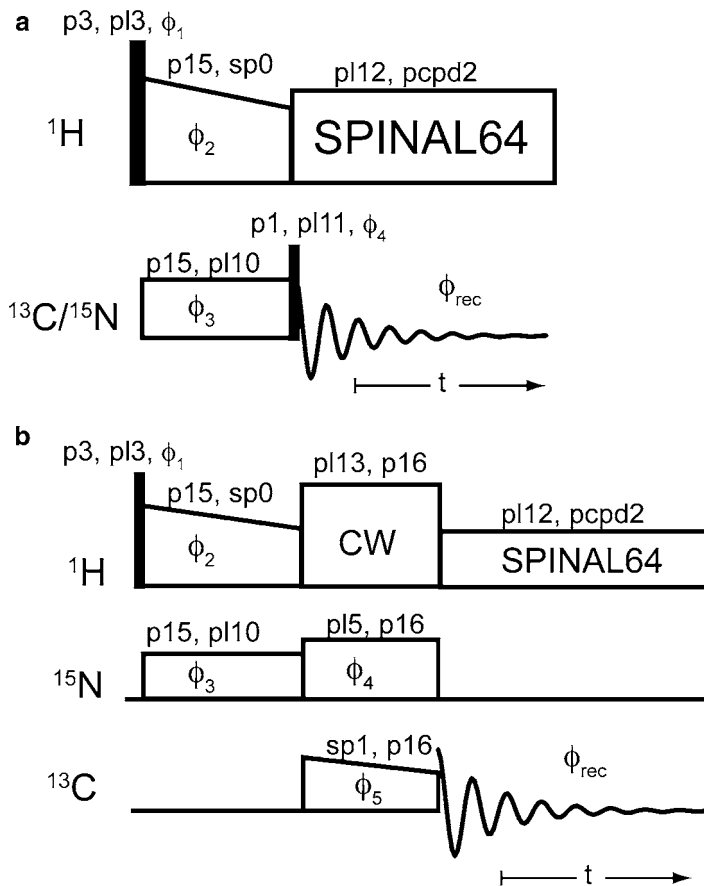
Fig. 2. General experimental schemes used for basic calibration of power levels, and CP conditions. Each pulse is characterized by three parameters: pulse length denoted p1, p2, etc., its power level pl1, pl2, etc., and pulse phase, ϕ_j. For shaped pulses the parameters that control power levels are sp0, sp1, etc. (**a**) CPMAS experiment for calibrating cross-polarization conditions and pulse lengths. Phase tables are as follows ($x = 0$, $y = 1$, $-x = 2$, $-y = 3$): $\phi_1 = 1, 3$; $\phi_2 = 0$; $\phi_3 = 0$; $\phi_4 = 1$; $\phi_{rec} = 0, 2$. (**b**) Double cross polarization (^{1}H-^{15}N-^{13}C) pulse sequence used to calibrate ^{15}N-^{13}C CP condition. Phase tables are as follows: $\phi_1 = 0$; $\phi_2 = 3$; $\phi_3 = 0, 2$; $\phi_4 = 0$; $\phi_5 = 0$; $\phi_{rec} = 0, 2$.

If the line width increases, reverse the direction by turning the adjustment knob. Continue the procedure until a line width of ~50 Hz is achieved.

3.2. Setting the 3D Experiment on a Protein

Obtaining spectroscopic assignment requires at least three 3D experiments. The experimental schemes shown in Fig. 3 consist of a few common basic polarization transfer steps arranged in different orders. Each step has to be calibrated as discussed below.

An important consideration when choosing the spinning frequency of an experiment in a uniformly labeled protein is to avoid matching rotational resonance (R^2) conditions (21, 22), $\delta_{Ci} - \delta_{Cj} \approx n \cdot v_r$, $n = 1, 2$, where δ_{Ci} and δ_{Cj} are isotropic chemical
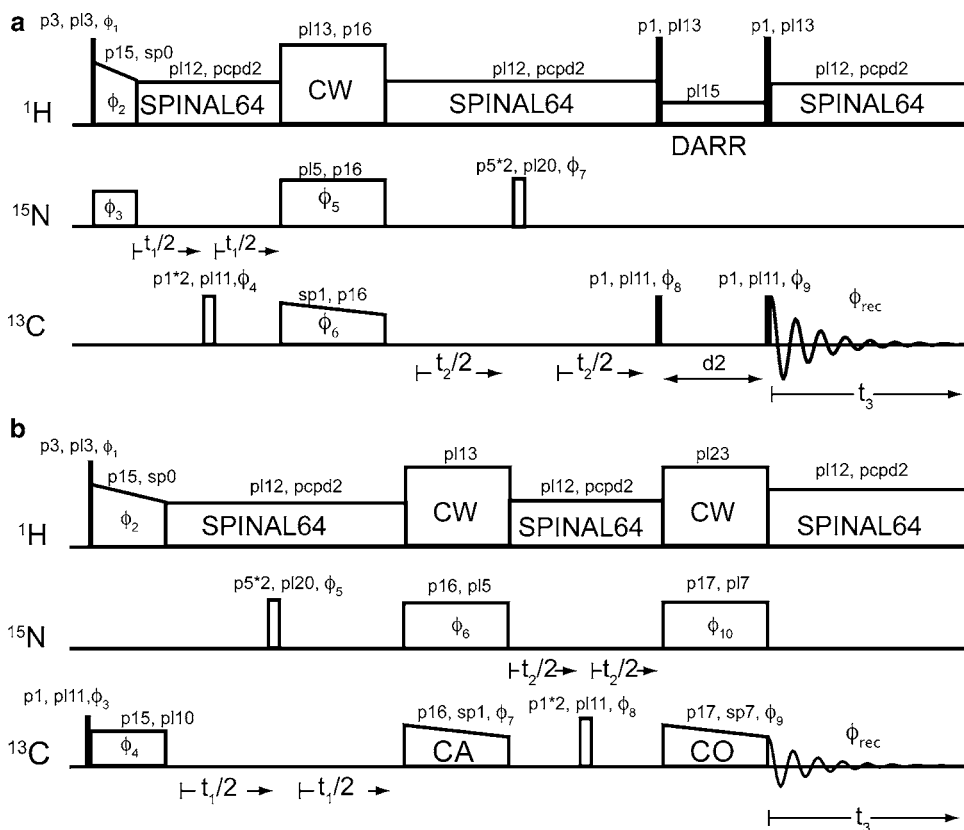
Fig. 3. Experimental schemes used to recording (**a**) 3D NCACX and NCOCX and (**b**) CANCO 3D correlations. Phase tables in (**a**) are $\phi_1 = 0$; $\phi_2 = 3$; $\phi_3 = 0, 2$; $\phi_4 = 1$; $\phi_5 = 0$; $\phi_6 = 0$; $\phi_7 = 1$; $\phi_8 = 1,1,3,3$; $\phi_9 = 3$; $\phi_{rec} = 0,2,2,0$. Phase sensitive detection in t_1 and t_2 dimensions is accomplished using TPPI scheme, by incrementing phases ϕ_5 and ϕ_8, respectively. Phase tables in (**b**) are $\phi_1 = 1$; $\phi_2 = 0,2$; $\phi_3 = 1, 3$; $\phi_4 = 0$; $\phi_5 = 0$; $\phi_6 = 0,0,2,2$; $\phi_7 = 0$; $\phi_8 = 0$; $\phi_9 = 0$; $\phi_{10} = 0$; $\phi_{rec} = 0,2,2,0$. Phase sensitive detection in t_1 and t_2 dimensions is accomplished using TPPI scheme, by incrementing phases ϕ_4 and ϕ_{10}, respectively.

shifts of spatially proximate nuclei, and v_r is the spinning rate. The most detrimental conditions are between directly bonded carbonyl and CA carbons. The isotropic chemical shift difference between these nuclei is in the range of 105–135 ppm, defining the range of potential R^2 conditions, which should be avoided. Most routine applications at 800 MHz in our lab are conducted at 14.3 kHz.

3.2.1. Basic Calibrations of Power Levels, Polarization Transfer Steps, and Decoupling

1. *Proton power calibration.* Using the pulse sequence in Fig. 2a, determine a proton power level that corresponds to 100 kHz. This is accomplished by setting the length of the pulse p3 to 5 μs and pl11 to 120 dB (no irradiation). Set p1 to 5 μs vary the power level pl3 until the signal disappears. This power level is equivalent to 100 kHz RF strength.

2. *Carbon power calibration.* In the pulse sequence of Fig. 2a, set p3 to 2.5 μs, pl3 to the value corresponding to 100 kHz as

determined in the previous step, and p1 to 5 μs. Set the carbon carrier frequency to 65 ppm. Vary the power level pl11 to minimize the signal around 65 ppm (see Note 3). The power level at which this is achieved corresponds to 50 kHz ^{13}C RF strength.

3. *HC CP optimization*. In Fig. 2a, set the ^{13}C carrier frequency o1p to 120 ppm, p15 to 2 ms, the carbon CP power level pl10 to the calibrated value of 50 kHz (step 2). Use ~7 kHz wide linear ramp for the proton CP shaped pulse. Observe carbon spectrum as a function of sp0 varying from ~80 kHz to ~50 kHz. Find the proton power level that yields the highest carbonyl intensity. The set of values (pl10, sp0) represents the optimal HC CP condition and will later be used in the 3D experiment (see Note 4).

4. *Decoupling*. For decoupling, use SPINAL64 (23) with decoupling power pl12 set to an equivalent of 83 kHz and pcpd2 to 5.6 μs (see Note 5).

5. *Setup nitrogen detection*. Set the nitrogen carrier frequency o1p to 120 ppm, p15 to 2 ms, p1 to 7 μs, and pl11 to 120 dB. Set sp0, pl10 to approximately match the Hartmann–Hahn CP condition. As a good starting point, set pl10 to ~35 kHz determined, for example, on standard samples during the spectrometer setup, and vary sp0 of the linearly ramped pulse (ramp width ~7 kHz) from ~40 kHz to ~60 kHz (see Note 6). Pick the values giving the highest intensity.

6. *Nitrogen power calibration*. Repeat step 2 to calibrate the nitrogen power. With p1 = 7 μs, vary pl11 to achieve a zero signal. The power level at which the signal is minimal corresponds to ~35.7 kHz nitrogen RF strength.

7. *HN CP optimization*. Repeat step 5. We recommend determining HN CP conditions for higher values of nitrogen RF strengths. Set the nitrogen power level at 45 kHz and 50 kHz, and vary sp0 in each case from ~45 kHz to 70 kHz to achieve optimal HN CP performance.

8. *N-CA CP optimization*. Use the experimental setup of Fig. 2b. Set o1p = 60 ppm, o3p = 120 ppm, and parameters (p3, pl3) and (pl10, sp0) according to what was found in steps 1 and 7, respectively. At spinning frequencies ~12–14 kHz, a good starting point for optimization is to choose pl5 equivalent to $2.5 \times \nu_r$, and optimize the sp1 value around $1.5 \times \nu_r$ (see Note 7). Use p16 of 5–6 ms, and linear ramp of ~2 kHz width.

9. *Decoupling during N-CA CP*. Repeat the same experiment to optimize decoupling power during ^{15}N-^{13}C CP. Choose optimal conditions of step 8, and run a series of experiments as a function of pl13. Vary pl13 from ~100 kHz to 80 kHz. The value resulting in the highest signal is the optimal value.

10. *N-CO CP optimization.* Use the same pulse sequence of Fig. 2b. Set o1p = 175 ppm, o3p = 120 ppm, and (p3, pl3) and (pl10, sp0) as determined in steps 1 and 7, respectively. At spinning frequencies of ~12–14 kHz, a good starting condition for optimization is to choose pl5 equivalent to $2.5 \times v_r$ (as in step 8), and optimize sp1 around $3.5 \times v_r$.

11. *Decoupling during N-CO CP.* Repeat the same experiment to optimize decoupling power during N-CO CP. Choose optimal conditions of step 10, and rerun the experiment as a function of pl13, as in step 9.

3.2.2. Setting the 3D NCOCX Experiment

Use the pulse sequence in Fig. 3a. Since every block of the polarization transfer path has been calibrated in previous steps, we can now setup the NCOCX experiment, which consists of six basic steps: (1) HN CP, (2) nitrogen t_1 chemical shift evolution, (3) band selective NCO CP (24), (4) CO t_2 chemical shift evolution, (5) CO-CX (CX = CA, CB, etc.) transfer, and (6) CX chemical shift t_3 evolution. All CP conditions have already been optimized (see Notes 8 and 9).

1. Set carrier frequencies for carbon, proton, and nitrogen to 120 ppm, 5 ppm, and 120 ppm, respectively.

2. Use the sp0, pl10 parameters optimized in step 7 of Subheading 3.2.1 to set HN CP.

3. Set the spectral width to 50 ppm during ^{15}N chemical shift evolution. Either TPPI (time proportional phase incrementation) or States-TPPI can be used for phase sensitive detection in the indirect dimension. For a homogeneously reconstituted sample, the number of points should be chosen to achieve the length of the indirect dimension equal to ~12 ms.

4. For N-CO CP transfer, the carbon carrier frequency should be switched to ~175 ppm in the pulse sequence. Use the sp1, pl5 parameters optimized in step 10 of Subheading 3.2.1 to set NCO CP and the decoupling power pl13 as optimized in step 11.

5. With the carbon carrier frequency set at 175 ppm during CO chemical shift evolution, set the spectral width to $0.76 \times v_r$ to avoid an overlap between the folded chemical shift anisotropy spinning side bands and the main spectrum. For a 14.3 kHz spinning rate and an 800 MHz field, this translates into ~50 ppm. Either TPPI or States-TPPI should be used for phase sensitive acquisition of the indirect dimension. The number of points should be set to achieve the total acquisition length of ~7 ms.

6. To set DARR (25) mixing, the power level pl15 should be set equal to the spinning frequency. Two spectra should be recorded with mixing times of 50 ms and 100 ms (see Note 9).

7. For decoupling during direct and indirect chemical shift evolution, use SPINAL64 with decoupling parameters pcpd2, pl12 optimized in step 4 of Subheading 3.2.1. Keep the direct acquisition time on the order of 20 ms (see Notes 12 and 13).

3.2.3. Setting the 3D NCACX Experiment

Generally, the protocol for setting an NCACX chemical shift correlation experiment is similar to that described for NCOCX. The experiment uses the same pulse sequence of Fig. 3a, but the experimental parameters for NCA CP, the indirect chemical shift evolution of CA, and the CACX mixing are different. Follow the procedure outlined in Subheading 3.2.2 with some modifications:

In step 4, the carbon carrier frequency should be switched to ~55 ppm in the pulse sequence. Use the sp1, pl5 parameters optimized in step 8 of Subheading 3.2.1 to set the NCA CP and the decoupling power pl13 as optimized in step 9 of Subheading 3.2.1.

In step 5, keep the carbon carrier frequency at 55 ppm during CA chemical shift evolution. Set the spectral width to 40 ppm (see Note 10). Either TPPI or States-TPPI should be used for phase sensitive acquisition in the indirect dimension. The number of points should be set to achieve the total acquisition length of ~6–7 ms.

In step 6, two spectra should be recorded with mixing times of 20 ms and 50 ms (see Notes 11–13).

3.2.4. Setting the CANCO Experiment

The experimental pulse sequence in Fig. 3b consists of (1) excitation of carbons by HC CP and $\pi/2$-pulse excitation, (2) CA t_1 chemical shift evolution, (3) CA-N CP, (4) nitrogen t_2 chemical shift evolution, (5) NCO CP, and (6) CO t_3 chemical shift evolution.

1. Set carbon, proton, and nitrogen carrier frequencies (o1p, o2p, o3p) to 175 ppm, 5 ppm, and 120 ppm, respectively.

2. For CA excitation and the subsequent chemical shift evolution and CA-N CP, change the carbon carrier frequency to 55 ppm in the pulse sequence. Set p3, pl3 to the values optimized in step 1 of Subheading 3.2.1.

3. Set pl11, p1 to the values optimized in step 2 of Subheading 3.2.1.

4. Set the HC CP parameters sp0, pl10, p15, to the values optimized in step 3 of Subheading 3.2.1.

5. For CA chemical shift evolution, keep the CA carrier at 55 ppm; set spectral width to 40 ppm; and choose the number of indirect acquisition points to result in 6–7 ms acquisition length (see Note 11).

6. Choose CA-N CP parameters sp1, pl5, p16, pl13 as optimized in steps 8 and 9 of Subheading 3.2.1.

7. Set the spectral width to 50 ppm during ^{15}N chemical shift evolution. Either TPPI or TPPI can be used for phase sensitive detection in the indirect dimension. The number of points should be chosen to achieve the length of the indirect dimension equal to ~12 ms.

8. For NCO CP transfer, reset the carbon carrier frequency back to 175 ppm, and set the CP parameters sp7, pl7, p17, pl23 to the values calibrated in steps 10 and 11 of Subheading 3.2.1.

9. For indirect chemical shift evolutions and for direct detection, use SPINAL64 decoupling with power level pl12 and pulse length pcpd2 set to those calibrated in step 4 of Subheading 3.2.1 (see Notes 12 and 13).

3.2.5. Data Processing   In preparation for further analysis, the 3D NCACX, NCOCX, and CANCO spectra should be processed to find the best compromise between the requirements of high resolution and optimal signal-to-noise. The sensitivity of the measurements depends on many factors, including the characteristic features of the studied molecule (i.e., size, amount of material, overall mobility, etc.) and the NMR spectrometer (i.e., magnetic field strength, type and state of probe, etc.). Processing should be optimized for each spectrum. From our experience with PR and ASR, good results are obtained through the combined application of Lorentizian-to-Gaussian broadening. Typical parameters, for a homogeneous sample with sub-ppm nitrogen line width and ~0.5 ppm carbon line width, are 0.2 ppm of Lorentzian line narrowing and 0.5 ppm of Gaussian line broadening on ^{15}N; 0.2 ppm of Lorentzian line narrowing and 0.4 ppm of Gaussian line broadening on ^{13}CO and ^{13}CA, which allow for improving both the resolution, especially at the baseline, and the signal-to-noise (Fig. 4).

4. Notes

1. We recommend to begin with well-behaved microcrystalline protein, such as SH3, GB1, etc. A good practice is to record a full data set on a model protein, process the data, and then analyze spectra to yield assignments. This helps to identify potential problems with the experimental setup and to learn the software tools.

2. We assume in the following that a structurally homogeneous protein sample is available.

3. Usually, it is impossible to reach zero, so one has to achieve the minimum signal intensity.
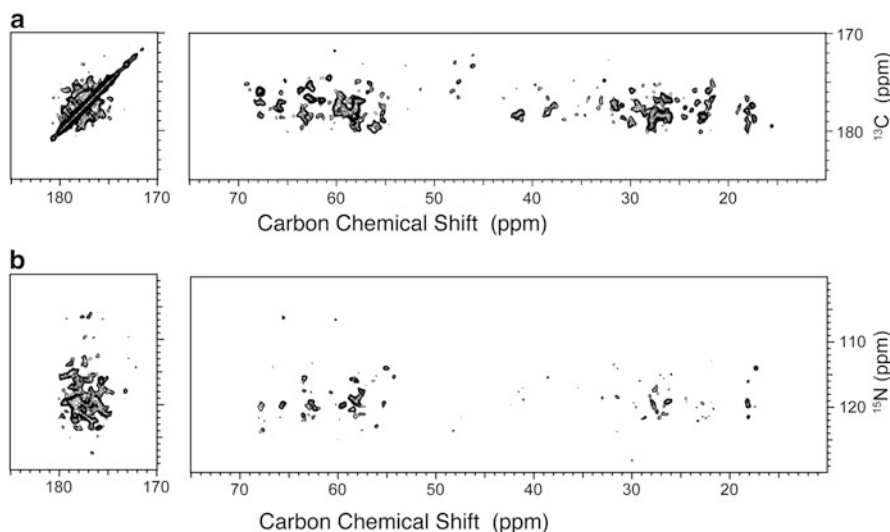
Fig. 4. Representative CO-CX and N-CX 2D spectra collected on ASR using the NCOCX pulse sequence in Fig. 2a and typical parameters discussed in the protocol. The data were collected at 800 MHz proton frequency, with eight scans, 1.8 s recycle delay, spinning frequency of 14.3 kHz. DARR mixing time was 100 ms. (**a**) CO-CX plane ($t_1 = 0$), total acquisition time is 40 min. (**b**) N-CX plane ($t_2 = 0$), total acquisition time is 40 min. Data were processed with Lorentzian-to-Gaussian window functions. The first contour is cut at $4 \times \sigma$, with the contour multiplication factor of 1.1.

4. Carbonyl carbons are chosen for optimization of HC CP because their Hartmann–Hahn condition is narrower.

5. In our experience, the efficiency of decoupling in proteins does not greatly depend on the choice of pcpd2, as long as it is close to $0.9{*}t_\pi$, where t_π is the $180°$-pulse length corresponding to the decoupling power level pl12 (23). Further optimization can be performed by observing ^{13}C spectra for different values of pcpd2.

6. Avoid matching the nitrogen power level to multiple spinning frequencies.

7. Typical efficiencies of NCA and NCO DCP in an immobilized protein should be on the order of 50 % or higher (intensity of DCP signal obtained using the pulse sequence of Fig. 2b divided by the intensity of the direct CP obtained using the pulse sequence in Fig. 2a). In reality, these can vary greatly depending on the state of the sample and precise experimental conditions. For example, NCA and NCO DCP efficiencies in proteorhodopsin were around 30 and 50 %, respectively.

8. Care should be exercised when high power equipment is used. In particular, one should avoid a long period of high power irradiation and of high duty cycles. Please consult your manufacturer for specifications of the probe. Do not exceed the recommended maximum power levels and duty cycle.

9. Choice of DARR mixing times given here is based on our experience with relatively large proteorhodopsin and *Anabaena* Sensory Rhodopsin proteins, which result in spectra of poorer sensitivity. In PR and ASR, DARR mixing of 50 ms results in detectable one- and two-bond transfers (CO-CA, CA-CB). A longer 100 ms mixing time was required to observe three- and four-bond correlations in the amino acids with long aliphatic side chains, such as valines, leucines, etc. One should also expect interresidue correlations, for example, CO[i]-CA [i + 1]. In smaller systems such as the first immunoglobulin binding domain of protein G, where signal-to-noise is much higher, such long mixing times can easily result in long-range (e.g., interresidue) transfers. Thus, the durations of the DARR mixing times should be optimized experimentally.

10. To cover the whole range of CA resonances without spectral folding or aliasing, 40 ppm is sufficient. If NCA transfer results in significant intensity for CB resonances during the optimization of N-CA transfer (as would be evident from significant intensity buildup in the ~20–30 ppm region), the spectral width would need to be set to a larger value (e.g., 50 ppm) to avoid folding of CB resonances and their possible overlap with CA peaks.

11. A DARR mixing time of 20 ms is required for the observation of one- and two-bond CA-CX transfers (e.g., CA-CB, CA-CO, and CA-CG) and 50 ms for longer transfers. Longer mixing frequently results in interresidue correlations CA[i]-CO[i–1], although they are weak. In lower molecular weight systems, shorter mixing times can be used to minimize interresidue transfers (see also Note 9).

12. We recommend using dummy scans at the beginning of each experiment in order to place the probe in a steady state condition.

13. Most 3D SSNMR experiments are relatively long and last a few days. Drift of the magnetic field is often linear. In such cases, we recommend using linear drift compensation. We also recommend letting the system stabilize for ~12 h after He fill and for a few hours after nitrogen fill.

References

1. McDermott A, Polenova T (2007) Solid state NMR: new tools for insight into enzyme function. Curr Opin Struct Biol 17:617–622

2. Ramamoorthy A (2009) Beyond NMR spectra of antimicrobial peptides: dynamical images at atomic resolution and functional insights. Solid State Nucl Magn Reson 35:201–207

3. Renault M, Cukkemane A, Baldus M (2010) Solid-state NMR spectroscopy on complex biomolecules. Angew Chem Int Ed Engl 49:8346–8357

4. Rienstra CM, Tucker-Kellogg L, Jaroniec CP, Hohwy M, Reif B, McMahon MT, Tidor B, Lozano-Perez T, Griffin RG (2002) De novo determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy. Proc Natl Acad Sci USA 99:10260–10265

5. Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. Nature 420:98–102

6. Pauli J, Baldus M, van Rossum B, de Groot H, Oschkinat H (2001) Backbone and side-chain C-13 and N-15 signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. Chembiochem 2:272–281

7. Igumenova TI, McDermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ (2004) Assignments of carbon NMR resonances for microcrystalline ubiquitin. J Am Chem Soc 126:6720–6727

8. Bockmann A, Lange A, Galinier A, Luca S, Giraud N, Juy M, Heise H, Montserret R, Penin F, Baldus M (2003) Solid state NMR sequential resonance assignments and conformational analysis of the 2×10.4 kDa dimeric form of the Bacillus subtilis protein Crh. J Biomol NMR 27:323–339

9. Li Y, Berthold DA, Gennis RB, Rienstra CM (2008) Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. Protein Sci 17:199–204

10. Schuetz A, Wasmer C, Habenstein B, Verel R, Greenwald J, Riek R, Bockmann A, Meier BH (2010) Protocols for the sequential solid-state NMR spectroscopic assignment of a uniformly labeled 25 kDa protein: HET-s(1-227). Chembiochem 11:1543–1551

11. Sperling LJ, Berthold DA, Sasser TL, Jeisy-Scott V, Rienstra CM (2010) Assignment strategies for large proteins by magic-angle spinning NMR: the 21-kDa disulfide-bond-forming enzyme DsbA. J Mol Biol 399:268–282

12. Helmus JJ, Surewicz K, Nadaud PS, Surewicz WK, Jaroniec CP (2008) Molecular conformation and dynamics of the Y145Stop variant of human prion protein. Proc Natl Acad Sci USA 105:6284–6289

13. Shi L, Kawamura I, Jung KH, Brown LS, Ladizhansky V (2011) Conformation of a seven-helical transmembrane photosensor in the lipid environment. Angew Chem Int Ed Engl 50(6):1302–1305

14. Tycko R, Hu KN (2010) A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning. J Magn Reson 205:304–314

15. Moseley HN, Sperling LJ, Rienstra CM (2010) Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of beta1 immunoglobulin binding domain of protein G (GB1). J Biomol NMR 48:123–128

16. Shi L, Ahmed MAM, Zhang W, Whited G, Brown LS, Ladizhansky V (2009) Three-dimensional solid-state NMR study of a seven-helical integral membrane proton pump—structural insights. J Mol Biol 386:1078–1093

17. Shi L, Lake EM, Ahmed MA, Brown LS, Ladizhansky V (2009) Solid-state NMR study of proteorhodopsin in the lipid environment: secondary structure and dynamics. Biochim Biophys Acta 1788:2563–2574

18. Sperling LJ, Nieuwkoop AJ, Lipton AS, Berthold DA, Rienstra CM (2010) High resolution NMR spectroscopy of nanocrystalline proteins at ultra-high magnetic field. J Biomol NMR 46:149–155

19. Pines A, Gibby MG, Waugh JS (1973) Proton-enhanced NMR of dilute spins in solids. J Chem Phys 59:569–590

20. Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. J Magn Reson 162:479–486

21. Andrew ER, Bradbury A, Eades RG, Wynn VT (1966) Resonant rotational broadening of nuclear magnetic resonance spectra. Phys Lett 21:505–506

22. Raleigh DP, Levitt MH, Griffin RG (1988) Rotational resonance in solid-state NMR. Chem Phys Lett 146:71–76

23. Fung BM, Khitrin AK, Ermolaev K (2000) An improved broadband decoupling sequence for liquid crystals and solids. J Magn Reson 142:97–101

24. Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. Mol Phys 95:1197–1207

25. Takegoshi K, Nakamura S, Terao T (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 344:631–637

Chapter 13

Wide-Line NMR and Protein Hydration

K. Tompa, M. Bokor, and P. Tompa

Abstract

In this chapter, the reader is introduced to the basics of wide-line NMR, with particular focus on the following: (1) basic theoretical and experimental NMR elements, necessary before switching the spectrometer and designing the experiment, (2) models/theories for the interpretation of measured data, (3) definition of wide-line NMR spectrometry, the description of the measurement and evaluation variants, useful hints for the novice, (4) advice on selecting the solvent, which is not a trivial task, (5) a note of warning that not all data are acceptable in spite of the statistical confidence. Finally, we wrap up the chapter with the results on two proteins (a globular and an intrinsically disordered).

Key words: Wide-line NMR spectrometry, Relaxation times, Protein hydration, Ubiquitin (UBQ), ERD10

1. Introduction

1.1. Preamble (Preambulum)

Despite rapid advance in experimental and theoretical studies on intrinsically disordered proteins (IDPs) (1–7), even basic issues such as definition of structural disorder, its clear-cut experimental demonstration and detailed structural characterization are fraught with a high level of uncertainty. Whereas a wide range of experimental approaches can be used for their characterization, none of the methods provide the complete solution to all static and dynamic aspects of the structural ensemble of IDPs. It is also of note that structural interpretation of experimental results is model dependent even in the case of globular proteins. The situation might be more serious in the case of IDPs, which are extremely heterogeneous in nature and are much more sensitive to environmental conditions.

An important point with IDPs is that a static structural picture is of little relevance due to their extreme level of dynamics. Whereas the structure of ordered proteins can be characterized at first approximation by the spatial coordinates of their atoms, IDPs are

in constant fluctuation and some parts of their atoms do have NMR-visible mobility at ambient (and even at lower) temperatures. Because their function is often related to diverse interactions with their environment, a detailed description of their surface (interfacial region) and its interaction with water molecules and other small cosolvent molecules is also necessary.

In this chapter, we outline wide-line NMR as a technique to tackle these aspects of IDPs. Traditionally, spectroscopic techniques build on the energy/frequency-dependence of the spectral response (Fourier transformed, time-dependent signal in the case of pulse excitation). Spectrometry builds on measuring several other parameters too, such as the intensity of global response, shape, width, and momentum of spectral lines, shifts and relaxation times. The intensity of the NMR signal provides information on the number of contributing nuclei, whereas chemical shifts, one of the basic quantities of high-resolution NMR, inform on electronic structural properties (on the local diamagnetic response to the applied magnetic field). The distances and angular dependent parameters of geometry are provided by the dipole–dipole interaction, i.e., they can be obtained from the line-width (second moment) of the signal or from the coupling constant in the relaxation times. In our interpretation, wide-line spectrometry means the experimental investigation of the dominant nonzero dipolar contribution to the NMR characteristics. Also having to consider dynamic aspects of structure, it is not sufficient to determine one parameter, but several need to be combined to provide a complete description in the phase space of statistical physics.

We briefly introduce the reader to the methods of wide-line NMR spectroscopy/spectrometry with special attention to the investigation of IDPs. Our special goal is to provide a complex method for describing the molecular mobility of proteins, their interactions, heterogeneity of binding, and motion of interfacial water molecules. Our aim is to demonstrate the differences between ordered and disordered states of proteins and to point out that the disordered state can be characterized by virtue of the binding heterogeneity of the protein surface by means of interfacial water molecules as atomic probes. Obviously, the extent of the chapter is limited, and therefore, the results on only two examples (UBQ and ERD10, dissolved in water) are presented. The effects and problems of buffered solutions and applied models are mentioned only for further investigation and discussions.

1.2. Elements of NMR: Minimum Know-How

1.2.1. Resonance Condition, Applied Fields, and the Units Used

Nuclear magnetic resonance (NMR) is the resonance absorption of electromagnetic waves by nuclear magnetic moments (nuclear spins) located in a time-independent magnetic field. The well-known "resonance condition," namely, the formula

$$\omega = \gamma B, \tag{1}$$

expresses the relation between the main scalar quantities in question, where ω is the frequency of the applied radiofrequency field, γ is the gyromagnetic ratio of the resonant nucleus (for protons $\gamma = 2.675222099 \ 10^8/\text{s/T}$), and B is the magnitude of the magnetic induction (field) at the position of the nucleus. The local field can be an "external field" (B_0) generated by a superconducting or electromagnet, or an "internal field" (B_{loc}) of hyperfine origin generated by magnetic moments or/and moving electric charges (currents). In measuring the field, NMR serves as a magnetometer or field stabilizer and in the latter case, we know the method as NMR spectroscopy/spectrometry—this being a powerful tool among the spectroscopic methods of materials research. NMR is suitable for identifying and qualifying structural units forming molecules and condensed matters rather than determining their long range structure. The combination of NMR and X-ray (and neutron) diffraction provides a more complete description of structure because diffraction methods are determined by long-range ordering and periodicity. Moreover, NMR is able to inform us about the local atomic or molecular movements. To help understand the physical basis of NMR, we first provide a concise dictionary of the physical quantities used.

In the radiofrequency range in which the spectrometers operate, there are a great number of photons in the radiation field even for the smallest power applied. In this case, induced emission and absorption dominate, and spontaneous emission is negligible. Consequently, the applied electromagnetic field can be described classically in the form of $B_1 \cos(\omega t + \phi)$, where B_1 is the amplitude, ϕ is the phase of the radiofrequency field, and t is the time. Both B_1 and ϕ are known exactly, without the limitation of a suitable Heisenberg's uncertainty relation (between photon number and phase), and the phase has information of equivalent value/content with the amplitude.

At this stage, it is necessary to say something about the units and dimensions used in the NMR literature. In this chapter and in the articles that have recently been published, SI units are used, but this is not the case in the older NMR literature, and the multitude of units used often poses difficulties for beginners. The crucial quantity, the magnetic field, is described by two *vectors*: magnetic field intensity **H** and magnetic induction **B** (see Note 1).

The relations between **B** and **H** vectors are

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{2}$$

in vacuum, and

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(1 + \chi)\mathbf{H}, \tag{3}$$

inside a magnetized sample (under the investigation), where $\mathbf{M}$ is the magnetization and χ is the susceptibility of the sample, and μ_0 is the vacuum permeability (see Note 1).

Since the discovery of NMR (8, 9), more than 60 years have passed and many scientists of various fields have worked in developing and using the method. Physicists tend to be at the frontiers of theoretical, methodological developments, and condensed state physical research, but chemists, biochemists, biophysicists, and nowadays physicians in the field of MRI (magnetic resonance imaging) are the people who are particularly concerned with applications of NMR in molecular structural research and in human diagnosis. The literature cited here is aimed at helping the reader to gain a deeper insight into the subject. First of all, let us mention the two "bibles" of the NMR literature written by Abragam (10) and Slichter (11) both giving a sound theoretical basis. The experimental principles are given in refs. 12–14, but not the constantly developing technical details. The principles and an excellent introduction to the Fourier-transformed NMR are given by Ernst et al. in ref. 15 and by Ferrar and Becker in ref. 16. Finally, the students and beginners in this field are recommended to consult the book of Hennel and Klinowski (17), a "simple but exact" introduction of NMR.

1.2.2. Nuclear Probes and Nuclear Paramagnetism

From the hyperfine structure of atomic spectra it is known that an atomic nucleus has its own angular momentum ($\mathbf{J}$) and its own intrinsic magnetic moment ($\boldsymbol{\mu}$). The two vectors $\boldsymbol{\mu}$ and $\mathbf{J}$ may be taken as "parallel." By introducing now the dimensionless nuclear spin vector $\mathbf{I}$, the angular momentum vector, $\mathbf{J} = \hbar\mathbf{I}$, thus the nuclear magnetic moment $\boldsymbol{\mu}$ can be expressed in the form

$$\boldsymbol{\mu} = \gamma\hbar\mathbf{I}, \tag{4}$$

where $\hbar = 1.05457151628 \ 10^{-34}$ J s (Joule·secundum) is Planck's constant divided by 2π, and γ, the gyromagnetic ratio is tabulated by nuclear physicists. Equation 4 expresses that the two vectors are "parallel," meaning that either they are parallel ($\gamma > 0$), or are antiparallel ($\gamma < 0$) and this fact simplifies the description of nuclear spin motion in magnetic fields. Different atomic nuclei can have different angular momentum and magnetic moment, including 0. Nuclei with an even number of protons and neutrons have $I = 0$, and these nuclei are not accessible by NMR. A number of important nuclei belong to this category, e.g., ^{12}C, ^{14}N, and ^{16}O. Nuclear physics offers us these quantities for all the stable isotopes which we are interested in (see Note 2).

If a macroscopic sample consisting of N_0 identical nuclear spins is immersed in a magnetic field $\mathbf{B}_0$ ($\mathbf{B}_0 \parallel z$, the axis of the laboratory frame of reference, LF) for an extended period, it will reach a state of equilibrium, where the distributions of nuclear spins on the

Zeeman energy levels (in the case of protons, there are two!) are described by Boltzmann's law. The law states that the population

$$N_m = c \, \exp(-E_m/k_B T), \tag{5}$$

where c is a proportionality constant, k_B is Boltzmann's constant, T is the absolute temperature of the sample, and $\sum N_m = N_0$.

In the state of thermal equilibrium the net magnetization of the sample is

$$\mathbf{M}_0 = \frac{\gamma^2 \hbar^2 I(I+1) N_0}{3 k_B T} \mathbf{B}_0, \tag{6}$$

where the magnetization vector $\mathbf{M}_0$ has the same direction as the magnetic field irrespective of the sign of γ. Equation 6 is known as the Curie-Langevin formula, and it enables N_0 (that is the quantity of participating nuclei) to be measured.

1.2.3. Classical Treatment: Vector Model

Many details of experimental NMR are easily understandable by using the classical description of the motion of the magnetic moment (nuclear spin) in static and/or time-dependent magnetic field(s). This description is equally valid for a single nuclear spin and for the $\mathbf{M}_0$ magnetization of the sample.

We have introduced here the rotating frame of reference (RF) with axes x', y', z', where $z' \parallel z \parallel \mathbf{B}_0$ that is z' coincides with the z axis of the LF and it rotates with an angular velocity $\boldsymbol{\Omega}$ with respect to LF. The equations of motion of $\boldsymbol{\mu}$ or $\mathbf{M}$ in both LF and RF have the same structure (11) (the real form is given in the Bloch equation, see Subheading 2.4), but the magnetic field in RF is

$$\mathbf{B}_{\text{eff}} = \mathbf{B}_0 + \boldsymbol{\Omega}/\gamma. \tag{7}$$

In a special case, where $\boldsymbol{\Omega} = -\gamma \mathbf{B}_0$, $\mathbf{B}_{\text{eff}} = 0$, the magnetic moment vector is constant in RF, i.e., it rotates at an angular velocity $\boldsymbol{\Omega}$ with respect to the LF. This is the so-called Larmor precession, and the two frequencies $|\boldsymbol{\Omega}|$ and ω_0 are equal.

We fix the x' axis of RF to the active rotating component of the radiofrequency field. In this case the effective field in RF in the case of resonance is

$$\mathbf{B}_{\text{eff}} = \mathbf{B}_1, \tag{8}$$

and the nuclear spin precesses in RF by an angular frequency

$$\omega_1 = \gamma B_1. \tag{9}$$

The motion of the spin or magnetization in LF is the combination of the two precessions. Equation 9 is the basis of the selection of the pulse length in NMR spectroscopy (see Subheading 2.2).

After a certain time spent in the radiofrequency field, the populations on the two Zeeman levels must be equal because the transition probabilities of absorption and induced emission are equal. In contrary in real samples, there is a net magnetization even in the case of radiofrequency excitation. The value of that is smaller than $\mathbf{M}_0$ (Eq. 6) but it is not 0. There must therefore exist a mechanism for inducing transitions between the Zeeman levels, which arises because of the coupling of the nuclear spins to the "lattice." The term lattice does not mean a crystal lattice but a thermal reservoir, which can take up (or give) energy from (or to) the spin system. This process can be characterized by the spin–lattice relaxation time T_1. The quantity T_1 is related to the microscopic details of both the lattice and the nuclear spin system. The lattice (and not the "spin system"!) is the part of the sample that is interesting for the biophysicists.

We have mentioned several quantities describing the motion of nuclear magnetization in magnetic fields and the connection of the spin system with the other degrees of freedom in the sample. Bloch (9) combined all the quantities in a relatively simple vector differential equation (see Note 3), which has the components

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{M}_z = \gamma[\mathbf{M} \times \mathbf{B}]_z + (\mathbf{M}_0 - \mathbf{M}_z)/T_1$$

$$r\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{M}_{x,y} = \gamma[\mathbf{M} \times \mathbf{B}]_{x,y} - \mathbf{M}_{x,y}/T_2 \,, \tag{10}$$

where the vector product components come from the equation of motion for noninteracting spins and the kinetic terms represent the interactions of spins with the lattice and with the other spins. T_1 involves energy exchange between the spins and the lattice. Bloch also realized that the internal interactions within the spin system conserve the spin energy and are represented by T_2, called the spin–spin relaxation time. The two relaxation times represent two time-windows for mapping the dynamical processes inside the spin–lattice system.

Generally speaking, the solution of these equations is a complicated task, but the so-called "steady state" solution is simple in a low $\mathbf{B}_1$ field and is often cited (see for instance in ref. 11). The form of the solution gives the absorption and dispersion modes of the $\mathbf{M}_{x,y}$ magnetization. These quantities represent the NMR signals (absorption and dispersion, respectively) which can be detected using an out-of-phase $(\phi = 90°)$ and an in-phase $(\phi = 0°)$ detection mode. The components are the basic quantities in quadrature detection and Fourier transformation of time-domain NMR signals. From the experimental point of view, we have to remember that a repetition of an experiment within a time period shorter than $5\,T_1$ can give a misleading result.

1.3. Interactions and Consequences

1.3.1. Local Magnetic Fields, Magnetic Interactions

The *magnetic interaction* between nuclear magnetic moments and the local magnetic field plays important role in the spin orientation dependent interaction of the nucleus with its surroundings in NMR. The magnetic interaction has the form of ($\boldsymbol{\mu} \cdot \mathbf{B}$), that is the ($\boldsymbol{\mu} \cdot \mathbf{B}$) scalar product gives the energy of the magnetic interaction. The induction of magnetic field $\mathbf{B}$ as was expressed in Eqs. 2 and 3 could contain contributions of "external" and "internal" origin. The internal components are expressed by $\mathbf{M}$ or χ, and are generated by the magnetic moments of the neighboring electrons or nuclei and by electric currents of moving charges. The external components may be either the applied fields created and used for experimental purposes. Both fields can be either time independent (static) or could depend explicitly on time. We shall meet the exact forms of the magnetic interaction energy later in this chapter. The magnetic interaction energy in operator form is, that is the Hamiltonian

$$\boldsymbol{H} = (\boldsymbol{\mu} \cdot \mathbf{B}). \tag{11}$$

In the previous sections, we have briefly introduced the reader to the two basic interactions which play roles in NMR, namely we got to know the Zeeman interaction and the interaction with the radiofrequency field. Besides these external interactions, the internal interactions are simply characterized by local fields (the effect of electric field gradient is omitted here) or by T_1 and T_2 within the framework of the Bloch theory.

Now we give a unified description of all the interactions (necessary for us) separating the "local fields" of different physical origin, using the Hamiltonian formalism. This formalism expresses the interaction of nuclei with their surroundings (including the lattice) in a simple way. The interaction Hamiltonian can be expressed as

$$\boldsymbol{H} = \boldsymbol{H}_Z + \boldsymbol{H}_{rf} + \boldsymbol{H}_d + \boldsymbol{H}_\sigma + \boldsymbol{H}_J, \tag{12}$$

where the subscripts denote the relevant interactions: *Z* Zeeman, *rf* radiofrequency, *d* direct dipole–dipole interaction, σ chemical shielding, *J* indirect nuclear spin-nuclear spin interaction (see Note 4).

If a magnetic field gradient is in $\boldsymbol{H}_Z$, it transforms the local coordinates to frequency values and serves the MRI and gives a chance for the measurement of translational atomic movements. From our point of view, $\boldsymbol{H}_d$ is the most important Hamiltonian.

1.3.2. Dipole–Dipole Interaction

This is the most important interaction between the two nuclear spins *I* and *S*. It can be expressed as

$$\boldsymbol{H}_d = \gamma_I \gamma_S \hbar^2 \frac{[(\boldsymbol{I} \cdot \boldsymbol{S}) - 3(\boldsymbol{I} \cdot \mathbf{r})(\boldsymbol{S} \cdot \mathbf{r})/r^2]}{r^3} \frac{\mu_0}{4\pi}, \tag{13}$$

where I and S are the spin operators, $\mathbf{r}$ is the position vector between the two spins. In the case of $I = S$, we can speak about homonuclear coupling, and if $I \neq S$, the coupling is heteronuclear. For a multispin system one has to add similar terms as in Eq. 13 between all the spin pairs. Van Vleck (18) gave the elegant theory of dipole–dipole interaction and he calculated the NMR spectrum parameters (moments) originating from the dipolar interaction. The Hamiltonian in polar coordinates may be expressed as a sum of six terms, the so-called "dipolar alphabet," and from these six terms the first two, viz.

$$A = -I_z S_z (3 \cos \Theta - 1),$$

$$B = \frac{1}{4}[I_+ S_- + I_- S_+](3 \cos^2 \Theta - 1), \qquad (14)$$

contribute to the so-called "truncated" dipolar Hamiltonian,

$$H_d^0 = \gamma_I \gamma_S \hbar^2 \frac{[A + B]}{r^3} \frac{\mu_0}{4\pi}. \qquad (15)$$

For heteronuclear spin pairs (e.g., ^{1}H and ^{13}C), normally only the term A contributes to the observed spectra, whereas both A and B contribute to the spectra of homonuclear spin pairs (as the two protons in a water molecule), due to the $I_+ S_-$ and $I_- S_+$ "flip–flop" terms in operator B. The $I_\pm$ and $S_\pm$ operators are the absorption and emission operators originating from I_x, I_y, S_x, and S_y, respectively (11). The flip–flop terms are important in the spin relaxation processes. Consequently, the equations describing the resonances for dipolar coupled spins are different for homo- and heteronuclear cases. There are very important consequences of the Van Vleck theory, namely:

(a) The theory cannot describe the NMR line shape function in an exact analytical form.

(b) The odd moments are all 0, and the consequences are symmetrical line shape and zero line shift.

(c) But it can offer the even moments, and the second moment is basic from topological point of view, and has the following form in frequency units for an example in which all spins are located in equivalent positions (details see in ref. 11).

$$M_2 = \frac{3}{4}\gamma^4 \hbar^2 I(I + 1) \sum_k \frac{1 - 3 \cos^2 \theta_{jk}}{r_{jk}^6}. \qquad (16)$$

Each term is of the order of $(\gamma B_{\text{loc}}^k)^2$ where B_{loc}^k is the contribution of kth spin to the local field at spin j. The important point about Eq. 16 is that it gives a precise meaning to the concept of local field, which enables one to compare a theoretical quantity with experimental value. The formula for nonidentical spins contains a

double-sum (details see in ref. 11), but any variant of that contains direct topological elements as the r_{jk} distance between the two interacting spins and the θ_{jk} angle between the $\mathbf{r}_{jk}$ and $\mathbf{B}_0$ vectors. In rigid solids, both r_{jk} and θ_{jk} are time-independent quantities and the theoretical second moment is calculable on the basis of a topological model. In the case of motion of a spin either in real or in spin space, the Hamiltonian will be averaged, and the consequence is "motional" and/or "spin-flip" narrowing of the spectrum, respectively (see ref. 11 and references cited therein). The second-rank Legendre polynomial term $(3\cos^2\Theta - 1)$ is responsible for the "motional averaging" in liquids and for the narrowing of NMR spectra in the case of macroscopic sample rotation (magic angle spinning, MAS). The average value of $(3\cos^2\Theta - 1)$ is 0 in the case of isotropic rotation and the uniaxial spin-pair motion results in a reduced line width and second moment.

1.3.3. Classification of Hamiltonians and Consequences

Before closing the Subheading 1.3, let us summarize the peculiarities of the Hamiltonian terms in Eq. 12 following refs. (19, 20).

The Zeeman term (including the gradient term), the chemical shift and radiofrequency terms are linear in spin operators. The off-resonance Zeeman term and the heteronuclear direct dipole–dipole interaction belong to this category, too. The homonuclear direct dipole–dipole interactions are bilinear in spin operators.

Another classification comes from the effect of the Hamiltonian on the line shapes, which can be homogeneous or inhomogeneous. A typical example of an inhomogeneously broadened line is a resonance broadened by the inhomogeneity of $\mathbf{B}_0$. If the applied field has spatial inhomogeneity, otherwise identical spins in different parts of the sample will have slightly different resonance frequencies. As a result, the NMR line will be broader because it is a superposition of lines at different frequencies. All the Hamiltonian terms, linear in spin operators, give similar broadening. On the other hand, a homogeneous line could come from the homonuclear dipole–dipole coupling and lifetime broadening. An important consequence of the classification is that the existence of different echoes depends on the existing Hamiltonian(s) and consequently on the nature of broadening.

1.3.4. Relaxation Times and Characteristic Parameters

Nuclear relaxation rate formulae for a single-phase (single-fraction) spin system are presented from the local field fluctuation model (11). Note, that the relaxation rate is the inverse of the relaxation time, $R_i = 1/T_i$. The longitudinal and transversal relaxation rates (R_1 and R_2, respectively) for a homogeneous sample (identical/single spin system) can be written as ((21–23) and references cited therein)

$$R_1 = \gamma_I^2 \sigma_1^2 \tau_1 / (1 + \omega_0^2 \tau_1^2), \tag{17a}$$

$$R_2 = \gamma_I^2 (\sigma_0^2 \tau_0 + \sigma_1^2 \tau_1 / 2(1 + \omega_0^2 \tau_1^2)), \tag{17b}$$

and

$$R_2 - R_1/2 = \gamma_I^2 \sigma_0^2 \tau_0. \qquad (17c)$$

The equations are derived for a statistical sequence of rectangular pulses of local fluctuation of magnetic induction $B_q(t) = \pm a_n$ with a mean jump time τ_q (Poisson process) and a statistical amplitude distribution of $\langle a_n \rangle = 0$ and $\langle a_n^2 \rangle = \sigma_q^2 = \langle B_{\text{loc}}^2 \rangle/3$.

Equation 17c provides a simple opportunity for the experimental determination of the secular contribution (see the definition in ref. 11) to the transversal relaxation rate $R_2' = R_2 - R_1/2$. The quantity is independent of ω_0 and the temperature dependence comes from that of τ_0. The mean jump time τ_q follows the Arrhenius law as

$$\tau_q = \tau_{q0} \exp(E/RT), \qquad (18)$$

where E is the activation energy of the molecular motion and R is the molar gas constant. The longitudinal relaxation rate has an extreme at $\omega_0 \tau_q = 1$. The interaction strength σ_q^2 can be determined from the maximum R_1 value. The ratios R_2/R_1 at the extreme of R_1 give substantial information on the homogeneity of the spin system. There are two important deviations, firstly, if a distribution of correlation time τ_q exists and secondly, if two substantially different correlation times τ_{q1} and τ_{q2} describe the otherwise homogeneous spin system. The consequences of the first one are that the R_1 vs. T curve around the maximum is not so sharp and the distance between R_1 and R_2 on the R_i vs. T map is increased with respect to τ_q without distribution. In the second case, there are two terms independent of ω_0 in the R_2 expression with two different σ_{qi}^2 and τ_{qi} values.

The decay of magnetization, that is the relaxation functions (longitudinal and transversal, $i = 1$ and 2, respectively) for a single-fraction spin system are of the form

$$M_i(t) = M_0 \exp(-R_i t), \qquad (19)$$

where t is the time and M_0 is the equilibrium nuclear magnetization. M_0 is proportional to the number of contributing nuclear spins and the exponent R_i is the relaxation rate. R_2 represents a narrower time window than R_1 for the low-frequency motions.

We recall the variant of the transversal relaxation function, which is valid in the case of inhomogeneous magnetic fields and for those samples where the translational diffusion of resonant nuclei is not negligible. The following formula (11, 16) is valid in the case of a uniaxial magnetic induction gradient $\partial B/\partial z$.

$$\ln[M_2(t)/M_0] = -\tau^2 F\left(\frac{\partial B}{\partial z}, D(t)\right) - t R_2, \qquad (20)$$

where $t = n2\tau$ is the time measured between the exciting pulses in the CPMG train in 2τ units and the first exponent is the F function of field gradient and the D diffusion constant characteristic to

translational diffusion. We reproduced the formula here in a short form because we would like to use it only for the demonstration of the existence or absence of the translational diffusion.

The relaxation functions for a two-phase system (phases *a* and *b*) at intermediate exchange rates between the phases are formally similar to that given in Eq. 19 but they are constituted of probabilities and relaxation rates having different complicated form and meaning (for the details see Eqs. 2.1–2.5 in ref. 23). The formalism in the case of *slow exchange*, where $\tau_{a,b} \gg 1/R_{ia,b}$ (that is for two-exponential relaxation) is however as simple as follows.

$$M_i(t) = p_a \exp(-R_{ia}t) + p_b \exp(-R_{ib}t), \tag{21}$$

where $p_a + p_b = 1$ are the abundance probabilities of the fractions and $R_{ia,b}$ denotes the relaxation rates.

2. Experimental Aspects of NMR Spectrometry

2.1. Experimental Aspects in General

This section deals with NMR spectrometry with different line shapes and in the field of pulsed NMR spectrometry the following items are mentioned: the free induction signal (FID) and its Fourier transformation, the Fourier spectrum of a relatively short radio frequency pulse, and the connection between the NMR spectrum (frequency-domain NMR signal) and the FID (time-domain NMR signal). In the following parts more complicated pulse sequences are briefly explained, e.g., spin echoes and pulse combinations for measuring the spin–spin and spin–lattice relaxation times. From the NMR methods, only those elements and variants are selected here which are necessary to understand the examples for demonstrating the present applications of protein research.

One thing that should be mentioned is the repetition time of the experiment. Before starting a new experiment, the nuclear spin system has to reach thermal equilibrium. The time constant of this evolution is $5\,T_1$. In some cases, T_1 can be very long, e.g., for ice that is important for us at low-temperature measurements, the order of magnitude of T_1 is about few tens of seconds (24) (depending on the magnitude of the static field). The electronic averaging of the NMR signals needs a great number of repetitions of the elementary experiment. Every step must start in thermal equilibrium.

It should also be mentioned that the high magnetic field and the low temperature increase the polarization of the nuclear spin system and consequently, the sensitivity of the method. When utilizing any version of NMR spectrometry, one needs a high-field magnet of good homogeneity (with known inhomogeneity) and stability; a radiofrequency transmitter of proper power, frequency,

and timing; a phase sensitive detector of good sensitivity; a cryo-system for controlling the sample temperature and a computer to arrange the whole business of collecting and evaluating the data.

The radiofrequency field $B_1\cos(\omega t + \phi)$ is produced by a coil with the sample inside and placed into the constant magnetic field in a plane perpendicular to $\mathbf{B}_0$. In the old cw NMR the amplitude of $B_1\cos(\omega t + \phi)$ is constant in time, i.e., we irradiate the sample continuously without timing the radiofrequency field. In this case, the irradiation field contains a component of single frequency. To measure the whole NMR spectrum, the sweep of the field $\mathbf{B}_0$ or of the frequency is necessary. At around the resonance, the nuclear spins absorb energy from the radiofrequency field, this absorption manifest itself as a loss in the coil, and we detect a change of voltage at the ends of the coil as an NMR signal. One can select the radiofrequency phase to detect either the dispersion ($\phi = 0°$) or the absorption ($\phi = 90°$) modes. The absorption line shape can be approximated in liquids by a Lorentzian function,

$$L(\omega)/L(0) = 1/[1 + (\omega/\delta\omega)^2], \qquad (22)$$

where $\delta\omega$ is a constant, its value and the function itself are predicted by the out-of-phase steady-state solution of Bloch equations 10.

In solids, the line shape is often approximated by a Gaussian, i.e., in normalized form

$$G(\omega)/G(0) = \exp(-\omega^2/2M_2), \qquad (23)$$

where M_2 is the second moment of the spectrum, and the corresponding half-width at half-height equals $[2\ln 2 M_2]^{1/2}$. The important property of a Gaussian is that its Fourier transform is also a Gaussian. Generally in solids, the line shape is neither Lorentzian nor Gaussian, and sometimes the approximations proposed by Abragam (10) or the analytic function of Harper and Barnes (25) are used. The latter is formally a stretched exponential function with a variable exponent of values between one and two. The information theory of line shape is given in the paper of Powles and Carazza (26).

For a liquid–solid mixture, a composite Lorentzian–Gaussian line shape is predictable.

The application of radiofrequency pulses in NMR has started in 1949 by Torrey (27) and in 1950 by Hahn (28, 29) and it was in 1966, that the pulses were first utilized for the Fourier transformed NMR spectroscopy (30). When the pulse ends, the rotating x–y component of the nuclear magnetization induces an oscillating voltage of the resonance frequency in the receiver coil (which is often the same as the transmitter coil). This voltage can be observed as the free induction decay (FID). If more than one pulse is applied or the experiment is repeated many times, the phase difference between consecutive pulses remains constant, i.e., the pulses are coherent;

and the phase of the phase-sensitive receiver system is also fixed to the transmitter phase. Consequently, the information in the phase in the up-to-date pulse NMR spectrometry is of equal value to that in amplitude. In the different pulse NMR experiments, we must define the duration (length) and the phase of the applied pulses.

As we have seen in connection of the RF system (Eq. 9) the angle traversed by the nuclear magnetization during the duration of the pulse τ is

$$\alpha = -\gamma B_1 \tau. \tag{24}$$

The most frequently used pulses are the $\pi/2$ and π pulses, where the pulse angle α is made equal to $\pi/2$ or to π. If the $\pi/2$ pulse is denoted shortly by P, then P_x means a pulse where the phase of the rotating component of B_1 and x' (i.e., the x' axis of the RF) coincides. The meaning of P_{-x}, P_y, and P_{-y} are consequently clear.

The P_x pulse moves the magnetization from its equilibrium position along the $\mathbf{B}_0 \parallel z$ axis to the position coincident with y' (the immediate P_{-x} pulse reverses the motion). Upon the termination of the P_x pulse, the magnetization moves together with the y' axis of the RF, in the LF it precesses around $\mathbf{B}_0$ and its transverse component $\mathbf{M}_{x,y}$ induces voltage in the coil known as the FID, which is a time-domain function. The π-pulse moves the nuclear magnetization from its position to the antiparallel direction. A pulse of finite length (τ) and of a well defined frequency ω contains a continuous spectrum of many frequencies. This is because a truncated cosine function is no longer a cosine but a more complicated function, and this function contains components in the

$$\omega \pm 2\pi/\tau, \tag{25}$$

range in the frequency-domain. It is clear that the $\tau \to \infty$ limit gives back the pure single frequency cosine function. Otherwise, the shorter the τ, the broader is the frequency distribution of the pulse. The local fields could be different from one spin to another because of the different chemical shifts, dipole–dipole interaction, inhomogeneity of magnetic field, etc. Consequently, the resonance condition cannot normally be satisfied exactly for all spins. The shorter the τ and consequently the larger the B_1, the smaller is the inaccuracy. In other words, a selective pulse means a narrow frequency spectrum, a hard or nonselective pulse means a broad one.

Fourier transformation connects the frequency-domain (cw spectrum) and the time-domain (e.g., FID) results by the well known formalism

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \, \exp(-i\omega t)\mathrm{d}t, \tag{26}$$

and

$$f(t) = \int_{-\infty}^{\infty} F(\omega) \, \exp(i\omega t) \mathrm{d}\omega, \tag{27}$$

where $f(t)$ is the free induction decay, and it is often derived from two simultaneous measurements (quadrature detection), one of which determines its real part, the other its imaginary part. If one of these is 0 at all t, a single measurement is sufficient to determine the value of $f(t)$. It is proved that the Fourier-transform of the FID gives the spectrum, so the pulse and cw methods give equivalent results. The two functions 26, 27 in the frequency-domain have the following form, for the Lorentzian spectrum

$$f(t) = \exp(-t/T_2), \tag{28}$$

and for Gaussian line shape

$$f(t) = \exp\left(-\frac{M_2}{2} t^2\right), \tag{29}$$

where t is the time, T_2 is the spin–spin relaxation time, and M_2 is the second moment of the spectrum. More information can be found in ref. 26.

Both NMR experiments consist of exciting the spins by a radio-frequency field and observing the response to the excitation. In the old cw method, the spins are excited by monochromatic radiation, whose frequency is linearly swept through the whole spectral range. In the pulse method, the entire spectrum is excited simultaneously. The pulse method is much faster, but has a real drawback because of the dead time of the receiver system. The loss of the early part of the FID may lead to distortion of the tail of the spectrum. This drawback is serious for solids (e.g., in proteins or in ice). The echoes can help to decrease the drawback.

2.2. Relaxation Time Measurements and Echoes

2.2.1. T_1 Measurements

There are three popular pulse sequences for the measurements of spin–lattice relaxation time: the $2P_x$–t_0–P_x, the P_x–t_0–P_x, and the n (P_x)–t_0–P_x pulse sequences. In each of these sequences the first pulse (or group of pulses) prepares the spins and the second pulse measures the evolution of magnetization after a waiting period.

In the $2P_x$–t_0–P_x (inversion recovery) sequence, the first pulse inverts the spin population and the recovery goes from $-M_0$ to $+M_0$, so the range of magnetization is $2M_0$ during the measurements. On the other hand, the drawback of this sequence is that the $2P$ (or π) pulse is twice as long as the P_x pulse, and the frequency width of the preparation-pulse power spectrum is only half as wide as that of a P_x pulse, consequently. From this point of view the P_x–t_0–P_x sequence gives much more accurate line shapes and intensities for a broad spectrum. The P_x–t_0–P_x sequence is called saturation recovery, and a well-known variant of this is the saturation

comb-recovery sequence, where the preparation pulse is replaced by a series of P_x pulses producing the full saturation of Zeeman levels. In any of these methods, the sampling pulse is a P_x pulse, and a FID signal which follows the sampling pulse gives information about the evolution of the magnetization.

2.2.2. Two- and Three-Pulse Echoes

The excitation of the spin system by two radiofrequency pulses with a time delay t_0 shorter than the spin–lattice relaxation time produces echoes which have the form of two FIDs back-to-back, and the position of the echo maximum is "far" from the end of the second exciting pulse. So the echo is free from the consequence of the dead time distortion and contains all the information which is involved in the FID. The existence and the information content of the echo depend on the applied pulse sequence and on the state of the material in which the spin system is embedded.

Hahn (28, 29) as a PhD student was the first to produce echoes using a P–t_0–P–t_0–echo pulse sequence. Using the vector model, the Carr–Purcell sequence, i.e., the P–t_0–$2P$–t_0–echo sequence, can be followed very easily in the understanding of the echo evolution (31). The repetition of the second pulse can produce an echo train which is time consuming, and serves as a R_2 measuring method. A $90°$ phase shift between the P and $2P$ pulses gives the Carr–Purcell–Meiboom–Gill (CPMG) (32) method which is probably the most powerful in measuring short R_2 values. Both of these sequences (the Hahn and the Carr–Purcell) are effective in the motionally averaged state, i.e., in liquids or in solids in which there is a moving component.

Rigid spin systems, i.e., solids, cannot produce this type of echo, usually called spin-echo. It is possible to form echoes in solids by the sequence P_x–t_0–P_y–echo under the condition $t_0 \ll 1/R_2$. Unfortunately, these echoes, called "solid echoes" do not have such a simple physical explanation as does the spin echo. Anyhow, one has to remember that the solid-echo pulse combination refocuses only the spins which have nonzero dipole–dipole interaction. The method is useful because it overcomes the dead time problem. Original references are (33, 34), and a review is given in (35).

In a liquid-solid mixture both type of echoes (spin and solid echo) can in principle be detected providing information about both components.

2.3. The Applied Pulse Sequences, the NMR Responses, and the Types of Contributing Nuclei

(a) Preparation pulse, $P_x \Rightarrow$ response: FID signal, which is the inversely Fourier-transformed NMR spectrum. All the protons in the sample contribute to the FID signal. The amplitude of this response at time zero is proportional to $M_0 \propto N_0 B_0/T$. Similar terms can be applied for each component in the case of multi-component spin system.

Otherwise, all the following responses are proportional to the number (N_0) of the contributing resonant spins.

(b) P_x–t_0–P_x or $2P_x$–t_0–P_x pulse sequences give ⇒ nonequilibrium FID (for short t_0) and two methods for the longitudinal (spin–lattice) relaxation rate (R_1) measurements. R_1 connects the spins with the "lattice." The sample is considered here as a thermal reservoir and is denoted by the term lattice. This experiment serves as the basis of the generally used NMR dispersion $R_1(\omega)$ method (NMRD), that gives the correlation time τ_q and its temperature dependence, if $R_1(\omega,T)$ was measured.

(c) Preparation pulses, (P_x–t_0–P_x or P_x–t_0–P_y) and P_x–t_0–$2P_x$ or P_x–t_0–$2P_y$ ⇒ response: (Hahn-) Carr–Purcell-echo (H-CP) (28, 29, 31). The extension of the latter pulse sequence, as P_x–t_0–$2P_y$–$2t_0$–$2P_y$–$2t_0$–$2P_y$–$2t_0$–..., ⇒ response: CPMG (32) echo-train. Only those protons contribute to the response signal, for which the reduced dipolar-field (36) is smaller than the inhomogeneous contribution to the line widths, which is nearly zero because of the "motional narrowing" (11). In that case only the mobile protons are detected. The CPMG echo-train enables the measurements of transversal (spin–spin) relaxation rate R_2.

2.4. Practical Aspects for the Measurements

The intensity of the slow component of low-temperature ($T < 0°C$) wide-line ^1H NMR signals is the measure of the number of mobile protons or equivalently, the number of mobile water molecules. FID-signal, R_1 and R_2 relaxation rate measurements are done at temperatures in between +40°C and −70°C.

1. To measure the absolute quantity of water in the lyophilized (LPH) protein, a standard of known ^1H content can be used. The spin–lattice (longitudinal, R_1) relaxation rate is determined (see Note 5) and then knowing the R_1, the free induction signal (FID) is recorded (see Note 6).

2. The spin–lattice relaxation rate (see Note 5) is measured for the LPH protein and then knowing the R_1, the FID is recorded. The solid-echo signal is recorded with a t_0 set to the shortest possible. Both phase shifted (P_x–t_0–P_y–t_0–echo) and in-phase (P_x–t_0–P_x–t_0–echo) pulses are used (see Note 7). The P_x–t_0–P_y–t_0–echo sequence refocuses the signal of all the resonant nuclear spins. The P_x–t_0–P_x–t_0–echo sequence provides a control over the types of couplings between the nuclear spins since such an echo refocuses the signal of the inhomogeneously coupled nuclear spins only. The difference of these two types of echo signals gives the echo signal of the spins in homonuclear dipolar coupling. The spin-echo signal (P_x–t_0–$2P_y$–t_0–echo) is recorded with a t_0 set to be sufficiently long to give a well-shaped echo.

3. For the protein solution, the spin–lattice (longitudinal, R_1) relaxation rate is measured (see Note 5) and then knowing the R_1, the FID is recorded (see Note 6). R_1 should be measured at each temperature set and it is advisable to check for the one- or two-exponential nature of the magnetization decay curve (see Subheading 1.3.4). The zero-time extrapolated FID signal intensity detected at ambient temperatures serves as the 100% reference point for establishing the low-temperature mobile water fractions. It is important to have a few FID recordings at different T values above 0°C before and after freezing the protein solution. These data and the ones collected after the freeze–thaw cycle will serve as controls concerning the occasional damage caused by the freezing process.

4. NMR experiments are made at temperature intervals determined by the characteristics of the actual protein solution (see Note 8). The temperature is lowered gradually and NMR signals are recorded as needed. The protein solution is cooled down to −70°C and it is advisable to wait there some 30 min to be sure the whole solution is frozen completely, even the hydration water of the protein. The temperature is then raised gradually and NMR signals are recorded as needed. Small steps in the temperature (few degrees at once) can be applied in order to be able to catch the point where the first amount of hydration water becomes mobile. When recording CPMG-echo trains (P_x–t_0–[$2P_y$–t_0–echo–t_0–]$_n$) in order to get the R_2 values, it is checked whether the decay of the echo amplitudes depends on t_0. If yes, this is the sign of the onset of translational diffusion for water molecules and the temperature is of interest where it happens.

2.5. Data Evaluation

1. The FID signal of the standard is fitted to get the zero-time intensity. For solid/crystalline standards, a simple Gaussian fitted to the first some 50 µs can be used in most cases (see Note 9). The long-time part of the FID can be fitted with an (stretched) exponential (see Note 10) to get the water belonging portion of the signal at zero time.

2. The water content of the LPH protein is calculated by comparing the zero-time extrapolated signal intensities obtained. The intensity values should be corrected for temperature and pulse length beforehand (see Note 11).

3. The FID signal obtained is Fourier-transformed. The spectrum is made of a narrow component, which represents the water molecules and a wide component belonging to the protein protons. The narrow line corresponds to the long FID component while the wide line is yielded from the fast and medium components. The second moments of these spectral lines inform us about the motional states of the ^{1}H nuclei involved.
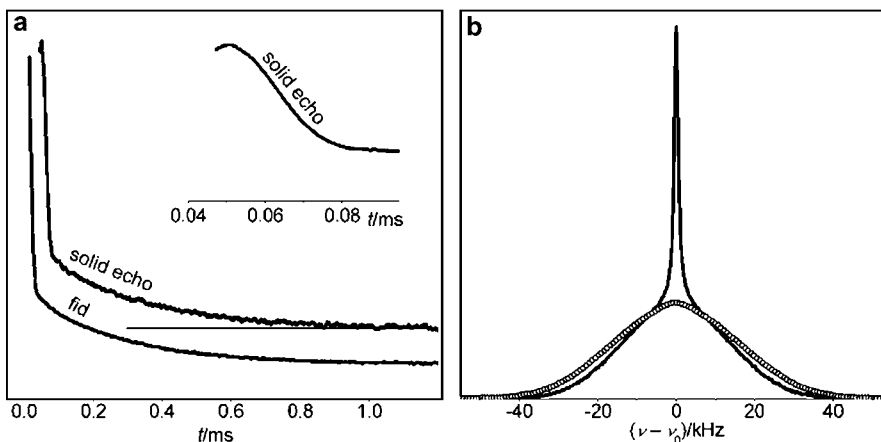
Fig. 1. (**a**) ^{1}H NMR FID and solid-echo signals for LPH ubiquitin at room temperature. The *insert* shows the central part of the solid echo. (**b**) ^{1}H NMR spectra for LPH ubiquitin at +24.8°C (*solid line*) and −74.5°C (*circles*) (37).

4. For the evaluation of the mobile water fraction of the FID signal measured for protein solutions, the following recipe can be used. The above-zero-temperature FID signal is used for setting the level where all of the ^{1}H nuclei in the protein solution are detected. The zero-time extrapolated and corrected (for temperature, P_x pulse length, receiver gain, and resonance frequency, whichever relevant) FID signal intensities measured above 0°C solve as the 100% level. The low-temperature (i.e., below 0°C) FID signals are made of three component as general (see Fig. 1). The component coming from the mobile water fraction has the longest time decay constant (see Note 12). This component belongs to the water molecules in the sample, the amount of which is the desired result, so the fitting range is set to times longer than 100–200 μs (see Note 13). A stretched exponential function can be used for fitting (see Note 10).

5. It is useful to get the Fourier-transformed ^{1}H NMR spectra for a few representative temperatures. The width of the spectra above 0°C can be taken as a measure of the inhomogeneity of the applied static magnetic induction B_0. Going down with the temperature, the spectral component corresponding to interfacial water widens to *fwhm* = 0.2–0.7 kHz. This means that the motional averaging is not as effective as in normal liquids, i.e., the motions are restricted. At the lowest temperatures, where there are no detectable molecular motions of water (and the signal intensity is very poor), the ^{1}H NMR signal comes from the moving end-groups of the amino acid residues, and the estimated value of the scale of 10 kHz corresponds roughly to that measured on a LPH protein sample.

6. The magnetization decay curves measured to obtain the relaxation rates R_1 are usually mono-exponential in the whole temperature range studied. The R_1 data measured at $T > 0°C$ are analyzed by applying the Arrhenius expression (Eq. 18) and we get the activation energy as a result. This activation energy means an averaged value for the motion of the water molecules. The R_1 data measured at $T < 0°C$ can be analyzed in the frame of the fluctuating local field model (see Eq. 17a) and the correlation time constant τ_{10}, the activation energy E and the value of σ_1^2 (see Note 14) can be obtained. This model is applicable in the temperature range, where the mobile water fraction is constant. When the mobile water fraction changes with the temperature, we have a situation, where water molecules involved in interactions of different strengths (or even of different nature, see Note 15) are the subjects of relaxation. The fluctuating local field model for a single relaxation channel (see Eq. 17a) cannot handle such situations.

7. If the echo amplitude series of the CMPG-trains, which were detected at the same temperature but with different t_0 values, decay with different apparent time constants, this is the clear sign that there is translational diffusion present. The amplitude series are then fitted simultaneously with Eq. 20 to get the true R_2 without the effects of the translational diffusion. Whether there is translational diffusion or not, it is important to check the amplitude series for the number of R_2 components. Depending on the actual protein/solvent system and the experimental settings, there can be one to three R_2 components. The fastest one with the smallest weight corresponds to the protein protons. The translational diffusion is irrelevant for this component. The next, slower R_2 component belongs to the water molecules located at the surface of the protein molecule. The slowest R_2 component comes from the bulk water (if there is any at the actual temperature).

2.6. Materials

NMR-signal intensity standard can be any material of known ^{1}H-content, e.g., liquid water, metal-hydrogen system, crystalline adamantane, etc. The standard is used also for setting the NMR spectrometer.

Protein solutions—sample preparation for NMR experiments. When calculating the mass of the lyophilized protein needed, its water content should be taken into the consideration, as established previously (see Subheading 2.4).

It is always useful to study solutions prepared with pure water only (if applicable), i.e., no buffering agent and/or added salt. The hydration results obtained for such protein solutions are devoid of complex effects due to the cosolutes. The reason of using two solvents is (37) that NMR parameters which characterize the

proteins of different structure depend on the solvent too, and our conclusion was that the protein–water accessible molecular surface and interfacial region also is solvent dependent.

The aqueous protein solutions were prepared by dissolving the proper amounts of protein in water and in Tris buffer, which contained 50 mM Tris and 1 mM EDTA at pH 7 (Tris: tris(hydroxymethyl)-aminoethane; EDTA: ethylenediamine-tetraacetic acid).

70–100 μl aqueous solution contained in closed teflon capsules was used.

For determining the amount and dynamical properties of water bound per unit protein, and having noted that the concentration measurement of IDPs is error-prone due to their unusual reactions with colorimetric dyes, we directly measured the amount of protein dissolved by determining the mass of samples lyophilized from distilled water. This measure provided the absolute concentration of the protein, which could be directly used for calculating the absolute average concentration of its constituent amino acids.

For each solution composition, NMR measurements on 3–5 samples prepared independently is necessary. Reproducible data within the given statistical errors are necessary for confidential results. Measurements are also necessary for a protein-free solvent (buffer) sample.

3. Examples

We explain in detail the individual steps necessary to carry out the technique through illustrative examples of proteins belonging to the classes globular and intrinsically disordered. The globular protein ubiquitin (UBQ) is a small, more or less spherical protein found in all eukaryotic cells where it labels misfolded/damaged proteins destined for degradation. The early responsive to dehydration (ERD) 10 is a highly hydrophilic (38) and intrinsically disordered protein, which is expressed in plants in certain very actively dividing tissues and ubiquitously under drying conditions (39). ERD10 is a typical representative of IDPs (40, 41).

3.1. Lyophilized Proteins   The room-temperature time-domain FID and echo signals of ^{1}H NMR (Fig. 1a) measured on lyophilized UBQ samples comprise three components of different time dependence and relative weight. The initial ($t = 0$) NMR amplitudes of the components are proportional to the number of ^{1}H nuclei in the relevant parts of the sample. The relative weights are: 85% fast, 3% medium, and 12% slow component. The fast component comes from the least mobile parts of the LPH UBQ sample. The middle component is of ^{1}H nuclei for which the dipolar interaction is averaged out much less effectively by motions than for the slow component,

but for a higher degree than for the fast component. The fast and the medium components are of the ^{1}H nuclei belonging to the protein. The protein molecules do not undergo translation or reorientation as a whole, only some reorientation motions of molecular groups of the amino acid residues are active in this state, the remaining parts of the protein molecules (mainly the backbone) are almost immobile (but not completely rigid) from NMR's point of view. The slow component is the response of mobile protons in the interfacial region, i.e., it comes from the water content of the LPH protein sample; this phase represents reorienting water molecules.

We calibrate the amplitude of the slow component with our standard and get finally the mass of water in our LPH protein sample. And finally, given the mass of the sample as known, we get the water content, which was $h = 0.145$ (12.7 wt%) for our manufacturer-delivered UBQ (see Note 16) (37).

The Fourier-transformed spectra of the time-domain components give more insight into the dynamics of the UBQ molecules and the water molecules in the LPH sample (Fig. 1b).

The fast time-domain NMR-signal component transforms into a broad spectrum line of *fwhm* = 32.0 kHz and the middle component corresponds to a narrower, 7.7 kHz wide line. The second moments (M_2, see Eq. 23) are 10.1×10^{-8} T^2 for the broad (see Note 17) and 0.4×10^{-8} T^2 for the narrower line. The M_2 value of the broad line refers to rotation around a symmetry axis performed by high-symmetry molecular groups (e.g., methyl), which becomes obvious when the second moments for a rotating and an immobile methyl group is compared (see Note 18). The intramolecular M_2 contribution of immobile methyl groups is 22.8×10^{-8} T^2, which is reduced by a factor of 0.25 by the rotation about the C_3 symmetry axis of the group (42). The middle component probably comes from even more mobile parts of the protein.

The time-domain long tail corresponds to a narrow line in the frequency domain (*fwhm* = 1.2 kHz, $M_2 = 0.1 \times 10^{-8}$ T^2). The water-origin of the signal is proved by the temperature dependence of the signal parameters as the spectrum width and intensity (Fig. 1b)—the narrow line measured at +23°C is absent in the spectrum measured at −70°C. The full width at half maximum of the narrow line is greater than that of hydration shell protons measured at low temperatures. This means that the mobility of the water molecules in the LPH protein is more restricted (wider spectrum) than in the first hydration shell at low temperatures (narrower spectrum). The water protons in a rigid (static) water molecule give a line of *fwhm* ~ 40 kHz and a second moment of ~23×10^{-8} T^2. The room-temperature time-domain FID and echo signals for ERD10 comprise also three, namely a fast, a medium, and a slow component as it was found for ubiquitin. The relative weights are: 87% fast, 6% medium, and 7% long

component. The *fwhm*s are 34.2 kHz for the fast, 7.4 kHz for the medium, and 2.6 kHz for the long components. The second moments are 11.4×10^{-8} T^2 for the fast, 0.5×10^{-8} T^2 for the medium, and 0.4×10^{-8} T^2 for the long components. The water content was found thus $h = 0.44 \pm 0.09$ (4.5 ± 0.6 wt%) for our LPH sample of ERD10 (see Note 16).

These two proteins belonging to different protein classes show very similar spectra in their lyophilized state. The only remarkable difference could be found in their broad spectrum components, which contain information on the proteins themselves. The difference between UBQ and ERD10 seems to be characteristic of the other globular and ID proteins too (43), and we use that as a working hypothesis in the fast room temperature distinction between them.

3.2. Protein Solutions: NMR-Signal Intensities and Spectra

The measured mobile water fractions $x = n(\text{mobile H})/n(\text{total H})$ (see Subheading 2.5, step 4) are given for UBQ in double distilled water solutions (and for comparison for pure water too) in Fig. 2a (37). The same results in the form of hydration ($h = $ g water/ g protein) are given in Fig. 2b. The mobile water fraction x is zero for up to 0°C in the water sample, i.e., molecular motion starts at the melting point of ice in the whole sample. The x vs. T data show a step for UBQ dissolved in water at about -46 ± 1°C, which suggests that detectable molecular motion starts at this temperature. This motion is probably the rotation of water molecules located at the surface of the protein, i.e., in the interfacial region. This means that the thermal excitation of "bound" (44) water molecules needs a lower energy in the interfacial region than the "free" water in ice! The amplitude of the NMR signal remains unchanged up to the melting point of bulk water (see Note 19), where the whole sample melts, i.e., x becomes equal to one ($h = 24.4$). The measured number of mobile water molecules in the interfacial region equals the hydration level corresponding to $h = 0.4$ ($x = 0.016$), required for protein activity. The temperature independent nature of x also demonstrates that there is no change at the molecular surface and consequently, there is no phase transformation in the protein structure at -46°C $< T < -5$°C.

The measured mobile water fractions are given for ERD10 in double distilled water solutions (and for comparison for water too) in Fig. 3a. The same results in the form of hydration ($h = $ g water/ g protein) are given in Fig. 3b. The x vs. T data show a step for ERD10 at about -46 ± 1°C in water (Fig. 3a), the detectable molecular motion starts at this temperature. This means that the thermal excitation (44) of water molecules "bound" to proteins needs a lower energy in the interfacial region than in hexagonal ice, and the transition temperatures for the present IDP molecule is the same as the UBQ but different (higher!) than for the other globular one. The amplitude of the NMR signal does not remain
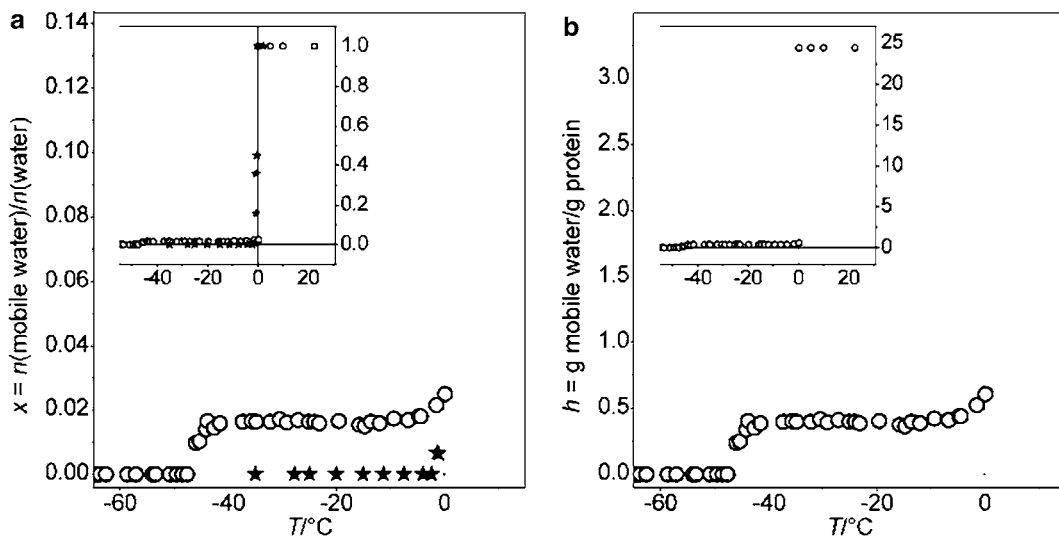
Fig. 2. (**a**) Mobile water fractions $x = n$(mobile H)/n(total H) determined from ^{1}H NMR signal intensities for 41 mg/cm^3 ubiquitin dissolved in water (*circles*) and pure water (*stars*). (**b**) Hydration (g water/g protein) calculated from the mobile water fractions for the ubiquitin solution presented in *panel* (**a**). The *inserted graphs* show the data on extended ordinates.
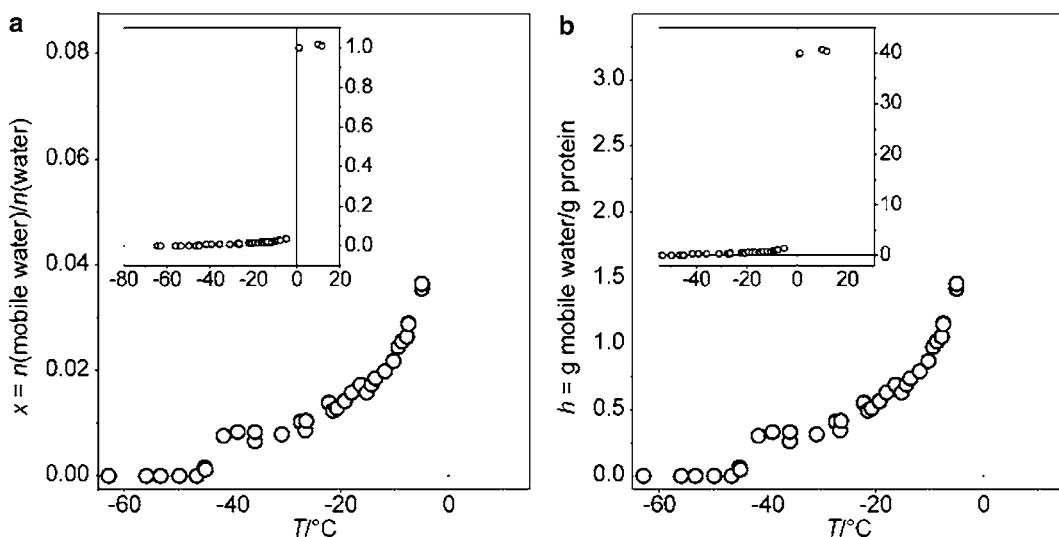
Fig. 3. (**a**) Mobile water fractions $x = n$ (mobile H)/n (total H) determined from ^{1}H NMR signal intensities for 25 mg/cm^3 ERD10 dissolved in water. (**b**) Hydration (g water/g protein) calculated from the mobile water fractions for the ERD10 solution presented in *panel* (**a**). The *inserted graphs* show the data on extended ordinates.

unchanged up to the melting point, where the whole sample melts, but there is a step-by-step increase (the number of mobile water molecules changes).

There are qualitative and quantitative differences between the globular and the disordered protein molecules as far as their protein–water interfacial behaviors are concerned. The increase of the

number of mobile water molecules can be described by exponential functions of temperature difference from 0°C for IDPs, while the exponent is nearly zero for globular ones. This fact reveals the heterogeneity of the IDP surface with regards to the distribution of interaction energy between water molecules and protein surface, as opposed to globular proteins, which are homogeneous in this regard.

3.3. Protein Solutions: Relaxation Rates

For UBQ dissolved in pure water, we found the activation energy of 20 ± 3 kJ/mol by the analysis of R_1 data (see Subheading 2.5, step 6) at $T > 0°C$ (Fig. 4), which is a typical value for water or aqueous solutions. In the range of subzero temperatures where the mobile water fraction is constant (Fig. 2), we applied the fluctuating local field model (see Subheading 2.4, step 3 and Fig. 4). We obtained $\langle B_{loc}^2 \rangle = (27.2 \pm 0.9) \times 10^{-8}$ T^2 which is comparable to the second moment of a static water molecule (see Note 14), and $\tau_1 = (7 \pm 8) \times 10^{-13}$ s $\exp(16 \pm 2$ kJ/mol/RT). The very high R_2/R_1 ratio at the R_1 maximum (Fig. 4) indicates that there works a second relaxation channel also, which is characterized by a much longer correlation time (11). For the UBQ dissolved in Tris-buffer, although we found the same activation energy above 0°C as for the UBQ/water system, but the analysis of the subzero temperature range yielded unreal parameter values $\langle B_{loc}^2 \rangle = (30.7 \pm 0.7) \times 10^{-8}$ T^2; $\tau_1 = (6 \pm 3) \times 10^{-17}$ s $\exp(35 \pm 1$ kJ/mol/RT). These extreme values show that the single-channel (sc) fluctuating local field model cannot be applied here as it was expected, taking into consideration the continuously changing mobile water fraction (37) and the NaCl cosolute.
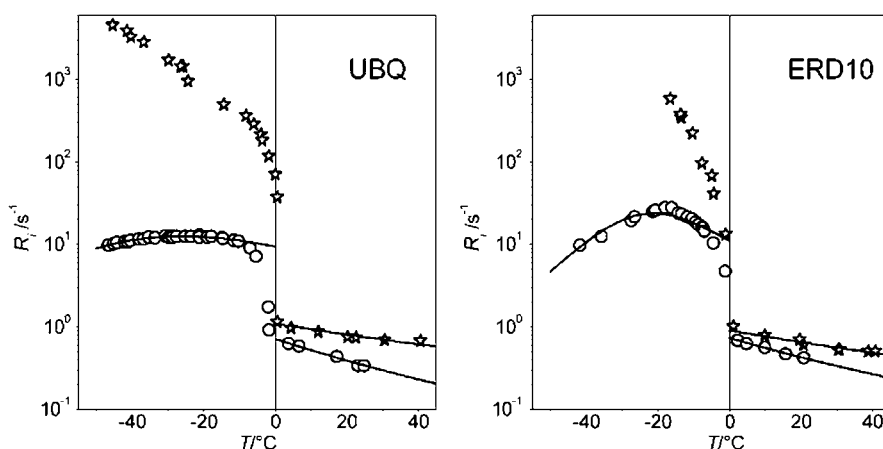
Fig. 4. ^{1}H nuclear spin relaxation rates for 41 mg/cm^3 ubiquitin (*left panel*) and for 25 mg/cm^3 ERD10 (*right panel*) dissolved in water: spin–lattice relaxation rate (R_1, *circles*) and bulk-water spin–spin relaxation rate (R_2, *stars*). The *solid line* corresponds to the fluctuating local field model. The spin–lattice relaxation (*circles*) and the spin–spin relaxation (*stars*) obey the Arrhenius law above 0 °C.

For the protein ERD10 dissolved in pure water, we found the activation energy of 18 ± 2 kJ/mol by the analysis of R_1 data (see Subheading 2.5, step 6) at $T > 0°C$ (Fig. 4), which is again a typical value for water or aqueous solutions. In the range of $-45°C < T < -25°C$, where the mobile water fraction has the smallest slope (Fig. 2), we applied the fluctuating local field model (Fig. 4). We obtained unreal parameter values as $\left\langle B_{loc}^2 \right\rangle = (51 \pm 9) \times 10^{-8}\,T^2$ and $\tau_1 = (3 \pm 15) \times 10^{-17}\,s\ \exp(37 \pm 9\,kJ/mol/RT)$. The sc fluctuating local field model fails in this case because of the continuously changing mobile water fractions (Fig. 3). The very high R_2/R_1 ratio at the R_1 maximum (Fig. 4) indicates that there works a second relaxation channel also, which is characterized by a much longer correlation time (11). For the ERD10 dissolved in Tris-buffer, we found a slightly higher activation energy above $0°C$ than for the ERD10/water system, $E = 25 \pm 1$ kJ/mol (41). The analysis of the subzero temperature range yielded similarly unreal parameter values as for ERD10/water $\left\langle B_{loc}^2 \right\rangle = (66 \pm 3) \times 10^{-8}\,T^2; \tau_1 = (6 \pm 5) \times 10^{-16}\,s\ \exp(30 \pm 2\,kJ/mol/RT)$. These extreme values show again that the fluctuating local field model cannot be applied here as it was expected taking into consideration the continuously changing mobile water fraction (41) and the NaCl cosolute.

For bulk-water R_2 components, for the proteins UBQ and ERD10 dissolved in either pure water or in Tris-buffer at $T > 0°C$, we found activation energies which are much lower (10 ± 1 and 11 ± 2 kJ/mol, respectively) than the E measured by the R_1 values (Fig. 4).

CPMG echo trains could be detected at $-26°C$ as lowest temperature for UBQ dissolved in water and translational diffusion (see Subheading 2.5, step 7) became detectable at the melting of the bulk water ($0°C$). For UBQ dissolved in Tris-buffer, CPMG echo trains were observable at temperatures as low as $-52°C$ and translational diffusion was present even at $-10°C$.

For ERD10 dissolved in water, the CPMG echo trains were detectable from $-20°C$ on and translational diffusion started with the melting of bulk water, at about $-1°C$. When the solvent was Tris-buffer, we found the corresponding temperature limits as $-33°C$ and $-20°C$, respectively.

The particularly important conclusion could be drawn from the data cited in this section that the solvent has a major role in the homo/heterogeneity of the protein–water interactions. It could be seen that the interaction of a globular protein with the water molecules showed a narrow distribution in energy when the solvent was pure water. The situation is completely changed when we dissolve the same globular protein in a salt-containing buffer. The presence of the ionic cosolutes results in energetically very heterogeneous protein–water interactions. The energies of the interactions between the intrinsically disordered proteins and the water molecules show a

rather broad distribution even when the solvent is pure water. The salt content of the solvent increases this heterogeneity even more. The markedly different trends of mobile water fraction vs. temperature data produces by the water-only solutions of globular and intrinsically disordered proteins can be used in practice for the distinction between these protein classes.

4. Notes

1. The definition of $\mathbf{H}$ is based on the generation of a magnetic field by an electric current. The unit of $\mathbf{H}$ is therefore A/m (ampere/meter). On the other hand, the definition of $\mathbf{B}$ is connected with the generation of 1 V (volt) electrical potential in a conducting coil (whose cross section is measured in m^2, $meter^2$) placed in a time dependent magnetic field. The unit of $\mathbf{B}$ is, therefore, 1 tesla (T) = 1 V s/m^2. In the older NMR literature, $\mathbf{H}$ is often expressed in oersted, and $\mathbf{B}$ in gauss, and the conversion factors are 1 Oe = $10^3/4\pi$ A/m and 1 G = 10^{-4} T. The advantage of these units is that the value of the vacuum permeability $\mu_0 = 1$ G/Oe. The same quantity in SI units is $\mu_0 = 4\pi \times 10^{-7}$ N/A^2.

2. In quantum mechanics $\mathbf{J}$, $\mathbf{I}$, and $\boldsymbol{\mu}$ are (vector) operators and their eigenvalues, i.e., their measurable values, can be expressed with the help of the spin quantum number I. All these quantities are used later in this chapter.

3. In quantum mechanics, the Liouville–von Neumann equation or more precisely its reduced form corresponds to the Bloch equations (see in (15) and the literature cited there) and the Redfield's relaxation superoperator describes the lattice and the spin–lattice interaction in principle. The relevant form of the equation of motion depends on the dominant interactions included in the Hamiltonian.

4. Metallic shift (H_K), bulk susceptibility shift (H_s), quadrupolar interaction (H_q) is missing from the line. Nevertheless H_K is important in metallic samples, H_s in extended objects with heterogeneous magnetic properties and in the case of bulk paramagnetism. H_σ will be mentioned because the anisotropic part of that contains indirect structural elements and together with H_J, they are the substantial interactions in high-resolution NMR spectroscopy. H_σ is a tensor quantity, and its eigenvalues depend on the orientation of the sample with respect to $\mathbf{B}_0$ in solids (or in anisotropic medium). The motion of electrons around the resonant nucleus generates a magnetic field which modifies the applied field $\mathbf{B}_0$. The induced magnetic field is proportional to the strength of the external field and the

electrons generally shield the nucleus from the applied field. In liquids, only the isotropic average is measurable, but in solids the chemical shielding varies with the orientation of the sample in the magnetic field. In a single crystal, this leads to a variation in the chemical shift as a function of orientation. In a powder, all possible orientations of the molecules or of the crystallites contribute to the observed NMR signal, and an asymmetric powder line shape results. In liquids or in the case of sample rotation, the isotropic average of the proper quantity is measurable. In our further discussion, H_σ and H_J play no role at all, and we shall omit them.

5. The $T_1 = 1/R_1$ value is needed to avoid the saturation effect when recording the FID. When using water as a standard, it is practical to dissolve in it some $CuSO_4$ to shorten the T_1 of water. Adamantane has T_1 of 1–2 s at room temperature and $\omega_0/2\pi = 83$ MHz. The best method is to record the time-domain signal intensity at a given time and to use these data for obtaining the actual T_1 value. The integral area of the Fourier-transformed spectrum is burdened with spin–spin (transversal) relaxation effects.

6. One needs to check the length of the P_x pulse from time to time in order to get reliable intensity data.

7. The short inter-pulse delay (short relative to the length of the FID signal) is needed to get signal of the dipolar-coupled homonuclear spin system refocused.

8. If the solution contains salt (e.g., NaCl) and the salt-to-protein molar ratio is not low enough, a hysteresis phenomenon may occur (40). This is caused by the eutectic phase made of the "free" salt (i.e., not bound to the protein) and water which will freeze and melt with hysteresis. In this case it is useful to have data both in cooling and heating directions.

9. In the case of adamantane, it is useful to apply a composite function made up of a damped sine (sine cardinal) multiplied with a Gaussian plus (spin-doublet of CH_2 groups) a stretched exponential ("lonely" spins of CH groups): $h^*\exp(-(A^*(t-t_0))^2/2)^*\sin(B^*(t-t_0))/(B^*(t-t_0)) + g^*\exp(-(C^*(t-t_0))^d/2)$. The parameter t_0 is set to the half of the length of the P_x pulse. The zero-time intensity equals with $h + g$.

10. The stretched exponential function is $A^*\exp(-1^*(|t-t_0|/T_{eff})^c)$. The parameter t_0 is set to the half of the length of the P_x pulse.

11. This step is only needed when different temperatures and/or pulse lengths were applied. To correct for the temperature, multiply the zero-time intensity value with the absolute temperature. To correct for the pulse length, multiply the zero-time intensity value with the lengths of the P_x pulse.

12. The half width at half maximum (*hwhm*), i.e., the time when the actual intensity is the half of the zero-time value is a good indicator for discriminating between the various FID components. The *hwhm* is at least 100 μs for the mobile water FID-component. This is a somewhat arbitrary limit it can be refined according to the actual circumstances.

13. At this time, the FID components for the ice and the protein itself are decayed fully and give no detectable contributions.

14. The parameter $\sigma_q^2 = \langle B_{\text{loc}}^2 \rangle / 3$ informs us about the magnitude of the fluctuating local field. In the case of homonuclear dipolar interaction, the average internuclear distance, r can be calculated from it as $\langle B_{\text{loc}}^2 \rangle = \gamma_I^2 \left(\frac{\mu_0}{4\pi}\right)^2 I(I+1) r^{-6}$. $\langle B_{\text{loc}}^2 \rangle$ corresponds to the second moment of the static spin-pair (e.g., ^{1}H nuclei of a water molecule).

15. If the solvent is not pure water, i.e., it contains other solutes as well, water molecules can experience fluctuating local fields not of dipolar origin only. The common additives, sodium and chloride ions, e.g., have quadrupolar nuclear moments and generate local fields of distinctly different nature than the dipolar interactions accordingly.

16. The water content of the LPH protein is well below the limit required for biological functioning or for full water coverage. It is estimated that $h = 0.2$–0.4 is sufficient to cover most of the protein surface with a single layer of water molecules and to fully activate the protein functionality (45, 46).

17. The M_2 of the broad line can also be estimated from the second derivative of the relevant solid echo component (32, 34) with the same result.

18. The second moment values obtained were found to correspond to rotating methyl and possibly to reorienting methylene groups. This is in accordance with the conclusions of Russo et al. (47) who used elastic and quasielastic neutron scattering experiments to investigate the dynamics of methyl groups in a protein-model hydrophobic peptide in solution. Their results proved that a hydrophobic side-chain requires more than a single layer of solvent to attain the liquid-like dynamical regime. They showed that when hydrophobic side chains are hydrated by a single water layer, the only allowed motions are attributed to simple rotations of methyl groups. The result (37) is also supported by the MDS of Curtis et al. (48) which proved that water is required for the methyl groups to exhibit the full range of motion characteristic of the liquid-like protein state at room temperature. Their study provided direct evidence that both the amplitudes and distribution of motion are influenced by the solvent and is not exclusively an intrinsic property of the protein molecule itself.

19. The constant concentration of mobile water molecules measured by us in water solvent is 187 ± 2 mol H_2O/mol UBQ at $-45°C < T < -5°C$; this hydration value can be compared with the number estimated by the sum of the hydration values of the individual amino acids needed to build the protein molecule. Kuntz proposed amino acid hydrations deduced from cw NMR experiments for polypeptides at $-35°C$ (49). The application of these numbers give 200 ± 30 mol H_2O/mol UBQ. Two other works which cite Kuntz's data give somewhat different amino acid hydrations and they result in 193 ± 29 (50) and 222 ± 47 mol H_2O/mol UBQ (32). The good agreement between our directly measured hydration and the estimated ones is surprising.

Acknowledgments

References

1. Tompa P (2009) Structure and function of intrinsically disordered proteins. University of Cambridge, UK

2. Uversky VN, Longhi S (eds) (2010) Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation. Wiley, Hoboken, NJ

3. Saito N, Kobayashi Y (2001) The physical foundation of protein architecture. World Scientific, Singapore

4. Antzutkin ON (2001) Molecular structure determination: applications in biology. In: Duer MJ (ed) Solid-state NMR spectroscopy: principles and applications. Wiley-Blackwell, Oxford

5. Shimonson T (2003) Electrostatics and dynamics of proteins. Rep Prog Phys 66:737–787

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

7. Halle B (2004) Protein hydration dynamics in solution: a critical survey. Philos Trans R Soc Lond B: Biol Sci 359:1207–1224

8. Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. Phys Rev 69:37–38

9. Bloch F, Hansen WW, Packard M (1946) Nuclear induction. Phys Rev 69:127

10. Abragam A (1961) The principles of nuclear magnetism. Clarendon, Oxford

11. Slichter CP (1990) Principles of magnetic resonance. Springer, Berlin

12. Fukushima E, Roeder SBW (1993) Experimental pulse NMR: a nuts and bolts approach. Westview, Boulder, CO

13. Andrew ER (1955) Nuclear magnetic resonance. Cambridge University Press, Cambridge

14. Mehring M (1976) High resolution NMR spectroscopy in solids. In: Diehl P, Fluck E, Kosfeld R (eds) NMR: basic principles and progress, vol 11. Springer, Berlin

15. Ernst R, Bodenhausen G, Wokaun A (1991) Principles of nuclear magnetic resonance in one and two dimensions. Clarendon, Oxford

16. Ferrar TC, Becker ED (1971) Pulse and Fourier transform NMR. Academic, New York, London

17. Hennel JW, Klinowski J (1993) Fundamentals of nuclear magnetic resonance. Longman Scientific and Technical, Harlow

18. Van Vleck JH (1948) The dipolar broadening of magnetic resonance lines in crystals. Phys Rev 74:1168–1183

19. Power WP, Wasylishen RE (1991) NMR studies of isolated spin pairs in the solid state. In: Webb GA (ed) Annual reports on NMR spectroscopy 23:1–84. Academic, London

20. Cory DG (1992) Solid state NMR imaging. In: Webb GA (ed) Annual reports on NMR spectroscopy 24:87–180. Academic, London

21. Noack F (1971) Nuclear magnetic relaxation spectroscopy. In: Diehl P, Fluck E, Kosfeld F (eds) NMR basic principles and progress 3:83–144. Springer, Berlin

22. Spiess HW (1978) Rotation of molecules and nuclear spin relaxation. In: Diehl P, Fluck E, Kosfeld F (eds) Dynamic NMR spectroscopy, NMR basic principles and progress 15:54–214. Springer, Berlin

23. Pfeifer H (1972) Nuclear magnetic resonance and relaxation of molecules absorbed on solids. In: Diehl P, Fluck E, Kosfeld F (eds) NMR basic principles and progress 7:53–153. Springer, Berlin

24. Barnaal D, Kopp M, Lowe IJ (1976) Study of HF doped ice by pulsed NMR. J Chem Phys 65:5495–5506

25. Harper WP, Barnes RG (1976) A useful nuclear magnetic resonance lineshape function for metallic solids. J Magn Res 21:507–508

26. Powles JG, Carazza B (1970) In: Coogan CK et al (eds) Magnetic resonance. Plenum, New York, NY, pp 133–161

27. Torrey HC (1949) Transient mutations in nuclear magnetic resonance. Phys Rev 76:1059–1068

28. Hahn EL (1950) Nuclear induction due to free larmor precession. Phys Rev 77:297–298

29. Hahn EL (1950) Spin echoes. Phys Rev 80:580–594

30. Ernst RR, Anderson WA (1966) Application of Fourier transform spectroscopy to magnetic resonance. Rev Sci Instrum 37:93–102

31. Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. Phys Rev 94:630–638

32. Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. Rev Sci Instrum 29:688–691

33. Lowe IJ (1957) Double pulse nuclear resonance in solids. Bull Am Phys Soc 2:344

34. Powles JG, Mansfield P (1962) Double-pulse nuclear-resonance transients in solids. Phys Lett 2:58–59

35. Mansfield P (1971) Pulsed NMR in solids. In: Emsley JW, Feeney J, Sutcliffe LH (eds) Progress in nuclear magnetic resonance spectroscopy, vol 8. Pergamon Press, Oxford, pp 41–101

36. Diakova G et al (2007) Changes in protein structure and dynamics as a function of hydration from ^{1}H second moments. J Magn Reson 189:166–172

37. Tompa K et al (2009) Interfacial water at protein surfaces: wide-line NMR and DSC characterization of hydration in ubiquitin solutions. Biophys J 96:2789–2798

38. Garay-Arroyo A et al (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. J Biol Chem 275:5668–5674

39. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Characterization of two cDNAs (ERD10 and ERD14) corresponding to genes that respond rapidly to dehydration stress in *Arabidopsis thaliana*. Plant Cell Physiol 35:225–231

40. Tompa P et al (2006) Protein–water and protein–buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. Biophys J 91:2243–2249

41. Bokor M et al (2005) NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. Biophys J 88:2030–2037

42. Grüner G, Tompa K (1968) Molekuláris mozgások vizsgálata szilárdtestekben NMR módszerrel. Kémiai Közlemények 30:315–356

43. Bokor M, Tompa P, Tompa K (2011) Wide-line NMR and relaxation characterization of interfacial water in protein solutions. In: 8th European Biophysics Congress, Budapest, Hungary, 23–27 August 2011

44. Cooke R, Kuntz JD (1974) The properties of water in biological systems. Annu Rev Biophys Bioeng 3:95–126

45. Gregory RB (1995) Protein–solvent interactions. CRC, New York

46. Teeter MM (1991) Water–protein interactions: theory and experiment. Annu Rev Biophys Biophys Chem 20:577–600

47. Russo D, Hura GL, Copley JRD (2007) Effects of hydration water on protein methyl group dynamics in solution. Phys Rev E 75:040902(R)

48. Curtis JE, Tarek M, Tobias DJ (2004) Methyl group dynamics as a probe of the protein dynamical transition. J Am Chem Soc 126:15928–15929

49. Kuntz ID (1971) Hydration of macromolecules. III. Hydration of polypeptides. J Am Chem Soc 93:514–516

50. Kuntz ID et al (1969) Hydration of macromolecules. Science 163:1329–1331

5-Fluoro-D,L-Tryptophan as a Dual NMR and Fluorescent Probe of α-Synuclein

Candace M. Pfefferkorn and Jennifer C. Lee

Abstract

Analysis of conventional proton nuclear magnetic resonance (NMR) experiments on intrinsically disordered proteins (IDPs) is challenging because of the highly flexible and multiple rapidly exchanging conformations typifying this class of proteins. One method to circumvent some of these difficulties is to incorporate nonnative fluorine (^{19}F) nuclei at specific sites within the polypeptide. ^{19}F NMR is particularly suitable for characterization of unfolded structures because ^{19}F chemical shifts are highly sensitive to local environments and conformations. Furthermore, the incorporation of fluorine analogs of fluorescent amino acids such as 5-fluoro-D,L-tryptophan (5FW) allows for complementary studies of protein microenvironment via fluorescence spectroscopy. Herein, we describe methods to produce, purify, characterize, and perform steady-state fluorescence and 1D NMR experiments on 5FW analogs of the IDP α-synuclein.

Key words: Fluorine, Parkinson's disease

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is an indispensible tool for protein structure determination. While well-folded proteins generally exhibit distinct proton (^{1}H) NMR resonances enabling assignment to specific nuclei, the innate flexibility and multiple conformations associated with intrinsically disordered proteins (IDPs) make assignments using conventional proton NMR methods a challenge (1). In these cases, performing NMR on IDP analogs that contain a fluorine (^{19}F) nucleus is particularly useful. ^{19}F NMR has several advantages over ^{1}H NMR and requires no additional technology (^{19}F spin = ½ nucleus). For example, ^{19}F resonances are exquisitely sensitive to changes in local environment with nine as opposed to one electron surrounding the fluorine nucleus resulting in a greatly enhanced chemical shift dynamic range for ^{19}F (up to 400 ppm) as compared to ^{1}H (up to 13 ppm)

(2–5). Furthermore, with advancements in biosynthetic methods for incorporation of fluorinated amino acids in proteins during bacterial growth (6), fluorine NMR can be applied to study a wide variety of previously difficult, large proteins and IDPs (3, 7, 8). A good candidate is the fluorinated tryptophan analog, 5-fluoro-D, L-tryptophan (5FW). Because Trp is typically in low abundance in most polypeptide sequences, minimal site-directed mutagenesis is required to make specific resonance assignments and most importantly, fluorescence spectroscopy can be employed as an additional site-specific probe of local environments and conformation.

In this chapter, we present methods to produce, purify, and characterize using ^{19}F NMR and fluorescence spectroscopy single 5FW mutants of an example IDP α-synuclein (α-syn). While several other fluorine containing amino acid analogs have been used in ^{19}F studies of α-syn (7, 9–11), we choose 5FW for its utility as a dual NMR and fluorescence probe (*vide supra*) (8). Specifically, we describe procedures to express 5FW-containing proteins in *Escherichia coli* by inhibiting the Shikimate pathway for aromatic amino acid synthesis with *N*-phosphonomethyl-glycine (glyphosate) (6). The importance of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), ultraviolet-visible (UV-Vis) spectroscopy, and electrospray ionization mass spectrometry (ESI-MS) in assessing protein purity is also emphasized. Furthermore, to highlight the utility of 5FW as a dual NMR and fluorescence probe of the local microenvironment and protein conformation of α-syn, experimental protocols for both techniques are presented.

2. Materials

2.1. Production of 5-Fluoro-D,L-tryptophan-Containing Proteins

1. QuickChange kit for site-directed mutagenesis (Stratagene).
2. Primers for α-syn single tryptophan variants.
3. α-Syn expression plasmid (pRK172) (12).
4. DNA sequencing service.
5. BL-21(DE3)pLysS *E. coli* cells.
6. Heat block (42 °C).
7. Super optimal broth with catabolite repression (SOC) medium.
8. Antibiotic stocks: 68 mg/mL chloramphenicol (CM) and 100 mg/mL ampicillin (AMP).
9. Luria Broth (LB) agar: 16 g/L tryptone,16 g/L NaCl, 8 g/L yeast extract, 24 g/L bacto agar.
10. Cell culture plates (100 × 15 mm style).
11. Cell spreader.

12. Sterile toothpick.

13. Luria Broth (LB) medium: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.0.

14. Temperature-controlled incubator shaker.

15. Centrifuge tubes (30 mL to 1 L).

16. Centrifuges for 30 mL to 1 L centrifuge tubes.

17. 5 mL sterile pipettes.

18. Disposable 1 cm path length cuvettes rated for visible spectroscopy.

19. *N*-(Phosphonomethyl)glycine (glyphosate).

20. 5FW.

21. Modified New Minimal Medium (NMM) (8, 13): 7.5 mM ammonium chloride, 8.5 mM NaCl, 55 mM potassium dihydrogenphosphate, 100 mM potassium hydrogenphosphate, 1 mM magnesium sulfate, 20 mM glucose, 1 mg/L Ca^{2+}, 1 mg/L Fe^{2+}, 1 μg/L Zn^{2+}, 10 mg/L thiamine, 10 mg/L uracil, 40 mg/L all amino acids except tryptophan and proline (see Note 1).

22. 1 M glucose.

23. Sterile Erlenmeyer flasks.

24. UV-Vis spectrometer.

25. Wash medium: 7.5 mM ammonium chloride, 8.5 mM NaCl, 55 mM potassium dihydrogenphosphate, 100 mM potassium hydrogenphosphate, 1 mM magnesium sulfate, 20 mM glucose.

26. 0.5 mM isopropyl-β-ᴅ-thiogalactopyranoside (IPTG).

2.2. Protein Purification

1. Lysis buffer: 100 mM Tris, 300 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.2).

2. 100 mM phenylmethylsulfonyl fluoride (PMSF) (see Note 2).

3. Benchtop probe homogenizer.

4. Stirring hot plate.

5. pH meter with calomel pH probe.

6. Concentrated HCl.

7. Dialysis buffer: 20 mM Tris, 1 mM EDTA (pH 8.0).

8. Dialysis tubing with <10 kDa molecular weight cut off (see Note 3).

9. Dialysis tubing closures.

10. Chromatography buffers: 20 mM Tris and 20 mM Tris, 1 M NaCl (pH 8.0) (see Note 4).

11. Fast protein liquid chromatography (FPLC) system equipped with fraction collector and a UV (280 nm) detector.

12. HiPrep DEAE FF and MonoQ anionic exchange columns (GE Healthcare).

13. Gel electrophoresis machine equipped with gel developer.

14. UV-Vis spectrometer.

15. 1 cm Path length quartz cuvette for absorption.

16. Mass spectrometry facility with ESI-MS.

17. Protein concentration device with molecular weight cutoff <10 kDa.

2.3. Spectroscopic Characterization of ^{19}F α-Synucleins

1. PD-10 desalting columns (GE Healthcare).

2. Buffer for spectroscopy: 10 mM sodium phosphate, 100 mM NaCl (pH 7.4) (see Note 4).

3. *N*-Acetyl-tryptophanamide (NATA).

4. 5FW.

5. YM-100 centrifugal filter units.

6. Benchtop ultracentrifuge.

7. UV-Vis spectrometer.

8. Spectrofluorometer equipped with temperature controlled cuvette holder.

9. 1 cm Path length quartz cuvette for fluorescence.

2.4. ^{19}F NMR

1. NMR spectrometer.

2. D$_2$O.

3. Trifluoroacetic acid.

4. Data processing software.

3. Methods

3.1. Production of 5-Fluoro-D,L-tryptophan Proteins

1. Since WT α-syn does not contain any native Trp residues, site-directed mutagenesis must be performed to produce single Trp-containing α-syn plasmids at native aromatic sites (F4W, Y39W, F94W). This can be done using the QuickChange kit, the wild type α-syn expression plasmid (pRK172) (12), and appropriate PCR primers (see Note 5).

2. Confirm desired mutations by DNA sequencing.

3. Transform BL21(DE3)pLysS *E. coli* cells with appropriate pRK172 constructs by gently pipetting 1 μL DNA into 25 μL BL21(DE3)pLysS cells. Incubate tube containing DNA–cell mixture on ice for 15 min and then place into heat block at

42 °C for 80 s followed by incubation on ice for 2 min. Add 200 μL SOC medium (see Note 6).

4. Pipette 15 μL of transformed cells onto LB agar cell culture plates containing CM (34 μg/mL) and AMP (100 μg/mL) antibiotics. Uniformly spread the cells using a sterile cell spreader (see Note 7).

5. Incubate the cell culture plates at 37 °C for 8–12 h or until distinct cell colonies are observed.

6. Using a sterilized toothpick, innoculate a sterilized 125 mL Erlenmeyer flask containing 25 mL of LB media, CM (34 μg/mL), and AMP (100 μg/mL). Incubate for 8–12 h at 37 °C shaking at 170 rpm.

7. Pour the cells into a sterile 50 mL centrifuge tube and centrifuge at $2,900 \times g$ (4 °C) for 10 min.

8. Pour off supernatant and resuspend pellet with 20 mL LB containing CM (34 μg/mL) and AMP (100 μg/mL).

9. Add resuspended pellet to a sterilized 4 L Erlenmeyer flask containing 1 L LB media containing CM (34 μg/mL), AMP (100 μg/mL), and glucose (20 mM).

10. Incubate 1 L culture at 30 °C (shaking 170 rpm) until the optical density (O.D.) at 600 nm $\approx$ 0.6–0.8 (see Note 8).

11. Pour cell culture into presterilized centrifuge tubes and centrifuge at $5,000 \times g$ (4 °C) for 30 min.

12. Wash harvested cells with Wash medium (Subheading 2.1, step 25).

13. Resuspend harvested cells in 20 mL of media (Subheading 2.1, step 21) and then use 10 mL to inoculate each 1 L media containing glyphosate (1 g/L), 5FW (40 mg/mL), CM (34 μg/mL), and AMP (100 μg/mL) in 4 L sterilized Erlenmeyer flask.

14. Incubate culture at 30 °C (shaking 170 rpm) for 30 min.

15. Induce protein expression by adding IPTG to a total concentration of 0.5–1 mM.

16. Continue incubating at 30 °C (170 rpm) for 5–6 h.

17. Pour cell culture into centrifuge tubes and centrifuge at $5,000 \times g$ (4 °C) for 30 min.

18. Remove supernatant and collect cell pellet (see Note 9).

3.2. Protein Purification

1. Prepare a boiling water bath (see Note 10).

2. Resuspend 2 L pellet in 50 mL lysis buffer (Subheading 2.2, step 1; 25 mL per 1 L culture) with freshly added PMSF (0.5 mM).

3. In medium size (150 mL) centrifuge bottle, homogenize pellet using a benchtop probe homogenizer (see Note 11).

4. Put small stir bar in tube with homogenate and purge with gentle nitrogen stream for 15 min at room temperature.

5. Seal bottle and place in boiling water bath for 15 min.

6. Centrifuge sample for 30 min at $20,000 \times g$ (4 °C).

7. Collect supernatant in beaker with stir bar.

8. While stirring on a stir plate, add concentrated HCl dropwise until pH reaches 3.5 (see Note 12).

9. Place beaker with sample on stir plate for 15 min at 4 °C.

10. Place sample in centrifuge tubes and centrifuge at $20,000 \times g$ (4 °C) for 1 h.

11. Collect supernatant.

12. Prepare two 4 L beakers of dialysis buffer (Subheading 2.2, step 7) and add a large stir bar to each beaker.

13. Place sample into dialysis tubing and carefully clamp each tube end with dialysis tubing closures.

14. Add fresh PMSF to a total concentration of 0.5 mM to the first beaker containing dialysis buffer and gently place sample in dialysis tubing into beaker. Incubate for 6–12 h on stir plate at 4 °C.

15. Add freshly prepared PMSF to 0.5 mM to the second beaker of dialysis buffer. Remove sample from first dialysis beaker and gently place in the second beaker. Incubate for 6–12 h on stir plate at 4 °C.

16. Collect sample in small beaker and load onto a preequilibrated (20 mM Tris, pH 8.0) HiPrep 16/10 DEAE FF anionic exchange column (see Note 13).

17. Elute protein using a linear gradient of 20 mM Tris and 1 M NaCl. Collect 8 mL fractions of eluent. α-Syn protein will elute upon exposure to 100–300 mM NaCl (see Note 14).

18. Evaluate fractions corresponding to peaks in the O.D. at 280 nm by SDS-PAGE (see Note 15).

19. Evaluate fractions corresponding to peaks in the O.D. at 280 nm by UV-Vis spectroscopy (see Note 16).

20. Combine fractions containing pure or semipure α-syn and concentrate using a protein concentration device with a MWCO < 10 kDa.

21. Exchange buffer via dialysis according to steps 13–15.

22. Load sample onto a preequilibrated (20 mM Tris, pH 8.0) MonoQ anionic exchange column.
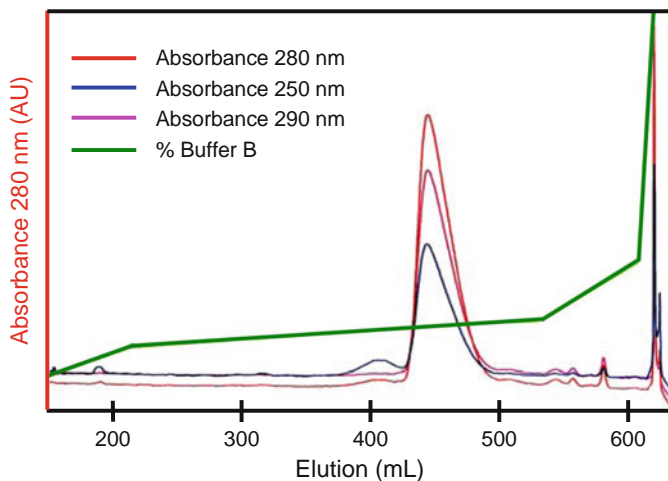
Fig. 1. Purification of α-synuclein (α-syn) using a MonoQ anionic exchange column. The elution of α-syn protein from a MonoQ column is shown with absorbance at 280, 250, and 290 nm denoted by *red*, *blue*, and *pink traces*, respectively. The *green trace* represents the linear gradient of Buffer B (20 mM Tris, 1 M NaCl, pH 8).

23. Elute, evaluate, and concentrate purified α-syn protein according to steps 17–20. An example MonoQ chromatogram is shown in Fig. 1. Purified α-syn corresponds to the large singular peak between 425 and 500 mL with O.D. at 280 nm greater than that of other wavelengths (250 and 290 nm) recorded. α-Syn elutes between 100 and 300 mM NaCl corresponding to ~10 to 30 % Buffer B (20 mM Tris, 1 M NaCl) (see Note 17).

24. Evaluate protein purity using SDS-PAGE.

25. Combine pure fractions and concentrate to desired volume.

26. Use UV-Vis absorbance spectroscopy to determine the concentration of purified protein (see Note 18).

27. Confirm appropriate molecular weight of protein by performing ESI-MS on purified protein (see Note 19).

28. Use protein right away for experiments or flash freeze and store at −80 °C for later use (see Note 20).

3.3. Spectroscopic Characterization of ¹⁹F Containing Synucleins

1. Use a PD-10 column to buffer exchange up to 2.5 mL purified protein into the buffer for spectroscopy (10 mM NaPi, 100 mM NaCl buffer, pH 7.4).

2. Determine the concentration of desalted protein using UV-Vis spectroscopy (see Note 18).

3. Prepare a sample of 10 μM fluorinated protein by diluting the concentrated protein stock with the buffer for spectroscopy.

4. Prepare a 10 μM sample of the tryptophan model complex *N*-acetyl-tryptophanamide (NATA) using the buffer for
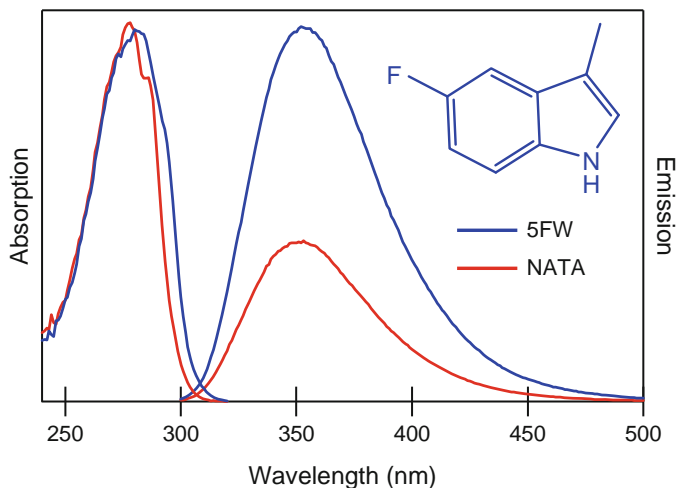
Fig. 2. Absorption and emission spectra of 10 μM tryptophan model complex, NATA, and 5FW (20 mM sodium phosphate buffer, pH 7.4, 25 °C). *Inset*: Chemical structure of 5FW.

spectroscopy. The molar extinction coefficient at 280 nm for NATA is 5,500/M/cm.

5. Prepare a 10 μM sample of the 5FW using the buffer for spectroscopy. The molar extinction coefficient at 280 nm for 5FW is 5,700/M/cm.

6. Take a UV-Vis spectrum of all 10 μM samples (200–500 nm).

7. Substitution of the fluorine atom on the tryptophan indole ring causes a characteristic red shift in the absorption spectrum as compared to native tryptophan. This shift (~5 nm) is most obvious upon comparison of the red edge of the NATA and 5FW absorption traces (see Fig. 2). To confirm high incorporation of the 5FW into α-syn protein, compare of the red edge of the UV-Vis absorption spectrum of fluorinated protein sample to that of both NATA and 5FW in buffer alone. If >90 % of the protein has 5FW incorporated, the absorption of the protein sample should overlay with that of 5-fluortyptophan in buffer alone.

8. To evaluate the local environment of 5FW incorporated in α-syn protein samples, collect a fluorescence emission spectra by placing a fluorescence quartz cuvette containing sample in a temperature controlled spectrofluorometer cuvette holder (see Note 21).

9. Set spectrofluorometer to the following parameters: $T = 25$ °C, excitation = 295 nm; slit width 1 nm, emission = 300–500 nm; slit width = 1 nm, integration time = 0.25 s (see Note 22).

10. Collect the steady-state emission spectrum for all protein samples and buffer. The emission spectra for NATA and 5FW alone in spectroscopy buffer are displayed in Fig. 2 (see Note 23).

3.4. ^{19}F NMR

1. Desalt protein samples and determine protein concentration according to Subheading 3.3, steps 1–2.

2. Prepare 500 μL samples of 50 μM fluorinated protein samples by diluting the concentrated protein stock with buffer for spectroscopy and 50 μL D$_2$O (10 %, v/v).

3. Place samples into appropriate NMR tubes and then into the NMR sample chamber holder.

4. Use trifluoroacetic acid as an external standard (0 ppm).

5. Set NMR spectrometer to collect spectra in 1,500 transient blocks over 2–11 h at 15 °C with the following parameters: relaxation delay of 1.0 s, pulse width of 7.6 μs, and sweep width of 20 kHz.

6. Collect data for all fluorinated proteins and 50 μM 5FW alone according to Subheading 3.4, steps 1–5.

7. Use data processing software to analyze data.

8. Evaluate the local environment of 5FW at each of the sites in α-syn tested by comparing the ^{19}F NMR spectra of fluorinated α-syn variants to that of 5FW in buffer alone. The ^{19}F NMR spectra of fluorine containing α-synucleins and free 5FW are given in Fig. 3. The ^{19}F NMR spectra reveal that all α-synucleins exhibit ^{19}F resonances similar to the model complex (5FW) confirming the absence of secondary and tertiary structure for this intrinsically disordered protein. Though modest, small differences between the ^{19}F resonances at the different sites reflect minor differences in the three protein microenvironments.

4. Notes

1. NMM media is prepared as the following. Autoclave 0.9 L water in 4 L Erlenmeyer flasks and add other reagents as described. Prepare 10× salt solution (KH$_2$PO$_4$ 0.55 M; K$_2$HPO$_4$ 1 M; NaCl 85 mM) adjust pH to 7.2 with NaOH, filter sterilize, and add 100 mL to autoclaved water. Autoclave individual stocks of ammonium chloride (1 M), magnesium sulfate (1 M), and glucose (1 M) and add to a final concentration of 7.5, 1, and 20 mM, respectively. Stocks of Ca^{2+} (1 g/L) and Fe^{2+} (1 g/L) and Zn^{2+} (1 mg/L) were filter sterilized and added to a final concentration of 1 mg/L for Ca^{2+}, Fe^{2+}, and 1 μg/L for Zn^{2+}. Stocks of thiamine (2 g/L) and uracil (10 g/L)
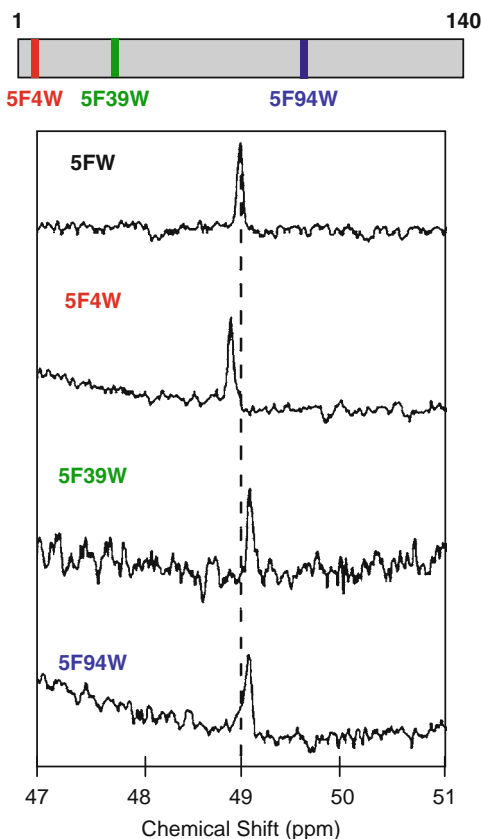
Fig. 3. The 470 MHz ^{19}F NMR spectra of free 5FW and 5FW containing α-synucleins. (*Top*) Schematic representation of the α-syn primary sequence with 5FW mutations *highlighted*. (*Bottom*) Single ^{19}F resonances for all three synuclein variants (−48.9 ppm (5FW4), −49.1 ppm (5FW39), −49.1 ppm (5FW94)) and the model complex (−49.0 ppm). All spectra were collected at 15 °C to avoid protein aggregation.

were filter sterilized and added to a final concentration of 10 mg/L. All amino acids except tryptophan and proline (40 mg/L each) were weighed, premixed, added as a solid, and shaken well (0.72 g/L media).

2. A fresh stock solution of PMSF (100 mM) should be prepared in ethanol or isopropanol daily and kept at 4 °C until use. PMSF, a serine protease inhibitor, is a cytotoxic chemical and should be handled with gloves at all times and disposed appropriately.

3. Some dialysis tubing must be treated prior to use. Please follow instructions from the dialysis tubing manufacturer for specific details.

4. Buffers should be filtered through a 0.22 μm pore size filter prior to use.

5. A detailed protocol for performing site-directed mutagenesis reactions and designing appropriate primers can be found in the QuickChange kit.

6. If using small tubes (0.5 mL) expose DNA/cells to heat block for 40 s.

7. LB agar plates containing appropriate antibiotics can be made by diluting stocks of CM (68 mg/mL) and AMP (100 mg/mL) to final concentrations of 34 and 100 μg/mL, respectively into presterilized LB agar medium. If using an autoclave to sterilize LB, first let media cool such that is warm to the touch prior to adding antibiotics. Pour warm LB with antibiotics into cell culture plates (100 × 15 mm size) until cell culture plates are approximately half full. Cover the plates and let cool to room temperature before use. The plates can be stored for ~1 month at 4 °C for future use.

8. To check O.D., remove ~1 mL of culture from Erlenmeyer flask using a sterile 5 mL pipette and place into a disposable cuvette. Place ~1 mL of culture media that is not inoculated in a separate cuvette. Use a spectrometer to measure O.D. at 600 nm for media alone. Repeat for inoculated media. Subtract O.D. of inoculated media from that of media alone. If O.D. of inoculated media is between 0.6 and 0.8, proceed to step 11 of Subheading 3.1. If O.D. is <0.6 continue to incubate culture and retest O.D. as needed.

9. Pellet can be used immediately or stored at −20 °C for later use.

10. To prepare the boiling water bath, place a 1 L beaker filled $\frac{3}{4}$ with water and a stir bar on a stirring hot plate.

11. While thorough homogenization is essential and no clumps should be visible, avoid excess homogenization resulting in foam.

12. As the solution pH will change dramatically with the addition of each drop of concentrated HCl, make sure pH has completely stabilized before adding additional drops. At pH 3.5, a significant amount of protein will have visibly precipitated.

13. Refer to FPLC manual and chromatography column information for specific details regarding proper column equilibration and general use.

14. Make sure FPLC is equipped to monitor and record the O.D. of eluted sample at 280 nm. Fractions containing α-syn protein can be identified by a peak in the O.D. at 280 nm.

15. Sample from each fraction thought to contain α-syn protein should be analyzed on an SDS-PAGE gel. Protein corresponding to α-syn monomer will run near ~15,000 kDa. Use of a silver staining method for gel development will ensure all fractions containing a small amount of α-syn will be visible. Combine

pure protein for use (only α-syn band is observed) or combine
semipure α-syn fractions (multiple protein bands including α-syn
are observed) for further purification.

16. A UV-Vis spectrum should be collected for sample from each
fraction thought to contain α-syn protein. The UV-Vis spectra
should be collected between 200 and 500 nm using a 1 cm
pathlength quartz cuvette. Purified proteins will exhibit a single
maximum peak in O.D. around 280 nm with a pronounced
right shoulder corresponding to the presence of 5FW. If frac-
tions contain DNA or other contaminants, additional peaks will
be observed. Combine pure protein for use (only single α-syn
peak observed) or combine semipure α-syn fractions (peaks for
both α-syn and other contaminants are observed) for further
purification.

17. If there are many peaks in O.D at 280 nm corresponding to the
α-syn containing region (100–300 mM NaCl) after purification
using both the 16/10 HiPrep DEAE FF and MonoQ columns,
repeat steps 20–23 of Subheading 3.2 until the MonoQ chro-
matogram resembles Fig. 1 and SDS-PAGE gels and UV-Vis
spectra reflect >95 % purity.

18. If using a 1 cm pathlength quartz cuvette, the concentration of
protein can be calculated according to Beer's Law by dividing
the O.D. at 280 nm by the estimated molar extinction coeffi-
cient at 280 nm ($\varepsilon_{280\ nm}$). For 5FW4 and 5FW94 proteins
$\varepsilon_{280\ nm} = 6,980/M/cm$ and for 5FW39 protein
$\varepsilon_{280\ nm} = 5,700/M/cm$ (in these variants C-terminal Tyr had
been mutated to Phe). When calculating protein concentra-
tions make sure that there is no contribution from buffer
background. To do this, subtract the O.D. of buffer alone at
280 nm from that of the protein.

19. Generally ~15 μL of 50–100 μM α-syn is sufficient to perform
ESI-MS. Speak with your core facility manager for specific
requirements and procedures. Expected molecular weights for
fluorinated α-syn variants are 14,469 for 5FW4/Y125F/
Y133F/Y136F and 5FW94/Y125F/Y133F/Y136F, and
14,451 for 5FW39/Y125F/Y133F/Y136F.

20. To flash freeze, place liquid nitrogen in a small dewar and drop
tubes containing protein into the dewar. Once protein samples
have frozen, remove carefully with tweezers and place in
−80 °C freezer.

21. If using a small volume quartz cuvette, make sure that the
excitation beam is aligned to propagate through the 1 cm
pathlength with adjacent quartz windows aligned parallel to
fluorescence emission detector.

22. Slit widths and integration time may need to be adjusted to ensure adequate signal detection depending on the specific spectrofluorometer used.

23. The spectrofluorometer manual should be consulted to ensure that measured counts are within the linear range of the detector. For all measurements the O.D. of the sample should be no greater than 0.1 at the excitation wavelength. For all samples, a buffer blank spectrum should be recorded to ensure that no contaminants are contributing to the emission spectrum. The 5FW emission spectrum in Fig. 2 represents a solvent exposed chromophore. The spectral properties (quantum yield and wavelength of maximum emission) will change if the 5FW microenvironment changes.

References

1. Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Unfolded Proteins 62:311–340

2. Ropson IJ, Frieden C (1992) Dynamic NMR spectral-analysis and protein folding—identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by F-19 NMR. Proc Natl Acad Sci U S A 89:7222–7226

3. Gerig JT (1994) Fluorine NMR of proteins. Prog Nucl Magn Reson Spectrosc 26:293–370

4. Lian CY et al (1994) F-19 nuclear-magnetic-resonance spectroscopic study of fluorophenylalanine-labeled and fluorotryptophan-labeled avian egg-white lysozymes. Biochemistry 33:5238–5245

5. Harris RK (1983) Nuclear magnetic resonance spectroscopy: A physicochemical view. Pitman, London

6. Kim HW, Perez JA, Ferguson SJ, Campbell ID (1990) The specific incorporation of labeled aromatic-amino-acids into proteins through growth of bacteria in the presence of glyphosate: Application to fluorotryptophan labeling to the H+-ATPase of *Escherichia coli* and NMR-studies. FEBS Lett 272:34–36

7. Li CG et al (2010) Protein F-19 NMR in *Escherichia coli*. J Am Chem Soc 132:321–327

8. Winkler GR, Harkins SB, Lee JC, Gray HB (2006) Alpha-synuclein structures probed by 5-fluorotryptophan fluorescence and F-19 NMR spectroscopy. J Phys Chem B 110:7058–7061

9. Wang GF, Li CG, Pielak GJ (2010) F-19 NMR studies of alpha-synuclein-membrane interactions. Protein Sci 19:1686–1691

10. Wang GF, Li CG, Pielak GJ (2010) Probing the micelle-bound aggregation-prone state of alpha-synuclein with F-19 NMR spectroscopy. ChemBioChem 11:1993–1996

11. Li CG et al (2009) F-19 NMR studies of alpha-synuclein conformation and fibrillation. Biochemistry 48:8578–8584

12. Jakes R, Spillantini MG, Goedert M (1994) Identification of 2 distinct synucleins from human brain. FEBS Lett 345:27–32

13. Budisa N et al (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. Eur J Biochem 230:788–796

Alpha Proton Detection Based Backbone Assignment of Intrinsically Disordered Proteins

Perttu Permi and Maarit Hellman

Abstract

Assignment of NMR resonance frequencies to a particular atom in the molecule establishes a vital step for any detailed structural study. Approaches for sequential assignment typically involve amide proton detection, which may become suboptimal in case of intrinsically disordered proteins (IDPs) at high pH and/or temperature. Here we describe an alternative approach: assignment protocol based on alpha proton detected triple-resonance experiments, which offer several advantages over well-established experiments relying on amide proton detection. Our experiments are suitable for studies of IDPs at any pH and enable sequential assignment of proline-rich segments.

Key words: Alpha proton detection, Backbone assignment, Intrinsically disordered proteins, NMR spectroscopy

1. Introduction

Discovery of means for bacterial overexpression of ^{15}N, ^{13}C enriched recombinant proteins has been a major breakthrough for structural characterization of proteins by NMR spectroscopy. During past 20 years, assignment protocols utilizing so-called amide proton (HN) detected triple-resonance experiments have found the widest audience (*for review see*, e.g., (1, 2)). Impetus for the success of this approach stems from direct and indirect advantages offered by the NH detection scheme, e.g., wide dispersion of ^{15}N, ^{1}HN correlation map, applicability to ^{15}N, ^{13}C, ^{2}H labeled samples together with transverse relaxation optimized spectroscopy (TROSY) for studies of larger proteins (3, 4), robust and high-quality water suppression (5, 6), and straightforward extension to dynamical characterization of proteins at amide site in terms of ^{15}NT_1, T_2 relaxation times, steady-state $\{^1$H$\}$-^{15}N heteronuclear

NOEs or proton–deuterium exchange rates (7). In addition to collection of data on protein dynamics, studies of protein–ligand interactions have also been carried out using two-dimensional ^{15}N, ^1HN correlation experiment (^{15}N-HSQC).

The assignment strategy relying on HN detection is based on linking residues i and $i − 1$ $^{13}C\alpha/^{13}C\beta$ (HNCACB/CBCA(CO) NH experiments) or $^{13}C'$ (HN(CA)CO/HNCO experiments) resonance frequencies to the ^{15}N, ^1H chemical shifts of residue i (*for review* see, e.g., (1, 2)). This protocol then establishes sequential assignment of polypeptide chains, which are uniformly labeled with ^{15}N and ^{13}C isotopes. Resonance assignment especially in the case of intrinsically disordered proteins (IDPs) may become troublesome due to exchange broadened NH resonances at high pH and/or temperature. Moreover, IDPs are often rich in N-substituted proline residues, whose assignment using the NH detection scheme is difficult.

Here we show an assignment protocol resting on alpha proton (HA)-detected triple-resonance experiments referred to as (HCA) CON(CA)H, iH(CA)NCO, H(CA)CON. These experiments are built on archaic triple-resonance experiments that were introduced and flourished in the early days of NMR studies of isotopically enriched proteins as well as their more recent derivatives (8–14). HA-detection approach offers several advantages over the HN-detection protocol by enabling, e.g., use of alkaline conditions, or more exactly any pH, for NMR studies as well as assignment of proline-rich segments that are abundant in IDPs.

2. Materials

1. The uniformly ^{15}N and ^{13}C labeled protein dissolved in appropriate buffer and pH so that H_2O–D_2O ratio equals to 95 %:5 %, or alternatively to 100 % D_2O. Here ^{15}N and ^{13}C labeled GB1 (immunoglobulin-binding domain B1 of streptococcal protein G) and intrinsically disordered protein, CT16 (cancer/testis antigen protein, 112 residues), have been used when exemplified NMR experiments and assignment procedure. Protein concentrations used in the studies were 1 mM (CT16) and 1.5 mM (GB1).

2. Data collection necessitates access to high-field NMR spectrometer, equipped with at least three radiofrequency (rf) channels, a triple-resonance ($^1H/^{13}C/^{15}N$) probehead, or preferably with a cryogenically cooled probehead, and an actively shielded z-gradient system.

3. A set of three-dimensional HA-detected ^1HA, $^{13}C'$ and ^{15}N correlation experiments (15, 16). Pulse sequence codes for

Varian Unity Inova spectrometers are available at Web site http://www.biocenter.helsinki.fi/bi/nmr/permi/seqlib.html or upon request from authors.

4. Data processing and spectrum analysis can be carried out using various data processing and data analysis software packages, e.g., VNMR and Sparky.

3. Methods

HA-detection based backbone assignment of IDPs employs three triple-resonance experiments: (HCA)CON(CA)H, iH(CA)NCO and H(CA)CON. In short, these three-dimensional experiments correlate $^{13}C'(i-1)$, $^{15}N(i)$, $^{1}HA(i)$, and $^{15}N(i)$, $^{13}C'(i)$, $^{1}HA(i)$, and $^{15}N(i+1)$, $^{13}C'(i+1)$, $^{1}HA(i)$ spins in ω_1, ω_2 and ω_3 domains, respectively. Motivation of using the $^{13}C'$ spin instead of $^{13}C^{\alpha}$ spin for the sequential assignment of IDPs stems from vaguer dependence of $^{13}C'$ chemical shift on residue type (17). Indeed, $^{13}C^{\alpha}$ chemical shifts tend to cluster between identical residues, which render the assignment highly ambiguous and hence very tedious. Here we first outline practical approach for optimal data collection with the HA-detected experiments (see Subheadings 3.1–3.3) and subsequently describe an assignment protocol based on these data sets (see Subheading 3.4). All three HA-detected experiments shown here provide adequate water suppression and, indeed, the assignment of CT16 was carried out in 93 %/7 % H_2O/D_2O. However, use of pure D_2O as a solvent improves the quality of data and enables assignment of HA resonances that are completely underneath the residual water signal. It is beneficial to decouple scalar interaction between ^2H and ^{15}N spins in D_2O using the ^2H composite pulse decoupling (CPD). The data of all three spectra were collected on 1 mM $^{15}N/^{13}C$-labeled cancer testis (CT16) antigen at 25 °C. Spectrometer was Varian Unity Inova 800 MHz system, equipped with a cryogenically cooled $^1H/^{15}N/^{13}C$ probehead and actively shielded z-axis gradient.

3.1. (HCA)CON(CA)H

The (HCA)CON(CA)H pulse scheme labels the chemical shifts of $^{13}C'(i-1)$, $^{15}N(i)$, $^{1}HA(i)$ spins (Fig. 1). The magnetization transfer pathway in this "out-and-stay"-type experiment can be simplified using the following notation

$$^1H^{\alpha}(i-1) \rightarrow {}^{13}C^{\alpha}(i-1) \rightarrow {}^{13}C'(i-1)(t_1) \rightarrow {}^{15}N(i)(t_2)$$
$$\rightarrow {}^{13}C^{\alpha}(i) \rightarrow {}^{1}H^{\alpha}(i)(t_3).$$
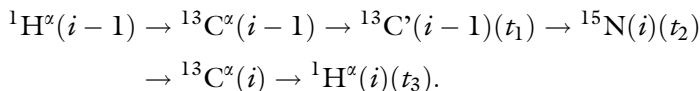
Chemical shift of $^{13}C'(i-1)$ spin is being labeled during the first indirectly detected (t_1) period, which is implemented in a
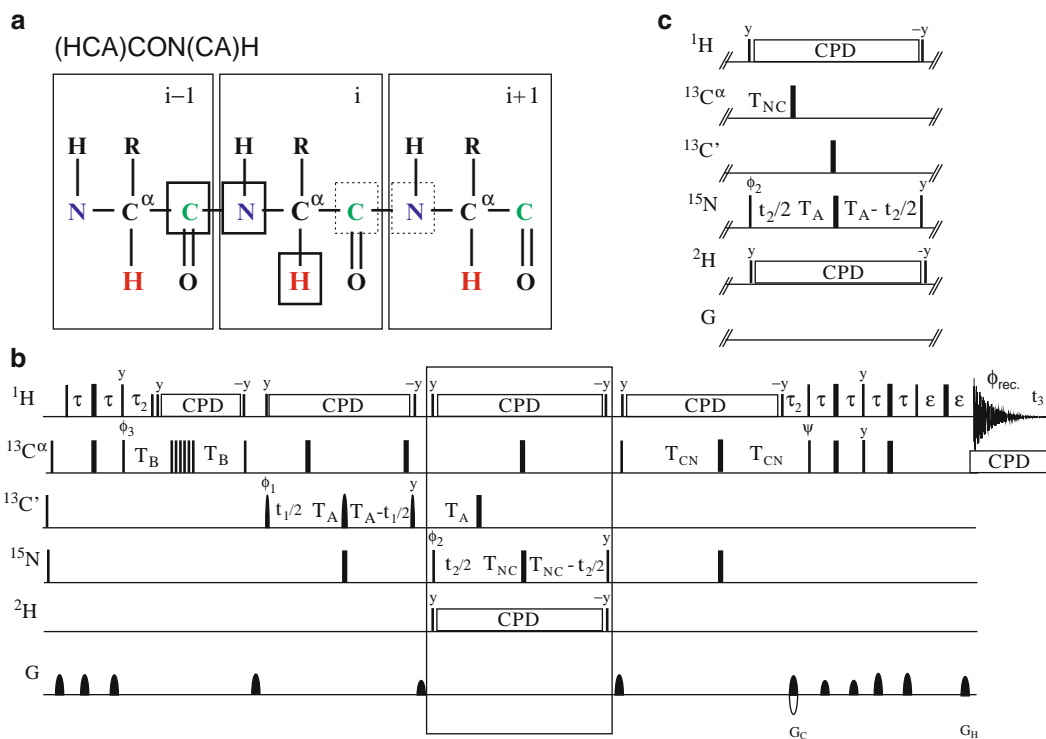
Fig. 1. Schematic representation of backbone atoms that are correlated in (HCA)CON(CA)H experiment (**a**). *Solid rectangles* highlight $^1H^\alpha(i)$, $^{13}C'(i-1)$, and $^{15}N(i)$ spins that are detected using the pulse sequence of (HCA)CON(CA)H experiment in (**b**) and the delay setting $T_{NC} = 25–26$ ms. *Rectangles* with *dotted lines* highlight $^1H^\alpha(i)$, $^{13}C'(i)$, and $^{15}N(i+1)$ spins that are additionally correlated in an alternative implementation of (HCA)CON(CA)H experiment in (**c**) using the delay $T_{NC} = 12–15$ ms. All rectangular rf pulses that correspond to 90° and 180° flip angles are shown with *narrow* and *wide bars*, respectively. All pulses are applied with phase *x* unless otherwise indicated. The ^{13}C carrier is set initially to the middle of $^{13}C^\alpha$ region (58 ppm), shifted to carbonyl carbon region (176 ppm) before the 90° $^{13}C'$ pulse with phase ϕ_1, and shifted back to 58 ppm after the 90° ^{15}N pulse with phase *y*. Rectangular 90° and 180° pulses for ^{13}C were applied with durations of 40.8 μs (90°) and 36.4 μs (180°) at 800 MHz, respectively. This provides minimal mutual excitation of $^{13}C^\alpha$ and $^{13}C'$ spins (9). The cascade of rectangular pulses on $^{13}C^\alpha$ denotes a composite pulse for ultra-broadband inversion with durations defined by pwC*(β_i/90), where β_i is a flip angle for individual pulses in the cascade, i.e., 158.0, 171.2, 342.8, 145.5, 81.2, 85.3 (20). 180° $^{13}C'$ pulse in the middle of t_1 period had the shape of one-lobe sinc profile with duration of 60.4 μs. The waltz-16 composite pulse decoupling (CPD) sequence (21) with strength of 4.8 kHz was employed to decouple ^1H spins. The adiabatic WURST field (22) was used to decouple ^{13}C during acquisition. Delay durations: $\tau = 1/(4J_{HC}) \sim 1.7$ ms; $\tau_2 = 2.4$ ms (for observing both glycine and nonglycine residues) or 3.4 ms (optimized for nonglycine residues); ε = duration of G_H + field recovery ~0.4 ms; $T_B = 1/(6J_{C\alpha C'}) \sim 3.4$ ms; $T_A = 1/(4J_{C'N}) \sim 16.5$ ms; $T_{NC} \sim 25$ ms (**b**) of 12–15 ms (**c**); $T_{CN} \sim 14$ ms; $t_{1,max} < 2.0*T_A$; $t_{2,max} < 2.0*T_{NC}$ (**b**) $t_{2,max} < 2.0*T_A$ (**c**). Frequency discrimination in $^{13}C'$ and ^{15}N dimensions is obtained using the States-TPPI protocol (23) applied to ϕ and ϕ_2, respectively. Phase cycling: $\phi = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\psi = x, -x$; $\phi_{rec.} = x, 2(-x), x, -x, 2(x), -x$. Gradient strengths and durations: $G_C = 13$ k G/cm (1.6 ms), $G_H = 13$ k G/cm (0.4 ms).

constant-time manner and incorporated into the delay $2T_A$. Therefore, number of complex data points (t_1 increments) per spectral width, which defines the maximum t_1 value, and hence the attainable resolution, is limited by the duration of $2T_A \sim 28–33$ ms (see Note 1). Analogously, the $^{15}N(i)$ chemical shift is being frequency labeled during the second indirectly detected period (t_2) and the

experimental resolution is limited by the delay $2T_{NC}$ (implementation shown in Fig. 1b) or $2T_A$ (Fig. 1c).

1. Setting of delays. Typically delays are 1.7; 2.4; 14–16.5; 3.4; 25; and 12.5–14 ms for τ, τ_2, T_A, T_B, T_{NC}, and T_{CN}, respectively. Of these, τ_2 and T_{NC} have significant influence on observed peak patterns and will be discussed below.

   (a) All residues establish a CH spin system between $^1H^\alpha$ and $^{13}C^\alpha$ spins, the glycine being the only exception as it carries two α-protons, which form a CH_2 spin moiety. If the delay τ_2 is set to 3.4 ms, that is the optimal value for the CH spin system, detection of glycines or residues following the glycine will be hindered. Thus, setting $\tau_2 = 2.4$ ms warrants detection of glycine residues with a modest cost (10–15 %) in sensitivity of CH spin systems. Glycines will also exhibit 180° phase difference with respect to other amino acid residues owing to the sign change during the final ^{15}N–$^{13}C^\alpha$ refocusing period $2T_{CN}$. This feature can be useful during the assignment procedure (*vide infra*).

   (b) The delay $2T_{NC}$ has critical influence on signal intensity as well as sign of emerging cross peaks. Using the implementation shown in Fig. 1b and by setting the delay T_{NC} to 25 ms, the intensity of $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ cross peak will be maximized while the intensity of $^{13}C'(i)$, $^{15}N(i+1)$, $^1HA(i)$ correlation is minimized. In practice, this delay setting suppresses the $^{13}C'(i)$, $^{15}N(i+1)$, $^1HA(i)$ cross peak and hence only $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ correlations will be observed as exemplified for GB1 (Fig. 2a, c and e). It is also possible to equalize the intensities of both $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ and $^{13}C'(i)$, $^{15}N(i+1)$, $^1HA(i)$ correlations by setting T_{NC} to 12.5–15 ms as in the implementation shown in Fig. 1c. Although this delay tuning inverts the sign of $^{13}C'(i)$, $^{15}N(i+1)$, $^1HA(i)$ cross peaks with respect to $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ correlations and facilitates their recognition, it also diminishes the attainable intensity of $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ correlations on IDPs or even small globular proteins/domains such as GB1 that have favorably long transverse relaxation times (compare Fig. 2b, d and f to Fig. 2b, c and e). This implementation may also lead to partial, or even complete, cancellation of cross peaks in case of incidental overlap (Fig. 2f). Thus, owing to extensive cross peak overlap present in IDPs, we prefer obtaining solely $^{13}C'(i-1)$, $^{15}N(i)$, $^1HA(i)$ correlations from the (HCA)CON(CA)H experiment, and gathering complementary information for the sequential assignment either from H(CA)CON or (HCA)NCO(CA)H experiments that provide $^{13}C'(i)$, $^{15}N(i+1)$, $^1HA(i)$ correlations (*vide infra*).
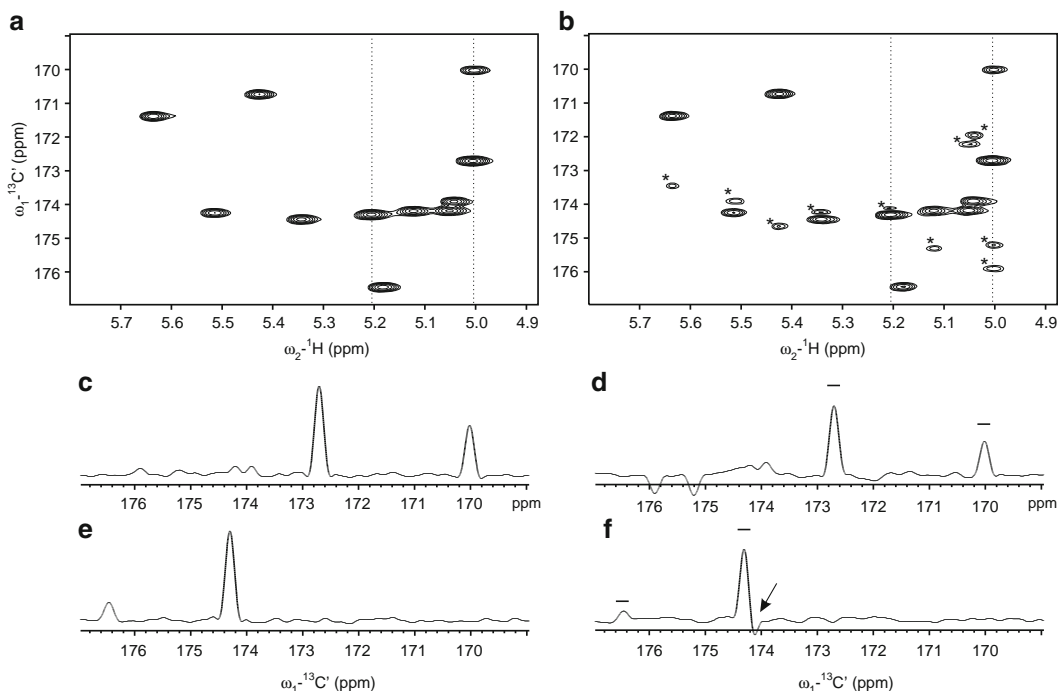
Fig. 2. Illustrative expansions of (HCA)CON(CA)H spectrum of GB1 are shown. Expansion in (**a**) shows two-dimensional ^{13}C′-^{1}H$^{\alpha}$ correlation spectrum with solely ^{13}C′$_{i-1}$–^{1}H$_i^{\alpha}$ cross peaks, whereas both ^{13}C′$_{i-1}$–^{1}H$_i^{\alpha}$ and ^{13}C′$_i$–^{1}H$_i^{\alpha}$ correlations in spectrum (**b**) are observed. The corresponding 1D projections taken at ^{1}H chemical shift of 5.01 and 5.21 ppm are shown underneath the two-dimensional spectra. *Horizontal lines* above resonance lines in 1D traces (**d**) and (**f**) highlight sensitivity improvement with respect to spectra shown in (**c**) and (**e**). Note also partial cancellation of ^{13}C′$_i$–^{1}H$_i^{\alpha}$ cross peak in the spectrum (**f**) indicated with an *arrow*. Both spectra were recorded as two-dimensional ^{13}C′-^{1}HA experiments, using eight transients per data point. Number of complex points in ^{13}C′ and ^{1}HA dimensions were 84 and 768, corresponding to acquisition time of 28 and 64 ms. Both spectra were apodized with squared cosine bell window functions and zero-filled to 2,048 × 4,096 data matrices prior to Fourier transform.

2. Placing the transmitter offsets. Make sure that the ^{1}H, ^{13}C′, ^{13}C$^{\alpha}$, and ^{15}N (and ^{2}H in D_2O) transmitters are placed into appropriate frequencies, i.e., ~4.7, 176, 58, and 118 ppm, respectively (see Table 1).

3. Calibrate ^{1}H, ^{13}C, and ^{15}N pulse widths (see Note 2).

4. Set number of t_1 and t_2 increments for three-dimensional experiment, and use States-TPPI quadrature detection protocol in t_1 and t_2. Initiate the experiment using at least two transients per data point.

The data collection was carried out using two transients per data point yielding overall experimental time of 22 h. Other relevant experimental and processing parameters for (HCA)CON(CA)H spectrum are listed in Tables 1 and 2.

Table 1
Experimental parameters for the (HCA)CON(CA)H, H(CA)CON, and iH(CA)NCO schemes

	Offset (ppm)	sw (Hz)	np	ni	ni$_2$	AT (ms)	
(HCA)CON (CA)H	4.65	8,000	768			64	^{1}H
	58						^{13}C
	176	2,600		86		33.1	^{13}C′
	118	3,000			128	42.7	^{15}N
H(CA)CON	4.65	8,000	768			64	^{1}H
	58						^{13}C
	176	2,600		86		33.1	^{13}C′
	118	3,000			128	42.7	^{15}N
iH(CA)NCO	4.65	5,000	320			64	^{1}H
	58						^{13}C
	176	2,600		48		18.5	^{13}C′
	118	3,000			128	42.7	^{15}N

Offset transmitter offset, *sw* spectral width, *ni* number of complex points in t_1 domain, *ni$_2$* number of complex points in t_2 domain, *np* number of complex points in acquisition domain, *AT* acquisition time

Table 2
Processing parameters for the (HCA)CON(CA)H, H(CA)CON, and iH(CA)NCO schemes

	sb	sbs	sb1	sbs1	sb2	sbs2	ZF	
(HCA)CON (CA)H	−np/sw	−np/sw					2,048	^{1}H
			−ni/sw1	−ni/sw1			256	^{13}C′
					−ni2/sw2	−ni2/sw2	256	^{15}N
H(CA)CON	−np/sw	−np/sw					2,048	^{1}H
			−ni/sw1	−ni/sw1			256	^{13}C′
					−ni2/sw2	−ni2/sw2	256	^{15}N
iH(CA)NCO	−np/sw	−np/sw					2,048	^{1}H
			−ni/sw1	−ni/sw1			256	^{13}C′
					−ni2/sw2	−ni2/sw2	256	^{15}N

sb sinebell in acquisition dimension, *sbs* shifted sinebell in acquisition dimension, *sb1* sinebell in t_1 domain, *sbs1* shifted sinebell in t_1 domain, *sb2* sinebell in t_2 domain, *sbs2* shifted sinebell in t_2 domain, *ZF* zero filling

3.2. H(CA)CON

The H(CA)CON scheme labels the chemical shifts of ^{13}C′(i), ^{15}N $(i + 1)$, ^{1}HA(i) spins exclusively (Fig. 3), yielding complementary information to the (HCA)CON(CA)H experiment described above. Magnetization transfer pathway in this "out-and-back"-type experiment is as follows:
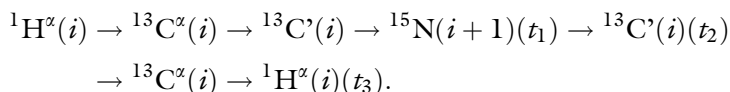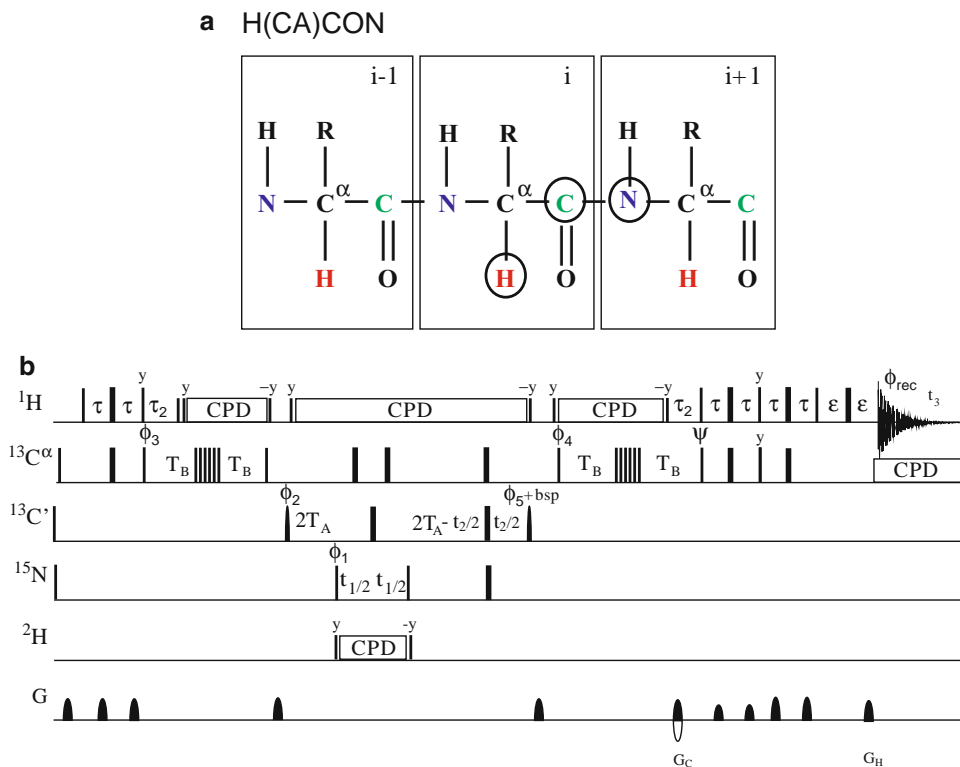
$$^1H^\alpha(i) \to {}^{13}C^\alpha(i) \to {}^{13}C'(i) \to {}^{15}N(i+1)(t_1) \to {}^{13}C'(i)(t_2)$$
$$\to {}^{13}C^\alpha(i) \to {}^1H^\alpha(i)(t_3).$$

a   H(CA)CON

Fig. 3. Schematic representation of backbone atoms that are correlated in the H(CA)CON experiment (**a**). $^1H^\alpha(i)$, $^{13}C'(i)$, and $^{15}N(i + 1)$ spins, whose resonance frequencies are correlated in the H(CA)CON experiment in (**b**), are *encircled*. The ^{13}C carrier is set initially to the middle of $^{13}C^\alpha$ region (58 ppm), shifted to carbonyl carbon region (176 ppm) before the 90° $^{13}C'$ pulse with phase ϕ_2, and shifted back to 58 ppm prior to the 90° $^{13}C^\alpha$ pulse with ϕ_4. Rectangular 90° and 180° pulses for ^{13}C were applied with durations of 40.8 μs (90°) and 36.4 μs (180°) at 800 MHz, respectively. Two selective 180° $^{13}C'$ pulses in the middle of t_1 and t_2 periods had the shape of one-lobe sinc profile with duration of 60.4 μs. Delay durations: $t_{2,max} < 4.0 * T_A$. Phase cycling: $\phi = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = x$, $\phi_5 = x + bsp$; $\psi = x$, $\phi_{rec.} = x, 2(-x), x, -x, 2(x), -x$. Because a selective 180° pulse for $^{13}C^\alpha$ in the middle of delay $4T_A$ induces a Bloch–Siegert shift to $^{13}C'$ magnetization, a careful adjustment of phase (bsp) of the last $^{13}C'$ 90° (phase ϕ_5) pulse is necessary.

For IDPs, effort has been made to maximize the attainable resolution. To this end, whole $4T_A$ period that is the $^{13}C'$ (i)—$^{15}N(i + 1)$ transfer step is utilized for the frequency labeling of $^{13}C'$ shifts (see Note 3). Resolution in ^{15}N (ω_1) domain is limited only by the availability spectrometer time owing to the real-time implementation.

Thanks to rather uncomplicated magnetization transfer pathway, setting up the H(CA)CON experiment for optimal performance is relatively straightforward. The least efficient magnetization transfer step, a round trip from the $^{13}C'(i)$ spin to $^{15}N(i + 1)$ spin and back, involves a relatively long period $(1/J_{C'N} \sim 66$ ms) during which the signal decays rapidly especially at the highest magnetic field strengths where the $^{13}C'$ coherence is susceptible to rapid transverse relaxation of $^{13}C'$ magnetization (18). This should not

pose to a serious problem on IDPs, where the transverse relaxation rate of $^{13}C'$ spin is slow due to elevated internal dynamics of the polypeptide chain.

1. Setting of delays. Typical delay settings are 1.7; 2.4; 12–16.5; and 3.4 ms for τ, τ_2, T_A, and T_B, respectively. Delay τ_2 can be adjusted as described for (HCA)CON(CA)H (see Subheading 3.1).

2. Placing the transmitter offsets. Make sure that the ^1H, $^{13}C'$, $^{13}C^\alpha$, and ^{15}N (and ^2H in D_2O) transmitters are placed into appropriate frequencies, i.e., ~4.7, 176, 58, and 118 ppm, respectively (see also Table 1).

3. Calibrate ^1H, ^{13}C, and ^{15}N pulse widths (see Note 2).

4. Adjust Bloch–Siegert phase of the last 90° $^{13}C'$ pulse (labeled as ϕ_5 + bsp in Fig. 3b) (see Note 4). This is a critical step as it has direct influence on signal-to-noise ratio.

5. Set number of t_1 and t_2 increments for three-dimensional experiment and use States-TPPI quadrature detection protocol in t_1 and t_2. Initiate the experiment using at least two transients per data point.

The data collection was carried out using two transients per data point yielding overall experimental time of 22 h. Other relevant experimental and processing parameters for H(CA)CON spectrum are listed in Tables 1 and 2.

3.3. iH(CA)NCO

The intraresidual iH(CA)NCO experiment correlates solely intraresidual $^{15}N(i)$, $^{13}C'(i)$, $^1HA(i)$ spins as the name stands for. The magnetization transfer pathway can be shortly described as follows

$$^1H^\alpha(i) \rightarrow {}^{13}C^\alpha(i) \rightarrow {}^{13}C'(i) \rightarrow {}^{15}N(i)(t_1) \rightarrow {}^{13}C'(i)(t_2)$$
$$\rightarrow {}^{13}C^\alpha(i) \rightarrow {}^1H^\alpha(i)(t_3).$$
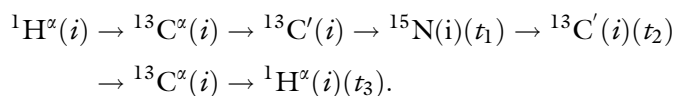
Intraresidual editing involves simultaneous transfer of magnetization from $^{13}C^\alpha$ to $^{13}C'$ and to both adjacent ^{15}N spins in a nested manner, making quenching of passive scalar interactions more challenging (19). Consequently, the resolution in $^{13}C'$ dimension is limited by the length of $2T'_C$ (see Note 5). Resolution in ^{15}N ω_1-domain is limited only by experimental time. This experiment is the most sophisticated of all three HA-detected experiments and requires a bit more careful adjustment of parameters.

1. Setting of delays. Typical delay settings are 1.7; 2.4; 9.5; 0–2; 16–17; and 12.5–14 ms for τ, τ_2, T_C, T_{CC}, T_A, and T_{CN}, respectively.

2. Placing the transmitter. Make sure that the ^1H, $^{13}C'$, $^{13}C^\alpha$, and ^{15}N (and ^2H in D_2O) transmitters are placed into appropriate
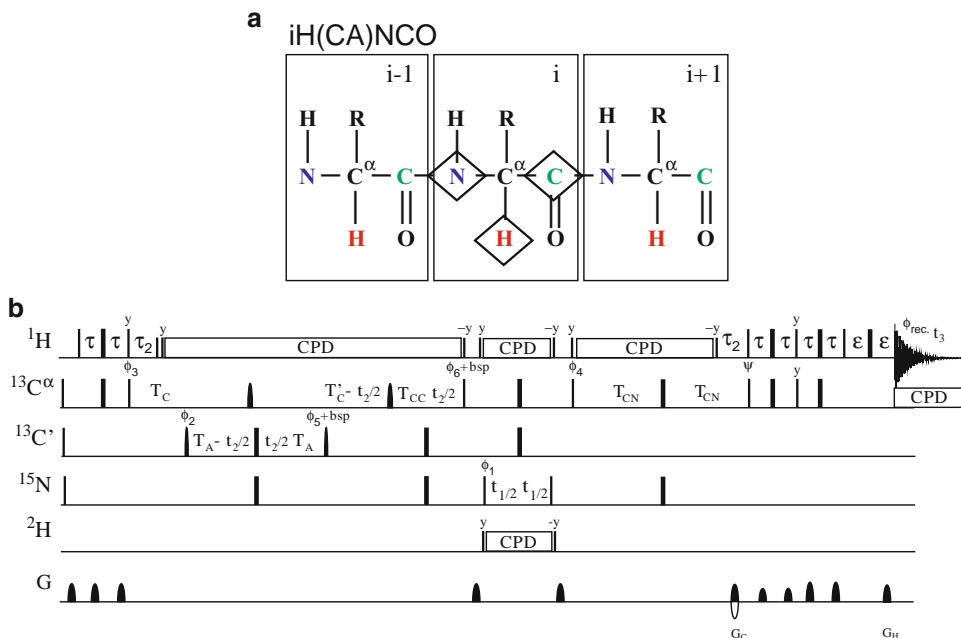
Fig. 4. Schematic representation of backbone atoms that are correlated in iH(CA)NCO experiment (**a**). $^1H^\alpha(i)$, $^{13}C'(i)$ and ^{15}N (*i*) spins, whose resonance frequencies are correlated in the iH(CA)NCO experiment in (**b**) are *highlighted*. The ^{13}C carrier is set initially to the middle of $^{13}C^\alpha$ region (58 ppm), shifted to carbonyl carbon region (176 ppm) before the 90° $^{13}C'$ pulse with phase ϕ_2, and shifted back to 58 ppm prior to 90° $^{13}C^\alpha$ pulse with phase ϕ_6. All rectangular 90° pulses for $^{13}C^\alpha$ (58 ppm) and 180° pulses for $^{13}C'$ (176 ppm) were applied with durations of 40.8 μs (90°) and 36.4 μs (180°) at 800 MHz, respectively. Two selective 90° pulses for $^{13}C'$ have the shape of center lobe of a sinc function and duration of 66.8 μs at 800 MHz. Three phase modulated 180° pulses, applied off-resonance for $^{13}C^\alpha$ during t_1 and t_2 periods had the shape of one-lobe sinc profile and duration of 60.4 μs. Delay durations: $T_C = 1/(2J_{C\alpha C'}) \sim 9.5$ ms; $T_A = 1/(4J_{C'N}) \sim 16$–$17$ ms; $T'_C = T_C + T_{CC}$; $T_{CC} = 1/(J_{C\alpha C\beta}) - 1/(4J_{C'N}) - 1/(2J_{C\alpha C'}) \sim 0$–$2$ ms; $T_{CN} \sim 14$ ms; $t_{2,max} < 2.0*T'_C$. Phase cycling: $\phi = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = x$; $\phi_5 = y + $ bsp; $\phi_6 = y + $ bsp; $\psi = x$, $\phi_{rec.} = x, 2(-x), x, -x, 2(x), -x$. Because selective 180° pulses on $^{13}C^\alpha$ and $^{13}C'$ induce Bloch–Siegert shifts to $^{13}C'$ and $^{13}C^\alpha$ magnetization, respectively, a careful adjustment of phases of $^{13}C'$ (ϕ_5) and $^{13}C^\alpha$ (ϕ_6) pulses is necessary.

frequencies, i.e., ~4.7, 176, 58, and 118 ppm, respectively (see also Table 1).

3. Calibrate ^1H, ^{13}C, and ^{15}N pulse widths (see Note 2).

4. Adjust the Bloch–Siegert phase of the last 90° $^{13}C'$ pulse (labeled as $\phi_5 + $ bsp in Fig. 4b) and the last $^{13}C^\alpha$ pulse prior to ^{15}N evolution period (labeled as $\phi_6 + $ bsp in Fig. 4b) (see Note 4).

5. Set number of t_1 and t_2 increments for three-dimensional experiment, and use States-TPPI quadrature detection protocol in t_1 and t_2. Initiate the experiment using at least two transients per data point.

The data collection was carried out using two transients per data point yielding overall experimental time of 14 h. Other relevant experimental and processing parameters for iH(CA)NCO spectrum are listed in Tables 1 and 2.

3.4. Backbone Assignment Protocol Using iH(CA)NCO, (HCA)CON(CA)H, and H(CA)CON Spectra

Next, we describe assignment procedure, which utilizes iH(CA)NCO, (HCA)CON(CA)H, and H(CA)CON spectra described above.

(Follow illustration in Fig. 5 and see example in Fig. 6).

1. During the assignment of this set of spectra, the easiest way to is to select ^{1}H onto the x-axis, ^{13}C and ^{15}N onto the y- and/or z-axis. For example in Sparky (or other spectral analysis program).

   (a) Plot iH(CA)NCO, (HCA)CON(CA)H, and H(CA)CON so that ^{1}H appears at x-axis, ^{13}C' at y-axis and ^{15}N at z-axis (Figs. 5 and 6). From now on this is called the $^{15}N_i$-plane set (see Note 5). Synchronize the ^{15}N-plane set to show the same ^{15}N-plane in all three spectra.

   (b) Duplicate spectra of iH(CA)NCO and H(CA)CON and adjust axes so that ^{1}H appears at x-axis, at ^{15}N y-axis and ^{13}C' at z-axis (Fig. 5). This is from here on called the $^{13}C_{i-1}$-plane set (see Note 6). Synchronize the ^{13}C-plane set to show the same ^{13}C'-plane in both spectra.
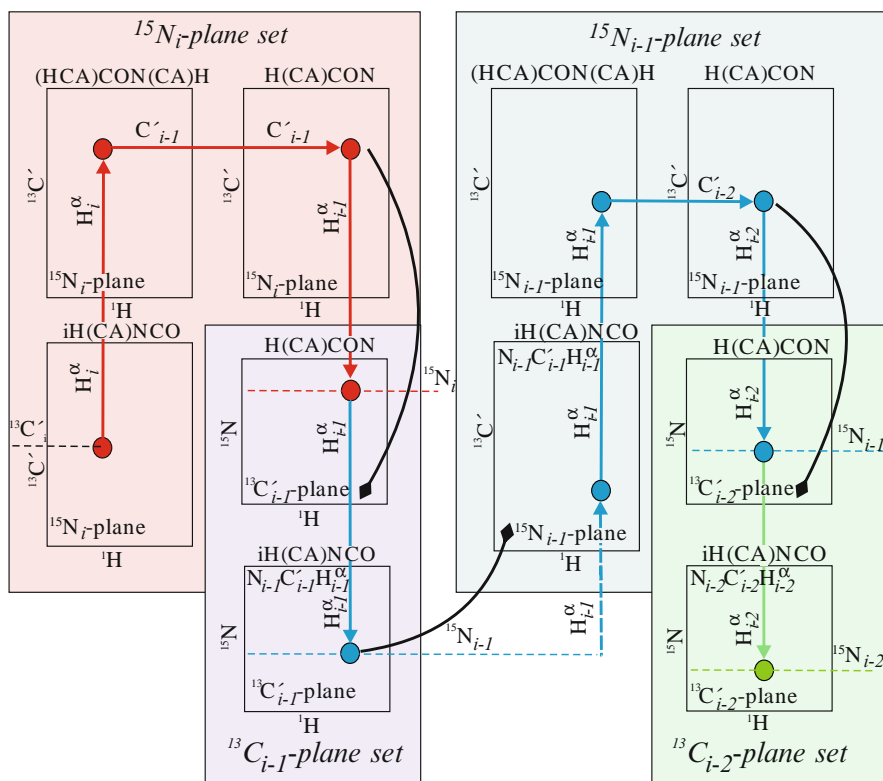
Fig. 5. Sequential assignment based on iH(CA)NCO (15), and (HCA)CON(CA)H (16) and H(CA)CON (15) spectra. Cross peaks resonating at the nitrogen frequency of the residues i, $i − 1$, and $i − 2$ are color-coded with *red*, *blue*, and *green*, respectively. $^{15}N_i$-plane set, $^{13}C_{i-1}$-plane set, $^{13}C_{i-2}$-plane set, and $^{15}N_{i-1}$-plane set are *highlighted* with *pink*, *violet*, *light blue*, and *light green*, respectively (also described in Subheading 3).
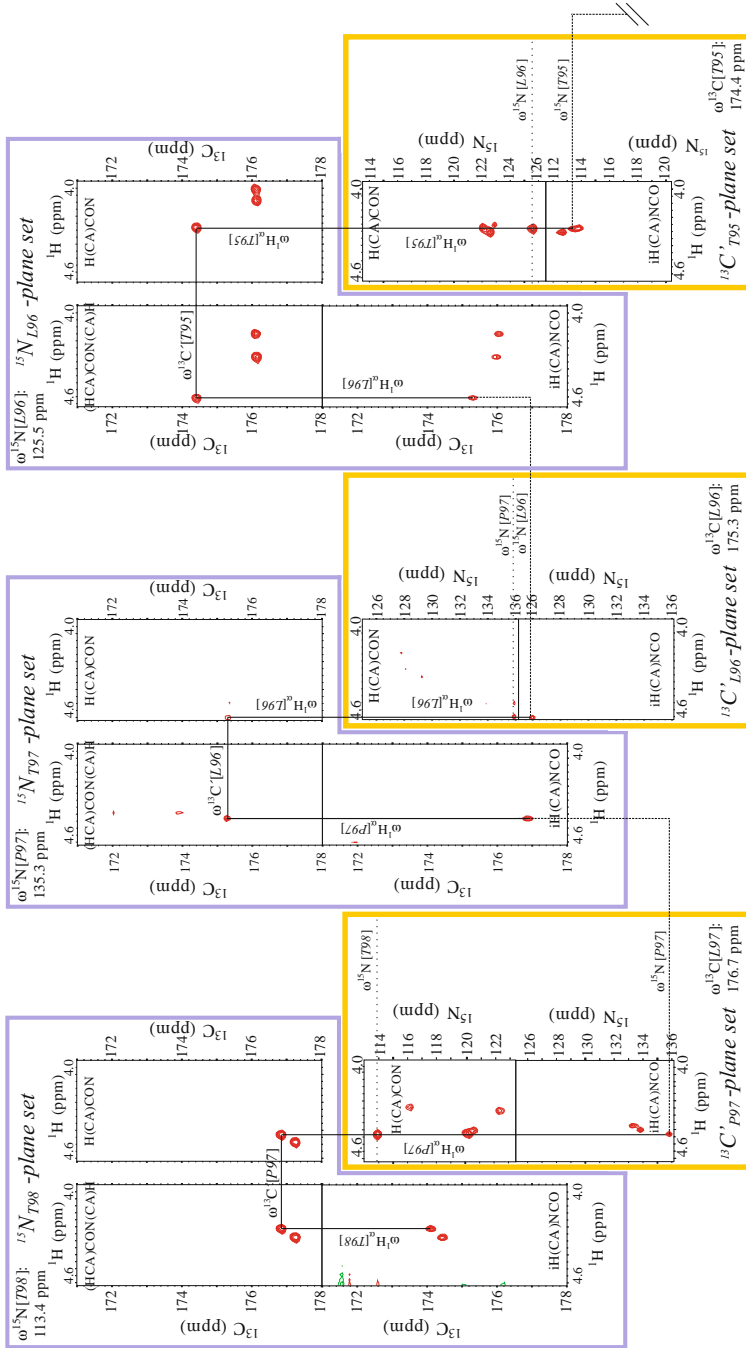
Fig. 6. Exemplification of backbone assignment procedure, based on iH(CA)NCO, (HCA)CON(CA)H, and H(CA)CON spectra. "Sequential walk" is by using residues R92-T98 of CT16 protein. Positive peaks are marked as *red* and negative as *green*. $^{15}N_i$-*plane sets* are framed with *violet* and $^{13}C_i$-*plane sets* with *orange* (see also Fig 5).
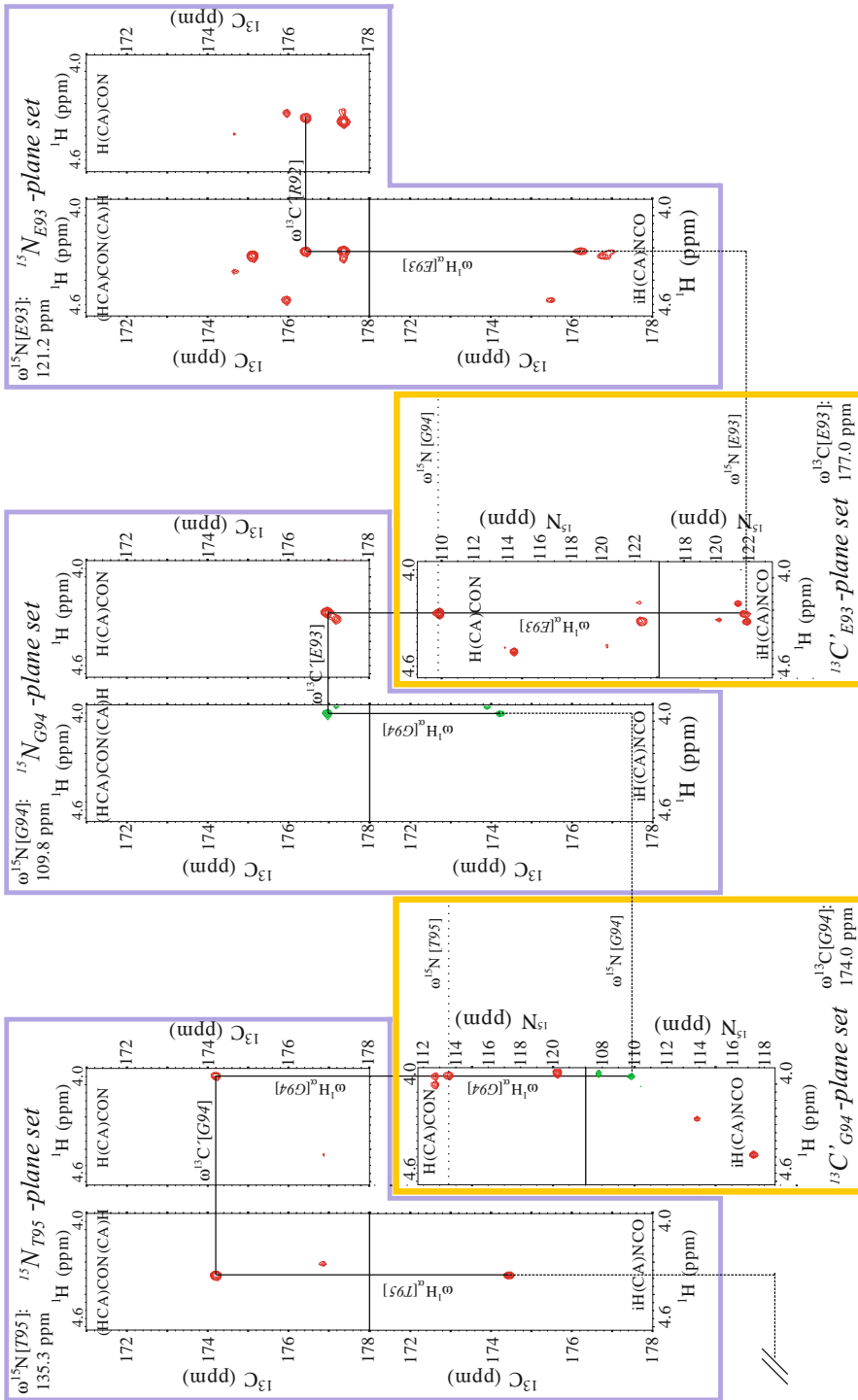
Fig. 6. (continued)

2. Start "sequential walk" from an amino acid residue with characteristic sign properties and/or chemical shift values.

   (a) Cross peaks corresponding to glycines show negative sign with respect to remaining residues in iH(CA)NCO and (HCA)CON(CA)H spectra.

   (b) On average, the smallest $^{13}C'$ chemical shifts belongs to glycines (~173 ppm), and largest for alanines (~177 ppm) (values taken from BMRB; http://www.bmrb.wisc.edu/ref_info/statful.htm).

   (c) On average, smallest ^{15}N chemical shifts belongs to glycines (~109 ppm), and largest for prolines (~140 ppm) (values taken from BMRB; http://www.bmrb.wisc.edu/ref_info/statful.htm).

   (d) For IDP's, proline residues have deviating $^{13}C'$ chemical shifts (our observation) between 176 and 177 ppm.

3. Select the $^{15}N_i$-*plane set*. Pick one peak from the iH(CA)NCO spectrum. At this point you have chemical shifts of ^{15}N, $^{13}C'$, and $^1H^\alpha$ of the residue i.

4. "Walk" on the $^{15}N_i$-*plane set* from iH(CA)NCO into (HCA)CON(CA)H along $^1H^\alpha{}_i$ chemical shift path. At this point you have chemical shifts of ^{15}N, $^{13}C'$, and $^1H^\alpha$ of the residue i and in addition $^{13}C'$ shift of the residue $i-1$.

5. Follow path on the $^{15}N_i$-*plane set* marked by $^{13}C'_i$ chemical shift from (HCA)CON(CA)H into the H(CA)CON spectrum. You have chemical shifts of ^{15}N, $^{13}C'$ and $^1H^\alpha$ of the residue i, chemical shift for $^{13}C'$ of the residue $i-1$ and in addition chemical shift of $^1H^\alpha$ of the residue $i-1$.

6. While still analyzing the H(CA)CON spectrum, select $^{13}C'$ chemical shift of the residue $i-1$ resulting in the $^{13}C'_{i-1}$-*plane set* (see Note 7).

7. "Walk" through the ^1H path into the iH(CA)NCO spectrum. As a result, you will get intraresidual peak of the residue $i-1$.

8. Jump into the $^{15}N_{i-1}$-*plane set*, viewing planes in all three spectra with chemical shift of $^{15}N_{i-1}$.

Repeat steps 4–8 as long as possible.

4. Notes

1. Attainable resolution in Hertz is defined by $1/AT$, where AT is acquisition time. Number of complex points (increments) in an indirectly detected dimension defines the maximum acquisition time $t_{i,\max} = \mathrm{ni}_i/\mathrm{sw}_i$ ($i = 1, 2, 3, \ldots$), where ni_i and sw_i are number of increments and spectral width, respectively.

2. This can be carried out in a "quick-and-dirty way" for ^1H pulse width in the ^{13}C-HSQC or ^{15}N-HSQC experiments. ^{15}N and especially ^{13}C pulse widths should then optimized using the particular experiment itself, i.e., by maximizing the signal intensity in the 1D spectrum by arraying ^{13}C pulse width.

3. Number of complex data points in t_2 is limited by $ni_2 < 4$ T_A*sw_2, where ni_2 and sw_2 refer to the number of t_2 increments and spectral width in ^{13}C dimension, respectively.

4. Without Bloch–Siegert phase shift the pulse ϕ_5 should be applied along the x- and y-axis in H(CA)CON and iH(CA)NCO experiments, respectively. However, additional angular rotation induced by the $180°$ $^{13}C^\alpha$ pulse changes this phase. Optimization of the Bloch–Siegert phase can be carried out by arraying the phase of pulse ϕ_5 from $0°$ to $180°$ in $10°$ steps and taking the maximum signal. The same procedure should be applied for optimizing the phase ϕ_6 of the $90°$ $^{13}C^\alpha$ pulse in iH(CA)NCO.

5. $2T'_C = 2(T_C + T_{CC})$, where T_C is ~9.5 ms and T_{CC} ~ 0–2 ms.

6. i is residue number, indicating its location at the primary structure, e.g., $i − 1$ is residue previous to i, $i − 2$ previous residue to $i − 1$ and so on. Similarly, subscript $i + 1$ behind the nuclei symbol, is used as an indication of amino acid locating after i, and so on.

7. You are viewing same correlation peak, only swapping $^{13}C'$ and ^{15}N axes.

Acknowledgment

References

1. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158

2. Permi P, Annila A (2004) Coherence transfer in proteins. Prog Nucl Magn Reson Spectrosc 44:97–137

3. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994) A suite of triple-resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^2H labeled proteins with high-sensitivity. J Am Chem Soc 116:11655–11666

4. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12731

5. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

6. Grzesiek S, Bax A (1993) The importance of not saturating H_2O in protein NMR—application to sensitivity enhancement and NOE measurements. J Am Chem Soc 115:12593–12594

7. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src Homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33:5984–6003

8. Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of ^1H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 29:4659–4667

9. Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514

10. Powers R, Gronenborn AM, Clore GM, Bax A (1991) Three-dimensional triple-resonance NMR of $^{13}C/^{15}N$-enriched proteins using constant-time evolution. J Magn Reson 94:209–213

11. Wang A, Grzesiek S, Tschudin R, Lodi PJ, Bax A (1995) Sequential backbone assignment of isotopically enriched proteins in D_2O by deuterium-decoupled HA(CA)N and HA(CACO)N. J Biomol NMR 5:376–382

12. Bottomley MJ, Macias MJ, Liu Z, Sattler M (1999) A novel NMR experiment for the sequential assignment of proline residues and proline stretches in $^{13}C/^{15}N$-labeled proteins. J Biomol NMR 13:381–385

13. Kanelis V, Donaldson L, Muhandiram DR, Rotin D, Forman-Kay JD, Kay LE (2000) Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. J Biomol NMR 16:253–259

14. Sayers EW, Torchia DA (2001) Use of the carbonyl chemical shift to relieve degeneracies in triple-resonance assignment experiments. J Magn Reson 153:246–253

15. Mäntylahti S, Aitio O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. J Biomol NMR 47:171–181

16. Mäntylahti S, Hellman M, Permi P (2010) Extension of the HA-detection scheme: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. Unpublished manuscript

17. Yao J, Dyson HJ, Wright PE (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. FEBS Lett 419:285–289

18. Permi P, Annila A (2001) A new approach for obtaining sequential assignment of large proteins. J Biomol NMR 20:127–133

19. Mäntylahti S, Tossavainen H, Hellman M, Permi P (2009) An intraresidual i(HCA)CO(CA)NH experiment for the assignment of main-chain resonances in 15N, 13C labeled proteins. J Biomol NMR 45:301–310

20. Shaka AJ (1985) Composite pulses for ultra-broadband spin inversion. Chem Phys Lett 120:201–205

21. Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broad-band decoupling—Waltz-16. J Magn Reson 52:335–338

22. Kupće E, Wagner G (1995) Wideband homonuclear decoupling in protein spectra. J Magn Reson 109A:329–333

23. Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR-spectra without phase cycling—application to the study of hydrogen-exchange in proteins. J Magn Reson 85:393–399

Part III

Vibrational Spectroscopy

Chapter 16

Fourier Transform Infrared Spectroscopy of Intrinsically Disordered Proteins: Measurement Procedures and Data Analyses

Antonino Natalello, Diletta Ami, and Silvia Maria Doglia

Abstract

A peculiar property of intrinsically disordered proteins (IDPs), or of intrinsically disordered domains, is the absence of a well-defined three dimensional structure under native conditions. Moreover, IDPs usually acquire a specific structure in the presence of different interactors. In this framework, Fourier transform infrared (FTIR) spectroscopy is a powerful tool to assess the disordered character of a protein and to study its induced folding. In this chapter, we will show the detailed experimental procedures to measure the FTIR spectra of protein samples and the spectral analyses required to obtain information on the protein secondary structures and aggregation.

Key words: α-synuclein, Amyloid, Attenuated total reflection, Fourier transform infrared spectroscopy, Induced folding, Intrinsically disordered proteins, Protein aggregation, Secondary structures

1. Introduction

Intrinsically disordered proteins (IDPs) are characterized by the lack of a defined three dimensional structure under physiological conditions in vitro. The binding of an IDP with its different interactors is often associated with the acquisition and/or stabilization of secondary and/or tertiary structures (1, 2). The peculiar structural properties of IDPs result in several functional advantages as, for instance, the capability of these proteins to dynamically interact with a large number of different partners. These aspects are crucial in determining the complex functional role of IDPs that can be found at the crossroad of different physiological and pathological pathways. In the last decade, a growing number of proteins and of protein domains has been recognized to be intrinsically disordered, leading to an enormous interest on IDPs. Indeed, they are involved

in several biological processes as cell cycle and gene transcription regulation (3). Moreover, IDPs have been found to have a crucial role in diseases, as in tumors and in amyloidoses (2). For instance, α-synuclein is a well-known IDP whose aggregation is involved in the Parkinson' disease (4).

From the above considerations, it is, therefore, desirable to have methods able to characterize the structural properties of IDPs and to study their induced folding in different environmental conditions (5). Among the available spectroscopic techniques (5), Fourier transform infrared (FTIR) spectroscopy has been proved to be a powerful tool for the study of protein conformation, dynamics, and aggregation (6–12). In particular, the Amide I band (1,700–1,600 cm^{-1}), mainly due to the CO stretching vibration of the peptide bond, is the most used in protein structural studies. This band is sensitive to the backbone secondary structures and to the formation of intermolecular β-sheets in protein aggregates. Indeed, FTIR spectroscopy allows to examine not only proteins in solution but also highly scattering protein aggregates. A critical point to obtain the FTIR absorption spectrum of a protein in water solution is the subtraction of the solvent spectrum, since water displays a very high absorption in the Amide I spectral region. Under these conditions, the required protein concentration is of the order of 1–10 mg/ml. However, it is possible to decrease the protein concentration down to 0.3 mg/ml using deuterated water (D_2O) as resuspension solvent. Nevertheless, we should underline that for FTIR analyses the total amount of protein is limited by the low volume required for each measurement, namely 5–20 μl. It is also possible to further decrease the protein concentration performing the FTIR analysis in attenuated total reflection mode (ATR), measuring the protein as a film on the ATR plate.

We should note that FTIR spectroscopy has been successfully applied in IDP characterization for the study of their native secondary structures, induced folding, and aggregation—reviewed in ref. 9. Several reviews can be found in literature, where the FTIR basic technique and its application to protein studies are illustrated (6–12).

Here, we report the detailed practical procedures to obtain high quality and reproducible FTIR spectra of protein samples. In particular, we described in detail, first, the method that can be employed to measure the FTIR spectrum in the transmission mode of the well-known IDP α-synuclein, in deuterated buffer solution. Then, we examine the measurements of proteins in H_2O buffer in transmission and in ATR. Finally, we will describe the procedures of spectral analysis, most employed to get information on the protein secondary structures.

2. Materials

2.1. Chemicals, Buffers, and Proteins

Deuterium oxide, sodium phosphate monobasic, sodium phosphate dibasic, sodium chloride, and both deuterated and non-deuterated hydrochloric acids and sodium hydroxide were purchased from Sigma-Aldrich (St. Louis, MO, USA).

A stock solution of sodium phosphate buffer at pH 7.4 was obtained mixing sodium phosphate monobasic 0.1 M with sodium phosphate dibasic 0.1 M (~1:4 of volumes), adjusting the final pH with hydrochloric acid and sodium hydroxide. For α-synuclein FTIR analysis, a final concentration of 20 mM phosphate buffer has been used (13). To obtain a deuterated buffer, a non-deuterated buffer is lyophilized, resuspended in D_2O, and the final pH adjusted by deuterated hydrochloric acid and sodium hydroxide. Alternatively, deuterated sodium phosphate buffer can be prepared as the non-deuterated one, using D_2O (instead of H_2O) to resuspend the sodium phosphate monobasic and the sodium phosphate dibasic lyophilized powders. We should recall that the pH values measured on the pH meter can be converted in pD values by adding +0.4 (14).

Recombinant wild-type human α-synuclein was expressed in *Escherichia coli* and purified as previously described (15). The protein solution was dialyzed against Milli-Q water and aliquots were lyophilized and stored at −80 °C.

Recombinant *Candida rugosa* lipase 1 (CRL1) was expressed and secreted by *Pichia pastoris* and purified as previously described (16).

Bovine β-lactoglobulin (BLG) was purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification.

2.2. FTIR Spectrometers

The FTIR absorption spectra of α-synuclein and BLG were collected using a Varian 670-IR spectrometer (Varian Australia Pty Ltd, Mulgrave VIC, AU), equipped with a nitrogen-cooled mercury cadmium telluride (MCT) detector, under accurate dry-air purging (see Note 1). The following instrumental parameters were employed: 2 cm^{-1} spectral resolution, 25 kHz scan speed, 1,000 scan coaddition, and triangular apodization.

The FTIR absorption spectra of CRL1 were collected using an FTS-40A spectrometer (Bio-Rad, Digilab Division, Cambridge, MA, USA) equipped with a deuterated triglycine sulfate (DTGS) detector and an air dryer purging system under the following conditions: 2 cm^{-1} resolution, 5 kHz scan speed, 256 scan coaddition, and triangular apodization.

2.3. Transmission Cell

For FTIR measurements in the transmission mode, a sample solution was placed in a temperature-controlled transmission cell (Wilmad, Buena, NJ, USA) (see Fig. 1), with BaF_2 windows
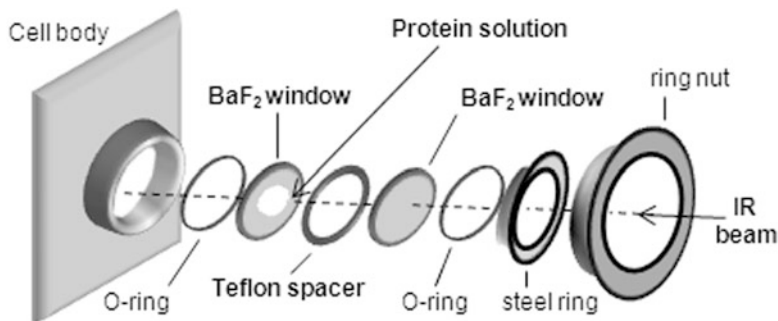
Fig. 1. Scheme of a transmission cell made by two BaF$_2$ windows. The optical path is determined by the thickness of the teflon spacer.

(see Note 2) (Ø 13 mm × 2 mm dimensions; Korth kristalle GmbH, Kiel, Germany), and an optical path made by a teflon spacer (Wilmad, Buena, NJ, USA) of about 6–15 μm for H$_2$O or of 50–150 μm for D$_2$O measurements.

2.4. ATR Device

For ATR measurements, the Golden Gate device (Specac Ltd, UK) with a single reflection diamond crystal plate was employed.

2.5. Softwares

The Resolutions-Pro software (Varian Australia Pty Ltd, Mulgrave VIC, AU) was employed for spectral collection and Fourier self deconvolution (FSD) analysis. All the other spectral analyses were performed using the GRAMS/32 software (Galactic Industries Corporation, Salem, NH, USA).

3. Methods

3.1. FTIR Spectrum of α-Synuclein in Deuterated Buffer Measured in Transmission Mode

3.1.1. Measurements

In this section we reported the protocol to obtain the FTIR spectrum in transmission mode of a protein in a deuterated buffer solution, taking as an example the IDP α-synuclein (13).

1. Resuspend the lyophilized α-synuclein in the proper volume of deuterated sodium phosphate buffer to reach a given protein concentration. In the example reported here, α-synuclein was measured at a concentration of 400 μM in 10 mM of deuterated sodium phosphate buffer at pH 7.4 (13).

2. Collect the background spectrum using the empty cell (Fig. 1), placed in the spectrometer sample compartment, under the same instrumental parameters that will be employed for the protein solution measurement: 2 cm^{-1} spectral resolution, 25 kHz scan speed, triangular apodization, and 1,000 scan coaddition.
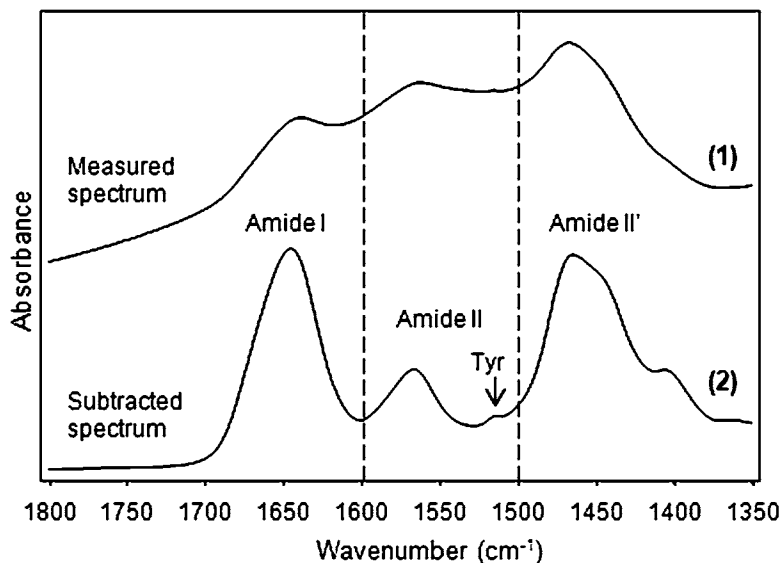
Fig. 2. FTIR spectra of α-synuclein in D_2O buffer measured in transmission mode. FTIR absorption spectrum of 400 μM α-synuclein, in 10 mM deuterated sodium phosphate buffer at pH 7.4, (spectrum 1) as measured and (spectrum 2) after the subtraction of the D_2O buffer spectrum.

3. Set the temperature of the transmission cell. Load 20 μl of the protein sample in the transmission cell, wait for the accurate purging of the sample compartment (~10 min in our system) and for the thermal equilibrium of the sample, before collecting the absorption spectrum. The solution volume placed in the cell depends on the cell window diameter and the optical path (namely, on the teflon spacer employed). Here, we used a 150 μm teflon spacer.

   The absorption spectrum of α-synuclein 400 μM in 10 mM of deuterated sodium phosphate buffer at pH 7.4 is reported in Fig. 2, in the 1,800–1,350 cm^{-1} spectral region.

4. To measure the buffer spectrum required for the subsequent subtraction (see later), repeat step 2 for the background acquisition and step 3 for the sample measurement, in this case loading 20 μl of the buffer solution in the transmission cell.

3.1.2. Basic Spectral Analyses

1. *Buffer subtraction*. The IR absorption of the protein was obtained by the subtraction of the buffer spectrum from that of the protein solution. The subtraction factor was adjusted to obtain a flat baseline around 1,800 cm^{-1} (where the protein does not absorb), removing also the D_2O bending band around 1,210 cm^{-1}. The FTIR spectrum of α-synuclein after buffer spectrum subtraction is also reported in Fig. 2. Three main spectral regions can be identified: the Amide I band, mainly due to the CO stretching vibration of the peptide bond; the residual Amide II band, mainly due to the backbone
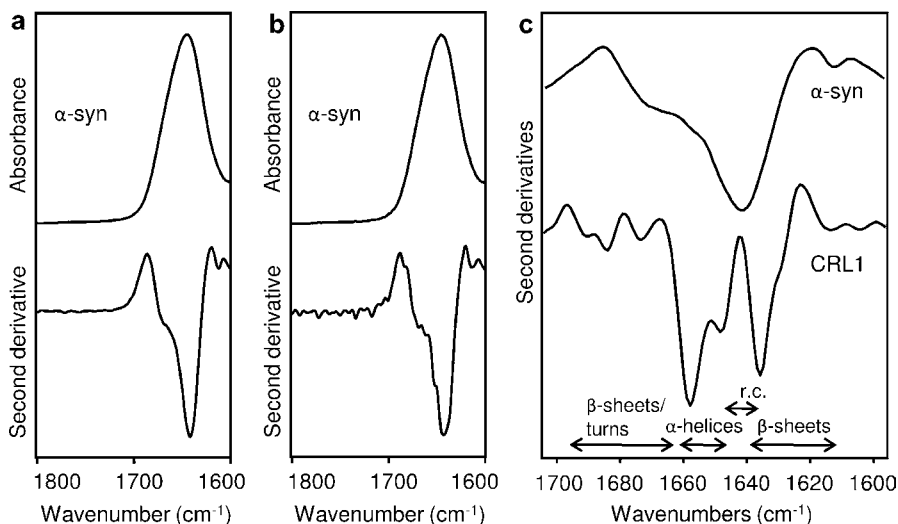
Fig. 3. FTIR spectra of α-synuclein: vapor subtraction and second derivative analysis. FTIR absorption spectra of α-synuclein already corrected for the buffer spectrum (**a**, **b**) and their second derivatives (**a–c**). While the two absorption spectra (**a** and **b**: top traces) look very similar, the second derivative analysis of spectrum in (**b**) discloses the presence of water vapor in the 1,720–1,800 cm^{-1} range. The residual water vapor leads to additional spectral components in the Amide I band not due to the protein secondary structures (**b**). In (**c**) the second derivative of α-synuclein (α-syn) is compared with that of the well-folded protein Lipase 1 of *C. rugosa* (CRL1). The Amide I band assignment to the protein secondary structures is also given in (**c**).

NH in-plain bending and CN stretching, with the contributions of other vibrations; and the Amide II′ (see Note 3). The absorption of the tyrosine (Tyr) side chain is also evident (see Note 4). In the particular case of α-synuclein in deuterated buffer, the Amide I band displays a maximum around 1,645 cm^{-1} (see Subheading 3.4 for band assignment).

2. *Water vapor subtraction*. After buffer subtraction, also the subtraction of the water vapor spectrum (see Note 5) has to be performed, when necessary, since water vapor displays very sharp peaks in the Amide I region. The subtraction factor can be evaluated directly in the protein absorption spectrum examining the vapor peak at ~1,653 cm^{-1} (17). In addition, residual vapor can be detected in the second derivative spectra (see later, step 4a of Subheading 3.1.2) by checking the presence of residual vapor signal in the 1,720–1,850 cm^{-1} spectral region, where the absorption of vapor, but not that of the protein, occurs (Fig. 3).

3. *Spectral smoothing*. Several smoothing function can be applied before the calculation of the second derivative of the absorption spectrum. Here we used an 11-point binomial smoothing to obtain satisfactory results. A larger number of points might be required for spectra with higher noise level. Again, this can be verified inspecting the second derivative spectrum in the 1,800–1,720 cm^{-1} region (see later, step 4a).

4. *Resolution enhancement*. The Amide I band displays a broad envelop due to the absorption of the protein backbone in the different secondary structures that overlap one over the other. To disclose these components, two resolution enhancement methods can be applied to the measured absorption spectrum, namely the second derivative analysis (18) and the FSD (6, 19).

(a) *Second derivative analysis*. The second derivative of the measured spectrum can be performed following different algorithms (18). Here we used the 5-point Savitsky-Golay algorithm after the 11-point binomial smoothing of the spectrum. As shown in Fig. 3a, the band components in the measured spectrum appear as negative peaks at the same wavenumbers in the second derivative spectrum. An enhanced spectral resolution is achieved since in the second derivative the band half width is reduced to about a factor 2.7. It is important to note that peak height in the second derivative is proportional to the original height but inversely proportional to the square of the original half width (18). Consequently, sharp peaks, as those of water vapor, noise, and of interference fringes, are strongly intensified compared to the relative broad band components of the different protein secondary structures. Thus, spectra with high signal-to-noise ratio and accurate purging of the spectrometer are mandatory requisites for reliable second derivative analysis. The second derivative spectrum of α-synuclein is reported in Fig. 3c (redrawn from ref. 13), together with that of the native-folded protein lipase 1 of *C. rugosa* (CRL1) (redrawn from ref. 16). The comparison clearly highlights the disordered character of α-synuclein that displays few broad components, while CRL1 second derivative spectrum showed well-resolved peaks assigned to the ordered secondary structures of the protein (see Subheading 3.4 for band assignment).

(b) *Fourier self deconvolution*. The FSD approach achieves the enhancement of the spectrum resolution working on the interferogram of the measured spectrum. Therefore, a reverse Fourier transformation is applied to the measured spectrum and multiplied by an exponential increasing weighting function. An appropriate apodization function is subsequently applied. This treatment results in a decreased rate of the interferogram decay. Finally, the Fourier transform of this new interferogram leads to a deconvoluted spectrum whose band components are characterized by a reduced band width, obtaining in this way an enhanced spectral resolution. In this approach, three crucial parameters are required: (1) the K-factor that determines the extent of the band narrowing (its maximal value is related to

the signal-to-noise ratio of the measured spectrum and should be limited to values lower than 3); (2) the half width of the band components to be resolved (values in the range from 6.5 to 15 cm^{-1} are typically used); (3) the apodization function that takes into account the limited distance of the moving mirror in the interferometer. The choice of these parameters is crucial to avoid artifacts in the deconvoluted spectrum. A criterion to evaluate the goodness of the resulting deconvoluted spectrum can be the comparison with the second derivative of the measured spectrum: the same components at the same peak positions should be obtained.

3.2. FTIR Spectrum of CRL1 in H₂O Solution Measured in Transmission Mode

In this section we briefly described the procedure to obtain and analyze the FTIR spectrum in transmission mode of a protein in a non-deuterated buffer solution, taking as an example the well-folded protein CRL1 (16). To this aim, the protocol described above (Subheadings 3.1.1 and 3.1.2) can be used with some essential modifications required by the very high absorption of H_2O in the Amide I spectral region: (1) a higher protein concentration is required for measurements in H_2O buffer (1–15 mg/ml of protein) compared to D_2O buffer; (2) a shorter optical path, given by the teflon spacer (see Fig. 1 and step 3 of Subheading 3.1.1), should be employed (typically 6–15 μm).

As shown in Fig. 4a (redrawn from ref. 16), the spectrum of CRL1 in H_2O solution is very similar to that of H_2O, making the solvent subtraction a crucial point. The FTIR absorption spectrum of CRL1 (Fig. 4b) was, therefore, obtained by the subtraction of the H_2O spectrum (collected under identical conditions) to that of the protein solution. The subtraction factor was adjusted in order to obtain a flat baseline in the 1,800–2,300 cm^{-1} range, where the water combination band (around 2,135 cm^{-1}), but not protein absorption, occurs.

Basic spectral analyses (namely vapor subtraction, smoothing, FSD, and second derivative calculation) can be performed on the protein absorption spectrum as described above (Subheading 3.1.2).

3.3. FTIR Spectrum of BLG Measured in ATR as Protein Film

ATR/FTIR measurements for protein characterization are widely employed thanks to the sampling easiness of this approach. In particular, the sample (in solution or in solid form) is placed on an ATR element having a refractive index higher than that of the sample. An ATR device is schematized in Fig. 5. The infrared beam reaches the ATR element at the interface with the sample with an angle higher than the limit angle of total reflection. The beam is, therefore, totally reflected, reaching the detector after one or more reflections. At each reflection, the evanescent beam wave penetrates into the sample where it can be absorbed. The beam penetration depth depends on the refractive indices of the ATR element and
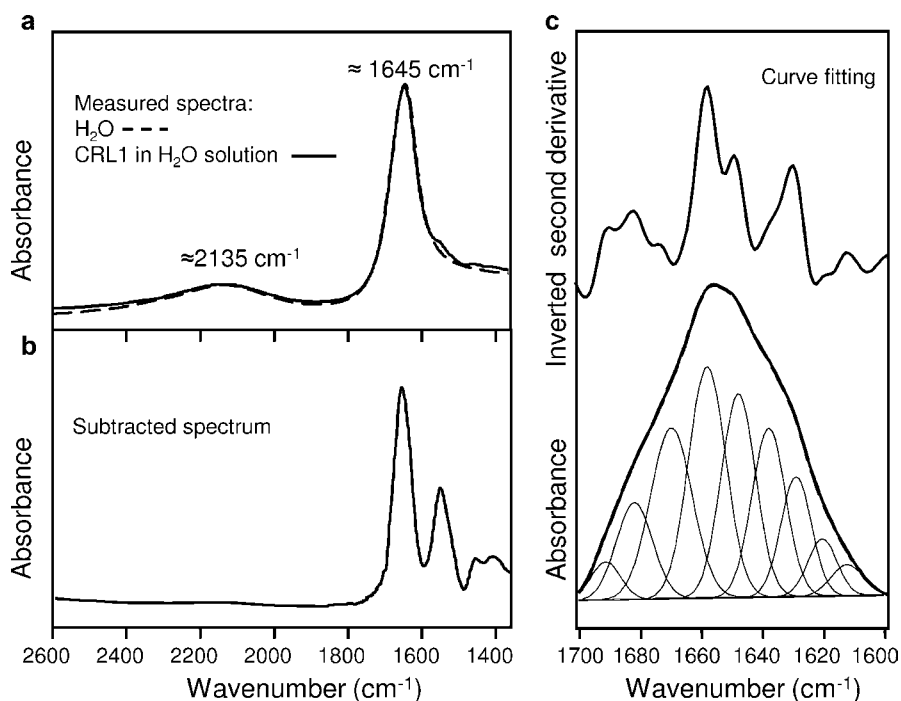
Fig. 4. FTIR spectra of CRL1 in H_2O solution measured in transmission mode. FTIR absorption spectra of H_2O (**a**), of 1.7 mM CRL1 in H_2O solution (**a**), and of the protein after the subtraction of the H_2O spectrum (**b**). The curve fitting in the Amide I band is given in (**c**): measured absorption spectrum (*continuous line*); curve-fitted spectrum (*dashed line*) into Gaussian components. The second derivative is also reported after multiplication by a factor-1 (inverted second derivative, top trace).

of the sample, on the incident angle, and on the beam wavelength ([20]). Consequently, the intensities of the band components in the spectrum can be corrected to compensate for the different penetration depth of the evanescent wave at the different wavenumbers ([20]).

In this section we reported a protocol to obtain the ATR/FTIR spectrum of a protein film taking as an example the BLG. The main advantage of protein film ATR/FTIR measurements is that only a few micrograms of protein are required. The great potential of this method for IDP characterizations is well highlighted by the interesting examples found in the literature on α-synuclein ([21], [22]) and other IDPs ([23], [24]).

3.3.1. Measurements and Basic Analyses

1. Collect the background spectrum, after the accurate cleaning of the ATR element, under the same instrumental parameters that will be employed for the samples. In the example reported here, we employed the Golden Gate ATR system (Specac Ltd, UK) with a single reflection diamond plate.

2. Depose 1–10 μl of the protein solution (depending on the protein concentration) on the ATR plate (Fig. [5], see Note 6)
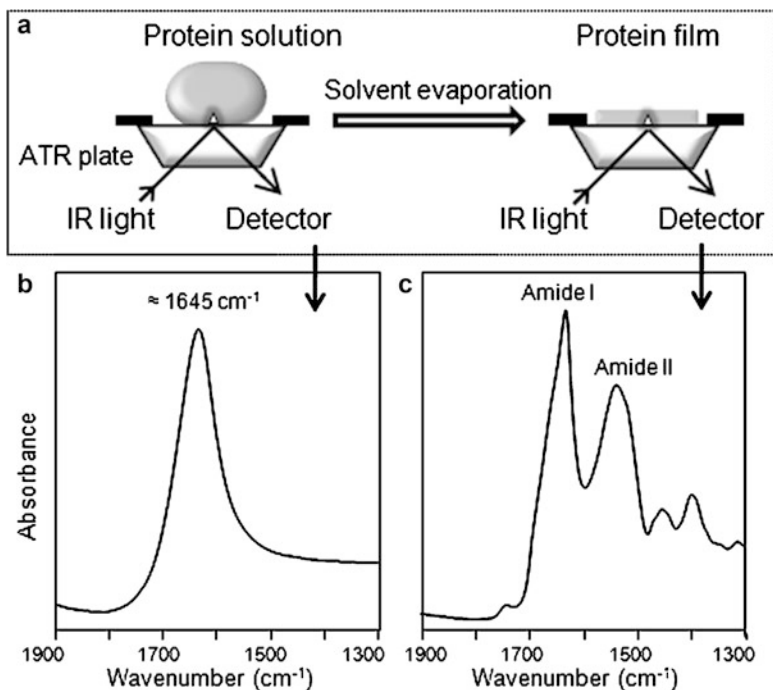
Fig. 5. ATR/FTIR spectrum of BLG in form of protein film. (**a**) Scheme of an ATR device and of the procedure to obtain a protein film. An aliquot of 5 μl of the BLG solution (0.5 mg/ml in H_2O) was deposed on the ATR plate (**a**). Its ATR/FTIR absorption spectrum (**b**) is dominated by the H_2O absorption around 1,645 cm^{-1}. After the solvent evaporation, a protein film was obtained (**a**). Its ATR/FTIR absorption spectrum (**c**) displays the typical protein bands in the Amide I and Amide II regions. In particular, the two components at ~1,628 and ~1,693 cm^{-1} can be assigned to the native β-sheet structures of the protein.

and allow the evaporation of the solvent (see Notes 7 and 8) to obtain a protein film.

3. Collect the spectrum of the protein film. The measured ATR/FTIR spectrum of BLG in the form of protein film is reported in Fig. 5. The main component in the Amide I region occurs at ~1,628 cm^{-1} that, together with the ~1,693 cm^{-1} band, can be assigned to the native β-sheet structures of the protein (25) (see Subheading 3.4 for band assignment).

4. Spectral analyses. The vapor spectrum subtraction (if required), smoothing, and resolution enhancement of the protein spectrum can be performed as described in Subheading 3.1.2.

3.4. Amide I Band Assignment to the Protein Secondary Structures

After the identification of the different components of the Amide I band, the next crucial step is their assignment to the protein secondary structures. Based on experimental FTIR studies of model compounds, peptides, and proteins with known structure and on theoretical analyses, it has been possible to correlate the peak position of the components to a specific protein secondary structure (6, 12). It should be noted that, often, band assignment is not

an easy task, since the absorption of the different secondary structures can partially overlap in the spectrum (i.e., random coil and α-helices overlap in the spectra of proteins in H_2O buffer) and the specific peak position of each secondary structure can vary depending on its interactions and peculiar properties (i.e., length of the structure, H-bonding with the solvent, etc.). Therefore, a critical evaluation of the band assignment is required. This can be obtained by combining different FTIR experiments (i.e., hydrogen/deuterium exchange, thermal and chemical protein denaturation) and by comparing the FTIR results with those of complementary techniques, as for instance circular dichroism.

Here, we briefly report the peak positions of the different secondary structures in the Amide I region.

Random coil. The FTIR response of random coil structures is characterized by a broad band centered around 1,654 cm^{-1} in H_2O (extremes 1,660–1,642 cm^{-1}), a spectral region where also the absorption of α-helix secondary structure occurs. However, upon H/D exchange in D_2O, only the band due to random coil downshifts to around 1,645 cm^{-1} (extremes 1,654–1,639 cm^{-1}), therefore allowing to discriminate between random coil and α-helix structures in the protein spectrum. As can be seen in Fig. 3, α-synuclein in D_2O buffer displays a broad absorption spectrum centered at 1,645 cm^{-1}, in agreement with its disordered character.

Moreover, both in H_2O and in D_2O, the absorption of open loops can be also found around 1,643 cm^{-1}. The comparison between the spectra of the protein in H_2O and in D_2O buffers can be, therefore, useful to evaluate the contribution of random coils, α-helix, and loops (26–28).

α-helices. The infrared response in the Amide I band of α-helices occurs in the region 1,660–1,648 cm^{-1} in H_2O buffer. Contrary to random coil structures, a downshift of only few wavenumbers takes place in D_2O buffer. The actual peak position depends on the peculiar structural properties of the helices. For instance, short and flexible helices absorb at higher wavenumbers (6) and α-helices forming H-bonding with the solvent (hydrated α-helices) shift toward lower wavenumbers, giving peaks in D_2O around 1,632–1,637 cm^{-1} (29–31).

β-sheets. The infrared response in the Amide I region of β-sheets structure is characterized by two absorption bands of different intensity. A low frequency band occurs in water around 1,633 cm^{-1} (extremes 1,640–1,623 cm^{-1}) and a high frequency band of lower intensity around 1,686 cm^{-1} (extremes 1,695–1,674 cm^{-1}). Both bands are downshifted in D_2O respectively to 1,638–1,615 cm^{-1} and 1,694–1,672 cm^{-1}. The actual band positions and intensities of β-sheet structures depend critically on the geometry of the structure, such as the strand twist angle, the number of strands per sheet (32), and the H-bond strength (10, 33). Moreover, a reduced

splitting between the two components and a reduced intensity of the high frequency band are expected for a parallel β-sheet (10). We should note that it is difficult to discriminate between parallel and antiparallel β-sheets structures as a similar infrared response has been observed for the two structures in native proteins (34, 35). Practically, a reliable parallel β-sheet assignment can be done when only the β-sheet band at low frequency is observed ((36) and references therein).

Protein aggregates. Intermolecular β-sheet structures in protein aggregates display shifted peak positions compared with those of native intramolecular β-sheets. In particular, peak extremes for the two β-sheets components occur, respectively, in the ranges: $1,630–1,620\ cm^{-1}$ and $1,698–1,692\ cm^{-1}$ in H_2O; $1,630–1,611\ cm^{-1}$ and $1,690–1,680\ cm^{-1}$ in D_2O.

β-turns and other secondary structures. The absorption of β-turns can be found around $1,686–1,660\ cm^{-1}$ in H_2O and $1,691–1,653\ cm^{-1}$ in D_2O (12). The infrared response of other secondary structures also occurs in the wide Amide I range (7, 10, 12, 18).

3.5. Protein Secondary Structure Evaluation

After the identification of the different components of the Amide I band and their assignment to the protein secondary structures, it is possible to follow their relative variations induced by changes in the environmental conditions. This can be obtained following the changes in peak intensities and peak positions that occur in the second derivative spectra during the process under investigation, as thermal or chemical unfolding, interactions with other partners, and exposure to different environmental conditions.

Moreover, it is also possible to evaluate the percentage of each secondary structure component by the curve fitting of the measured spectrum (see Note 9). The crucial point of this analysis is the selection of the input parameters: baseline, number of components, peak position, band width, and band intensity, as well as the choice of the analytical function to approximate each component.

Here we report the protocol that we follow for this analysis (16, 37).

1. To identify the number and peak positions of the input components, perform the second derivative and the FSD analyses on the measured spectrum (see step 4 of Subheading 3.1.2).

2. Set the input parameters for the curve fitting as follow: linear baseline; number and peak positions taken from second derivative and FSD analyses; the initial band heights of the main components can be set at 90 % of that of the measured spectrum, and at 70 % for the other components; band width can be estimated from second derivative and FSD analyses; select a Gaussian function to approximate each component (Lorentzian or mixed Lorentzian–Gaussian functions can be also employed).

3. Perform a first fitting of the measured spectrum fixing the band positions and letting free the other parameters to adjust iteratively in order to find the best set of fitting functions.

4. Perform a second fitting of the measured spectrum using the result of the first fitting as input parameters and letting free all the parameters to adjust iteratively in order to find the final set of the best fitting Gaussians.

5. The fractional area of each Gaussian component over the total area (given by the sum of all components) represents the percentage of the corresponding secondary structure.

An example of Gaussian curve fitting of the CRL1 spectrum obtained employing the above-described approach is given in Fig. 4.

We should note that also other approaches have been developed for quantitative secondary structure evaluation, (see, for instance, ref. 38 and references therein, and ref. 39).

4. Notes

1. The accurate purging of the spectrometer and of the sample compartment is an essential condition to obtain reproducible FTIR spectra. Indeed, water vapor displays several sharp peaks in the Amide I region. For this reason, residual vapor-interference can seriously affect the second derivative absorption spectra introducing additional intense components (see also Fig. 3, steps 2 and 4a of Subheading 3.1.2).

2. We should underline that not only the cleaning but also the optical polishing of the infrared windows is crucial for the success of the measurements. For instance, bubble formation is reduced in well-polished windows. In particular, we suggest to use a diamond polishing paste (size 1–9 μm).

3. The Amide II band is very sensitive to the hydrogen/deuterium (H/D) exchange. During H/D exchange the Amide II band decreases in intensity and a new band, called Amide II′, arises and grows in intensity around 1,490–1,460 cm^{-1}. Therefore, the kinetics of H/D exchange can be monitored by the time course of Amide II and Amide II′ intensity changes (20, 40).

4. Peaks due to amino acid side chain can be observed in the protein mid-infrared absorption spectrum within a wide range. Interestingly, the analyses of these bands can give useful information on the amino acid side chain conformation, such as exposure and H-bonding (41). A review of amino acid side chain absorptions in the mid-infrared can be found in ref. (42).

5. To measure the water vapor absorption spectrum, the water vapor level in the sample compartment can be increased—after

the background collection—reducing the rate of the dry-air flow of the purging apparatus.

6. At this point, it is also possible to measure the ATR/FTIR spectrum of the protein solution. Also in this case the protein spectrum can be obtained after the subtraction of the solvent ATR/FTIR spectrum collected under the same conditions. However, the measurement of the protein solution requires a higher protein concentration compared to that for the measurement of the protein film.

7. We suggest to speed the evaporation of the solvent by a gentle flow of nitrogen.

    If salt crystals interfere with the infrared response, rinse gently the protein on the ATR plate with deionizer water.

8. We should underline that dehydration could affect in some cases the protein secondary structures (36). Therefore, it is desirable to compare, in a control experiment, the FTIR results obtained in solution and in the form of film for the protein under investigation.

9. Curve fitting could be also performed on the Fourier self deconvoluted spectrum, where the different band components are well resolved, leading for this reason to a more efficient fitting. However, the distortion of the deconvoluted spectrum, compared to the measured one, and the uncertainty in the choice of FSD parameters raise the question on the reliability of the fitting results.

Acknowledgments

References

1. Turoverov KK, Kuznetsova IM, Uversky VN (2010) The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. Prog Biophys Mol Biol 102:73–84

2. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D-2 concept. Annu Rev Biophys 37:215–246

3. Xie HB, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN et al (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 6:1882–1898

4. Uversky VN (2007) Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation. J Neurochem 103:17–37

5. Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S (2006) Assessing protein disorder and induced folding. Proteins 62:24–45

6. Arrondo JLR, Muga A, Castresana J, Goni FM (1993) Quantitative studies of the structure of proteins in solution by fourier-transform infrared-spectroscopy. Prog Biophys Mol Biol 59:23–56

7. Arrondo JLR, Goni FM (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. Prog Biophys Mol Biol 72:367–405

8. Seshadri S, Khurana R, Fink AL (1999) Fourier transform infrared spectroscopy in analysis of protein deposits. Methods Enzymol 309:559–576

9. Natalello A, Doglia SM (2010) Intrinsically disordered proteins and induced folding studied by Fourier transform infrared spectroscopy. In: Uversky VN, Longhi S (eds) Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation. Wiley, Hoboken, NJ

10. Barth A, Zscherp C (2002) What vibrations tell us about proteins. Q Rev Biophys 35:369–430

11. Barth A (2007) Infrared spectroscopy of proteins. Biochim Biophys Acta 1767:1073–1101

12. Goormaghtigh E, Cabiaux V, Ruysschaert JM (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. III. Secondary structures. Subcell Biochem 23:405–450

13. Natalello A, Benetti F, Doglia SM, Legname G, Grandori R (2011) Compact conformations of alpha-synuclein induced by alcohols and copper. Proteins 79:611–621

14. Salomaa P, Schaleger LL, Long FA (1964) Solvent deuterium isotope effects on acid-base equilibria. J Am Chem Soc 86:1–7

15. Latawiec D, Herrera F, Bek A, Losasso V, Candotti M, Benetti F et al (2010) Modulation of alpha-synuclein aggregation by dopamine analogs. PLoS One 5:e9234

16. Natalello A, Ami D, Brocca S, Lotti M, Doglia SM (2005) Secondary structure, conformational stability and glycosylation of a recombinant *Candida rugosa* lipase studied by Fourier-transform infrared spectroscopy. Biochem J 385:511–517

17. Dong A, Huang P, Caughey WS (1990) Protein secondary structures in water from 2nd-derivative amide-I infrared-spectra. Biochemistry 29:3303–3308

18. Susi H, Byler DM (1986) Resolution-enhanced Fourier-transform infrared-spectroscopy of enzymes. Methods Enzymol 130:290–311

19. Kauppinen JK, Moffatt DJ, Mantsch HH, Cameron DG (1981) Fourier self-deconvolution—a method for resolving intrinsically overlapped bands. Appl Spectrosc 35:271–276

20. Goormaghtigh E, Raussens V, Ruysschaert J-M (1999) Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes. Biochim Biophys Acta 1422:105–185

21. Munishkina LA, Phelan C, Uversky VN, Fink AL (2003) Conformational behavior and aggregation of alpha-synuclein in organic solvents: modeling the effects of membranes. Biochemistry 42:2720–2730

22. Hong DP, Fink AL, Uversky VN (2008) Structural characteristics of alpha-synuclein oligomers stabilized by the flavonoid baicalein. J Mol Biol 383:214–223

23. Munishkina LA, Fink AL, Uversky VN (2004) Conformational prerequisites for formation of amyloid fibrils from histones. J Mol Biol 342:1305–1324

24. Denning DP, Uversky VN, Patel SS, Fink AL, Rexach M (2002) The *Saccharomyces cerevisiae* nucleoporin Nup2p is a natively unfolded protein. J Biol Chem 277:33447–33455

25. Dong A, Matsuura J, Manning MC, Carpenter JF (1998) Intermolecular beta-sheet results from trifluoroethanol-induced nonnative alpha-helical structure in beta-sheet predominant proteins: infrared and circular dichroism spectroscopic study. Arch Biochem Biophys 355:275–281

26. Roque A, Iloro I, Ponte I, Arrondo JLR, Suau P (2005) DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. J Biol Chem 280:32141–32147

27. Roque A, Ponte I, Suau P (2007) Macromolecular crowding induces a molten globule state in the C-terminal domain of histone H1. Biophys J 93:2170–2177

28. Roque A, Ponte I, Arrondo JLR, Suau P (2008) Phosphorylation of the carboxy-terminal domain of histone H1: effects on secondary structure and DNA condensation. Nucleic Acids Res 36:4719–4726

29. Martinez G, Millhauser G (1995) Ftir spectroscopy of alanine-based peptides – assignment of the amide I′ modes for random coil and helix. J Struct Biol 114:23–27

30. Javor S, Natalello A, Doglia SM, Reymond JL (2008) alpha-Helix stabilization within a peptide dendrimer. J Am Chem Soc 130:17248–17249

31. Walsh STR, Cheng RP, Wright WW, Alonso DOV, Daggett V, Vanderkooi JM et al (2003) The hydration of amides in helices; a

comprehensive picture from molecular dynamics, IR, and NMR. Protein Sci 12:520–531

32. Zandomeneghi G, Krebs MRH, McCammon MG, Fandrich M (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. Protein Sci 13:3314–3321

33. Torii H, Tatsumi T, Tasumi M (1998) Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and resonance Raman intensities of N-methylacetamide. J Raman Spectrosc 29:537–546

34. Khurana R, Fink AL (2000) Do parallel beta-helix proteins have a unique Fourier transform infrared spectrum? Biophys J 78:994–1000

35. Susi H, Byler DM (1987) Fourier-transform infrared study of proteins with parallel beta-chains. Arch Biochem Biophys 258:465–469

36. Natalello A, Prokorov VV, Tagliavini F, Morbin M, Forloni G, Beeg M et al (2008) Conformational plasticity of the Gerstmann-Straussler-Scheinker disease peptide as indicated by its multiple aggregation pathways. J Mol Biol 381:1349–1361

37. Vila R, Ponte I, Collado M, Arrondo JLR, Suau P (2001) Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA. J Biol Chem 276:30898–30903

38. Haris PI, Severcan F (1999) FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media. J Mol Catal B Enzym 7:207–221

39. Hering JA, Innocent PR, Haris PI (2002) Automatic amide I frequency selection for rapid quantification of protein secondary structure from Fourier transform infrared spectra of proteins. Proteomics 2:839–849

40. Dong AC, Hyslop RM, Pringle DL (1996) Differences in conformational dynamics of ribonucleases A and S as observed by infrared spectroscopy and hydrogen-deuterium exchange. Arch Biochem Biophys 333:275–281

41. Natalello A, Frana AM, Relini A, Apicella A, Invernizzi G, Casari C et al (2011) A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation. PLoS One 6(4):e18789

42. Barth A (2000) The infrared absorption of amino acid side chains. Prog Biophys Mol Biol 74:141–173

Monitoring Structural Transitions in IDPs by Vibrational Spectroscopy of Cyanylated Cysteine

Hailiu Yang, Johnny Habchi, Sonia Longhi, and Casey H. Londergan

Abstract

The fast intrinsic time scale of infrared absorption and the sensitivity of molecular vibrational frequencies to their environments can be applied with site-specificity by introducing the artificial amino acid β-thiocyanatoalanine, or cyanylated cysteine, into chosen sites within intrinsically disordered proteins. This amino acid can be inserted through native chemical ligation at single cysteines introduced via site-directed mutagenesis. The CN stretching band of cyanylated cysteine is sensitive to local changes in both structural content and solvent exposure. This dual sensitivity makes cyanylated cysteine an especially useful probe of binding-induced structural transitions in IDPs. The general strategy of creating single-site cysteine mutations and chemically modifying them to create the vibrational chromophore, as well as observation, processing and analysis of the CN stretching band, is presented.

Key words: Infrared spectroscopy, β-Thiocyanatoalanine, Solvent dynamics, Native chemical ligation, Intrinsically disordered proteins, Induced folding, Structural transitions

1. Introduction

Intrinsically disordered proteins (IDPs) present a particularly difficult structural characterization challenge, especially since they typically escape direct analysis by the most conventional techniques in structural biology (particularly X-ray crystallography). Infrared radiation has a much higher frequency than the signal radiation in radio- and microwave-based magnetic techniques, and because the detecting radiation is fast and the period of most molecular vibrations is quite short (tens to hundreds of fs), vibrational spectroscopy should be particularly useful in providing a snapshot of a dynamic protein's conformational distribution.

Unfortunately, the width and frequency overlap of most naturally occurring protein-based functional groups means that infrared spectroscopy is conventionally used only as a probe of global secondary

structure without any site-specific information. Much recent work (1–7) has focused on the incorporation of artificial functional groups into selected amino acid side chains. When these new functional groups appear in a clear region of the biomolecular vibrational spectrum, new site-specific probes with intrinsic fs-ps time resolution of the local structure are introduced. Artificial side chain incorporation is most typically implemented using total peptide synthesis (2) or semisynthesis (8), with notable recent exceptions involving either stop codon suppression (9, 10) or native chemical ligation to natural amino acids (4, 11, 12).

A simple approach to a vibrationally active artificial amino acid is the cyanylation of cysteine, which converts the native cysteine thiol group to a covalently attached thiocyanate. The CN stretching band of aliphatic thiocyanate appears in an otherwise clear spectral window near 2,160 cm^{-1} and is a reasonably strong infrared absorber (2, 13). Since IDPs are generally depleted in cysteines, many IDP sequences do not natively contain any cysteine residues (14). As long as there is a suitable expression system for a protein of interest, introduction of cysteine at a selected site followed by cyanylation of cysteine is a rather straightforward approach to creation of a site-specific vibrational probe moiety in IDP systems of arbitrary size. The CN stretching absorption frequency is sensitive to the local presence or absence of water (15), as well as to the local electric field presented by the nearby structure (4). Its linewidth is dynamically sensitive (12) to fluctuations of the local solvent and structure on the fs-ps time scale (which is the time scale of H-bond formation and breaking and dipolar reorientation). The "local" structure here means a length scale of only several Å, making this vibration a very site-specific probe.

We have recently shown (1) that the CN stretching band of cyanylated cysteine introduced in an IDP reports partner-induced structural transitions in two ways: via frequency shift due to changes in water exclusion and via lineshape change due to dynamical changes associated with structure formation.

Because the S-bound nitrile moiety is easy to incorporate and sensitive to structural changes (including order/disorder transitions), site-specific introduction of cyanylated cysteine is expected to become a widely used probe for mapping binding-induced structural transitions in IDPs (or IDP regions of partially structured proteins). A graphic of the general strategy for using this probe to map structural transitions is shown in Fig. 1. This chapter summarizes protocols for the introduction of cysteine at sites of interest, the native chemical ligation of these cysteines to produce the infrared probe group, and observation and interpretation of the probe vibration via infrared spectroscopy.
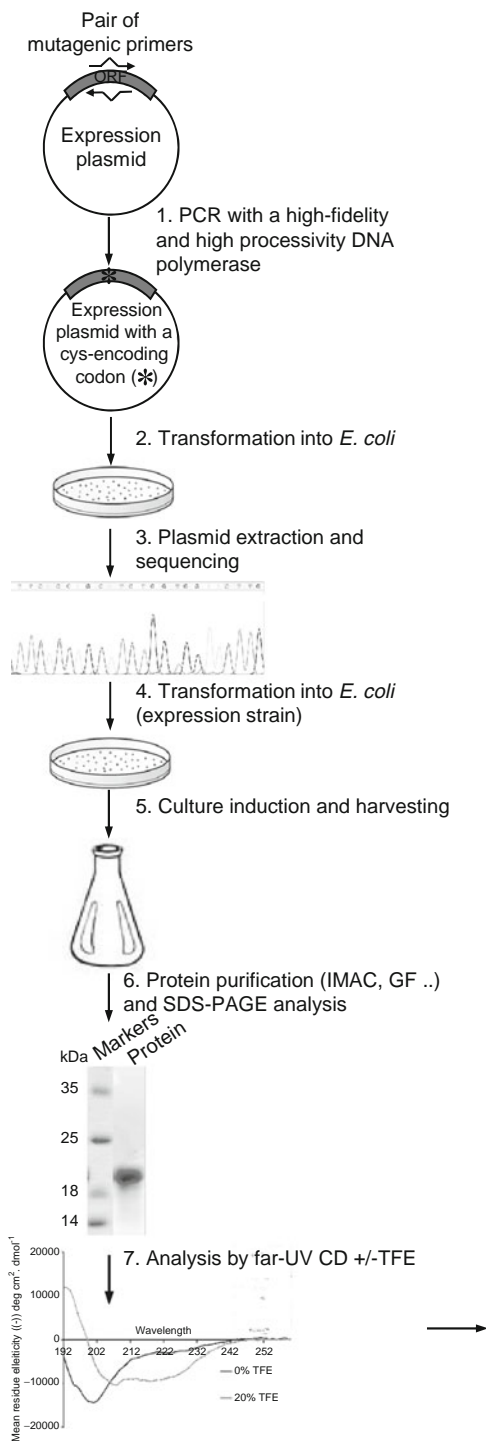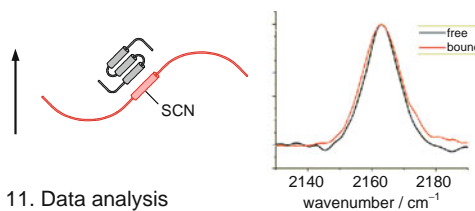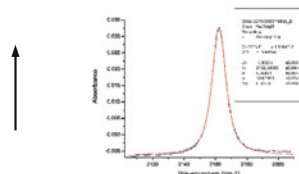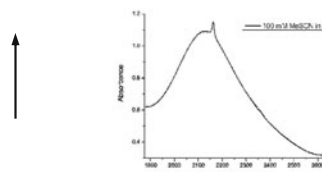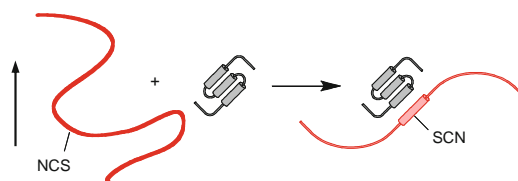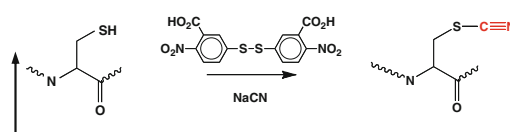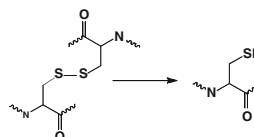
Fig. 1. Schematic outline of the strategy for vibrational spectroscopy of cyanylated cysteines.

2. Materials

2.1. In Vitro Site-Directed Cysteine Substitution Mutagenesis

2.1.1. Polymerase Chain Reaction Mutagenesis for Cysteine Substitution and Transformation of Escherichia coli Competent Cells

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a resistivity of 18 M MΩ/cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations when disposing of waste materials.

1. Expression plasmid template encoding the protein of interest with a purification tag.
2. Synthetic, mutagenic oligonucleotide primers (sense and reverse) bearing a cysteine-encoding TGC/T codon (Operon).
3. *Pfu* turbo thermostable DNA polymerase (Stratagene) for PCR amplification (or a similar, high fidelity and highly processive thermostable DNA polymerase).
4. dNTPs mixtures (2.5 mM each).
5. 10× reaction buffer.
6. *Dpn* I restriction enzyme (NE Biolabs).
7. Agarose gel electrophoresis equipment and TBE running buffer.
8. Polymerase chain reaction (PCR) thermocycler.
9. Ethidium bromide stock at 10 mg/mL.
10. DNA ladder (Invitrogen).
11. UV transilluminator for visualizing DNA in agarose gels stained with ethidium bromide.
12. Agarose gel loading buffer 6×.
13. *E. coli* TAM1 chemically competent cells (Active Motifs).
14. Petri agar plates with appropriate medium for bacterial growth (e.g., Luria Bertani or 2YT), sterile super optimal broth with catabolite repression (SOC), and appropriate antibiotics stock solutions.
15. Sterile 10 mL culture tubes and 1.5 mL Eppendorfs.
16. Incubator set at 45 °C.
17. Incubator set at 37 °C.

2.1.2. DNA Plasmid Extraction

1. Shaking incubator set at 37 °C, sterile culture tubes, culture medium, and antibiotics.
2. Macherey-Nagel (or equivalent) kit for plasmid DNA extraction.
3. Microfuge and 1.5 mL Eppendorfs.
4. Ethidium bromide stock at 10 mg/mL.
5. DNA ladder (Invitrogen).

6. Agarose gel electrophoresis equipment (Bio-Rad) and TBE running buffer.

7. UV transilluminator for visualizing DNA in agarose gels stained with ethidium bromide.

2.2. Protein Expression and Purification

2.2.1. Transformation of E. coli Expression Strain with the Mutated Plasmid Construct

1. Chemically competent *E. coli* Rosetta [DE3] pLysS strain (Novagen) for recombinant proteins expression.

2. Incubator set at 45 °C.

3. Sterile SOC medium and appropriate antibiotics stock solutions.

4. Sterile 10 mL culture tubes and 1.5 mL Eppendorfs.

5. LB (or 2YT) agar plates supplemented with appropriate antibiotics.

6. Shaking incubator set at 37 °C.

2.2.2. Cell Growth and Protein Expression

1. Rosetta [DE3] pLysS cells containing the protein expression vector that produces the cysteine-substituted protein.

2. Appropriate medium for cell growth and protein expression (e.g., Luria Bertani).

3. Sterile appropriate antibiotics stock solutions.

4. Sterile 250-mL and 3-L baffled-bottom flasks.

5. Inducer (such as Isopropyl-thio-β-D-galactopyranoside (IPTG) at 0.5 M).

6. Shaking incubator set at 37 °C.

7. Perkin UV-vis absorption spectrophotometer.

8. High-speed centrifuge (Sorvall).

2.2.3. 6× His-Tagged Protein Purification

1. Cellular pellets from induced cultures expressing the recombinant cysteine-substituted protein.

2. Buffers: 50 mM Tris–HCl pH 8.0, 300 mM NaCl, and 1 mM phenyl methyl sulphonyl fluoride (PMSF) supplemented with either 10 mM imidazole for cell lysis, or 20 mM imidazole for resin wash or 250 mM for protein elution.

3. Stock solutions of lyzozyme (50 mg/mL), DNase (2 mg/mL), and $MgSO_4$ (2 M).

4. Complete EDTA-free protease inhibitor cocktail (Roche).

5. Sonicator, Sorvall centrifuge, microfuge, and Perkin UV-vis absorption spectrophotometer.

6. Chelating Sepharose Fast Flow Resin (GE Healthcare) pre-loaded with Ni^{2+} ions.

7. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel, 2× Laemmli SDS-PAGE sample buffer,

running buffer, molecular mass markers (GE Healthcare), electrophoresis cell, and power supply (Bio-Rad).

8. Staining and destaining solutions for sodium dodecyl sulfate-polyacrylamide gels.

9. ÄKTA Explorer Chromatography System (GE Healthcare).

10. 0.22-μm polyethersulfone filters (Millipore).

11. Pre-packed Superdex 200 or 75 HR 16/60 columns (GE Healthcare).

12. Degassed and filtered 10 mM sodium phosphate pH 7 buffer, either supplemented or not with NaCl (50–500 mM) for protein elution from gel filtration columns.

13. ÄKTA collecting tubes, 1.5 mL Eppendorfs.

14. Centricon Plus-20 concentrators (Millipore) with appropriate molecular cutoff.

15. UV-Vis absorption spectrophotometer and Bradford reagent (Bio-RAD).

2.3. Protein Characterization Using Far-UV Circular Dichroism

1. Stock solution of the protein of interest.

2. 2,2,2-Trifluoroethanol (TFE) (Roth)

3. Jasco 810 dichrograph spectropolarimeter equipped with a Peltier thermoregulation system.

4. 1-mm thick quartz cells.

5. PD-10 desalting columns containing 8.3 mL of Sedaphex™ G-25 Medium (GE Healthcare) for buffer exchange.

6. 10 mM sodium phosphate buffer at pH 7.

7. SpectraManager program for spectra smoothing and analysis.

2.4. Modification and Infrared Spectroscopy of Single-Cysteine Variants

2.4.1. Concentration of Cyanylated Cysteine Variants to Infrared Concentrations and Mixing with Binding Partner

1. Buffers: 50 mM Tris–HCl, 200 mM NaCl, pH 7.0, and 100 mM Phosphate pH 7.0, degassed using a 0.20 μm GNWP nylon membrane (Millipore).

2. V-570 UV/Vis/NIR Spectrophotometer (JASCO) with 1 mm quartz spectrophotometer cell (Starna Cells Inc.).

3. 95+ % A.C.S. reagent sodium cyanide (Aldrich).

4. 99 % Titration dithiothreitol (DTT) (Aldrich).

5. 99 % 5,5′-Dithiobis(2-nitro-benzoic acid) (DTNB) (Acros).

6. AB 15 pH meter (Fischer Scientific).

7. 1.0 mL gastight syringe (Hamilton).

8. Superdex™ 75 10/300 GL size exclusion column (GE Healthcare).

9. AKTAPurifier chromatography system (GE Healthcare) with in-line pH monitor and $\lambda = 254$ nm UV filter (or $\lambda = 280$ nm for Trp-containing proteins).

10. Unicorn Manager 11 Software for AKTAPurifier.

11. Amicon Ultra-4 4 mL vol cellulose centrifugal filter cartridges of the appropriate MW cutoff(Millipore).

12. Sephadex™ G-25 M PD-10 columns (GE Healthcare).

13. Biomax 0.5 mL membrane centrifugal filter cartridges of the appropriate MW cutoff (Millipore).

14. VWR Galaxy 16 microfuge (VWR).

15. Swinging bucket 15 mL tube centrifuge.

2.4.2. Concentration of Cyanylated Cysteine Variants to Infrared Concentrations and Mixing with Binding Partner

1. Stock solution of the protein of interest.

2. Vivaspin500 0.5 mL polyethersulphone centrifugal filter cartridges of the appropriate MW cutoff (GE Healthcare).

3. Biomax 0.5 mL membrane centrifugal filter cartridges of the appropriate MW cutoff (Millipore).

4. VWR Galaxy 16 microfuge.

2.4.3. Infrared Spectroscopy of Cyanylated Single-Cysteine Variants

1. ~20 µM pathlength Biocell (spacer-free CaF_2 transmission cell) nested inside BioJack pressure and temperature collar (Bio-Tools, Jupiter, FL, USA).

2. Vertex 70 FTIR Spectrometer (Bruker Optics) with photovoltaic MCT detector and continuous purge of N_2 free of H_2O and CO_2.

3. OPUS 5.5 software (Bruker Optics).

4. Origin 8 software (OriginLab).

3. Methods

Vibrational spectroscopy of cyanylated cysteine is a recently developed approach to monitor and map disorder-to-order transitions that IDPs may undergo in the presence of their binding partner(s). Deciphering these folding events by this approach implies the generation and the study of a collection of variants of the IDP of interest, each of them bearing a single cysteine and hence a single CN label grafted at a specific position. The higher the number of variants, the more detailed the study will be, as information about the CN probe grafted at numerous protein sites will be collected. Typically, the interaction site can be mapped and information about the boundaries of the region undergoing structural transitions can be obtained. In addition, this approach is also expected to provide hints about possible pre-configuration in solution of molecular recognition elements (MoREs) (16–19), thereby allowing the molecular mechanism of complex formation to be inferred and plausible binding models to be proposed.

3.1. Generation of Single-Cysteine Variants

3.1.1. Design and Construction of a Cysteine Substituted Protein

The first step towards vibrational spectroscopy studies using the cyanylated cysteine side chain is the analysis of the amino acid sequence of the protein of interest. For implementation of this approach, the sequence should contain a single cysteine residue. An advantage of working with IDPs is that cysteine is usually under-represented (14). However, in the case that a cysteine residue occurs naturally within the protein, two situations can be distinguished: (1) the cysteine residue does not occur in the region of interest, and thus site-directed mutagenesis is required to replace the sulfhydryl group by an hydroxyl group of a serine leading to a minimal perturbation in the protein, followed by an additional site-directed mutagenesis step aimed at introducing the unique cysteine residue at the desired position; or, (2) the cysteine residue occurs at a suitable position and site-directed mutagenesis is not required. Since the β-thiocyanatoalanine artificial side chain is quite flexible, CN transition dipoles on neighboring artificial side chains do not couple to each other in predictable ways, so the clearest implementation of this approach is to use a single label at a time (in contrast to EPR studies, in which mutual relaxation between two site-directed spin labels can be used to map inter-label distances). If site-directed mutagenesis is required, whenever possible, isosteric substitutions (i.e., serine to cysteine) are introduced so as to introduce minimal structural perturbation in the expressed protein. However, it has been shown recently (Dalton, Londergan et al., unpublished work) that cyanylated cysteine can be successfully substituted for a wide range of amino acids without destructive perturbation of binding interfaces between proteins. This comes from its small size and nonnatural, relatively noninteracting thiocyanate functional group, which does not form 1:1 hydrogen-bonded complexes with any naturally occurring protein functional groups. The most immediately problematic substitutions in IDPs are at charged residues, which play a central role in maintaining the solubility and level of compaction of the IDP.

Numerous mutagenesis methods have been developed based upon PCR. The simplest and most broadly applicable protocol is currently the QuickChange™ Site-Directed Mutagenesis System (QCM) developed by Stratagene. With this approach, the mutation is introduced in a single PCR with one pair of complementary primers containing the mutation of interest. However, this method is restricted to primer pairs of 25–45 bases in length with melting temperature (T_m) > 78°C (see Note 1). Otherwise, primer dimer formation will become more favorable compared to the primer-template annealing, in particular for primer pairs with multiple mismatches, and the protocol has to be modified case by case, or more steps have to be carried out, e.g., by using an alternative two-stage PCR protocol.

3.1.2. PCR Mutagenesis Reaction and Transformation of E. coli Competent Cells with PCR Product

The following site-directed mutagenesis protocol is adapted to the use of *Pfu* turbo thermostable DNA polymerase (see http://www.stratagene.com). It allows the introduction of a cysteine residue by modifying the DNA sequence using a single step of PCR. For optimal results, PCR primers should be designed so as to have similar T_m (see Note 1).

1. The PCR reaction mix is prepared as follows (see Note 2): 1 μL template DNA ($\approx$10–20 ng/μL), 5 μL thermostable DNA polymerase 10× reaction buffer, 4 μL dNTP mix stock solution at 10 mM, 1.5 μL of each primer (39-mer each) at 10 μM (see Note 1), 1 μL *Pfu* turbo DNA polymerase (2.5 U/μL), and sterile H_2O up to 50 μL. If the thermocycler is devoid of a heated lid, overlay each reaction mixture with a few drops of mineral oil.

2. The reaction is placed in the PCR thermal cycler with the following program (see Note 3): an initial denaturation step of 1 min at 95 °C, followed by 18 cycles (see Note 4) consisting each of a denaturation step at 95 °C for 30 s, an annealing step at 50 °C for 1 min, and an elongation step at 68 °C for 10 min for a template of approximately 5 kb (the elongation rate of *Pfu* turbo DNA polymerase is of 0.5 kb/min).

3. The amplification can be checked by electrophoresis of 5 μL of the product on a 1 % agarose gel (see Note 5).

4. Add 1 μL of *Dpn* I restriction enzyme (10 U/μL) directly to the amplification reaction.

5. Gently and thoroughly mix the reaction mixture by pipetting the solution up and down several times, then immediately incubate the reaction at 37 °C for 1 h to digest the parental (i.e., the non-mutated) plasmid DNA.

6. An additional electrophoresis step for PCR yield estimation can be carried out (see Notes 5 and 6).

7. Transform TAM1 chemically competent cells using standard protocols (20).

8. Plate the cells onto agar plates containing the appropriate antibiotic and incubate at 37 °C for 18 h.

9. Peak the brightest colonies, culture them overnight in liquid medium at 37 °C with agitation (250 rpm), and extract plasmid DNA using the Macherey-Nagel kit for DNA extraction (according to manufacturer's instructions).

10. Check plasmids by agarose gel electrophoresis.

11. Sequence candidate clones (GATC Biotech, Genome Express) to ensure sequence integrity and proper reading frame.

3.1.3. Cell Growth and Protein Expression

The introduction of a foreign DNA into cells is a standard procedure that is routinely applied to produce recombinant proteins for biochemical studies. Different expression systems (i.e., bacteria, mammals, insects, yeasts, …) have been described and experimentally validated for the production of recombinant proteins. Although all available expression systems are used for roughly the same applications (i.e., structural analysis, functional assays, protein interactions, …), each one possesses its own advantages as well as its potential drawbacks. For instance, although expression in bacteria is commonly and widely used (scalable, low cost, compatible with different cloning systems, …), protein solubility can constitute a serious concern, especially when dealing with eukaryotic proteins or with proteins undergoing posttranslational modifications required for proper protein folding. Therefore, using the best expression system for your protein is the key to your success. In general, solubility, functionality, speed, and yield are often the most important factors to be considered when choosing an expression system. Moreover, the growth conditions should be properly controlled. Culture conditions and the induction of expression have profound effects on the way the recombinant proteins are produced. It is therefore advisable to empirically establish optimal conditions with small-scale cultures before purification on a large-scale is attempted. Moreover, other problems with protein expression may occur, including cell toxicity (see Note 7), protein instability (see Note 8), formation of inclusion bodies (see Note 9), improper processing or posttranslational modification, and inefficient translation.

As a general guideline, expressing the cysteine-substituted protein of interest using the same heterologous system used for the expression of the *wt* protein is advised.

The following protocol focuses on bacterial (*E. coli*) expression. Although the bacterial heterologous system is not applicable to all kinds of proteins, especially eukaryotic ones with intricate folds, it is still the most commonly used owing to its low cost, rapidity, and easiness of implementation, as well illustrated by the fact that 70 % of structures deposited within the PDB correspond to proteins expressed in *E. coli*. With careful choice of host strains, vectors, and growth conditions, many recombinant proteins can expressed at high levels in *E. coli*, and tuning of these parameters often allows increasing expression and solubility levels (21).

1. Transform chemically competent Rosetta [DE3] cells, already containing or not the pLysS plasmid (see Note 10), with the expression vector encoding the protein of interest bearing a cysteine at the desired position, and then select transformants on LB agar plates containing the appropriate antibiotics (see Note 11). Incubate the plate overnight at 37 °C.

2. Inoculate 0.5 mL of LB medium containing the appropriate antibiotics in a culture tube with a single colony from the plate, and incubate for 3–4 h at 37 °C with shaking (250 rpm).

3. Withdraw an aliquot (250 μL) of the culture, add glycerol to a final concentration of 30 % and store the cell stock at −80 °C.

4. Inoculate 50 mL of the same medium in a 250-mL baffled-bottom flask with the rest of the liquid culture and incubate overnight at 37 °C with shaking until saturation is reached.

5. The next morning, inoculate 1 L culture of the same medium (see Note 12) with the appropriate antibiotics in a 3-L baffled-bottom flask with the saturated overnight culture.

6. Grow the cells at 37 °C with shaking to mid-log phase ($OD_{600nm} \approx 0.6$–0.8).

7. Add IPTG to a final concentration of 0.2–1 mM. Incubate for 3–4 additional hours (see Note 13).

8. Recover the cells by centrifugation ($5,000 \times g$). Freeze the cell pellets at −20 °C.

3.1.4. Cell Lysis and Protein Purification

This section describes a smooth path to purification of the unmodified cysteine-containing protein. Very often, more than one purification step is necessary to reach the desired purity. Therefore, the key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements, and combine them in a logical way to maximize yield and minimize the number of steps required. Various purification tags are available (such as Trx, NusA, GST, Strep, 6×-His, MBP, GB1 ...) and can be used to both improve protein solubility and allow purification by affinity chromatography of recombinant proteins (22). Unlikely larger tags (GST, MBP, Trx ...) that generally have to be cleaved off by specific proteases recognizing cleavage sites located between the tag and the protein of interest, the 6× His tag does not have necessarily to be removed, as its very small size in most cases does not affect the protein properties.

The purification procedure described in this section concerns 6×-His tagged proteins (see Note 14) and comprises two steps of chromatography. The protein is first bound to a resin (either Ni-NTA or Sepharose resin preloaded with Ni^{2+} ions) in an immobilized metal affinity chromatography (IMAC) step. Subsequently, the eluted protein is loaded onto a size exclusion chromatography (SEC) column.

IMAC purification may be achieved with either a batch or a column procedure. In the procedure described below, the IMAC step is performed in batch. This strategy enhances binding of 6×-His tagged proteins and is therefore recommended when the tag is not fully accessible or when the protein of interest is present at very low concentrations. Even when the tag is not completely accessible, the protein will bind as long as more than two histidine residues are available to interact with the nickel ions. A big advantage of the IMAC approach is that the affinity of histidine for the Ni resin is greater than that of antibody–antigen or enzyme–substrate pairs.

Consequently, the addition of low concentrations of imidazole in the binding and washing buffers is very effective in preventing contaminating proteins from binding to the resin. The imidazole ring is part of the structure of histidine and is the moiety responsible for binding to the nickel ions. When the imidazole concentration is increased, the $6\times$-His tagged protein will dissociate from the resin because it can no longer compete for binding sites on the resin.

After IMAC, we recommend to carry out a gel filtration chromatography (or SEC) step so as to remove contaminants and aggregates, but also to estimate the monodispersity and oligomeric state of the protein. SEC separates molecules according to differences in their size (i.e., in their hydrodynamic radius). As such, this step also enables estimation of the impact of the cysteine substitution onto the compactness of the protein. Unlike affinity chromatography, molecules do not bind to the column matrix and consequently conditions can be varied to suit the type of sample or the requirements for further purification, analysis, or storage.

Below is provided a general protocol for the purification of a $6\times$ His tagged protein, where all steps, except gel filtration, are performed at 4 °C. Note that the cysteine substitution may result in a protein form behaving slightly differently with respect to the *wt* protein, thereby rendering it necessary to adjust and optimize the purification protocol on a case-by-case basis.

1. Thaw the cellular pellets on ice and suspend in 5 volumes (v/w) of ice-cold buffer A (50 mM Tris–HCl buffer pH 8, 300 mM NaCl, 10 mM imidazole, and 1 mM PMSF). Supplement buffer A with 0.1 mg/mL lysozyme, 10 μg/mL DNAse, 20 mM MgSO4 and protease inhibitor cocktail (Roche) (one tablet for 50 mL of bacterial lysate) (see Note 15).

2. After a 20-min incubation with gentle agitation, disrupt the cells by sonication.

3. Clarify the crude cell extract by high-speed centrifugation ($30,000 \times g$) followed by filtration (see Note 16).

4. Incubate the clarified supernatant for 1 h with gentle shaking with Chelating Sepharose Fast Flow resin preloaded with Ni^{2+} ions previously equilibrated in buffer A (2–4 mL resin *per* L of bacterial culture).

5. Wash the resin with 10 volumes of buffer A supplemented with 20 mM imidazole (see Note 17) and then elute the protein with 3 volumes of buffer A supplemented with 250 mM imidazole.

6. The presence of the desired product may be checked by SDS-PAGE analysis.

7. Load the protein onto a gel filtration pre-packed Superdex column (GE Healthcare) previously equilibrated in elution buffer (see Notes 18 and 19). A Superdex 200 or 75 HR

16/60 column can be used depending on whether the molecular mass of the protein of interest is higher or smaller than 70 kDa, respectively.

8. Monitor protein elution at 280 nm (see Note 20) and collect 1.5 mL fractions. Analyze by SDS-PAGE the fractions of interest and pool those exhibiting a sufficient purity level (see Note 21).

9. Concentrate the protein up to a concentration of 10 mg/mL (see Note 22), using Centricon Plus-20 concentrators of the appropriate cutoff.

10. Stock the protein at –20 °C (see Note 23).

3.1.5. Far-UV Circular Dichroism Analysis

A crucial point prior to further analysis is to ascertain that the cysteine substitution does not greatly affect the overall secondary structure content of the protein of interest. Cyanylated cysteine is in general a less perturbative variation than native cysteine, so well-behaved single-cysteine variants are expected to be excellent candidates for native chemical ligation at the introduced cysteine residue. It is therefore strongly recommended to record the far-UV circular dichroism (CD) spectrum of the cysteine variant and to compare it to the spectrum of the native protein (see Note 24). In addition, it is also suggested to record the CD spectrum of the variant in the presence of TFE so as to rule out possible dramatic impact of the cysteine substitution on the nascent folding abilities of the variant. The mixed buffer/TFE solvent is a secondary structure stabilizer that mimics the hydrophobic environment experienced by proteins in protein–protein interactions, and that is therefore widely used as a probe to unveil disordered regions having a propensity to undergo an induced folding (23). Depending on the protein structural propensities, TFE concentrations as high as 40 % can be required to observe structural transitions (24). In most cases however, a significant disorder-to-order transition takes place at 20 % TFE. It is recommended to record the CD spectrum of the cysteine variant in the presence of the TFE concentration that triggers the most dramatic structural transitions in the native protein, and then to compare the estimated secondary structure content of the variant to that of the *wt* protein.

1. Prepare a sample of both the *wt* and cysteine-substituted protein at 0.1 mg/mL in 10 mM sodium phosphate pH 7 either in the presence or in the absence of TFE (see Note 25).

2. Record the CD spectra at 20 °C on a spectropolarimeter (Jasco 810 dichrograph), equipped with a Peltier thermoregulation system, using 1-mm thick quartz cells. Record the spectra in the 185–260 nm range, with a scanning speed of 20 nm/min, and a data pitch (wavelength increment) of 0.2 nm. Average each spectrum from three scans.

3. Record the CD spectra of buffer (containing or not TFE) and correct each protein spectrum for buffer signal. After smoothing the spectra by using a third-order least square polynomial fit, calculate mean ellipticity values per residue ($[\Theta]$) as follows; $[\Theta] = 3{,}300 \ m \ \Delta A/(l \ c \ n)$, where l (path length in cm), n = number of residues, m = molecular mass in daltons, and c = protein concentration expressed in mg/mL.

4. Analyze the spectra using the DICHROWEB Web site (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (25, 26). Use the CDSSTR deconvolution method and the reference protein set 7 to estimate the content in α-helical and disordered structure of each protein sample.

3.2. Native Chemical Ligation and Infrared Spectroscopy of Single-Cysteine Variants

3.2.1. Cyanylation of Cysteine Variants

In performing the native chemical ligation of cyanide to the sulfur of cysteine, there are several important considerations. First, cyanylation cannot proceed if the cysteine is in the oxidized form. Therefore, a large excess of DTT is used to completely reduce all cysteines to thiols or thiolates prior to reaction. Since cyanylating reagents will react with any remaining DTT, thereby reducing the yield of the desired modified protein product, DTT must be thoroughly removed by chromatography before exposing the protein to the cyanylating reagents.

Several synthetic routes to cyanylated cysteine have been reported (11, 27), and all are effectively bio-orthogonal (if performed with adequate care) in that they do not use reagents or lead to products that can adversely affect the rest of the protein. All reactions that lead to the final β-thiocyanatoalanine product are considered to be equilibrium reactions and are not particularly fast kinetically. Thus, it is important to use a suitable excess of each modifying reagent and to work at a relatively high concentration to drive the reaction(s) to completion within a reasonable amount of time. The method reported here and most widely used in recent years is to sequentially add Ellman's reagent and a cyanide salt in a single reaction vessel (see Fig. 2 and details below). Since Ellman's reagent (DTNB) is poorly soluble in organic buffers, such as Tris–HCl and HEPES, it is easiest to dissolve DTNB in concentrated phosphate buffer (at least 100 mM) and add a small volume of this stock solution to the reaction vessel. After the addition of DTNB, a relatively large excess of the cyanide salt is added. Since the cyanide reacts with all species present in solution, including the protein mixed disulfide intermediate, the Ellman's reagent side product, and any remaining Ellman's reagent, *all sulfur species* in solution are converted to thiocyanate before the protein is separated from the reaction mixture.

It is important to consider that high concentrations of a cyanide salt in aqueous buffer solution near neutral pH lead to substantial quantities of HCN with an accompanying rise in the solution pH.
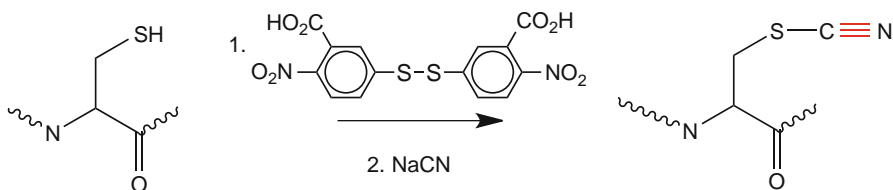
Fig. 2. Sequential approach to cyanylation of cysteine residues.

pH changes are problematic for the stability of the artificial cyanylated cysteine side chain, which can be viewed as an intermediate in the auto-scission of the protein at pH values far away from neutral (27). In general, the pH of the solution must be maintained between 6.0 and 8.0 throughout the modification reaction, purification, and all subsequent handling of the protein to prevent hydrolysis of the probe moiety. The probe is also sensitive to high temperatures; substantial probe loss has been observed at temperatures higher than 45 °C (12). The fragility of this probe in a protein context is balanced by its ease of introduction.

In IDPs, one must find a balance between a high reactive protein concentration and the onset concentration for aggregation; this will vary from system to system. However, since all residues in IDPs have a high degree of solution exposure, the single cysteine cyanylation generally proceeds quickly and with high yield. Hence, working with slightly more dilute solutions (50 μM to 1 mM) than in the subsequent infrared experiments is generally feasible. All side products, including the thiolate resulting from Ellman's reagent and NTCB (the cyanylated version of this thiolate), absorb in the visible region and their signals can be used to monitor the progress of the reaction by qualitatively inspecting the evolution of color or, if needed, quantitatively using visible spectroscopy. A successful reaction mixture turns yellow upon the addition of Ellman's reagent, and then bright yellow or orange when cyanide is introduced.

The protocol that follows is a slightly modified version of the previously described protocol for the measles virus intrinsically disordered N_{TAIL} domain (1). Except for minor variations in reactive concentration and buffers, this protocol is expected to be generally applicable to all single-cysteine variants of IDPs.

1. Filter 50 mM Tris–HCl + 200 mM NaCl through 0.20 μm GNWP nylon membrane and vacuum degas for at least 20 min with stirring.

2. Take a background scan for the V-570 UV/Vis photospectrometer using a 50 mM Tris–HCl, 200 mM NaCl buffer in a 1 mm quartz cell from $\lambda = 200$–500 nm.

3. Obtain approximately 1.2×10^{-7} mol of the protein to be cyanylated from a stock solution of purified, recombinant protein as obtained as described in Subheading 3.1.4.

4. Acquire the 200–500 nm spectrum of the sample protein in a 1 mm quartz cell using the UV/Vis spectrophotometer; for non-tryptophan containing proteins use the absorbance at 254 nm (instead of 280 nm) to quantify the protein concentration.

5. Dilute the protein sample to a total of 900 μL using 50 mM Tris–HCl, 200 mM NaCl buffer.

6. In a separate tube, dissolve 1,000 equivalents of DTT into 100 μL of the same buffer.

7. Equilibrate the Superdex™ 75 10/300 GL column on the AKTAPurifier pump system by running at least 40 mL of 50 mM Tris–HCl, 200 mM NaCl buffer at 0.8 mL/min (see Note 26)

8. Add the DTT solution prepared in step 6 into the protein sample and mix gently (avoid vigorous vortexing). Incubate for 20 min.

9. Inject the whole DTT-containing protein sample into the AKTA Purifier chromatography system using a 1-mL gastight syringe. Record the 280 nm (or 254 nm) absorption of the eluent. Collect 2-mL fractions from 8 min (6 mL) to 25 min (20 mL) (see Note 27).

10. Use UV absorption to select the appropriate protein fractions for concentration. DTT, which strongly absorbs at both 254 and 280 nm, elutes at approximately 25 mL. Be sure that there is a clear separation between the DTT peak and the protein peak (see Note 27).

11. To clean the column for future use, keep running 50 mM Tris–HCl, 200 mM NaCl buffer at 0.8 mL/min until the 280 (254) nm absorption shows that DTT has been completely purged from the column (typically, 3 column volumes are needed).

12. Concentrate the protein-containing fractions down to <1 mL total volume using Amicon Ultra-4 centrifugal filter cartridges ($<4,000 \times g$) and redetermine the protein concentration.

13. Prepare a 25 mM DTNB solution in 100 mM phosphate buffer. Add 8 molar equivalents of DTNB from this stock solution (should be a small volume) to the protein sample.

14. Incubate at room temperature for 20 min (see Note 28).

15. Prepare a 100 mM NaCN solution in 50 mM Tris–HCl, 200 mM NaCl buffer. Add 55 molar equivalents of NaCN from this stock solution to the reaction mixture.

16. Incubate at room temperature for 30 min (see Note 29). Immediately proceed to step 17 (see Note 30).

17. Set aside two Sephadex™ G-25 M PD-10 columns and wash both PD-10 columns with 3 × 4 mL of 50 mM Tris–HCl,

200 mM NaCl buffer. Evenly split the protein/DTNB/cyanide reaction mixture into two aliquots (approximately 550–600 μL each) and add one to each PD-10 column.

18. Collect the flow-through and elute the protein by adding 500 μL of 50 mM Tris–HCl, 200 mM NaCl buffer at a time to both columns and simultaneously collecting 500 μL fractions (see Note 31).

19. Using a 5 mm pathlength quartz cell, determine the presence of protein in each eluted fraction (see Note 32).

20. Concentrate the protein fractions down to a total volume of 500 μL using Amicon Ultra-4 filter cartridges and re-quantify protein.

21. Further concentrate the sample up to about 2 mM using a Biomax filter cartridge in micro-centrifuge ($<14,000 \times g$). This is the stock solution to be used for infrared spectroscopy and binding experiments. It should be flash frozen if possible and stored at $-80\,^\circ\mathrm{C}$ if not proceeding immediately to infrared spectroscopy.

3.2.2. Addition of Binding Partners

In using cyanylated cysteine to map binding-induced transitions, it is important to measure the spectrum of the probe vibration in both the bound and unbound states of the IDP of interest. Thus, two spectra are collected and compared: the spectrum of the cyanylated protein alone and the spectrum of the protein in the presence of a saturating concentration of the binding partner of interest. This experiment relies on a relatively high concentration of the cyanylated protein (near 2 mM), so even weakly bound (i.e., $K_D \sim \mu M$) complexes can reach saturation as long as a suitable excess of the binding partner is used.

1. Based on the concentration of the concentrated cyanylated protein sample (as obtained in step 21 of Subheading 3.2.1), obtain 1.5 equivalents of the binding partner in an identical buffer and concentrate the sample down to approximately 25 μL total volume by centrifuging Vivaspin500 filter cartridges in a micro-centrifuge ($<14,000 \times g$) (see Note 33).

2. Add the concentrated cyanylated protein solution from above into the same Vivaspin500 filter cartridge already containing the partner.

3. Concentrate the mixed solution to approximately 2 mM of the cyanylated protein and 3 mM of the binding partner.

3.2.3. Infrared Spectroscopy of the CN Stretching Band

The CN stretching band of the aliphatic thiocyanate moiety is a medium-strong absorber ($\varepsilon = 100$–250 cm^{-1}/M) (2, 13). However, when only one CN band is present in a large protein, it is instrumentally challenging to obtain a signal with a strong enough

signal-to-noise ratio to make quantitative conclusions about the band frequency and shape. An FTIR instrument with an HgCdTe (MCT) detector is required. The photovoltaic detector used here is nearly an order of magnitude more sensitive than a conventional MCT detector. The frequency of the SCN band (2,154–2,164 cm^{-1}) falls in a region largely devoid of other vibrations, including the normal modes of water (the H-O-H bend and the O-H stretch). This means that samples may be analyzed without needing to transfer the sample to D_2O, the standard medium in infrared experiments focusing on the amide vibration region (1,100–1,700 cm^{-1}).

The CN stretching band does, however, sit on top of a collective combination (libration plus bending) band of liquid H_2O that peaks at nearly the same frequency and absorbs between 1,900 and 2,400 cm^{-1} (see Fig. 3). This background absorption limits the sample pathlength that can be used in this frequency region to <50 μm, with >50 % of the infrared beam traversing the sample only at pathlength < 25 μm. Since the shape of this background absorption depends on the solute (including salt and protein) concentration, a good reference sample of identical buffer in exactly the same cell is needed. The collective water absorption also depends strongly on the temperature. Hence, in cases where the lab temperature may vary by more than about 1 °C, temperature control of the infrared cell should be implemented to minimize temperature variations between the protein-containing sample and the reference sample.

For best reproducibility between spectra, the infrared cell windows must be polished, optically flat, and made of CaF_2 or another hard infrared-transmissive material (i.e., sapphire or CdTe). The BioCell used in these experiments is an example of an ideal infrared liquid cell; the sample volume is cut into one of the two CaF_2 windows so that the cell pathlength has very little variability.

The CN stretching band of aliphatic thiocyanates is typically 8–15 cm^{-1} wide (full width at half maximum) and its frequency varies between 2,154 and 2,164 cm^{-1} in most solvents. These properties mean that relatively low spectral resolution (on the order of 2 cm^{-1}) can be used to decrease acquisition times. At such spectral resolutions, small variations on the order of >0.5 cm^{-1} in the frequency and >1.0 cm^{-1} in the bandwidth can be detected reproducibly. The CN band's shape has typically been analyzed (12, 13) by using a pseudo-Voigt lineshape (the linear combination of a Gaussian and a Lorentzian with the same width)

$$y(\tilde{v}) = y_0 + A\left(m_{\text{Lorentz}}\frac{2}{\pi}\frac{\text{FWHM}}{4(\tilde{v}-\tilde{v}_c)^2 + \text{FWHM}^2}\right.$$
$$\left. + (1 - m_{\text{Lorentz}})\frac{\sqrt{4\ln 2}}{\sqrt{\pi}\text{FWHM}}e^{-\frac{4\ln 2}{\text{FWHM}^2}(\tilde{v}-\tilde{v}_c)^2}\right) \qquad (1)$$
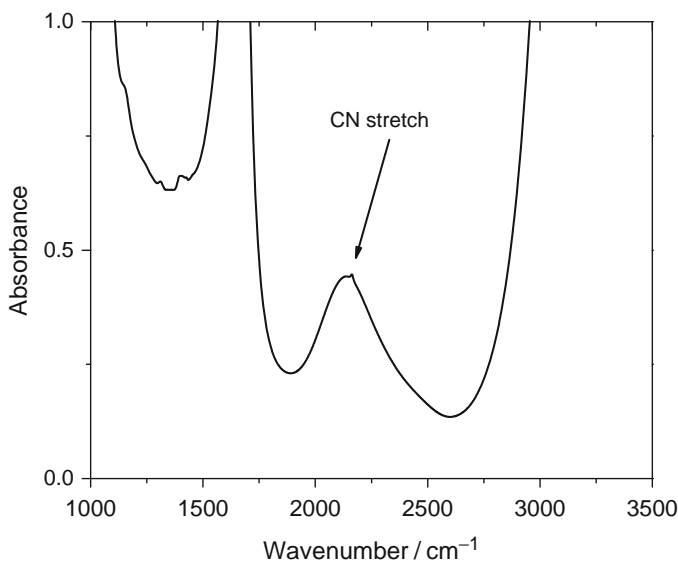
Fig. 3. The infrared absorption spectrum of methyl thiocyanate in water, showing the CN stretching band of thiocyanate on top of the collective absorption band of H_2O.

This band shape is nonphysical, since an actual symmetric band shape would be a Voigt profile (which is the convolution and not the sum of a Gaussian and a Lorentzian) and also since there is no particular reason that the CN band should be assumed to be symmetric about its mean. In the case of nearly symmetric band shapes, Eq. 1 is useful for analytical purposes since it yields a single width parameter. A more empirical and generalized lineshape analysis could include calculating the central moments of the CN distribution, or simply reporting its full width at half maximum, mean frequency, and mode frequency without making any assumptions about its underlying shape.

1. Collect a background scan (no sample or cell) on the Vertex 70 FTIR Spectrometer with 1,024 scans at 2 cm^{-1} resolution. Purge chamber for at least 20 min before scan.

2. Collect a spectrum (using the same parameters) of 50 mM Tris–HCl, 200 mM NaCl buffer in a 20 μm path length Biocell infrared cell. Place the BioJack-ed BioCell inside the FTIR sample chamber. Attach a thermocouple to the metal body of the BioJack, and connect to a temperature circulating bath if desired. Again, make sure the infrared chamber is purged for 20 min before starting the collection.

3. Transfer 20 μL of the concentrated cyanylated protein sample, with or without the binding partner, into the infrared cell (see Note 34).

4. Collect a spectrum (using the same instrumental parameters) of the protein sample.

5. Subtract the buffer spectrum from step 2 from the sample spectrum obtained in step 4 using OPUS software (see Note 35).

6. For each of the protein infrared spectra (bound and unbound), export the 2,000–2,300 cm^{-1} data as text to Origin. Remove the signal from 2,140 to 2,185 cm^{-1} and fit the remaining signal to a high-order polynomial. Then subtract this polynomial from the full signal to obtain the CN stretching band in the absence of all other background signals.

7. Analyze the CN stretching bands using a model of choice (see discussion above) and compare the spectra from bound and unbound species. The following changes lead to significant conclusions about transitions in the local structural environment:

    (a) CN frequencies of >2163.0 cm^{-1} mean that the probe group is exposed to solvent. In non-collapsed IDP regions, this should always be the frequency in the unbound state.

    (b) If the frequency shifts to the red from the unbound to the bound state, the probe group has moved to an environment that is at least partially water-excluded. This means that the artificial side chain is located along, and pointing towards, a binding interface. In this limit, linewidth differences may come from changes in the local electrostatic and/or H-bonding structural distribution.

    (c) If the frequency does *not* shift and the linewidth increases, then the probe is solution-exposed in all structures yet very near to a region that becomes more rigid, likely due to formation of secondary structure upon binding.

    (d) If a more complicated behavior is observed (i.e., shifting and significant broadening together), then more experiments should be performed at saturating concentrations of the binding partner to ascertain whether all of the IDP is bound. Multiple spectral subpopulations might be observed as a result of incomplete binding.

4. Notes

1. The following considerations should be made for designing mutagenic primers:

    (a) Both mutagenic primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid. The cysteine-encoding codon (i.e., TGC or TGT) should be chosen so as to minimize mismatches.

(b) Primers should be between 25 and 45 bases in length, with a melting temperature (T_m) of ≥ 78 °C. Primers longer than 45 bases may be used, but using longer primers increases the likelihood of secondary structure formation, which may affect the efficiency of the mutagenesis reaction.

(c) The desired mutation should be in the middle of the primer with ~10–15 bases of sequence perfectly matching the template on both sides.

(d) kThe primers optimally should have a minimum GC content of 40 % and should terminate in one or more C or G bases. The following formula is commonly used for estimating the T_m of primers.

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \% \text{ mismatch}$$

where N is the primer length in bases and values for %GC and % mismatch are whole numbers, N is the primer length in bases.

2. The PCR reaction can be modified in numerous ways to optimize results, depending on the nature of the template and primers and the DNA polymerase used for amplification.

3. PCR cycle conditions can also be varied. For example, the extension time should be increased for long DNA templates and that depending on the elongation rate of the DNA polymerase. The annealing temperature could be modified depending on the primers melting temperature (see Note 1) (i.e., the specificity of the annealing of the primers to the template increases with the temperature).

4. It is important to adhere to the 18-cycle limit when cycling the mutagenesis reactions. More than 18 cycles can have deleterious effects on the reaction efficiency and may introduce undesired mutations. A PCR with a control DNA plasmid template and a control couple of primers is advised.

5. A band may or may not be visualized at this stage. In either case, proceed with *Dpn* I digestion.

6. The DNA band intensity on the agarose gel electrophoresis should decrease after digestion with *Dpn* I owing to the degradation of the methylated, non-mutated DNA template (note that *Dpn* I only digests methylated DNA).

7. High transcription rates lead to slow growth, and this in turn is compounded by metabolic demands imposed by translation of the recombinant protein. Gene products that affect the host cell's growth rate at low concentrations are considered to be toxic. To reduce effects of protein toxicity on cell growth prior to induction, the level of basal transcription that occurs in the absence of induction "leakiness" should be repressed as much

as possible, and the number of generation before induction should be kept to a minimum.

8. Some proteins, particularly those that are smaller than 10 kDa, are not stable in *E. coli*, and may be degraded rapidly by proteases. This may be overcome by different means: reducing the growth temperature, inducing for a shorter period of time, adding low concentrations of PMSF (up to 1 mM) to the culture medium before induction.

9. In many cases, many gene products, when expressed at high levels in *E. coli*, accumulate as insoluble aggregates that lack functional activity. The formation of inclusion bodies is influenced by the nature of the protein, by the host cell, and by the level of expression resulting from the vector choice and the growth and induction conditions.

10. If the ORF encoding the target protein contains codons that are rarely used in *E. coli*, this can adversely affect the yield of the protein. In such cases, it is advisable to introduce an additional plasmid into the host cells that carries the cognate tRNA genes for rare codons. The pLysS plasmid contains such tRNA genes. It also encodes T7 lysozyme (a natural inhibitor of T7 RNA polymerase), which allows reducing the basal expression of genes under the control of the T7 promoter for better inhibition of expression under non-induced conditions. In addition, T7 lysozyme helps facilitating lysis of bacterial cells. The pLysS plasmid confers resistance to chloramphenicol and contains the p15A origin of replication, leading to compatibility with pUC- or pBR322-derived plasmids.

11. The working concentrations of mostly used antibiotics can be found in (20).

12. We recommend to perform a small-scale protein expression test to ensure that the cysteine substitution does not significantly impact the expression and/or solubility level of the protein of interest. For proteins well-expressed (i.e., at ≥ 5 mg/L), 1 L culture is sufficient to get protein amounts suitable to the ensuing spin labeling step. Otherwise, larger culture volumes are recommended.

13. Induction temperature and period can be modified based on a small-scale protein expression test. For example, it is known that decreasing the induction temperature in some cases increases the stability and the solubility of the protein without significantly decreasing the yield.

14. Primers are usually designed to introduce a 6× His tag encoding sequence at the 3′ end of the ORF (corresponding to the C terminus of the protein) so as to avoid purification of truncated forms arising from possible abortive translation. Note however, that N-terminal 6× His tags can also be used.

15. Protease inhibitor cocktails should be completely EDTA-free. This latter is a strong chelating agent that could bind to nickel ions and strips them from the NTA/Sepharose matrices. Resins become white in the absence of nickel ions.

16. Centrifugation of the disrupted cells for at least 30 min at 30,000 × g is recommended. Filtration through a 0.2-μm polyethersulfone is helpful to remove residual particulates prior to chromatography.

17. The washing step of the resin can be repeated several times and can be monitored by mixing 2 μL of washing buffer and 100 μL of Bradford reagent. In case a blue color (indicative of the presence of protein contaminants) appears, it is recommended to carry out additional washing steps.

18. Set an appropriate pressure and flow rate on the chromatography system so as to avoid damage to the column (adequate flow rates and pressure threshold are generally provided in the manufacturer's instructions).

19. For adequate resolution, the volume of the injected protein must not exceed 5 % of the total bed volume. Protein amounts as high as 30 mg can be injected, unless the protein is known to be aggregation-prone.

20. For proteins devoid of trp and tyr residues, the absorbance can be followed at 254 nm.

21. In most cases, the introduction of a cysteine causes the protein to at least partly dimerize. It is recommended to check by SDS-PAGE that the fractions of the first elution peak do correspond to a dimeric form of the protein of interest and not to a larger protein contaminant. The fractions of the monomer and of the dimer can be pooled in view of a DTT reduction step prior to spin labeling.

22. The protein concentration can be determined spectrophotometrically using either the Bradford reagent (Bio-Rad) or the absorbance at 280 nm. In case a protein assay is used, a calibration curve with a protein solution of known concentration has to be established. Protein concentrations can also be inferred from the absorbance at 280 nm using the Beer–Lambert relation. The absorption coefficient ε (mM^{-1}/cm) at 280 nm can be obtained either using the program ProtParam at the EXPASY server (http://www.expasy.ch/tools) or (preferentially) by direct analysis of the amino acid composition.

23. Some proteins require the addition of 10–20 % glycerol to prevent them from precipitating upon freezing.

24. It is strongly recommended to record the CD spectrum (either in the presence or in the absence of 20 % TFE) of the cyanylated form either so as to rule out possible dramatic effects brought

by the cyanylation itself. Substitution of cyanylated cysteine is not a particularly destabilizing force, but this is a necessary control when introducing any artificial amino acids into native proteins.

25. In case the cysteine-substituted protein is dissolved at a rather low concentration in a NaCl containing buffer, it is suggested to exchange the buffer to get rid of salt (which otherwise will give rise to strong absorption in the far-UV region) before far-UV CD analysis using a PD-10 desalting column.

26. Make sure the pressure of the Superdex™ 75 10/300 GL column does not exceed 1.8 MPa at any time; higher pressure may damage the column.

27. Typically, an IDP of 15 kDa elutes after 10–15 mL.

28. The color should change to yellow upon the addition of Ellman's reagent. To quantify intermediate yield, observe the concentration of the TNB thiolate product at $\lambda = 412$ nm.

29. Upon treatment with cyanide, the color should turn bright yellow or orange and stop changing after a few minutes. To monitor for reaction completion, observe the visible peaks from side products.

30. The protein should be immediately further purified after treatment with cyanide. Cyanylated proteins left in this solution degrade over time, possibly due to pH changes from the high cyanide concentration.

31. After adding the sample, allow the column to run until it stops dripping. Proceed to add 500 µL of elution buffer (50 mM Tris–HCl, 200 mM NaCl) at a time. After each addition, collect the elution from the column in a separate tube for further analysis. Keep collecting until the middle of the yellow DTNB band is eluting. If doing this experiment for the first time, collect fractions until the whole yellow DTNB band has been eluted to be safe.

32. The background UV scans for 50 mM Tris–HCl, 200 mM NaCl in a 1 mm cell and 5 mm cell are slightly different. For accurate results, a new background scan should be obtained for the 5 mm cell. However, for determining the presence of protein in the fractions, a new background scan is not needed. Complete separation of protein and reaction side products is typical for proteins > 10 kDa. If any fraction contains both protein and visible side products (as evidenced by any absorption at $\lambda > 400$ nm), repeat steps 17 and 18 with those fractions. Collect all fractions containing protein and no side products.

33. Note that at this stage, the molecular weight cutoff of the concentrator should be chosen at a fraction of the molecular mass of the smaller protein in the couple.

34. Place the sample as a drop on one of the BioCell windows. Carefully slide the other window over the first and gently rotate the windows versus each other to ensure complete coverage of the optical surface by the aqueous sample. Avoid bubbles and move the windows to help remove any bubbles. Inspect the windows for dust or other possible optical scatterers, and then place the windows inside the BioJack.

35. Between 2,000 and 2,300 cm^{-1}, a concentrated protein sample will often show less baseline absorbance than its corresponding buffer spectrum. Due to the excluded volume taken up by the protein, there is actually less H_2O in the same cell when protein is present. To adjust for this, it is reasonable to slightly under-subtract the buffer spectrum. The CN stretching peak should be obvious above the baseline throughout the subtraction procedure, and the goal is to obtain a relatively flat baseline (even if slightly slanted) around the CN stretching band.

References

1. Bischak CG, Longhi S, Snead DM, Constanzo S, Terrer E, Londergan CH (2010) Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines. Biophys J 99:1676–1683

2. Choi JH, Oh KI, Cho MH (2008) Azido-derivatized compounds as IR probes of local electrostatic environment: theoretical studies. J Chem Phys 129:11

3. Cremeens ME, Fujisaki H, Zhang Y, Zimmermann J, Sagle LB, Matsuda S, Dawson PE, Straub JE, Romesberg FE (2006) Efforts toward developing direct probes of protein dynamics. J Am Chem Soc 128:6028–6029

4. Fafarman AT, Webb LJ, Chuang JI, Boxer SG (2006) Site-specific conversion of cysteine thiols into thiocyanate creates an IR probe for electric fields in proteins. J Am Chem Soc 128:13356–13357

5. Getahun Z, Huang CY, Wang T, De Leon B, DeGrado WF, Gai F (2003) Using nitrile-derivatized amino acids as infrared probes of local environment. J Am Chem Soc 125:405–411

6. Taskent-Sezgin H, Chung JA, Banerjee PS, Nagarajan S, Dyer RB, Carrico I, Raleigh DP (2010) Azidohomoalanine: a conformationally sensitive IR probe of protein folding, protein structure, and electrostatics. Angew Chem Int Edit 49:7473–7475

7. Waegele MM, Tucker MJ, Gai F (2009) 5-Cyanotryptophan as an infrared probe of local hydration status of proteins. Chem Phys Lett 478:249–253

8. Sagle LB, Zimmermann J, Dawson PE, Romesberg FE (2004) A high-resolution probe of protein folding. J Am Chem Soc 126:3384–3385

9. Schultz KC, Supekova L, Ryu YH, Xie JM, Perera R, Schultz PG (2006) A genetically encoded infrared probe. J Am Chem Soc 128:13984–13985

10. Ye SX, Huber T, Vogel R, Sakmar TP (2009) FTIR analysis of GPCR activation using azido probes. Nat Chem Biol 5:397–399

11. Degani Y, Degani C (1980) Enzymes with asymmetrically arranged subunits. Trends Biochem Sci 5:337–341

12. Edelstein L, Stetz MA, McMahon HA, Londergan CH (2010) The effects of alpha-helical structure and cyanylated cysteine on each other. J Phys Chem B 114:4931–4936

13. Maienschein-Cline MG, Londergan CH (2007) The CN stretching band of aliphatic thiocyanate is sensitive to solvent dynamics and specific solvation. J Phys Chem A 111:10020–10025

14. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. Cell Res. doi:10.1038/cr.2009.87

15. McMahon HA, Alfieri KN, Clark KAA, Londergan CH (2010) Cyanylated cysteine: a

covalently attached vibrational probe of protein-lipid contacts. J Phys Chem Lett 1:850–855

16. Fuxreiter M, Tompa P, Simon I (2007) Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23:950–956

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

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

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

20. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratoty manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

21. Benoit I, Coutard B, Oubelaid R, Asther M, Bignon C (2007) Expression in *Escherichia coli*, refolding and crystallization of *Aspergillus niger* feruloyl esterase A using a serial factorial approach. Protein Expr Purif 55:166–174

22. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of

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

24. Van Hoy M, Leuther KK, Kodadek T, Johnston SA (1993) The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell 72:587–594

25. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32: W668–W673

26. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89:392–400

27. Doherty GM, Motherway R, Mayhew SG, Malthouse JP (1992) ^{13}C NMR of cyanylated flavodoxin from *Megasphaera elsdenii* and of thiocyanate model compounds. Biochemistry 31:7922–7930

28. Habchi J, Mamelli L, Darbon H, Longhi S (2010) Structural disorder within henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. PLoS One 5. doi:10.1371/journal.pone.0011684

affinity tags for protein purification. Protein Expr Purif 41:98–105

Chapter 18

Structure Analysis of Unfolded Peptides I: Vibrational Circular Dichroism Spectroscopy

Reinhard Schweitzer-Stenner, Jonathan B. Soffer, and Daniel Verbaro

Abstract

Vibrational circular dichroism (VCD) spectroscopy is an invaluable spectroscopic techniques utilized to exploit the optical strength of vibrational transitions for structure analysis. In this chapter, we describe the protocol for measuring and self-consistently analyzing VCD and the corresponding FT-IR spectra of short peptides. This process involves the decomposition of the IR spectrum as well as simulations of the amide I band profiles in both spectra based on structural models of the peptides investigated. This type of spectral analysis should be complemented with similar investigations of Raman spectra, which are described in the subsequent chapter. The structural analysis of short, unfolded peptides described in this chapter can easily be extended for the analysis of longer unfolded peptides or even proteins. This is particularly important in view of the demonstrated biological relevance of intrinsically disordered peptides and proteins (IDPs).

Key words: Infrared spectroscopy, Vibrational circular dichroism, Peptide purification, Spectral deconvolution

1. Introduction

Besides FT-infrared (FT-IR) spectroscopy, of the vibrations probed by circular dichroism (VCD) spectroscopy has emerged as another, even more powerful tool for the structural analysis of peptides and proteins (1, 2). Basically, circular dichroism is a measure of the difference between the extinction coefficients ε for left (L)- and right (R)-handed circularly polarized light (Fig. 1) (3):

$$\Delta\varepsilon = \varepsilon_L(\tilde{v}_L) - \varepsilon_R(\tilde{v}_R) \qquad (1)$$

In the case of VCD infrared rather than UV or visible light is used to probe the dichroisms of bands associated with transitions between different vibrational energy levels of the electronic ground state of a molecule. This serves as a measure of the chirality in the vicinity of the respective chromophore as defined by the normal

Right Circularly Polarized Light   Left Circularly Polarized Light
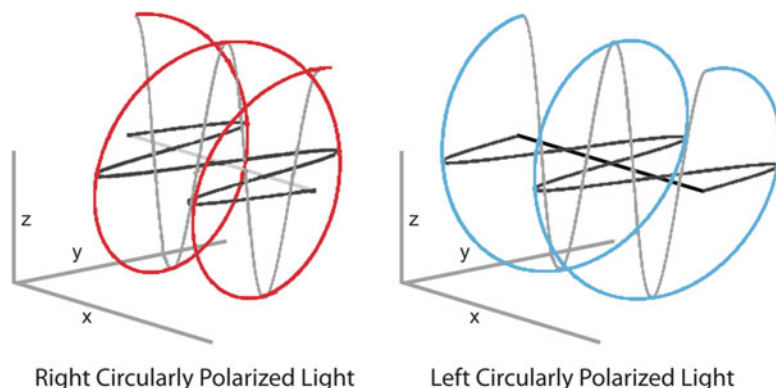
Fig. 1. Representation of right- and left-handed circular polarized light.

mode composition of the mode probed by the IR absorption. In general terms, the rotational strength associated with a transition is written as (4, 5):

$$R_j = \langle \vec{\mu}'_i \rangle \langle \vec{m}'_i \rangle \tag{2}$$

where $\langle \vec{\mu}'_i \rangle$ and $\langle \vec{m}'_j \rangle$ are the expectation values of the electronic and magnetic transition dipole moment operator associated with the i-th normal mode in the electronic ground state of the investigated molecule. Both depend on the normal coordinate Q_i of the respective mode i as described by:

$$\langle \vec{\mu}'_i \rangle = \langle g | \frac{\partial \vec{\mu}}{\partial Q_i} | g \rangle \langle 1_i | Q_i | 0_i \rangle$$
$$\langle \vec{m}'_i \rangle = \langle g | \frac{\partial \vec{m}}{\partial Q_i} | g \rangle \langle 1_i | Q_i | 0_i \rangle \tag{3}$$

where the first matrix element on the right hand side reflects the expectation value of the indicated dipole moment derivative in the electronic ground state |g>. The second matrix element accounts for the transition from the vibrational ground state into the first excited vibrational state of the i-th oscillator. Obviously, the rotational strength is zero if the magnetic moment is perpendicular to the electronic moment. This is generally the case if the chemical group the vibration is associated with is planar. Hence, all vibrations of the nearly planar peptide groups of a polypeptide (Fig. 2) should not show any intrinsic rotational strength. However, in reality one obtains structure sensitive signals for canonical amide modes. Figure 3 shows the VCD spectrum of the alanine dipeptide in aqueous solution (both H_2O and D_2O). The indicated signals are associated with the amide I, amide II, and III modes of the peptide, the normal modes of which are displayed in Fig. 4. An additional, comparatively strong negative signal appears in the region between 1,400 and 1,500 cm^{-1}, where bands are assignable
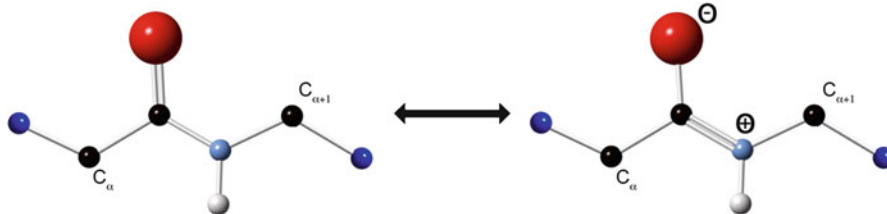
Fig. 2. Representation of the two mesomeric structures of the peptide bond. Real peptide bonds are a linear combination of the two wavefunctions representing these two states.
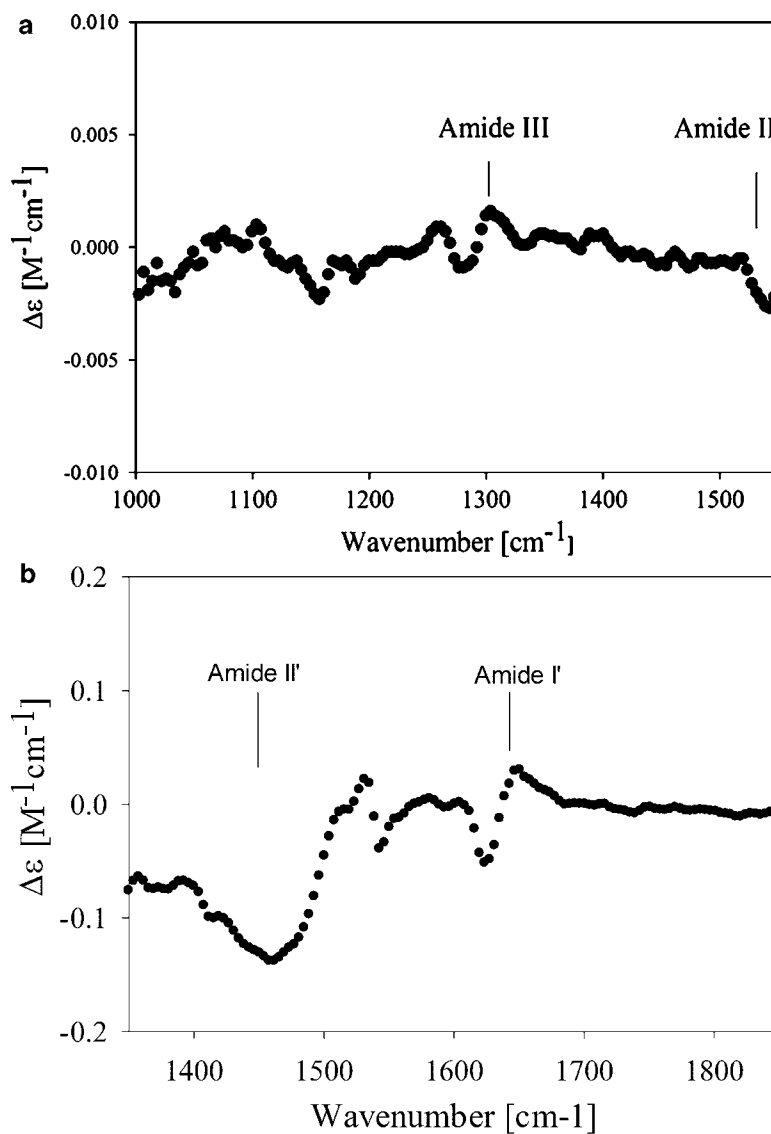
Fig. 3. VCD spectrum of the alanine dipeptide in aqueous solution, (**a**) in H_2O and (**b**) in D_2O.
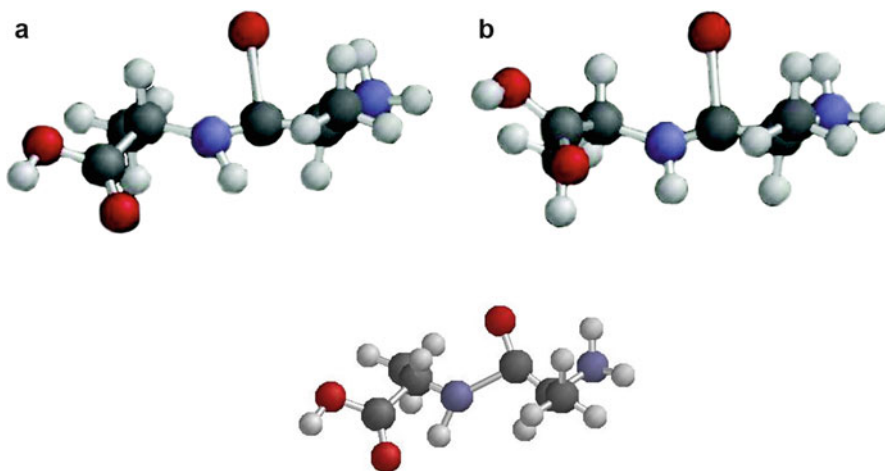
Fig. 4. Upper panel: snapshot of the amide I mode of (**a**) naturally abundant dialanine and (**b**) dialanine with deuterated amide groups in vacuo. The *left figure* shows the displacement of CO, NH, and C$_\alpha$H. In the *right figure*, only the carbonyl group is significantly displaced. *Lower panel*: Snapshot of the amide III mode of dialanine in vacuo. The eigenvector exhibits the characteristic in-phase vibration of NH ib and CN s combined with an in-plane vibration of C$_\alpha$H. The vibrations were obtained by performing a normal mode calculation based on the semiempirical force field, AM1, which is part of the TITAN software from Schrödinger, Inc.

to symmetric and antisymmetric methyl bending vibrations (6, 7). Note that the amide III band is shifted out of the displayed spectral region for alanine dipeptide in D_2O owing to the $H \leftrightarrow D$ exchange of the amide group. With the experimental conditions used for recording the spectra in Fig. 3, the VCD signal of amide I is only detectable if D_2O is used as a solvent. This band is traditionally denoted as amide I′. If H_2O is used as a solvent, the high absorptivity of the band arising from HOH bending modes at 1,636 cm^{-1} significantly distorts the signal. The amide I mode, which is the main focus of this chapter, is predominantly a CO stretching mode with an admixture of NH in-plane bending (ipb) and minor contributions from C$_\alpha$H bending and side chain vibrations. In D_2O, NH is replaced by ND which has a much lower intrinsic bending frequency and therefore does not mix with the CO stretching mode (amide I′). Amide I and I′ are both predominantly peptide modes, which should exhibit no and only weak rotational strength. This is indeed the case for peptides, which contain only a single peptide group (8). However, peptides with two peptide groups like the blocked alanine dipeptide show a pronounced negative couplet, i.e., a negative band at the low and a positive band at the high wavenumber position Fig. 3b. As is shown in more detail (Subheading 2) below, this optical activity is induced by excitonic coupling between adjacent amide I modes, which mixes the respective transition dipole moments. The situation is different for amide III. The composition of this mode is rather complex. Figure 4c shows the amide III mode of the C-terminal peptide group of the

alanine dipeptide. In addition to the NH ipb and CN stretching vibration of the peptide linkage, the respective eigenvector also contains contributions from $C_\alpha H$ ipb, C_α-C_β stretching, and $C_\beta H_3$ deformation modes. This admixture of non-backbone vibrations provides the dominant contribution to the weak negative VCD-signal at the amide III position in the spectrum.

The strength and the form of these VCD signals depend on the conformation adopted by the peptide backbone. This has been reported in detail in papers from the Keiderling group, who used DFT-based rotational strengths to model the VCD spectra of model peptides (9).

While performing ab-initio and/or DFT calculations for peptides in implicit as well as in explicit water is very useful to reveal the spectroscopic characteristics for the IR as well as for the VCD spectra of short peptides adopting well-defined conformations. These methods are less applicable for the analysis of unfolded peptides and proteins the backbone of which generally samples a substantial part of the Ramachandran plot (Fig. 5). The resulting manifold of structures is generally termed random coil, but for a variety of reasons the term statistical coil should be used instead (10). In this chapter, we show how a more empirically oriented analysis of the VCD and IR band profiles of the amide I mode can lead to a rather detailed exploration of the unfolded state of short peptides (11–13). Generally, this analysis is combined with a simulation of amide I profiles of polarized Raman spectra, which for methodological reasons are described in the next chapter. While this analysis is currently applied only to spectra of short peptides
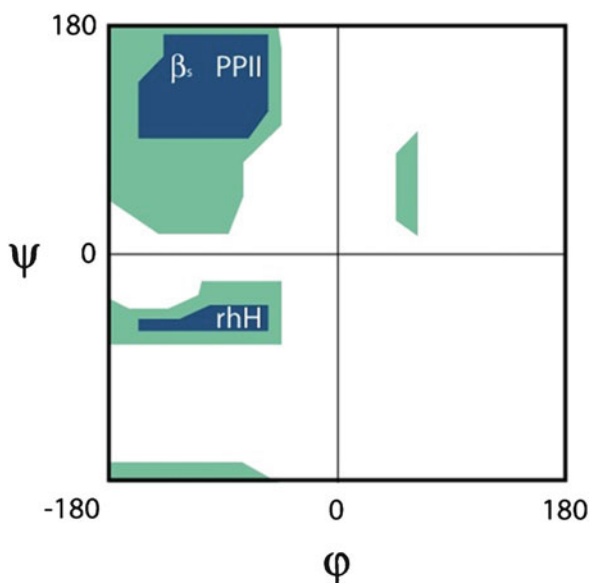
Fig. 5. Ramachandran plot of backbone dihedral angles in peptides.

(3–11 residues), it will be straightforward to extend it to longer peptides and even peptides if sufficient computer power is available. In what follows, we describe in detail the sample preparation, the instrumentation, the software for the acquisition of VCD and IR spectra (which are always measured with the same instrument using the same sample), and finally our analysis of the amide I profile, which are the main focus of this chapter.

2. Materials

2.1. Peptide Preparation   All peptides used are generally purchased from commercial sources (see Note 1) at greater than 95% purity. For purchased peptides the amount of the desired product in a sample can be determined from its gross weight and its analytical purity (see Note 2) (14). The commercially obtained peptide sample generally contains water, salts and small amounts of trifluoroacetic acid (TFA). The presence of TFA is the result of the HPLC process in which the free N terminus and side chains (Arg, Lys, and His in particular) will form trifluoroacetates, thus contaminating the peptide with small amounts of TFA. The presence of TFA gives rise to a band at around $1,674 \text{ cm}^{-1}$ which overlap with the amide I region of the IR spectra (14). In order to obtain get the best possible IR spectra, this contamitant must be removed.

2.2. Sample Preparation   The process of removing impurities involves multiple steps. For the first, preparative steps one needs a Nalgene Cryoware cryogenic sample vial (Rochester, NY), a Kimtech Kimwipe (Kimberly-Clark Inc. Roswell, GA), and a small rubber band. Figure 6 provides an overview of this process, which is described in much greater detail below.

1. Weigh out the peptide quantity needed into the cryogenic sample vial, a few milligrams should be sufficient (see Note 3).

2. Dissolve the peptide in approximately 300 μL of 0.1 M HCl, Fig. 6b.

3. With the cryogenic vial completely sealed, the sample is submerged into liquid nitrogen for about 2 min. The vial is kept at about a 45° angle with the lid pointed towards the side of the nitrogen container in case the gaseous nitrogen pushes the vial out of the grip of the tongs.

4. Remove the vial from the liquid nitrogen and remove cap.

5. Cut a small square off out of the corner of a Kimwipe and cover the top of the sample vial (see Note 4).
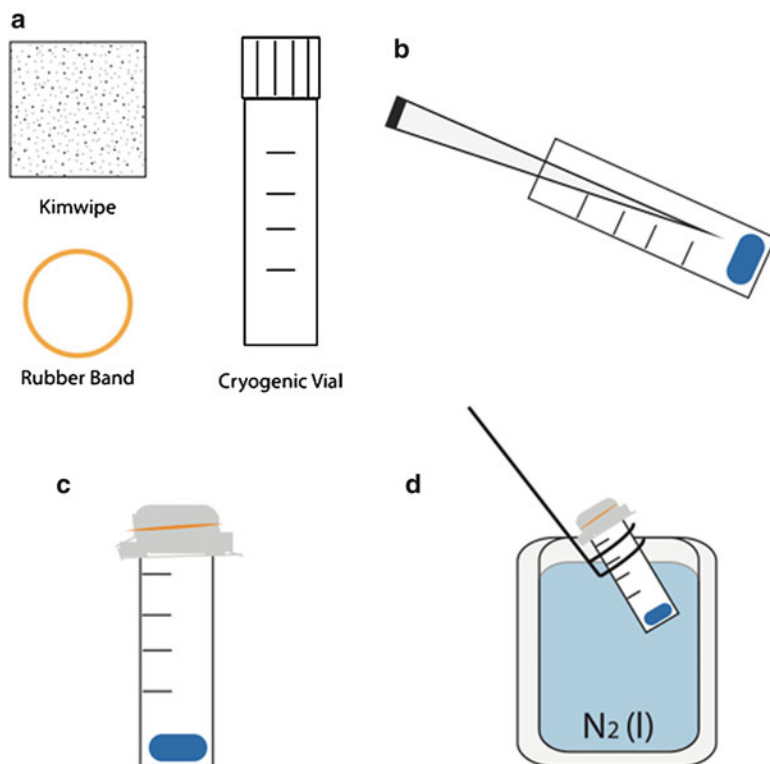
Fig. 6. Schematic representation of the freeze-drying procedure. (**a**) Materials and Devices, (**b**) dissolving the sample in HCl, (**c**) covering of the sample and (**d**) freezedrying in liquid N_2.

6. Hold the Kimwipe filter in place with a small rubber band, as shown in Fig. 6c.

7. Freeze the peptide sample again in liquid nitrogen, Fig. 6d, by carefully inserting the cryogenic sample vial into the liquid nitrogen (see Note 5).

8. Place the vial, positioning it as vertical as possible, into a round bottom flask and attach to the freeze dryer, and lock into place, as illustrated in Fig. 7a.

9. Place one nitrogen dewar with a small amount of nitrogen below the sample and another dewar filled with nitrogen onto the distillate receiver, as illustrated in Fig. 7.

10. Start the vacuum pump and close off the freeze dryer, Fig. 7b.

11. Freeze dry the sample for 6 h.

12. Repeat the freeze-drying process, steps 2–9, for a second time (see Note 6).

13. Redissolve the peptide, and titrate to neutral pH.

14. Repeat the freeze-drying process, steps 3–9, for a third time.
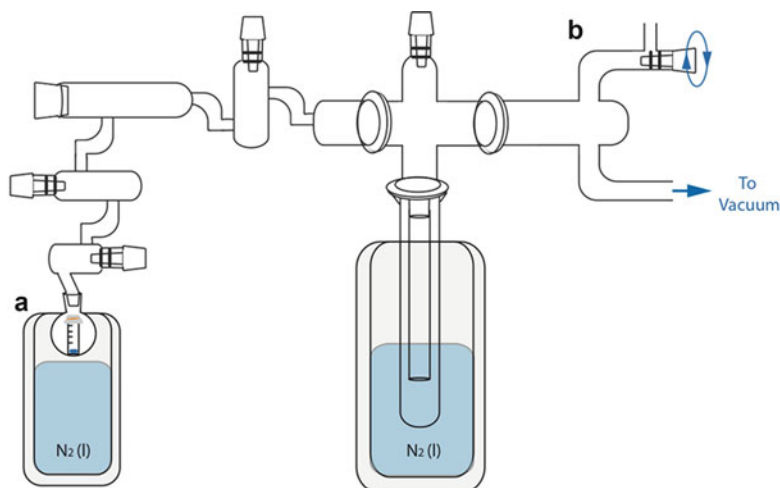
Fig. 7. Freeze-drying apparatus assembly.

15. Turn off the vacuum, releasing the pressure in the freeze dryer, Fig. 7b.

16. Remove the purified sample from the round bottom with flat-nosed tweezers.

17. Dissolve sample in D_2O at a concentration of around 2 mg/mL (see Note 7).

3. Methods

3.1. VCD Instrument and Description of Cells

3.1.1. Instrumentation

This section presents the acquisition process of the FTIR-VCD using the Bomem-BioTools Chiral*ir*™ (Quebec, Canada) integrated VCD analyzer. This instrument consists of a spectrometer, a digital signal processing (DSP) lock-in amplifier (Stanford Research Systems SRS810), dual channel filter (Stanford Research Systems SRS640), photoelastic modulator controller (Hinds PEM-90), and a standard desktop computer with an SEQ36 acquisition board running Bomem-GRAMS/32 software. The Chiral*ir*™ spectrometer houses the interferometer, IR source, filters, and optical components that allows for the modulation and transfer of light through the polarizer and lenses to the detector. This section primarily focuses on the spectrometer, which provides the double modulation of an infrared beam measured at a mercury-cadmium-telluride detector (MCT). Figure 8 shows the simplified schematic of the components housed within the aluminum casing of the spectrometer.
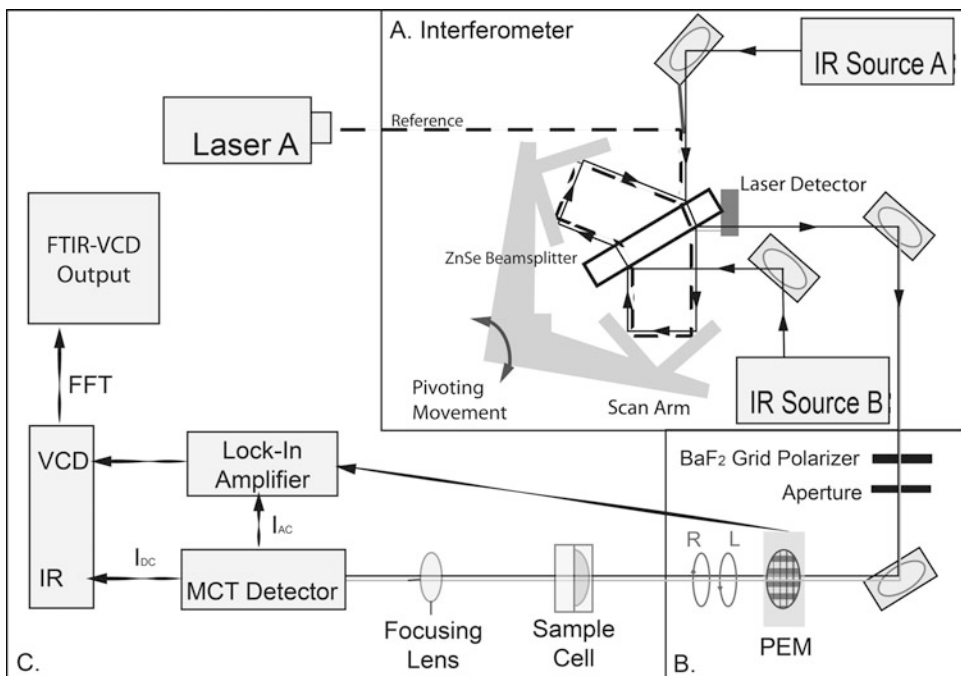
Fig. 8. Detailed optical diagram of our Bomem-BioTools ChiralIR™, with (**a**) showing the interferometer, (**b**) showing area that allows for the creation and modulation of circularly polarized light, and (**c**) the area of detection and analysis.

3.1.2. The Infrared Source and Interferometer

The optimal measurement of a VCD spectrum begins with the precise measurement of the IR spectrum. The measurement of infrared absorption requires a source of continuous infrared radiation, approaching that of a blackbody result, and a sensitive infrared transducer. There are three major types of infrared instruments commercially available dispersive grating spectrometers, nondispersive photometers and multiplex instruments that employ Fourier transform methods. Of the three the most commonly used are dispersive and multiplex instruments with the multiplex instruments being the primary choice due to its inherent advantages (15). A Fourier transform instrument in comparison with a dispersive instrument provides greater signal to noise, a high resolving power, and the important advantage of all elements of the source reaching the detector simultaneously (16). In a dispersive instrument, there are more optical components and slits to reduce the throughput. Fourier transform spectrometers are frequently used because spectra can be recorded in a rather short period of time. A double-beam system is used in the Chiral*IR* allowing the mirrors to direct the interferometer beam through the reference and sample, which are oscillated against the interferometer mirror so that the information present in both can be obtained at each mirror position. Due to the modulation characteristics of the interferometer, the

infrared emission and stray radiation from the sample compartments appear as a DC component to the interferogram and are inherently discriminated against.

To produce an interferogram, the infrared source must be split into two beams. The pathlength of one of these beams can be varied in a periodic fashion to yield an interference pattern. In the Chiral*ir*™ spectrometer, a zinc selenide (ZnSe) beam splitter with ABB Bomem wishbone design is used, Fig. 8a. This is a robuster set up than the traditional interferogram design in that there are two sets of mirrors that move by a pivoting motion. The Near-IR/Mid-IR ZnSe beam splitter is placed at the vertex of the right angle of the wishbone design at a 45° angle relative to the two mirrors. The light that passes through the beamsplitter is split into two 50/50 components, which is directed across the movable mirror of the scanning arm. These two components are shown as solid and dashed lines in Fig. 8a. The mirror on the scanning arm is moved over a distance of 2–20 μm to yield a spectral range of 200–4,000 cm^{-1}. This unpolarized light then leaves the interferometer, following the solid and dashed lines in Fig. 8a illustrating the two interfering beams.

3.1.3. The Creation and Modulation of Circularly Polarized Light

The unpolarized infrared radiation is then passed through a barium fluoride (BaF$_2$) grid polarizer, Fig. 8b, which convert the unpolarized beam into one with linear polarization. The latter then enters the ZeSn photoelastic modulator (PEM), where the linearly polarized light is modulated at the PEM frequency of 36.92 kHz. Its optical axis is oriented by 45° with respect to the vertical polarization of the incident light. The periodic modulation of the (optical) stress axis varies the phase difference $\delta\phi$ between the polarization components oriented parallel and perpendicular to this axis. As a result of the entire modulation cycle, the light oscillates between right and left circularly polarized light ($\delta\phi = \pm\pi/2$) with linear ($\delta\phi = 0,\pi$), and elliptically polarized light between these extreme states, as shown in Fig. 9. This modulated light leaves the PEM and
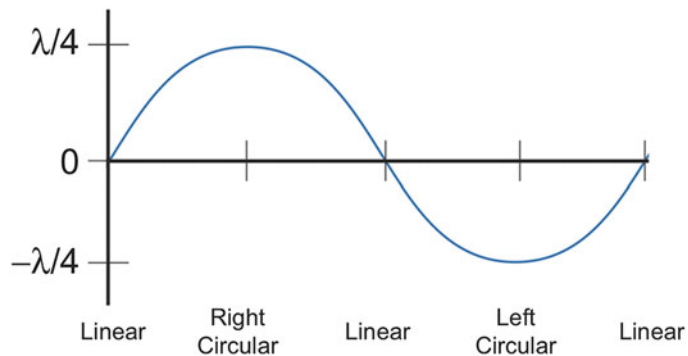
Fig. 9. Photoelastic modulator full sequence of quarter wave retardation.

enters the sample cell. The sample absorbs a certain amount of the modulated polarized light, before reaching the detector.

3.1.4. Detection and Determination of the FTIR-VCD

After passing through the sample cell, the transmitted light is focused by a lens onto the aperture of the detector. The Chiral*ir*™ contains a nitrogen cooled 2×2 mm HgCdTe detector (MCT), a common photodetector for infrared instruments. This detector is chosen due to its large spectral range and high quantum efficiency (15). An AC-coupled preamplifier then amplifies the sample signal, which is subsequently processed through two independent paths to determine the IR and VCD signal (Fig. 8c). These electronic components include a high and low pass filter (AC/DC Coupled).

The signal for the FTIR spectrum has to pass through a low-pass filter before it enters the computer to remove the AC signal. In the time domain, this yields an interferogram, which is then Fourier transformed by means of computer software, as illustrated in Fig. 10. The 20 kHz high pass filter allows an AC signal generated
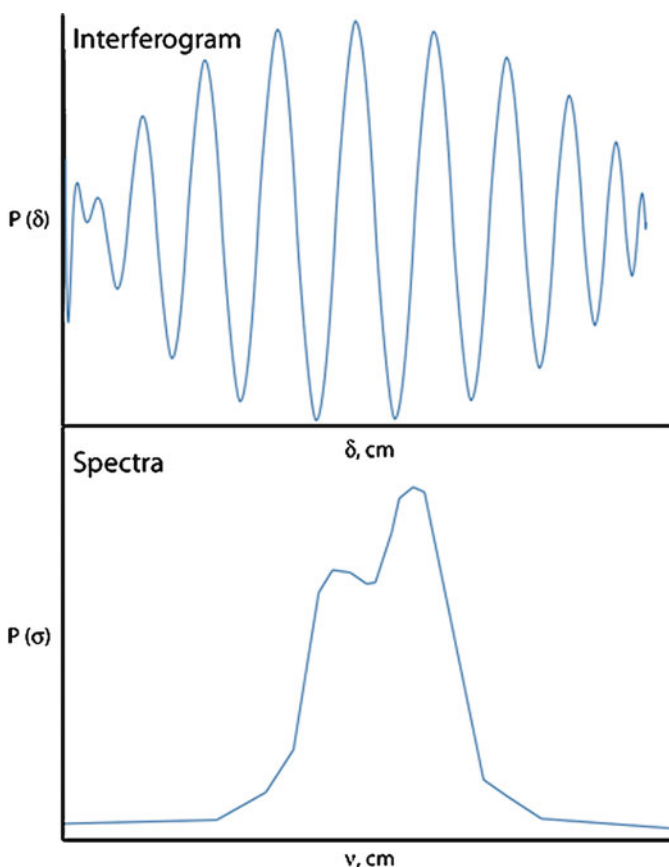
Fig. 10. Example of an interferogram (Interferogram) Fourier transformed into an IR spectrum (Spectra).

by the PEM to pass to the lock-in amplifier (LIA). The LIA is essential for the detection of the VCD signal in that it can facilitate the recording of a very small AC signal on an nV scale, using phase-sensitive detection (17). Phase-sensitive detection is used to single out the signal component at a specific referenced phase and frequency. Any signal, which exhibits a frequency and a phase different from the reference frequency (36.92 Hz) is filtered out and does therefore not contribute to the measured signal. The phase-sensitive detection produces two AC signals as an output, one with the difference frequency ($\omega_r - \omega_l$) and one with the sum frequency ($\omega_r + \omega_l$) (18). If both frequencies are identical, the function $\cos[(\omega_r - \omega_l)t + \phi_{sig} - \phi_{ref}]$ becomes time-independent (ϕ_{sig} and ϕ_{ref} are the phases of the signal and the reference, respectively) and the respective AC signal is converted to a DC signal.

The raw VCD is obtained through the demodulation of the doubly modulated signal at the PEM frequency by the LIA to form an interferogram of the modulated system. The computer then recombines these signals and takes a ratio of the two signals AC and DC signals where the final FTIR-VCD spectrum is obtained.

3.1.5. Sample Cell

The sample compartment of the Chiral*ir*™ is equipped with a 35 mm stainless steel slide mount, which fits a large variety of stress-free optics and clear aperture. The pathlength of the cell may be defined by the placement of the spacer, which is sandwiched between the two windows of the liquid cell (International Crystal Laboratories, Garfield, NJ). Alternatively, it may be placed in a CaF_2 cell with a predefined pathlength (BioTools Inc., Jupiter, FL). Cells with a wide range of pathlengths between 2 and 120 μm are commercially available. For our experiments, we frequently use 20 μm cell.

3.1.6. Sample Cell Setup

1. Begin by cleaning the cell using a small amount of water applied to a Kimwipe and rub the polished surface in a circular fashion to remove all debris that may have been left behind (see Note 8).

2. Allow the cell to air dry (see Note 9).

3. Using the sample that was prepared as described in Subheading 2.2, pipette the sample (see Note 10) into the recessed center of the cell, Fig. 11a (see Note 11).

4. Place the perfectly flat optically clear top plate onto the outer ring of the lower plate to seal the cell. This is best accomplished by sliding the top plate over the bottom plate slowly, as seen in Fig. 11b. If an air bubble begins to form, slide the plate backward and add an additional small volume so that the bubble is above the plane of the cell and continue to slide the top plate until the two plates are aligned (see Note 12).
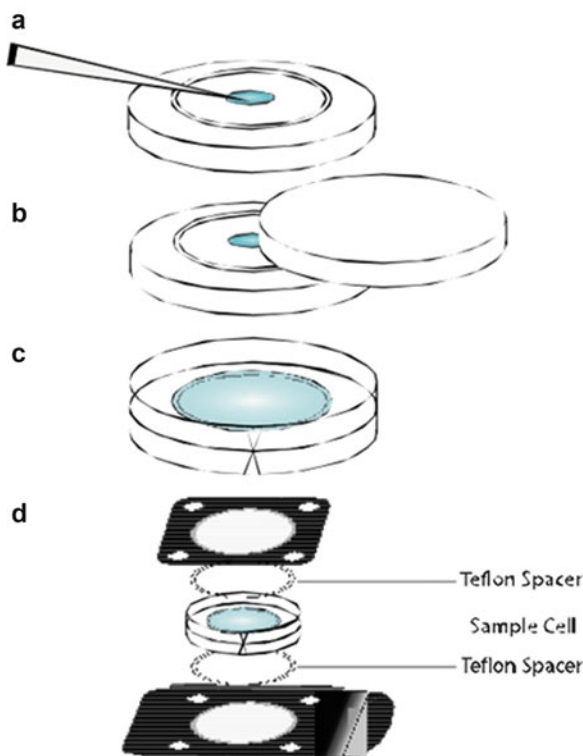
Fig. 11. Sample cell preparation for use in the Chiral*ir*™.

5. Align the upper and lower plate so the marking on the outer edge of the cell are aligned, Fig. 11c, thus minimizing the internal dichroism of the cell at that position to the manufacturer's specifications.

6. Allow the cell to sit for a few minutes on the bench top to allow the seal to set. The seal of the cell prevents the evaporation of water for about 24 h at room temperature.

7. The cell is then inserted into the appropriate cell holder for the experiment you wish to run, Fig. 11d.
   (a) Place the larger sized cell holder down on a flat surface.
   (b) Add the Teflon spacer, ensuring that the lip of the spacer goes into the fitted opening of the sample holder.
   (c) Place the CaF_2 sample cell into the base plate.
   (d) Add an additional Teflon spacer to the cell.
   (e) Carefully add the top plate to the cell (see Note 13).

8. Secure the cell, by adding the screws or nuts depending on cell type, to be at equal pressure around the cell. This sample is now ready for the FTIR-VCD acquisition process.

3.2. Loading the Instrument, Software, and Starting a Measurement

3.2.1. Proper Instrument Startup

1. Fill the MSL-8 nitrogen cooled dewar (InfraRed Assoc., Stuart, FL) of the detector, this should be done at around 30 mL increments. This process should be carried out very slowly, over the course of 30 min (see Note 14).

2. Begin to purge the instrument with nitrogen gas for at least 10 min prior to running your experiment; the flow through should be around 5–7 psi (see Note 15).

3. The spectrometer has a manual resolution knob located on the rear of the instrument; verify the status of the resolution. When setting up the experiment, one will have to verify that the resolution file matches the setting on the instrument. For the analysis of short peptides and proteins, one generally does not need to use a setting lower than 8 cm^{-1}.

3.2.2. Start Bomem-GRAMS Software and Checking the Signal

The BGRAMS/32 Chiral*ir*™ software may be accessed by a variety of methods. The software is most commonly loaded by either clicking on the shortcut key in the start menu of Windows™ (Programs/Galactic/GRAMS_AI), or from the BGRAMS application button on the desktop.

1. Upon startup the software will open the main BGRAMS screen. This initial screen provides for limited operation.

2. To get to the main Chiral*ir*™ screen, click on the Chiral*ir*™ extension in the shortcut (the last icon at the top is the VCD plugin) or click (Arithmetic/Do Program and then select \bgrams\chiralir).

3. Begin by running a background spectrum.

    (a) To acquire a background spectrum, select Collect from the menu and click on Collect (Collect/Collect) to display the Collect dialog box.

    (b) Select the following parameters:

    - Acquisition type: Background.
    - General Param: Number of scans: 10 or more.
    - File Name: Enter file name up to 144 characters.
    - Use Auto Gain: Select this option.

4. Click on OK Collect. During the acquisition of the background spectrum, each scan will be displayed on the screen; upon completion of the scans, the spectrum is displayed and saved in the user-specified location, as specified in step 3b.

5. Once the background spectrum has completed its acquisition, insert the sample into the Chiral*ir*™ slide mount, flat side down, and lock into place.

6. To acquire the IR spectrum of the sample, select Collect from the menu and click on Collect (Collect/Collect) to display the Collect dialog box.

(a) Select the following parameters:

- Acquisition type: Normal.
- Spectral Range Start and End: MIR $(4,000–400\ cm^{-1})$ or NIR $(14,000–4,000\ cm^{-1})$.
- General Param:
  - Initial delay: 0.
  - Number of scans: 10 or more.
- File name: Enter file name up to 144 characters.
- Use auto gain: Select this option.
- Data type: % trans.
- Background: Click on the absorb option and select the background file created in step 5.

7. Click on "OK Collect" to begin acquisition of your sample data. Upon completion, again the file will be automatically displayed and saved, as specified in step 6a.

8. Observe the acquired spectrum. As with all optical instruments, the absorbance values for the bands of interest should be within a range of 0.1–1 (see Note 16).

9. With the system now optimized, you can begin to acquire your FTIR-VCD spectrum.

3.2.3. Acquiring a VCD Spectrum

The VCD control screen is separated into two main tool bar groups: Measurements and Functions, Fig. 12. In the measurement tool bar, the Basic and Advanced buttons are used for the measurements of the Chiral*ir*™ analyzer, Fig. 12a. The basic measurement is useful for routine operations and time studies. The advanced option provides more flexibility allowing for slightly more custom experiments, such as kinetic experiments. The function tool bar allows for the viewing of spectra, as well as provides access to the diagnostic, configuration, and other functions, Fig. 12b.

1. To acquire a routine VCD spectrum, click on Basic from the measurement tool bar, Fig. 12a.

2. The Acquisition Parameters window will pop up to enter the desired experimental parameters.

   (a) Enter the following experimental parameters:

   - File Name: Enter the file name; up to 144 characters is acceptable.
   - File Memo: This will provide an additional description to the obtained file.
   - Duration: This will select the duration time of the acquisition (see Note 17).

3. Phase and calibration: The calibration may also be applied at the time of the acquisition. If "yes" is selected, the user must

Fig. 12. Main interface for operation of the Chiral*ir*™.
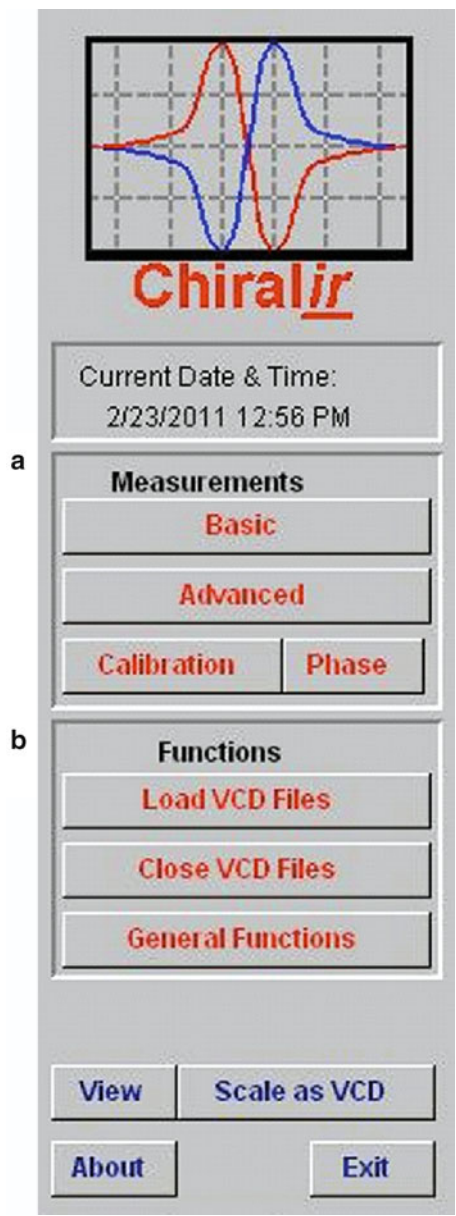
define the calibration file to be used. The phase file must be added to the acquisition parameters; this file must match the resolution that was selected on the back of the instrument, as mentioned in Subheading 3.1.4 (see Note 18).

4. Click *OK*

5. A window will then pop up informing you to "Prepare VCD for Background Acquisition."

    (a) Remove the sample from the slide mount.

    (b) Click OK.

6. Upon completion of the background, the instrument will automatically inform you to "Prepare VCD for Sample Acquisition."

    (a) Reinsert the sample cell into the slide mount, again with the flat edge down.

    (b) Click OK to begin the data acquisition of the FTIR-VCD spectrum.

    (c) Upon completion of the acquisition, the file will be automatically displayed and saved to the specified location as set in Subheading 4.3.2.

7. Export the file, under the file menu, as an ASCII text file (.txt) for use in external analysis software packages such as MultiFit and Microsoft's Excel.

3.3. Preparing Spectra for Final Analysis

3.3.1. Background Subtraction

A blank should be run as your background in a similar manner as with the previous section, repeating Subheading 4.3 steps 1–7. For short peptides, this usually involves the IR spectrum of D_2O. Once this spectrum has been acquired you have to subtract the background from the sample spectrum.

1. Begin by opening the MultiFit program.

2. Open both the IR of the background and the IR of the Sample. This is most easily accomplished by dragging the text files onto the main window of the MultiFit program. This program will be explained in greater detail in the analysis section, Subheading 3.4.

3. Double click on your IR spectrum, to have it as the active file. The active file in MultiFit is the black spectrum in the main screen.

4. Click on Spektrum/Chirurgie, to get to the surgery menu

5. Under Modifizierer: Click on the background spectrum

6. Under Operation: Click Differenz (HS-NS)

7. Click Ok; this will remove the background spectrum from the sample spectrum (see Note 20).

3.3.2. Baseline Correction

Once the background is removed from the sample the next step is the correction of the baseline. This is also a process that is very delicate due to the weak intensity of the signals, particularly in the VCD spectrum (c.f. Note 19).

1. Open the VCD spectrum in Excel and plot the data.

    (a) In Excel's File menu click Open.

    (b) Select the file, and click Ok.

(c) Click Next and verify that tab delimited is clicked.

(d) Click Finished.

(e) Plot the data.

2. Fit a line in the VCD spectrum so that the baseline is set to zero, this is usually a very minimal slightly sloping line.

3. Use the equation of the line to create the baseline in another cell column.

4. Subtract the baseline from the spectrum and save the file. To import back into MultiFit, you will need to create a two column table using the updated information and save as a text file (see Note 21).

5. After converting absorptivity and dichroism to ε and $\Delta\varepsilon$ (in units of M^{-1} cm^{-1}) both, the IR and VCD Spectra are prepared for spectral analysis using MultiFit, which will be explained in the next section.

3.4. Data Analysis   In this section, we use trialanine in H_2O/D_2O as a model system. Recent analyses of amide I profiles and NMR J-coupling constants have convincingly shown that this peptide predominantly samples PPII-like conformations (the corresponding mole fraction for these conformations lie between 0.8 and 0.9) (12, 13, 19). Here, we will simply show how the VCD and the IR spectra of trialanine can be self-consistently analyzed for exploring the conformational manifold adopted by the peptide.

Figure 13 displays the IR and VCD spectrum of trialanine in water and in D_2O in the region between 1,100 and 1,800 cm^{-1}. The former depicts the bands assignable to the classical backbone modes amide I, II, and III. In D_2O, the amide proton is replaced by a deuteron. Hence, the NH ipb mode is replaced by an ND ipb mode, which is now decoupled from CN s and skeletal modes. As a consequence, only the amide I and II bands appear in the high wavenumber region, which are now denoted as amide I′ and II′. The former is only slightly and the latter is more substantially downshifted by the deuteration of the amide group. The remaining bands in the spectra are assignable to combinations of CH_3 bending and skeletal modes, which are not particularly structure sensitive.

3.4.1. Amide I Band Spectral Decomposition   For the purpose of analyzing the amide I band, it is recommendable to perform a spectral decomposition into individual bands. In our group, this task is performed with the aforementioned program MultiFit, which has been developed in one of the former laboratories of the principal author of this article (see Note 22) (20). Figure 14 shows the translated command lines associated with the MultiFit program. The fitting of a spectrum is carried out with the sequence of windows shown in Fig. 15, which we will now explain step by step.
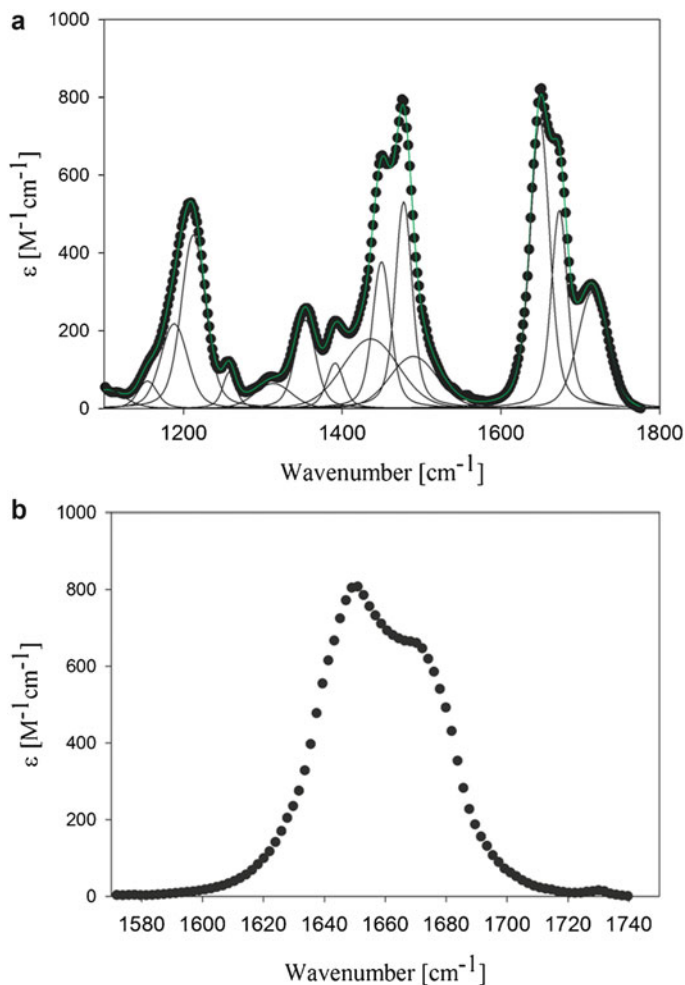
Fig. 13. (**a**) Decomposition of the FT-IR spectrum of AAA in D_2O between 1,200 and 1,800 cm^{-1} into Voigtian bands by MultiFit as described in the text, (**b**) Isolation of the amide I′ band profile of AAA in D_2O by subtracting the other bands of the spectrum obtained from the spectral decomposition.

1. Begin by clicking on "Peakfit" in the main menu, Fig. 15a, to see the peakfitting options, Fig. 15b.

2. Click on "Linienform" to bring up the fit-component option screen, Fig. 15c.

3. Select "Hinzufügen" (add), Fig. 15d, this window allows for the adding of bands to the proposed model to be used for fitting the experimental spectrum.

4. Concerning the profiles of the individual bands the program can convolute two different profiles, named "1. Faltungsfunktion" and "2. Faltungsfunktion" (First and second convolution function), as seen in Fig. 15d.
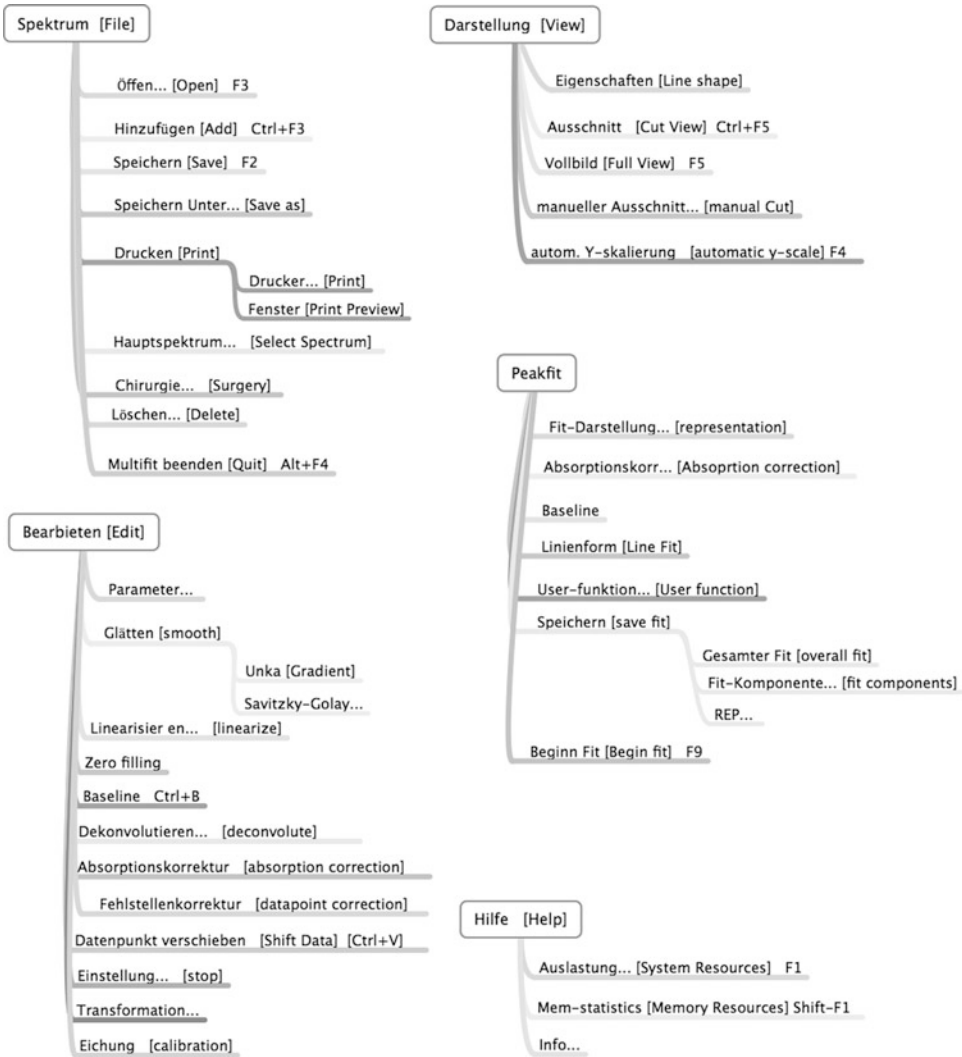
Fig. 14. List of translated MultiFit commands organized identically to that of the main menu.

(a) If one decides to use only a Lorentzian or a Gaussian profile, this choice can be made in the left rubric. In this case, one has to click "keine" (none) in the second one.

(b) For a Voigtian profile, one has to click Lorentzian in the left and Gaussian in the right rubric (see Note 23).

5. To define a spectroscopic model, guesses for the bandwidths have to be inserted into the program window shown in Fig. 15e (HWB). The respective value can be fixed, which becomes necessary if one carries out a self-consistent iterative procedure to simultaneously fit corresponding IR and Raman spectra.

6. Once this has been done, click on "Ok".
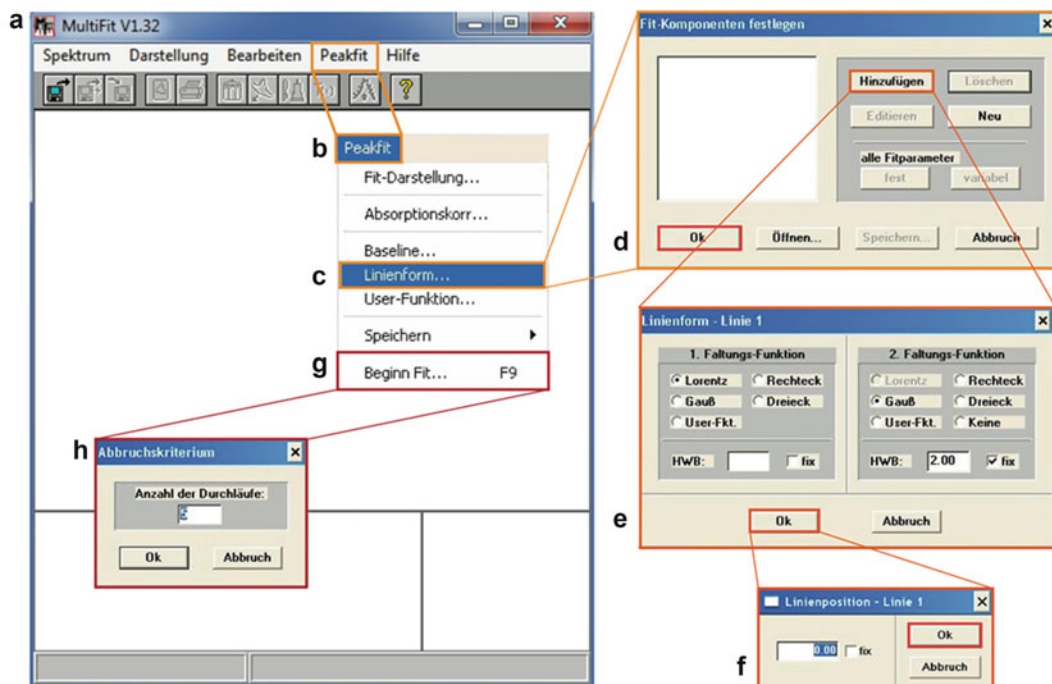
Fig. 15. Navigation of MultiFit's main screen for data analysis.

7. A new window opens, Fig. 15f, which prompts the operator to guess the value for the wavenumber position.

   (a) This can be fixed as well by clicking the box in the window.

8. Click Ok.

9. This process must then be repeated for adding more bands to the spectral model.

10. Once that process is completed and all of the bands are estimated, click "Ok", Fig. 15d.

11. Finally, one has to make a choice for baseline correction; this can be accomplished by clicking on "peakfit" and subsequently on "baseline", as seen in Fig. 15b. The program offers the choices of constant, linear, quadratic, and various hyperbolic baselines, which can all be combined with each other. The individual parameters of the baseline function are determined by the fit (see Note 24). A proper choice for the baseline of the spectrum is a prerequisite for an accurate spectral decomposition. It can best be achieved through a self-consistent fitting, which also involves the analysis of Raman spectra and which is explained in the Raman chapter of this book (see Note 25).

Figure 13 shows the deconvolution of the IR spectrum of cationic trialanine in D_2O. After subtracting all the bands not assignable to amide I′ in the region between 1,500 and 1,600 cm^{-1}, the

Table 1

Wavenumber positions ($\tilde{\nu}_i$, i = 1,2) and Gaussian halfwidths (Γ_{gi}) of amide I′ derived from a spectral decomposition of the amide I′ band profile of cationic AAA in D_2O

$\tilde{\nu}_1[cm^{-1}]$	$\tilde{\nu}_2[cm^{-1}]$	$\Gamma_{G1}[cm^{-1}]$	$\Gamma_{G2}[cm^{-1}]$
1,652	1,676	21.3	18.9
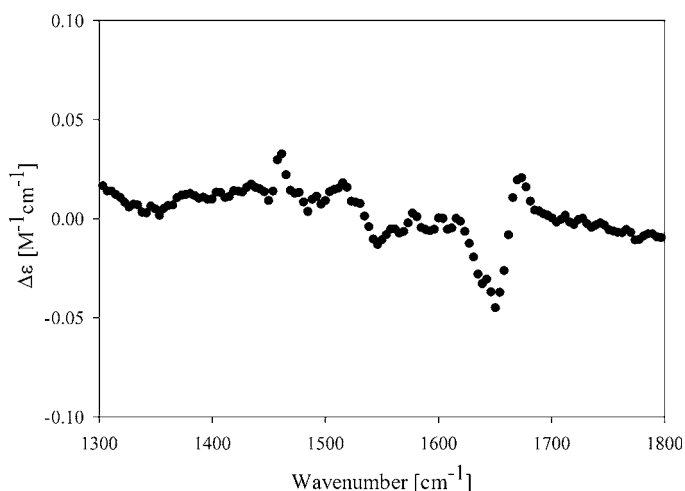
Fig. 16. VCD band profile of amide I′ of trialanine in D_2O measured at pH = 1.56.
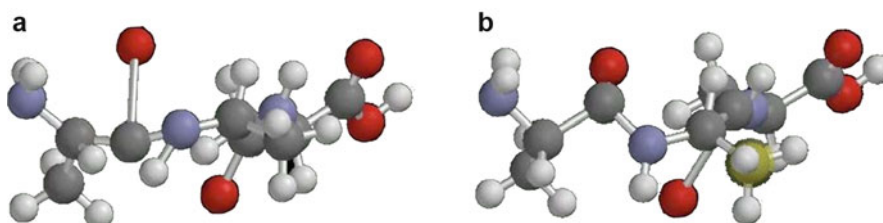
Fig. 17. Snapshot of the coupled amide I modes of trialanine in vacuo. The vibrations were obtained by performing a normal mode calculation based on a force field obtained with a DFT calculation at a BL3LYP 6-31G** level of theory. The program is part of the TITAN software from Schrödinger, Inc. The in-phase combination is depicted in the right and the out-of-phase combination in the left figure. Obtained from ref. 45 and modified.

amide I′ is isolated as shown in Fig. 17b. Two bands are clearly discernable. The spectral decomposition reveals two Voigtian bands AI_1 and AI_2, the spectral parameters of which are listed in Table 1. The respective Lorentzian halfwidth of 11 cm^{-1} used for our fit have been obtained from FTIR femtosecond experiments (21).

The AI_1-band is predominantly assignable to the C-terminal amide I′, whereas the eigenvector of AI_2 is dominated by N-terminal amide I mode (22). The difference between the integrated intensities of the two bands reflects excitonic mixing between the two amide I modes, which is discussed in more detail below (23).

Only the amide I shows a detectable VCD spectrum for trialanine in D_2O, as shown in Figs. 13 and 16. The amide I signal is the aforementioned relative intense negative couplet with a negative signal at AI_1 and a positive one at AI_2. The very existence of such a couplet demonstrates that the two modes are coupled. This coupling will be exploited in the following to determine the structural distribution of the central alanine residue of the peptide.

3.4.2. Analyzing the Amide I Based on an Excitonic Coupling Model

It is well established that the coupling between the two modes results from through-bond interactions (24). The result can be described classically just as two normal modes which contain amide I like contributions from both peptide groups (11, 25). A visual display of their eigenvectors is shown in Fig. 17. Roughly, the lower wavenumber mode could be ascribed as an out-of-phase combination of the two CO s oscillations, whereas the higher wavenumber mode is the respective in-phase combination.

An alternative way of describing this mode coupling for the sake of analyzing the IR and VCD band profiles of amide I′ is the excitonic coupling model, which is based on elementary quantum mechanics (21, 23, 26–29). The Hamiltonian for two coupled quantum mechanical oscillators is written as:

$$\hat{H} = \begin{pmatrix} \hat{H}_1 & \hat{H}_{12} \\ \hat{H}_{21} & \hat{H}_2 \end{pmatrix} \qquad (4)$$

where $\hat{H}_1$ and $\hat{H}_2$ are the Hamilton operators of the two isolated oscillators. The Hamiltonian is expressed in the basis of unperturbed oscillators $|1_1 0_2 >$ and $|0_1 1_2 >$, i.e., the quantum mechanical eigenvectors of a peptide with the C- and N-terminal amide I in the first excited state. The interaction operator $\hat{H}_{12}$ reads as (11):

$$\hat{H}_{ex} = \frac{\partial V}{\partial Q_1 \partial Q_2} Q_1 Q_2 \qquad (5)$$

where V is the total potential function of the molecule. Q_1 and Q_2 are the normal coordinates of the individual modes. The validity of this approach requires that the eigenvectors of the two individual modes do not have local modes in common. In this case, the normal modes of the coupled system can be transferred into the local normal modes of the peptides by a unitary transformation. This is indeed possible, for trialanine as it has been convincingly demonstrated by Cho and coworkers (28) as well as by Gorbunov et al. (27).

The solution of the Schrödinger equation for Eq. 4 is a linear combination of local oscillator functions (23):

$$\psi_- = \cos v \cdot |1_1 0_2\rangle - \sin v \cdot |0_1 1_2\rangle$$
$$\psi_+ = \sin v \cdot |1_1 0_2\rangle + \cos v \cdot |0_1 1_2\rangle \quad (6)$$

with:

$$v = \frac{1}{2} \arctan\left(\frac{2 \cdot \Delta}{\Omega_{A1} - \Omega_{A2}}\right) \quad (7)$$

where Ω_{A1}, Ω_{A2} are the wavenumber positions of the unperturbed amide I′ modes. The coupling value Δ can be written as the off-diagonal matrix element:

$$\Delta = \langle 1_1 0_2 | \hat{H}_{\mathrm{ex}} | 0_1 1_2 \rangle \quad (8)$$

All coupling energies are expressed in wavenumbers throughout this and the subsequent chapter. These coupling energies are generally obtained from Hartree-Fock and DFT calculations on short peptides (24, 27, 28, 30). Details are given below.

We now demonstrate how the amide I′ profiles of, for example, trialanine can be modeled. The amide I′ band profile can be rationalized in terms of a superposition of the transitions into two excitonic states ψ_+ and ψ_-:

$$\varepsilon(\tilde{v}) = A_\varepsilon \bar{\tilde{v}} \left[\frac{\cos v \langle \vec{\mu}'_{A1}\rangle - \sin v \langle \vec{\mu}'_{A2}\rangle}{\sigma_- \sqrt{2\pi}} e^{-\frac{(\tilde{v} - \Omega_-)^2}{2\sigma_-^2}} \right.$$
$$\left. + \frac{\sin v \langle \vec{\mu}'_{A1}\rangle + \cos v \langle \vec{\mu}'_{A2}\rangle}{\sigma_+ \sqrt{2\pi}} e^{-\frac{(\tilde{v} - \Omega_+)^2}{2\sigma_+^2}} \right] \quad (9)$$

where $A_\varepsilon = 1.0869 \times 10^{38}$ M^{-1} esu^{-2} cm^{-2} and $\bar{\tilde{v}}$ is the average wavenumber of the amide I′ region. The amount of the two transition dipole moments $\vec{\mu}_{A1}$ and $\vec{\mu}_{A2}$ can be assumed to be identical (2.7×10^{-19} esu cm) (31). The line profiles of the individual bands are described as Gaussian distribution with the half-halfwidths σ_+ and σ_-. In reality, as indicated above, the both profiles are Voigtians (see Note 23) (22). However, the corresponding Lorentzian are much smaller (ca by a factor of 2) than the corresponding halfwidths. Since the total halfwidth is approximately the square root of the sum of the squares of Lorentzian and Gaussian halfwidths, a pure Gaussian halfwidth is a reasonable approximation for our simulation. Using Voigtian rather than Gaussian profiles would significantly increase the time the simulation program needs to complete the calculation. The eigenenergies of the excitonic states are denoted as Ω_+ and Ω_-. They are identical with the wavenumber positions, which one derives by means of the above-mentioned spectral decomposition. However, these energies are not identical with the energies of the unperturbed amide I′ modes. The relation
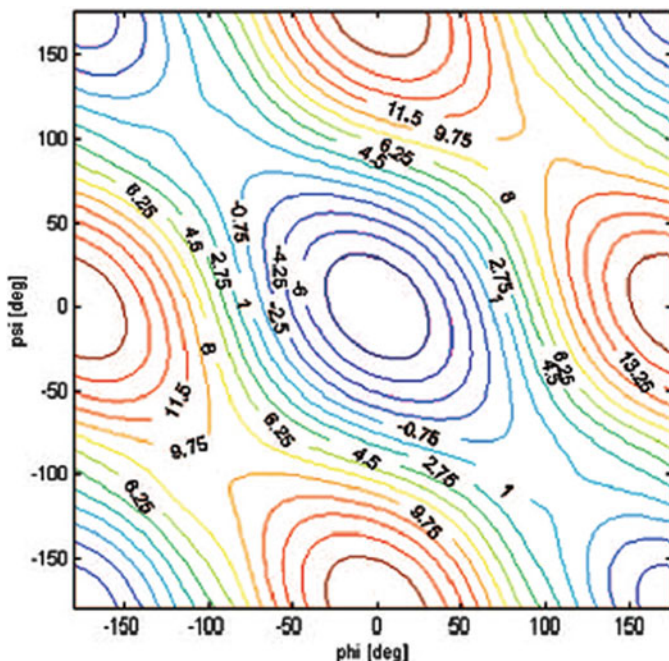
Fig. 18. Contour plot of the nearest neighbor-coupling constant as function of φ and ψ obtained with a heuristic model described in ref. (12), from where this figure has been taken and modified. The numbers in the figure reflect the coupling strength in cm^{-1}.

between energies of local oscillators and eigenenergies are given by (22, 32):

$$\Omega_{\pm} = \frac{\Omega_{A1} + \Omega_{A2}}{2} \pm \frac{1}{2} \sqrt{(\Omega_{A1} - \Omega_{A2})^2 + 4\Delta^2} \qquad (10)$$

where Ω_{A1} and Ω_{A2} are the wavenumbers of the local amide I′ modes.

In order to utilize Eq. 9, we first have to calculate the mixing factors θ. To this end, we need the expectation value Δ for excitonic coupling. This value is generally obtained from normal mode calculation based on ab initio calculated force fields. Contour maps for nearest neighbor coupling have been published by various groups (24, 27, 28). We found that the map of Torii and Tasumi accounts very well for amide I profiles of peptides which predominantly sample PPII (22, 33–35). In a recent paper, we reported a heuristic mathematical model by which we could reproduce the essential features of the Torii and Tasumi map. The result is shown in Fig. 18. The algorithm is described in detail in the paper of Schweitzer-Stenner et al., from which the relevant equations can be taken and programmed (36).

In order to carry out mathematical operations in Eq. 9, the orientations of the two dipole moments have to be expressed in
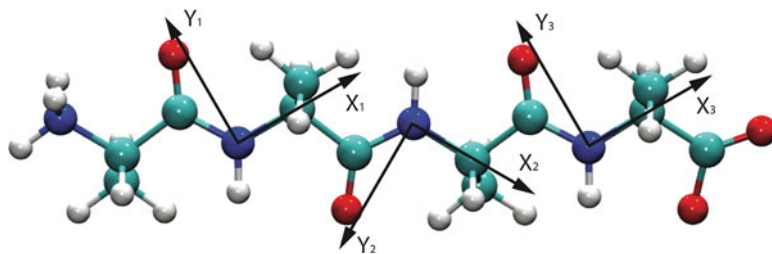
Fig. 19. Planar structure of tetra-alanine ($\phi = 180°$, $\psi = 180°$). Planar structure of tetra-alanine ($\phi = 180°$, $\psi = 180°$). The coordinate systems S1(x1, y1, z1), S2(x2, y2, z2), and S3(x3, y3, z3) were used to express the Raman tensors of the individual, uncoupled amide I modes and their transition dipole moments (the z-component for S2 has been omitted for the sake of clarity). The structure was obtained by using the program VMD.

the same coordinate system. To this end, we first define coordinate systems for the two peptide groups (Fig. 19). For practical reasons, the x-axes lay along the NC_α bonds with the zero point at the peptide nitrogens. The respective y-axes point towards the carbonyl oxygen. The z-axes are defined according to the three-finger rule. The components of the transition dipole moment of amide I′ in the N-terminal can then be written as follows:

$$\vec{\mu}'_{A1}(N) = \begin{pmatrix} \mu'_{A1} \cos \vartheta \\ \mu'_{A1} \sin \vartheta \\ 0 \end{pmatrix} \qquad (11)$$

where ϑ is the angle between the dipole moment and the x-axis. If one assumes that the dipole moment exhibits an orientation of 20° with respect to the carbonyl bond and that the NC_α bond lies in the peptide plane (in fact, there is a small inclination of 2°), one obtains −96.87° (23). Of course, Eq. 11 can also be used for the C-terminal dipole moment in the respective coordinate system. Now we have to rotate the N-terminal coordinate system into the C-terminal terminal system. This can be achieved by the following rotation (36):

$$\vec{\mu}'_{A1}(C) = \vec{\mu}'_{A1}(N) \cdot R(\varphi) \cdot R(-\xi) \cdot R(\psi) \cdot R(-\eta) \cdot R(\tau) \quad (12)$$

T indicates a transposed vector. R indicates a matrix, which describes the rotation by the rotational angle indicated in parenthesis. φ and ψ are the dihedral angles of the central (alanine) residue of the tripeptide, ξ (70°) is the supplementary angle to the bending angle $NC_\alpha C$, η (−65.81°) is the supplementary angle to $\angle C_\alpha CN$, and τ (59.04°) is the supplementary angle to $\angle CNC_\alpha$. There is no need to carry out the matrix multiplication in explicit terms, as it was done in earlier work (23). The equation can be programmed directly into MATLAB, after the matrix elements have been defined
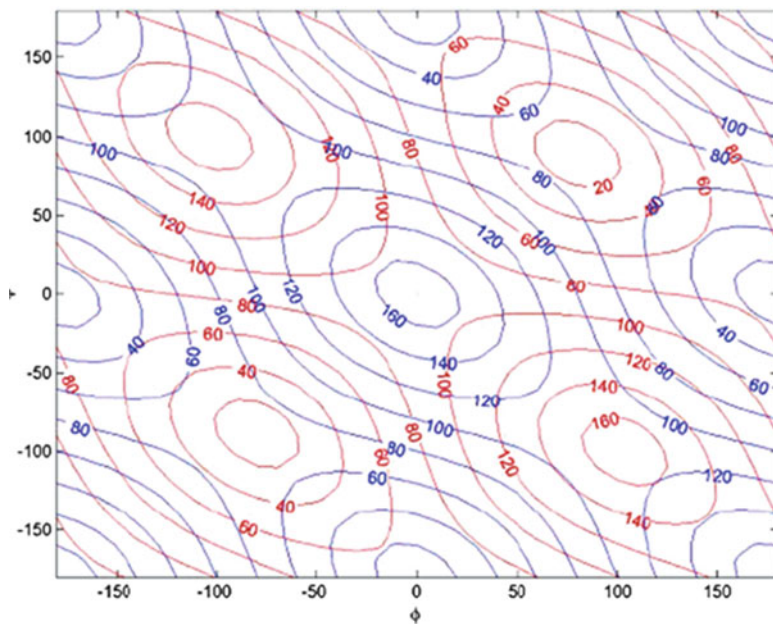
Fig. 20. Contour plots of the orientational angle θ (*gray*) between the transition dipole moments of amide I and of the angle θ′ (*black*) between the normals to the two peptide groups. Taken from ref. 32 and modified.

in the program. It should be noted that $R(\varphi)$ and $R(\psi)$ are carried out around the x-axes of the respective coordinate system, which coincides with $C_\alpha N$ for the former and $C_\alpha C$ for the latter. The remaining rotations are carried out around the z-axes of the respective coordinate systems. Representatives of these two types of rotations are written as:

$$R(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \tag{13a}$$

$$R(\xi) = \begin{pmatrix} \cos\xi & -\sin\xi & 0 \\ \sin\xi & \cos\xi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13b}$$

After carrying out the operation defined by Eq. 12, the respective dipole moment can be inserted into Eq. 9. An equation similar to Eq. 11 must be used to calculate the components of the transition dipole moment $\vec{\mu}'_{A2}$. Apparently, the angle between the two transition dipole moments depends on the dihedral angles φ and ψ, as it is shown by the contour plot in Fig. 20. The amide I′ IR distribution can be calculated for a given pair φ and ψ. Figure 21 illustrates how the profile changes as a function of ψ at a given φ-value $(-60°)$ and as a function of φ at a given ψ-value $(140°)$. Before we consider realistic distributions of conformations, we turn to the calculation of the VCD signal of amide I′.
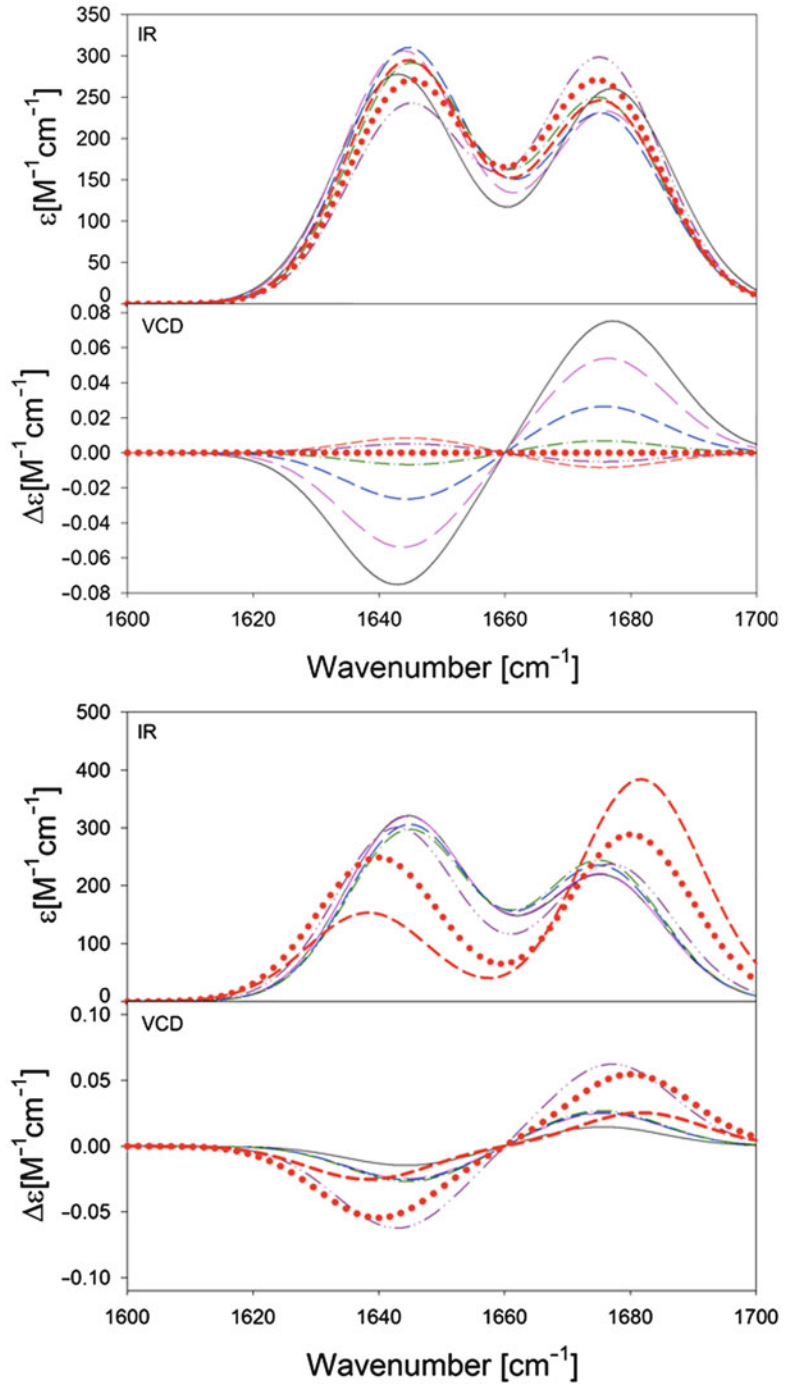
Fig. 21. IR band profile of amide I plotted (**a**) as a function of ψ for $\phi = -60°$ ($\psi = 170°$ (*solid*), 150° (*long dash*), 130° (*short dash*), 110° (*dash-dot*), 60° (*dash-dot-dot*), −30° (*dot-dot*)), and $(\phi, \psi) = (-30°, -30°)$; (**b**) as a function of ϕ for $\psi = 150°$, $\phi = -150°$ (*solid*), −130° (*long dash*), −110° (*short dash*), −90° (*dash-dot*), −60° (*dash-dot-dot*), −30° (*short dash*), and 30° (*short dash*, −30 and 30 respectively).

The rotational strength of the two amide I′ modes can be calculated by using the equation (22):

$$
\begin{aligned}
R_+ =& \langle \vec{\mu}'_{A1} \rangle \cdot \langle \vec{m}'_{A1} \rangle \cos^2 \nu + \langle \vec{\mu}'_{A2} \rangle \cdot \langle \vec{m}'_{A2} \rangle \sin^2 \nu \\
& + \left[\begin{array}{l} \dfrac{1}{2} (\langle \vec{\mu}'_{A1} \rangle \cdot \langle \vec{m}'_{A2} \rangle + \langle \vec{\mu}'_{A2} \rangle \cdot \langle \vec{m}'_{A1} \rangle) \sin(2\nu) \\[2mm] -\dfrac{1}{2} \sin(2\nu) \cdot \pi \cdot \vec{R}_{12} \cdot (\langle \vec{\mu}'_{A1} \rangle \times \langle \vec{\mu}_{A2} \rangle) \end{array} \right] \\
R_- =& \langle \vec{\mu}'_{A1} \rangle \cdot \langle \vec{m}'_{A1} \rangle \cos^2 \nu + \langle \vec{\mu}'_{A2} \rangle \cdot \langle \vec{m}'_{A2} \rangle \sin^2 \nu \\
& - \left[\begin{array}{l} \dfrac{1}{2} (\langle \vec{\mu}'_{A1} \rangle \cdot \langle \vec{m}'_{A2} \rangle + \langle \vec{\mu}'_{A2} \rangle \cdot \langle \vec{m}'_{A1} \rangle) \sin(2\nu) \\[2mm] -\dfrac{1}{2} \sin(2\nu) \cdot \pi \cdot \vec{R}_{12} \cdot (\langle \vec{\mu}'_{A1} \rangle \times \langle \vec{\mu}_{A2} \rangle) \end{array} \right]
\end{aligned} \tag{14}
$$

This equation considers two contributions to the rotational strength of individual amide I′ modes. The respective intrinsic contributions are accounted for by the dot products $\vec{\mu}' \cdot \vec{m}'$. As indicated above, they can be expected to be small owing to the planarity of the peptide groups. However, we have shown in the past that the C-terminal amide I′ modes of unblocked peptides exhibit some intrinsic rotational strength if the C-terminal carboxylate group is protonated (8, 22). The third and fourth terms in Eq. 14 reflect rotational strength induced by excitonic coupling between the two excited amide I′ vibrations which mixes, e.g., electronic and magnetic dipole strength of one amide I′ mode into the transition of the other mode. The contributions of these mixing terms depend of the relative orientation of electronic and magnetic transition dipole moments of the interacting oscillators and thus on the dihedral angles of central residue. In most cases, the fifth terms describe the predominant contributions to the rotational strengths of both modes. They contain cross products of the two electronic transition dipole moments and reflect the fact that a circular current in one peptide group can induce a magnetic dipole moment in the other one. Contributions described by these terms give rise to symmetric couplets (37).

In order to calculate the rotational strength with Eq. 14, we have to calculate the distance vector $\vec{R}_{12}$ by utilizing:

$$
\vec{R}_{12} = \sigma \vec{r}_{CO,1} + \vec{r}_{CN} + \vec{r}_{NC_\alpha} + \vec{r}_{C_\alpha C} + \sigma \vec{r}_{CO,2} \tag{15}
$$

where $\vec{r}_{CO,1}$ and $\vec{r}_{CO,2}$ are vectors pointing from the carbonyl carbons to the respective oxygens, $\vec{r}_{CN}$ is a vector pointing from the carbonyl carbon of the N-terminal peptide to the respective amide nitrogen, $\vec{r}_{NC_\alpha}$ points from the N-terminal nitrogen to the central C_α-atom, and $\vec{r}_{C_\alpha C}$ from the central C_α-atom to the C-terminal carbonyl carbon. The factor σ (0.5) ensures that $\vec{R}_{12}$ starts and ends at the position of the transition dipole moments.

All the vectors in Eq. 15 have to be expressed in the N-terminal coordinate system. This eventually leads to the following equation:

$$\vec{R}_{12} = \vec{r}_{O_{A1}N} + \vec{r}_{NO_{A2}} \tag{16}$$

with

$$\vec{r}_{O_{A1}N} = \begin{pmatrix} r_{CN} \cdot \cos \varepsilon + \sigma \cdot r_{CO,1} \cdot \cos \omega \\ r_{CN} \cdot \sin \varepsilon + \sigma \cdot r_{CO,1} \cdot \sin \omega \end{pmatrix} \tag{17}$$

and

$$\vec{r}_{NO_{A2}} = \begin{pmatrix} r_{NC_\alpha} + (r_{C_\alpha C} + r_{CD_x}) \cdot \cos \gamma - r_{CD_y} \cdot \sin \gamma \\ (r_{CD_y} \cdot \cos \gamma + (r_{C_\alpha C} + r_{CD_x}) \cdot \sin \gamma) \cdot \cos \varphi + r_{CD_z} \cdot \sin \varphi \\ r_{CD_z} \cdot \cos \varphi - (r_{CD_y} \cdot \cos \gamma + (r_{C_\alpha C} + r_{CD_x}) \cdot \sin \gamma) \cdot \sin \varphi \end{pmatrix} \tag{18}$$

where $r_{CO,1}$, r_{CN}, r_{NC_α} and $r_{C_\alpha C}$ are the lengths of the corresponding vectors defined above (1.22, 1.32, 1.42, and 1.53 Å), $\gamma = \pi - \xi$ and $r_{CD_j}(j = x, y, z)$ are the components of the vector from the C-terminal carbonyl carbon to the respective dipole moment, which is written as:

$$\vec{r}_{CD} = \begin{pmatrix} \sigma \cdot r_{CO,2} \cdot \cos \omega \\ \sigma \cdot r_{CO,2} \cdot \sin \omega \cdot \cos \psi \\ -\sigma \cdot r_{CO,2} \cdot \sin \omega \cdot \sin \psi \end{pmatrix} \tag{19}$$

where $r_{CO,2} = r_{CO,1}$ and $\omega = -59.08°$. After using Eqs. 16–19 to calculate $\vec{R}_{12}$ for a given pair of dihedral coordinates it can be inserted into Eq. 14 to calculate the rotational strength associated with the excitonic states ψ_+ and ψ_-. These can then finally be used to calculate the amide I′ VCD signal as follows:

$$\Delta\varepsilon(\tilde{\nu}) = A_{\Delta\varepsilon}\tilde{\nu}\left[\frac{R_+}{\sigma_+\sqrt{2\pi}} e^{-\frac{(\tilde{\nu}-\Omega_+)^2}{2\sigma_+^2}} + \frac{R_-}{\sigma_-\sqrt{2\pi}} e^{-\frac{(\tilde{\nu}-\Omega_-)^2}{2\sigma_-^2}} \right] \tag{20}$$

Finally, we have to account for the fact that the central residue adopts more than a single conformation. In fact, it samples different regions of the Ramachandran plot. Hence, the final ε and $\Delta\varepsilon$ profiles can only be obtained after the following conformational averaging:

$$\langle \zeta(\tilde{\nu}) \rangle = \frac{\int_{-\pi}^{\pi} \zeta(\tilde{\nu}, \varphi, \psi) F(\varphi, \psi) d\varphi \, d\psi}{\int_{-\pi}^{\pi} F(\varphi, \psi) d\varphi \, d\psi} \tag{21}$$

where $\zeta = \varepsilon, \Delta\varepsilon$. The distribution function is written as:

$$F = \sum_{j=1}^{n} \frac{\chi_j}{2\pi\sqrt{|\hat{V}_j|}} \exp\left[-0.5 \cdot \left(\vec{\rho} - \vec{\rho}_j^0\right)^T \hat{V}_j^{-1} \left(\vec{\rho} - \vec{\rho}_j^0\right) \right] \tag{22}$$
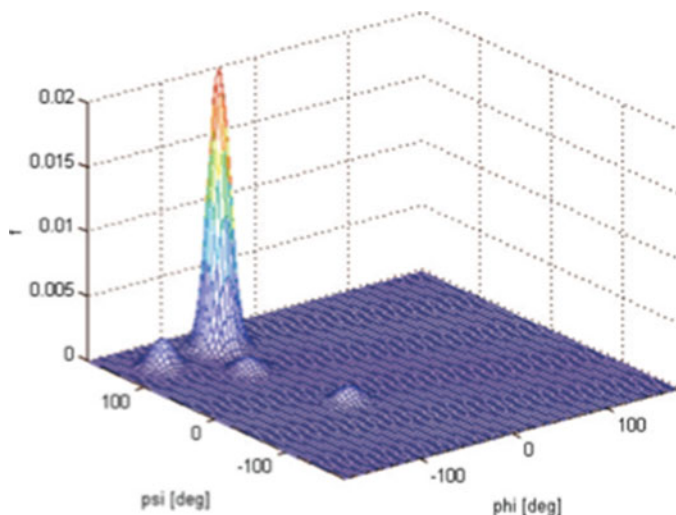
Fig. 22. Distribution of backbone conformations of the central residue used to fit the VCD and IR amide I profile of AAA in D_2O.

which is a superposition of two-dimensional Gaussian distributions with respect to ψ and φ weighted with the mole fractions χ_j of the respective sub-ensemble. The number of different sub-ensembles is denoted as n. The vector $\vec{\rho}$ is the position vector in the φ, ψ space:

$$\vec{\rho} = \begin{pmatrix} \varphi \\ \psi \end{pmatrix} \tag{23}$$

The corresponding vector $\vec{\rho}_j^0$ denotes the position of the maximum of the j-th distribution. The matrix $\hat{V}_j$ contains the information about the width of the j-th distribution:

$$\hat{V}_j = \begin{pmatrix} \sigma_{\varphi,j} & \sigma_{\varphi\psi,j} \\ \sigma_{\psi\varphi,j} & \sigma_{\psi,j} \end{pmatrix} \tag{24}$$

The diagonal elements of the matrix $\hat{V}_j$ are the half-halfwidths of the j-th distribution along the coordinates φ and ψ, and the corresponding off-diagonal element $\sigma_{\varphi\psi,j} = \sigma_{\psi\varphi,j}$ reflects correlations between variations along the two coordinates. If $\hat{V}_j$ is diagonal, the φ,ψ projection of the distribution is an ellipse with its main axis parallel to the φ and ψ axes. Correlation effects rotate the ellipse in the (φ,ψ) plane.

Now all the tools are in place to actually calculate the amide I′ profiles. Technically this can be done by programming the algorithm presented in this chapter into MATLAB or Mathematica (see Notes 26 and 27). Figure 22 shows the distribution which can be used to simulate the amide I′ profiles of amide I′ of zwitterionic trialanine in D_2O. It is dominated by a polyproline distribution centered at $\varphi_1^\circ = -69°$ and $\psi_1^\circ = 140°$ ($\chi_1 = 0.84$). The remaining fraction of the distribution contains β-strand ($\varphi_2^\circ = -136°$ and
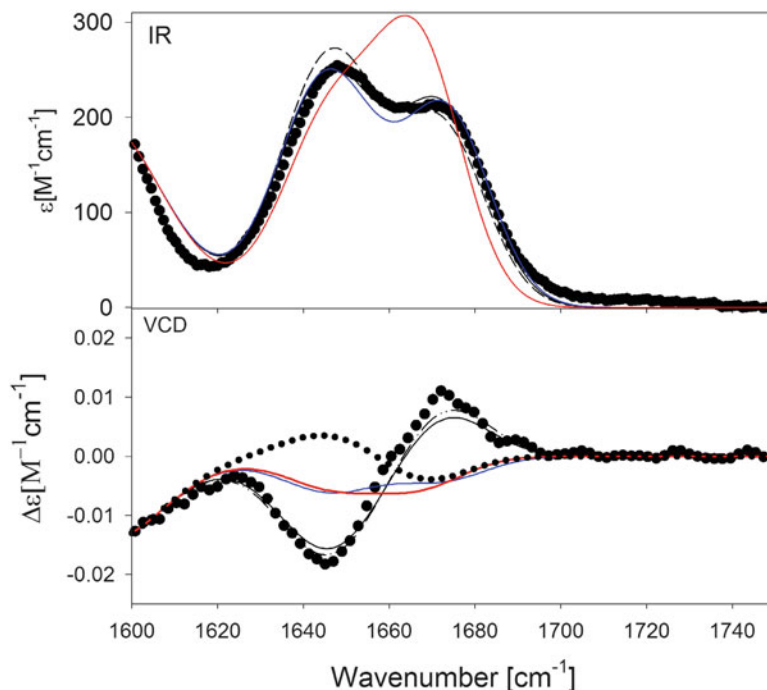
Fig. 23. Simulation of the IR and VCD amide I′ band profile of AAA in D$_2$O with (IR) the conformational distribution visualized in Fig. 22 (*solid line*) and distributions (VCD) with 100% PPII conformers (*dashed line*), 100% β-strand (*black*), and 100% right-handed helical (*gray*). The VCD spectra were all calculated assuming that the C-terminal amide I vibration is associated with a magnetic transition dipole moment of 5 × 10^{-24} esu cm. For comparison, the symmetric VCD (*dot-dot*) signal of a distribution containing only right-handed helical conformations was calculated without assuming an intrinsic magnetic moment.

$\psi°_2 = 132°$, $\chi_2 = 0.08$), right-handed helical conformations ($\varphi°_3 = -60°$ and $\psi°_3 = -32°$, $\chi_3 = 0.04$) and inverse γ-conformations ($\varphi°_3 = -85°$ and $\psi°_3 = -78°$, $\chi_3 = 0.04$).

The structural sensitivity of the two considered amide I′ profiles is illustrated in Fig. 23, which exhibits IR and VCD profiles of a tripeptide for a sample with 100% PPII, β-strand, and right helical conformations. Conformation means in this context a single two-dimensional Gaussian function, which is assignable to a minimum in the Gibbs energy space. Since PPII and β-strand are extended conformations (though to a different extent) in the upper left quadrant of the Ramachandran plot the IR profiles of these mixtures are very similar, but the β-strand gives rise to a more pronounced asymmetry in that the low wavenumber band becomes relatively more pronounced in this conformation. However, these profiles are qualitatively different from amide I′ profiles of helical structure for which the high wavenumber band is more intense. This reflects the fact that the angle between transition dipole moments is smaller than 90° for helical

conformations while it is substantially larger than 90° for β-strands (c.f. Fig. 20). If only a single conformation is present, the intensity ratio $R = \varepsilon_-/\varepsilon_+$ can be calculated by the following simple equation (23):

$$R = \frac{1 - \cos\theta \cdot \sin\nu}{\cos\theta \cdot \sin 2\nu + 1} \qquad (25)$$

where θ is the angle between the two transition dipole moments. For angles larger than 90°, the trigonometric term becomes negative. As a consequence, the numerator increases and the denominator decreases, both contributing to an increase of R.

As indicated above, analyzing and simulating the IR and VCD band profiles is only one part of the conformational analysis process. In the next chapter of this book, we outline how polarized Raman spectroscopy can be used in addition for determining structural distributions of peptides.

Thus far, we have formulated the theory for a tripeptide with two oscillators, since this is the system we will predominantly focus on in this chapter. However, an extension to longer peptides is straightforward. In this case, the Hamilton matrix has to contain contributions from non-nearest neighbor coupling:

$$H = \begin{pmatrix} \hat{H}_1 & \hat{H}_{12} & \hat{H}_{12} & . & . & . & . & . & \hat{H}_{1n} \\ \hat{H}_{21} & \hat{H}_2 & \hat{H}_{23} & \hat{H}_{23} & . & . & . & . & \hat{H}2n \\ \hat{H}_{31} & \hat{H}_{32} & \hat{H}_3 & \hat{H}_{34} & \hat{H}_{45} & & & & . \\ . & & & . & & & & & . \\ . & & & & . & & & & . \\ . & & & & & . & & & . \\ . & & & & & & . & & . \\ . & & & & & & & . & . \\ . & & & & & & & . & \hat{H}_{n-1n} \\ \hat{H}_{n1} & . & . & . & . & . & . & \hat{H}_{nn-1} & \hat{H}_n \end{pmatrix} \qquad (26)$$

For extended peptides (i.e., dihedral angles in the upper left quadrant of the Ramachandran plot) generally only nearest neighbor and second neighbor interactions are taken into account. For helical and turn conformations, interactions with the third and even the fourth neighbor might be non-negligible (24, 38). For β-sheets, nonlocal interactions enter the Hamiltonian, which can be comparatively strong for juxtaposed oscillators in neighboring strands (39–42). Generally, all non-nearest neighbor interactions can be modeled by transition dipole coupling, i.e., (25).

$$\hat{H}_{ij} = \left[\frac{\vec{\mu}'_i \cdot \vec{\mu}'_j}{\left|\vec{R}_{ij}\right|^3} - 3\frac{\left(\vec{\mu}'_i \cdot \vec{R}_{ij}\right)\left(\vec{\mu}'_j \cdot \vec{R}_{ij}\right)}{\left|\vec{R}_{ij}\right|^5} \right] Q_i Q_j \qquad (27)$$

where $\vec{R}_{ij}$ is the distance vector from dipole moment i to dipole moment j. The Hamiltonian (Eq. 26) can be diagonalized

numerically to obtain excitonic eigenfunctions and eigenenergies of the polypeptide. The former are written as the linear combination:

$$\psi_i = \sum_j \alpha_{ij} |1_j\rangle \prod_{\substack{k=1 \\ k \neq j}}^{N-1} |0_k\rangle \tag{28}$$

where we assumed that the total wavefunction of a system of uncoupled oscillators can be written as a product of oscillator wavefunctions. The eigenvector coefficients α_{ij} can be computationally obtained from the diagonalization of the Hamiltonian in Eq. 24. The summation runs over $N-1$ residues, if the peptide is unblocked, since the C-terminal residue is not associated with a peptide group. If the peptide is blocked at the N-terminal, the number of residues and amide I modes are identical, in this case the sum runs from 1 to N. If both sites are blocked, the sum runs from 1 to $N+1$. The coefficients α_{ij} can be obtained from a diagonalization of the Hamiltonian in Eq. 26. The MATLAB software offers a specific command for this purpose.

With respect to the IR band profile, the transition dipole moment $\langle \vec{\mu}'_{ex,i} \rangle$ for transitions into the i-th excitonic states is written as:

$$\langle \vec{\mu}'_{ex,i} \rangle = \sum_{j=1}^{N} \alpha_{ij} \langle \vec{\mu}'_j \rangle \tag{29}$$

In order to carry out this summation, all dipole moments of the polypeptide have to be transferred into the same common coordinate system at the C-terminal. The respective IR profile is then calculated as:

$$\varepsilon(\tilde{\nu}) = \frac{A_\varepsilon \bar{\bar{\tilde{\nu}}}}{\sqrt{2\pi}} \sum_{i=1}^{N-1} \frac{\langle \vec{\mu}_{ex,i} \rangle}{\sigma_i} \exp\left(-\frac{(\tilde{\nu} - \Omega_i)^2}{2\sigma_i^2} \right) \tag{30}$$

Generally, we assume the same half-halfwidth of ca. 11 cm^{-1} for all transitions. This parameter can slightly be changed to fine tune the fitting to the experimental data (see Notes 28 and 29 concerning the use of the correct transition dipole moment and the determination of the peptide concentration the sample, which is necessary to obtain correct ε-values).

The calculation of the VCD signal is less straightforward. For a chain with N oscillators, the rotational strength for the transition into the i-th excitonic state is written as (43):

$$R_i = \mathrm{Im}\left[\sum_{j=1}^{N-1} a_{ij} \langle \vec{\mu}'_j \rangle \sum_{k=1}^{N-1} a_{ik} \langle \vec{m}'_k \rangle - \frac{i\pi}{2} \left(\sum_{j=1}^{N-2} \sum_{k=2}^{N-1} \tilde{\nu}_{jk} \vec{R}_{jm} \cdot a_{ij} a_{ik} \langle \vec{\mu}'_j \rangle \times \langle \vec{\mu}'_k \rangle \right) \right] \tag{31}$$

Here, the product of $\left\langle \vec{\mu}'_j \right\rangle$ with $\langle \vec{m}'_k \rangle$ describes the intrinsic rotational strengths as well as contributions resulting from mixing electronic and magnetic transition dipole moments from different residues due to excitonic coupling (11). The multiplication of $\left\langle \vec{\mu}'_j \right\rangle$ with the last term on the right hand side yields terms which are proportional to the cross product of the electronic transition dipole moments associated with the j-th and the l-th (m-th) amide I oscillator. For a system with only two oscillators, this yields the last term in Eq. 14.

In order to calculate the distance $\vec{R}_{jk}$ between the j-th and the k-th oscillator, we have to proceed as follows. First, we use the formalism given by Eqs. 16–19 to calculate the distance between the dipole moments of the j-th and the $j + 1$-th residue. Subsequently, we apply:

$$\vec{R}'_{jj+1} = \vec{R}^T_{jj+1} \cdot R(-\varphi) \cdot R(-\xi) \cdot R(-\psi) \cdot R(-\eta) \cdot R(\tau) \quad (32)$$

which transforms this distance vector from the coordinate system of the j-th residue into that of the $j + 1$-th residue. Next, one calculates the distance between the $j + 1$-th and the $j + 2$-th residue and adds it to $\vec{R}_{jj+1}$. The resulting vector must then be transformed into the coordinate system of the $j + 2$-th residue. This procedure must be repeated until the C-terminal amide I oscillator has been reached.

Here, we demonstrate how this formalism has recently been used to simulate the VCD and IR profiles of the amide I mode of the hexapeptide A_5W (Fig. 23) (44). This peptide has five amide I oscillators. The calculation was carried out as follows. The central three amide I′ modes were assumed to all exhibit a local wavenumber of 1,652.5 cm^{-1}. End effects were considered for the terminal modes, which yields 1,648 cm^{-1} for the C-terminal and 1,675 cm^{-1} for the N-terminal mode. The transition dipole moment is 2.7×10^{-19} esu cm, as indicated above (31). The best simulation was obtained with a distribution model which considered contributions from polyproline II, β-strand, right-handed helical and inverse γ-turn-like conformations. To reduce the computer time for individual simulations, the respective Gaussian distributions were substituted by set of representative values at the position of the distribution maximum and its halfwidths along the φ and ψ coordinate. Hence, each Gaussian was represented by five points in the Ramachandran space. The result of the best simulation is shown in Fig. 24 The respective coordinates and the statistical weights of the utilized sub-distributions are reported by Verbaro et al. (44). It should be mentioned that the fine-tuning of the simulation was achieved by simultaneously reproducing the $^3J_{NHC\alpha H}$ coupling constants of the residues. This enabled us to differentiate between the populations of individual residues. The statistical weights in Table 2 suggest that the central alanine residues show a somewhat reduced PPII propensity, which further decreases in the vicinity of W, most likely due to nearest neighbor interactions.
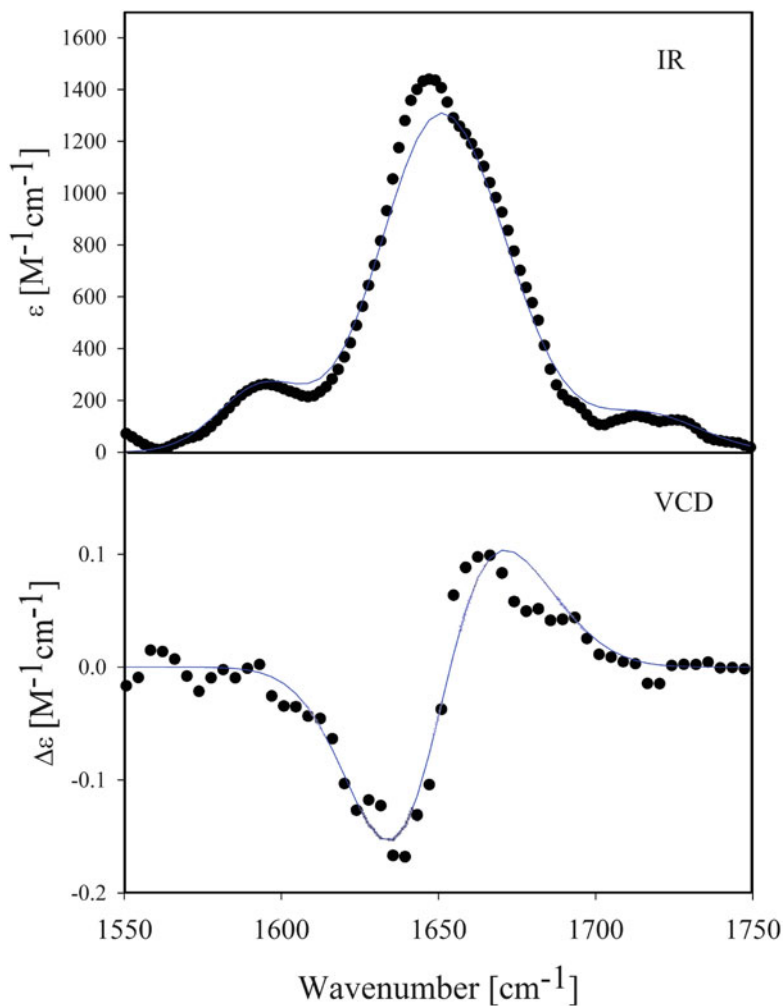
Fig. 24. FT-IR and VCD amide I′ profile of AAAAW in D_2O. The solid line results from a simulation described in the text (44).

Table 2
Fraction of conformations sampled by the indicated amino acid residues A_5W in D_2O as derived from the amide I′ profiles

	PPII $-70°$, $152°$	β-Strand $-120°$, $130°$	Helical $-60°$, $60°$	γ $-70°$, $53°$
A2	0.65	0.03	0.22	0.05
A3	0.65	0.03	0.2	0.06
A4	0.75	0.03	0.1	0.06
A5	0.75	0.14	0.05	0.03
W	0.7	0.3	0	0

Taken together this paragraph outlines an algorithm which can be used to calculate the IR and VCD profiles of amide I for polyeptides. A program using the equations presented herein can easily be written in MATLAB. Our current version is in principle available free of charge, but it is not very user friendly. We are planning some upgrading in this regard for the near future.

4. Notes

1. The most common sources are AnaSpec (Fremont, CA), BaChem (Torrance, CA), Celtek Bioscience (Nashville, TN), or a source used for isotopically labeled peptides is Cambridge Isotope Labs (Andover, MA).

2. The amount of the correct peptide content relative to all of the analytes present is the peptide purity. Possible impurities found in the content can include incompletely deprotected sequences, deletion sequences, truncated sequences, etc. This is determined through HPLC analysis with the detector set so the wavelength at which the peptide bond absorbs, 214 nm. The peptide purity does not take into account the water and salts that are present in the sample. The other factor that goes into the consideration of the absolute amount of a correct peptide is the peptide content. This is the percentage of the total amount of peptide that is present relative to everything else that might be present in the synthesis, which is determined through amino acid analysis.

3. The peptides are shipped at room temperature in sealed vials. For long-term storage, the peptide should preferably be stored at $-20°C$ or less. If possible, the peptide should also be in a sealed container in the presence of a desiccant; this will help to prevent oxidation, bacterial degradation, and the possibility of secondary structure formation. Once the peptide is removed from the freezer it is best to allow it to equilibrate to room temperature in the presence of the desiccant prior to opening and weighing of the sample. If not carefully performed, as the peptide warms there is the likeliness of condensate formation upon the opening of the vial, as peptides tend to be hygroscopic. This in effect will reduce the stability of the peptide under investigation.

4. The act of getting the rubber band onto the vial is something that is mastered over time. If one is new to the procedure, one may want to hold the entire Kimwipe over the sample vial and cut the excess off once the rubber band is in place.

5. One should be extremely careful as this process initially will cause the nitrogen to boil, sometimes causing the vial to pop out of the holder. It is best to use the wall of the dewar to inhibit the sample from coming out of the holder and dropping

into the liquid nitrogen. Proper safety precautions should be utilized when handling liquid nitrogen, cold temperature protective gloves and eyewear is necessary at all times.

6. This process ensures that the Cl$^-$ ions will exchange the acetate ions, which will then be removed with the solvent during the freeze-drying process.

7. D$_2$O is used instead of H$_2$O to remove the bending contributions to the amide I, this enables the IR and VCD to be obtained.

8. If there is still debris that is difficult to remove a small amount of ethanol can be used, but over time this will remove the polished finish of the cell. As an extreme last resort extremely dilute nitric acid can be used (2 drops in water) to clean the cell. One should keep in mind that CaF$_2$ is very slightly soluble in water (0.0016 g/100 mL H$_2$O at 20°C), and is insoluble in more acids and bases. It is soluble in ammonium salts (NH$_4$), and slightly soluble in dilute mineral acids.

9. The use of compressed air on the cell may cause the cell to crack. Avoiding anything that is extremely harsh dramatically prolongs the lifetime of the fragile cell.

10. This volume is dependent upon the pathlength chosen, for example a 100 μm cell can hold around 125 μL, and a 56 μm cell hold around 60 μL. From this information one can estimate the volume that any other cell can hold based on the pathlength chosen.

11. The center of the plate is recessed with a slight groove around the outer edge. This groove acts to both establish the pathlength as well as to help keep the sample from coming into contact with the seal that is created.

12. An alternate method is placing the lower plate into the palm of a hand and the top plate into the curl of fingers and pressing down. The act of filling a sample cell is something that takes experience to master, as it may take a few tries to ensure that there are no air bubbles that become trapped in the cavity as you press down.

13. If performing a temperature-dependent study, do not shock the CaF$_2$ cell with any large temperature changes, as it will crack the fragile salt plate. Handling of the cells and changing the conditions should always be performed gradually.

14. At the start of filling the dewar, you will hear the nitrogen boil off as the temperature of the detector cools; once you do not hear this sound any longer add more nitrogen to the dewar. As you continue this process, you will see a plume of nitrogen gas leaving the opening. Continue to fill the dewar until this plume is minimized, this should be just over 400 mL of liquid

nitrogen. Once full the dewar should have an 8 h hold time, for longer experiment you will have to refill the dewar using the same method.

15. The nitrogen gas enters the instrument at the point of the source to cool the instrument and prevent degradation of the housed optics.

16. Apply Beer–Lambert's law to optimize the system conditions; it is useful to have various path lengths available to fine-tune the intensity without sample concentration changes.

17. There are three preset times of 20, 40, and 60 min as well as a User-Defined setting in which you can add custom lengths of time for the duration of the acquisition. In order to increase the VCD signal of short peptide, one should try to maximize the time for data acquisition, using 12 h (720 min) as the standard acquisition time, filling the N_2-cooled dewar one additional time through the course of the experiment. The instrument scans in blocks of time, in minutes, in which the instrument will alternate between AC/DC scanning. Overall the instrument will dedicate 10% of the time to the DC scan and 90% of the time in the AC scan mode.

18. The phase files are aligned to each source; in selecting the correct one, one will see the extension PHA or PHB which correlates to source A and B, respectively. These items remain as the default items until the user changes the files selected.

19. For non-aggregated peptides, the VCD signal is four to six orders of magnitude smaller than the normal IR signal, so much greater care must be taken in the correction of the VCD signal. This must be done manually with MultiFit.

20. If one is not comfortable with MultiFit, this can easily be performed in Excel by subtracting the background from the sample.

21. MultiFit does not read any headers that are in the file and the numbers must be formatted with all columns having uniform significant figures. Also, if there is a space at the very end of the table, an error message will be displayed.

22. The MultiFit program was written in German, but it can easily be operated without a command of this language. A copy can be obtained from the authors free of charge. The program is only compatible with the Windows operating system.

23. We would like to emphasize here that a Voigtian profile is not just a sum of a Gaussian and a Lorentzian profile, an option frequently used and even incorporated in commercial fitting programs like Grams. Strictly speaking, such a profile has no physical basis and should not be used if one aims at obtaining correct spectral parameters. Other options of this program

(convolution with a spectrometer function) are more useful for experiments with diffraction spectrometers and will therefore be discussed in the section on Raman spectroscopy.

24. Alternatively, the program offers multiple options to subtract baselines prior to the fitting procedure. They can be obtained by clicking on "baseline", the options here are self-explanatory.

25. More details about MultiFit can be found in a short user manual, which will become available soon.

26. In order to construct the program efficiently, the distribution function, intrapeptide distances, coupling constants, and amide I profiles should be programmed in subroutines. It is advisable to store the main spectra and distributional parameters also in subroutines. The main program just calls the parameters, which are transferred to a subroutine for profile calculations. This subroutine is imbedded in a loop so that the profiles can be calculated for all dihedral angles sampled in the calculation. Generally, we sample the Ramachandran plot from $-180°$ to $180°$ in increments of $2°$ so that the entire distribution function contains 181×181 points. This binning is sufficient even for the PPII region in which the nearest neighbor coupling changes significantly even with small changes of dihedral angles. The subroutine calculating the amide I profiles calls sub-subroutines for calculating the respective values of the distribution function, nearest neighbor coupling, transition dipole coupling between second, third, etc., nearest neighbors, and distances between respective atoms of residues. Finally, the profiles obtained for all coordinates are superimposed and the resulting profile is plotted on the screen.

27. After having programmed, the algorithm several tests should be conducted. First of the contour plot for the nearest neighbor coupling and the total distribution function for various mixtures of sub-distributions should be visualized to check whether they were calculated correctly. Second, the distance between corresponding peptide group atoms should be calculated for a variety of single φ,ψ-values and compared with distances obtained from structures produced with molecular modeling programs. Third, the angle between transition dipole moments in adjacent residues should be calculated for single φ,ψ-values and compared with corresponding values obtained from published contour maps (c.f. Fig. 19). The same should be carried out for angles between normals of peptide linkages. Finally, one should check whether the program reproduces the following general features of amide I profiles. For extended structures with dihedral angles in the upper left quadrant of the Ramachandran plot, the first moment of the IR band profile, i.e.,

$$\vec{v}_1 = \frac{\int\limits_{-\infty}^{\infty} \tilde{v}\varepsilon(\tilde{v})d\tilde{v}}{\int\limits_{-\infty}^{\infty} \varepsilon(\tilde{v})d\tilde{v}} \tag{33}$$

should be higher than the peak wavenumber. The amide I VCD should be a negative couplet (negative signal at lower wavenumbers), which increases with increasing statistical weight for PPII. For right-handed helical conformations, the first moment of the band profile should be lower than the peak wavenumber and the VCD couplet should be positive. Switching from right-handed to a left-handed structure should just invert the VCD signal but should not change the IR band profile.

28. Care must be taken to determine the concentration of the peptide as accurately as possible. This is easy if the peptide contains an aromatic side chains like tryptophan for which the extinction coefficient is known. Alternatively, one can measure NMR spectra, for example three different concentrations determined from weighing the solute prior to dissolving it in water. If the real peptide concentrations are identical with those determined with the balance, the extrapolated plots of the NMR intensities versus concentration should go through the zero-point. Any offset indicates a systematic error, which can then be corrected for.

29. Normally, we take the transition dipole moments from the work of Measey et al. (31), who reported values for several AX and XA dipeptides in D_2O. Very often these values needed only minor changes to account for magnitude of the amide I signals. Since variations of the transition dipole moment affect both, IR and VCD band profiles in a similar way, such adjustments do not affect the validity of the simulation.

References

1. Keiderling TA (1996) Vibrational circular dichroism: application to conformational analysis of biomolecules. In: Fasman GD (ed) Circular dichroism and the conformational analysis of biomolecules. Plenumn, New York, p 555
2. Keiderling TA, Xu Q (2002) Unfolded proteins studied with IR and VCD spectra. Adv Protein Chem 62:111–161
3. Sutherland JC (2009) Measurement of circular dichroism and related spectroscopies with conventional and synchotron light sources: theory and instrumentation. In: Wallace BA, Janes RW (eds) Modern techniques for circular dichroism spectroscopy. IOS, Amsterdam, pp 19–73
4. Nafie LA (1997) Infrared and Raman vibrational optical activity. Annu Rev Phys Chem 48:357
5. Nafie LA, Dukor RK, Freedman TB (2002) Dichroism and optical activity in optical spectroscopy. In: Chalmers J, Griffiths P (eds) Handbook of vibrational spectroscopy. Wiley, New York, pp 731–744
6. Chen XG, Schweitzer-Stenner R, Asher SA, Mirkin NG, Krimm S (1995) Vibrational assignments of trans-N-methylacetamide and some of its deuterated isotopomers from band decomposition of IR, visible, and resonance Raman spectra. J Phys Chem 99:3074–3083

7. Han W-G, Jakanen KJ, Elstner M, Suhai S (1998) Theoretical study of aqueous N-Acetyl-L-alanine N-methylamide: structures and Raman, VCD, and ROA spectra. J Phys Chem B 102:2587–2602

8. Schweitzer-Stenner R, Gonzales W, Bourne JT, Feng JA, Marshall GA (2007) Conformational manifold of α-aminoisobutyric acid (Aib) containing alanine-based tripeptides in aqueous solution explored by vibrational spectroscopy, electronic circular dichroism spectroscopy, and molecular dynamics simulations. J Am Chem Soc 129:13095–13109

9. Bouŕ P, Keiderling TA (2005) Vibrational spectral simulation for peptides of mixed secondary structure: method comparisons with the Trpzip model hairpin. J Phys Chem B 123:23687–23697

10. Tanaka S, Scheraga HA (1976) Statistical mechanical treatment of protein conformation. II. A three-state model for specific-sequence copolymers of amino acids. Macromolecules 9:150–167

11. Schweitzer-Stenner R (2006) Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure. A critical review. Vib Spectrosc 42:98–117

12. Schweitzer-Stenner R (2009) Distribution of conformations sampled by the central amino acid residue in Tripeptides inferred from amide I band profiles and NMR scalar coupling constants. J Phys Chem B 113:2922–2932

13. Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R (2010) Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I band profiles and NMR scalar coupling constants. J Am Chem Soc 132:542

14. Chicz RM, Regnier FE (1990) High-performance liquid chromatography: effective protein purification by various chromatographic modes. Methods Enzymol 182:392–421

15. Ingle JD, Crouch SR (1988) Infrared spectrometry. In: Spectrochemical analysis. Prentice Hall, Upper Saddle River, NJ, p 404–437

16. Skoog DA, Holler FJ, Nieman TA (1998) An introduction to infrared spectrometry. In: Principles of Instrumental Analysis, 5th ed. Harcourt Brace and Company, Philadelphia, USA, p 380–428

17. (2000) Model SR810 DSP lock-in amplifier, 1.6 ed. Stanford Research Systems, Sunnyvale, CA

18. (1999) Series SR640 Dual-channel filters, 2.6 ed. Stanford Research Systems, Inc, Sunnyvale, CA

19. Graf J, Nguyen PH, Stock G, Schwalbe H (2007) Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/nmr study. J Am Chem Soc 129:1179–1189

20. Jentzen W, Unger E, Karvounis G, Shelnutt JA, Dreybrodt W, Schweitzer-Stenner R (1995) Conformational properties of nickel (II) octaethylporphyrin in solution. 1. Resonance excitation profiles and temperature dependence of structure-sensitive Raman lines. J Phys Chem 100:14184–14191

21. Woutersen S, Hamm P (2000) Structure determination of trialanine in water using polarized sensitive two-dimensional vibrational spectroscopy. J Phys Chem B 104:11316–11320

22. Eker F, Cao X, Nafie L, Schweitzer-Stenner R (2002) Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR and VCD spectroscopy study. J Am Chem Soc 124:14330–14341

23. Schweitzer-Stenner R (2002) Dihedral angles of tripeptides in solution determined by polarized Raman and FTIR spectroscopy. Biophys J 83:523–532

24. Torii H, Tasumi M (1998) Ab initio molecular orbital study of the amide i vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. J Raman Spectrosc 29:81–86

25. Krimm S, Bandekar J (1986) Vibrational spectroscopy of peptides and proteins. Adv Protein Chem 38:181

26. Woutersen S, Hamm P (2001) Isotope-edited two-dimensional vibrational spectroscopy of trialanine in aqueous solution. J Chem Phys 114:2727–2737

27. Gorbunov RD, Kosov DS, Stock G (2005) *Ab initio*-based exciton model of amide I vibrations in peptides: definition, conformational dependence and transferability. J Chem Phys 122:224904–224915

28. Ham S, Cha S, Choi J-H, Cho M (2003) Amide I modes of tripeptides: Hessian matrix reconstruction and isotope effects. J Chem Phys 119:1452–1461

29. Ham S, Cho M (2003) Amide I modes in the N-methylacetamide dimer and glycine dipeptide analog: diagonal force constants. J Chem Phys 118:6915–6922

30. Gorbunov RD, Nguyen PH, Kobus M, Stock G (2007) Quantum-classical description of the amide I vibrational spectrum of trialanine. J Chem Phys 126:054509

31. Measey T, Hagarman A, Eker F, Griebenow K, Schweitzer-Stenner R (2005) Side chain dependence of intensity and wavenumber position of amide I′ in IR and visible Raman spectra

of XA and AX dipeptides. J Phys Chem B 109:8195–8205

32. Eker F, Cao X, Nafie L, Griebenow K, Schweitzer-Stenner R (2003) The structure of alanine based tripeptides in water and dimethyl sulfoxide probed by vibrational spectroscopy. J Phys Chem B 107:358–365

33. Eker F, Griebenow K, Schweitzer-Stenner R (2004) A β_{1-28} fragment of the amyloid peptide predominantly adopts a polyproline II conformation in an acidic solution. Biochemistry 43:6893–6898

34. Schweitzer-Stenner R, Measey T, Hagarman A, Eker F, Griebenow K (2006) Salmon calcitonin an amyloid β: two peptides with amyloidogenic capacity adopt different conformational manifolds in their unfolded states. Biochemistry 45:2810–2819

35. Schweitzer-Stenner R, Measey T, Kakalis L, Jordan F, Pizzanelli S, Forte C, Griebenow K (2007) Conformations of alanine-based peptides in water probed by FTIR, Raman, vibrational circular dichroism, electronic circular dichroism, and NMR spectroscopy. Biochemistry 46:1587–1596

36. Schweitzer-Stenner R, Eker F, Perez A, Griebenow K, Cao X, Nafie LA (2003) The structure of tri-proline in water probed probed by polarized Raman, Fourier transform infrared, vibrational circular dichroism and electronic ultraviolet circular dichroism spectroscopy. Biopolymers 71:558–568

37. Holzwarth G, Chabay I (1972) Optical activity of vibrational transitions: a coupled oscillator model. J Chem Phys 57:1632–1638

38. Huang Q, Schweitzer-Stenner R (2004) Conformational analysis of tetrapeptides by exploiting the excitonic coupling between amide I modes. J Raman Spectrosc 53:586–591

39. Lee C, Cho MH (2004) Local amide I mode frequencies and coupling constants in multiple-stranded antiparallel beta-sheet polypeptides. J Phys Chem B 108:20397–20407

40. Kubelka J, Keiderling TA (2001) Differentiation of β-sheet-forming structures: Ab initio-based simulations of IR absorption and vibrational CD for model peptide and protein β-sheets. J Am Chem Soc 123:12048–12058

41. Kubelka J, Keiderling TA (2001) The anomalous infrared amide I intensity distribution in ^{13}C isotopically labeled peptide β-sheets comes from extended, multiple –stranded structures. An ab initio study. J Am Chem Soc 123:6142–6150

42. Schweitzer-Stenner R, Measey TJ (2010) Simulation of IR, Raman and VCD amide I band profiles of self-assembled peptides. Spectroscopy 24:25–36

43. Schweitzer-Stenner R (2004) Secondary structure analysis of polypeptides based on an excitonic coupling model to describe the band profile of amide I of IR, Raman and vibrational circular dichroism spectra. J Phys Chem B 108:16965–16975

44. Verbaro D, Ghosh I, Nau WM, Schweitzer-Stenner R (2010) Discrepancies between conformational distributions of a polyalanine peptide in solution obtained from molecular dynamics force fields and amide I′ band profiles. J Phys Chem B 114:17201–17208

45. Schweitzer-Stenner R, Measey TJ, Hagarman A, Dragomir I (2010) The structure of unfolded peptides and proteins explored by Raman and IR spectroscopies. In: Assessing Structures and Conformation of Intrinsically Disordered Proteins", Editors: S. Longhi and V.N. Uversky. Wiley & Sons, 171–224:2012

Structural Analysis of Unfolded Peptides by Raman Spectroscopy

Reinhard Schweitzer-Stenner, Jonathan B. Soffer, Siobhan Toal, and Daniel Verbaro

Abstract

Raman spectroscopy has positioned itself as an invaluable tool in the study of complex biological systems, consistently being used to obtain information illustrating a vast array of fundamental properties. Of primary interest, with respect to the focus of this chapter, are conformational changes of peptide backbones. For short peptides to larger biological systems this understanding can be extended to local hydrogen bonding interactions and the probing of other structural or organizational properties. With regard to unfolded peptides Raman spectroscopy can be used as a technique complementary to infrared (IR) and vibrational circular dichroism (VCD) spectroscopy. This chapter describes how high quality polarized Raman spectra of peptide can be recorded with a Raman microspectrometer and how the structure sensitive amide I band profiles of isotropic and anisotropic Raman scattering can be analyzed in conjunction with the respective IR and VCD profiles to obtain conformational distributions of short unfolded peptides.

Key words: Raman, Spectral decomposition, Nonresonance, Amide I, Amide III, Trialanine

1. Introduction

Raman spectroscopy has developed as a valuable tool for exploring the structure and dynamics of biomolecules over the last 40 years [1]. While resonance Raman, which involves the excitation of molecules within the wavenumber range of optical absorption of chromophores, is very frequently utilized to probe site-specific structural changes, nonresonance Raman spectroscopy can still be used to explore the secondary structure changes of proteins and peptides in a similar way that is most commonly achieved through IR spectroscopy, i.e., measuring and analyzing the band profiles of amide backbone modes [2], which have been introduced in the preceding chapter [3]. Raman spectroscopy has the unique advantage
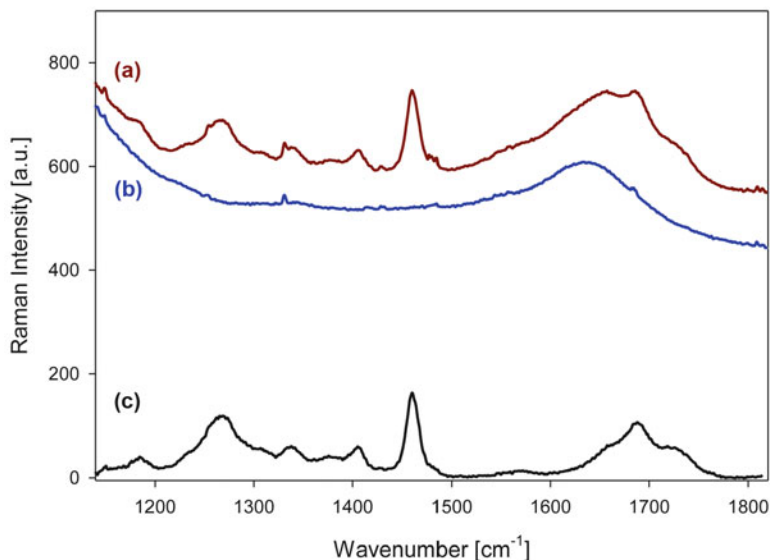
Fig. 1. The amide I profile of trialanine in H_2O (**a**) with water background (**b**) subtraction and (**c**) after water background subtraction.

that when compared with the intensity of amide I the relative intensity of the band arising from the HOH bending modes is much less pronounced than the corresponding band in the IR spectra (4). Hence the amide I profile as found in Raman spectra of peptides in H_2O can be determined after careful subtraction of the background spectrum. This is demonstrated by Fig. 1 for trialanine. Besides amide I, bands in the spectral region between 1,200 and 1,400 cm^{-1} can be further used for structure analysis, as it will be demonstrated below. What amide I is concerned, we will primarily focus on peptides in D_2O, first for the sake of comparison with the analysis of IR and VCD profiles as described in the preceding chapter and to eliminate vibrational mixing between amide I and HOH bending modes (5, 6).

Raman scattering is produced by the induced dipole moment, $\vec{\mu}$, induced by the electric field component $\vec{E}$ of an electromagnetic field generally provided by a laser beam:

$$\vec{\mu} = \hat{\alpha}\vec{E}, \tag{1}$$

where $\hat{\alpha}$ is a second rank tensor, the so-called polarizability tensor. The expression for the dipole moment can be expanded further into higher orders of the electric field, the contributions of which are accounted for by higher rank tensors (7). In order to be significant, these contributions require rather strong electric fields, which are not provided in the experiments discussed in this chapter. Hence, we focus on the linear approach.

Equation 1 can be developed further through the Taylor expansion of the polarizability tensor with respect to the normal coordinates, Q_j, of the investigated molecule, namely:

$$\hat{\alpha} = \hat{\alpha}_0 + \sum_j \frac{\partial \hat{\alpha}}{\partial Q} Q_j \qquad (2)$$

Now, we describe the electric field and the normal coordinates by means of periodic functions:

$$Q = Q_0 \cos(\omega_j t + \delta_j). \qquad (3)$$

$$\vec{E} = E_0 \cos(\omega_e t), \qquad (4)$$

where ω_e is the frequency of electromagnetic excitation and ω_j is the frequency of the molecular vibration with which dipole oscillation and radiation are modulated. With respect to the exciting wave, each vibration exhibits a phase, δ_j. Inserting Eqs. 3 and 4 into Eq. 5 yields:

$$\begin{aligned}
\vec{\mu} &= \left[\alpha_0 + \sum_j \frac{\partial \hat{\alpha}}{\partial Q_j} Q_{j0} \cos(\omega_j t + \delta_j) \right] \vec{E}_0 \cos(\omega_e t) \\
&= \alpha_0 \vec{E}_0 \cos(\omega_e t) + \sum_j \frac{1}{2} \frac{\partial \hat{\alpha}}{\partial Q_j} Q_{j0} \vec{E}_0 \\
&\quad \times \left\{ \cos[(\omega_e - \omega_j)t + \delta_j] + \cos[(\omega_e + \omega_j)t + \delta_j] \right\}
\end{aligned} \qquad (5)$$

where we just utilized elementary rules for the multiplication of trigonometric functions. The first term of the equation describes a dipole moment, which oscillates with the frequency of the exciting electromagnetic radiation. This is generally referred to as Rayleigh scattering. The second and third term describes oscillations whose frequencies are reduced or increased by the frequency of the considered vibrations. This is called Anti-Stokes and Stokes scattering, respectively. Altogether, this is called spontaneous Raman scattering. Figure 2 shows the relative intensities of these scattering types. This name honors one of the discoverers of this effect, Chandrasekhara Venkata Raman (8).

In what follows we will term the second summand of Eq. (2), i.e.

$$\hat{\beta}_j = \frac{\partial \hat{\alpha}}{\partial Q_j} Q_j, \qquad (6)$$

the Raman tensor of the jth vibration of a molecule.

Thus far, we have not specified a molecular coordinate system for the Raman tensor. Molecular coordinate system means that it is associated with a single molecule. With respect to any laboratory system, the molecular coordinate systems of a sample constitute a randomly oriented ensemble. An infinite number of coordinate
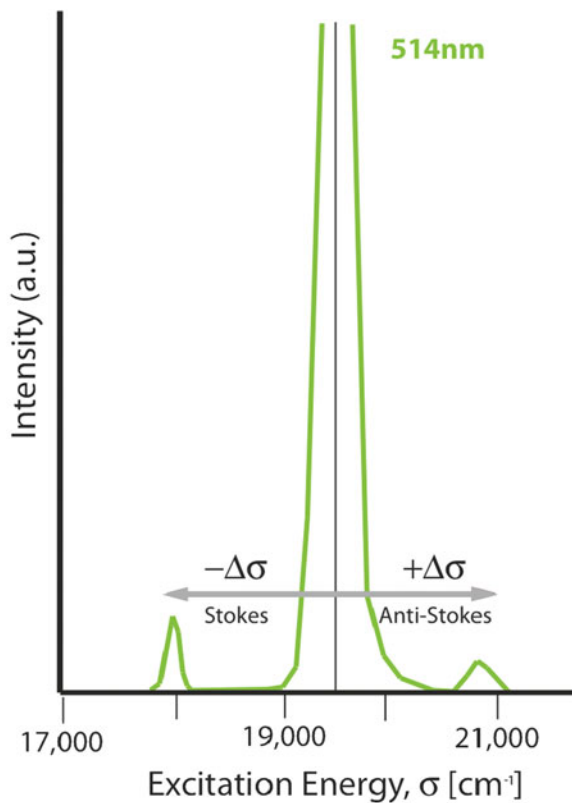
Fig. 2. Illustration of a spontaneous light scattering spectrum encompassing the intense Rayleigh line (*center*) and more weakly intense Stokes and Anti-Stokes scattering. The Anti-Stokes scattering is weaker because it depends on the thermal population of the first excited vibrational states.

systems can be envisaged for each molecule, which all could exhibit different tensor elements. However, since we do not consider resonance Raman scattering, the Raman scattering can be considered as symmetric. In this case one can always find an orthogonal transformation B, which transforms any tensor $\hat{\beta}$ into one represented by a diagonal tensor, i.e.,

$$\hat{\beta}_{\text{part}} = B^{\text{T}} \hat{\beta} B, \tag{7}$$

where *part* (principal axes of the Raman tensor) indicates that the tensor is expressed in terms of its principal axes.

Equations 6 and 7 describe the Raman tensor of a single molecule. If Raman measurements are performed on single crystals with known orientations of molecules and their *part*, a laboratory coordinate system can easily be defined and the *part* tensors of the unit cell molecules can be transformed into the
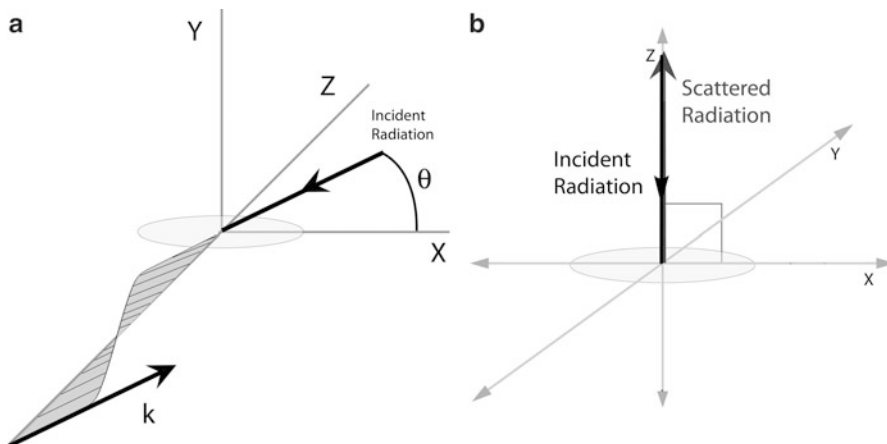
Fig. 3. Schematic representation of Raman scattering. *Left*: The scattering angle is 90°, the polarization vector of the incident light forms an angle θ with the direction of the scattered light. *Right*: Backscattering geometry of a Raman microspectrometer.

laboratory system (9). However, for molecules in solution the scattering intensity has to be calculated as an average over all possible orientations. This is a very lengthy procedure, which is described in detail in textbooks (7). Here, we will just summarize the results.

Let us assume that the electric field vector of the exciting radiation can be written as:

$$\vec{E} = \begin{pmatrix} E_0 \cos \theta \\ E_0 \sin \theta \\ 0 \end{pmatrix}, \tag{8}$$

where θ is the angle between the electric field vector of the incident light and the k-vector of the scattered light, as illustrated in Fig. 3. For the backscattering geometry, this angle is equal to $\pi/2$ irrespective of the direction of the electric field vector in the x, y-plane.

$$I_{x,j} = \frac{N_{v_j} \left[\omega_e \pm \omega_j \right]^4 E_0^2}{16 \varepsilon_0 c} \left[\beta_{s,j}^2 \sin^2 \theta + \frac{1}{45} \gamma_j^2 (3 + \sin^2 \theta) \right].$$

$$I_{y,j} = \frac{N_{v_j} \left[\omega_e \pm \omega_j \right]^4 E_0^2}{16 \varepsilon_0 c} \times \frac{1}{15} \gamma_j^2. \tag{9}$$

where N_{vJ} is the occupation number of the vth state of the Jth-oscillator. The tensor invariants, β_{sj} and γ_j^2 are the so-called tensor invariants and are written as:

$$\beta_{s,j}^2 = \frac{1}{9} Tr\left(\hat{\beta}_j\right)$$
$$\gamma_j^2 = \frac{1}{2}\left[\left(\beta_{xx,j} - \beta_{yy,j}\right)^2 + \left(\beta_{yy,j} - \beta_{zz,j}\right)^2 + \left(\beta_{zz,j} - \beta_{xx,j}\right)^2\right]$$
$$+ \frac{3}{4}\left[\left(\beta_{xy,j} + \beta_{yx,j}\right)^2 + \left(\beta_{yz,j} + \beta_{zy,j}\right)^2 + \left(\beta_{xz,j} + \beta_{zx,j}\right)^2\right].$$

$$(10)$$

$\beta_{s,j}^2$ and γ_i^2 describe the isotropic and anisotropic contribution to the Raman scattering. If all of the diagonal elements are identical in the *part*, the scattering is purely isotropic. The scattering is purely anisotropic, if, e.g., two part-elements have opposite signs but the same amount and if the third *part*-element is zero.

For reasons outlined below it is useful to individually analyze isotropic and anisotropic Raman scattering, which can be obtained from $I_{x,i}$ and $I_{y,i}$ as follows:

$$I_{\text{aniso}} = I_y.$$
$$I_{\text{iso}} = I_x - \frac{4}{3} I_y.$$

$$(11)$$

The depolarization ratio of the scattered light can thus be calculated as:

$$\rho_j(\theta) = \frac{I_{y,j}}{I_{x,j}} = \frac{3\gamma_j^2}{45\beta_{s,j}^2 \sin^2\theta + \gamma_j^2(3 + \sin^2\theta)}.$$

$$(12)$$

For $\theta = \pi/2$, the depolarization ratio reads as:

$$\rho_j(\theta) = \frac{I_{y,j}}{I_{x,j}} = \frac{3\gamma_j^2}{45\beta_{s,j}^2 + 4\gamma_j^2}.$$

$$(13)$$

This illustrates that the depolarization ratio is zero for isotropic scattering and 0.75 for pure anisotropic scattering. Generally, this can be used to identify the symmetry of modes if the molecule exhibits a high symmetry. Symmetric modes are generally more polarized than asymmetric modes (7). This is of course not the case for the modes of polypeptides. The peptide groups themselves exhibit C_s-symmetry, in which in-plane (A') and out-of-plane (A'') modes can still be distinguished. However, peptide modes generally mix with other skeletal (e.g., NC_α stretch, $C_\alpha Cs$ stretch, $C_\alpha H$ bending) and side chain modes. In this case, modes cannot longer be classified in terms of symmetry representations. However, Raman bands can still show vastly different depolarization ratios, as illustrated in Fig. 4. The bands of the peptide modes amide I, II, and III are clearly polarized ($\rho < 0.75$), whereas bands attributed to antisymmetric out-of-plane CH_3 modes and to the antisymmetric
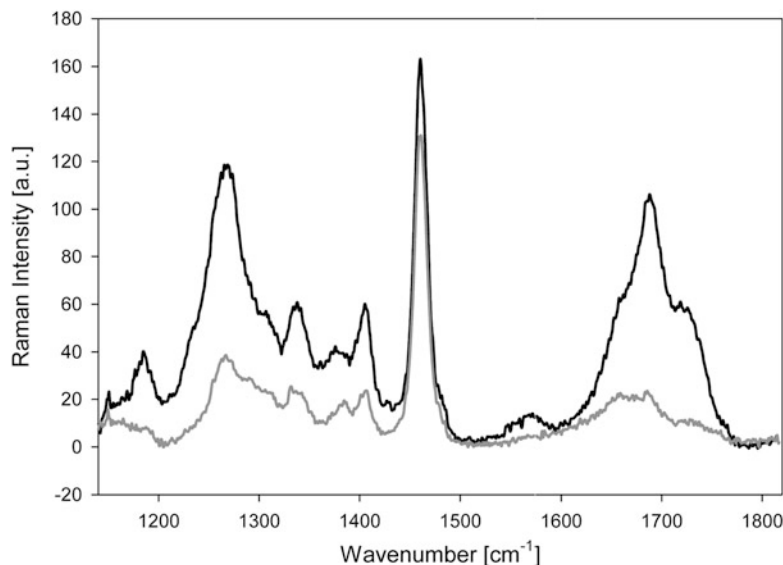
Fig. 4. Polarized Raman spectra of trialanine in H_2O between 1,200 and 1,800 cm^{-1}. The *x*-polarized spectrum is plotted in *black*, whereas the *y*-polarized spectrum is shown in *grey*.

COO^- stretch are depolarized ($0.5 < \rho < 0.75$). The latter generally shows a depolarization ratio close to 0.75. Interestingly, the depolarization ratio of amide II (1,572 cm^{-1}) and amide III (1,267 cm^{-1}) is close to 0.33, which suggests that their *part* is dominated by a single tensor element. We have argued in earlier papers that this reflects a dominant pre-resonance contribution from the far UV $\pi \rightarrow \pi^{\star}$ transition between the HOMO and LUMO of the peptide moiety to the Raman tensor (5, 10).

2. Materials

2.1. Peptide Preparation   The methods of sample preparation for use in the Raman spectrometer follows a very similar procedure to the methods mentioned in the previous chapter, Chapter 18, on vibrational circular dichroism, Subheadings 2.1 and 2.2 (3). For this method of analysis the protein must primarily be purified removing any remnants of TFA as well as any other impurities that may be present as result of the HPLC analysis. Overall the sample handling for a Raman spectrum is much simpler than that of infrared since the cell can be made of glass and the presence of water has a much weaker effect than the observed absorbance found in the ultraviolet region. This is particularly useful in the study of biological systems.

2.2. Sample Cell (Slide) Setup

1. Begin by cleaning a Hanging Drop Microscope Slide and coverslip (Fisher Scientific, Pittsburg, PA) by washing the slide with water followed by an ethanol rinse.

2. Dry the cell using a Kimtech Kimwipe (Kimberly-Clark Inc. Roswell, GA).

3. Using the sample that was prepared in the previous chapter, Subheading 2.1, pipette 50 μL of the 0.2 M sample into the recessed center of the cell, Fig. 5a.

4. Place the edge of coverslip onto the edge of the drop and carefully slide the coverslip into position, as shown in Fig. 5b.

5. At the point where air bubble formation begins to occur add a second 50 μL aliquot to the cell and continue to slide the cover slip into position, as seen in Fig. 5c. This will prevent air bubbles being trapped within the cell.
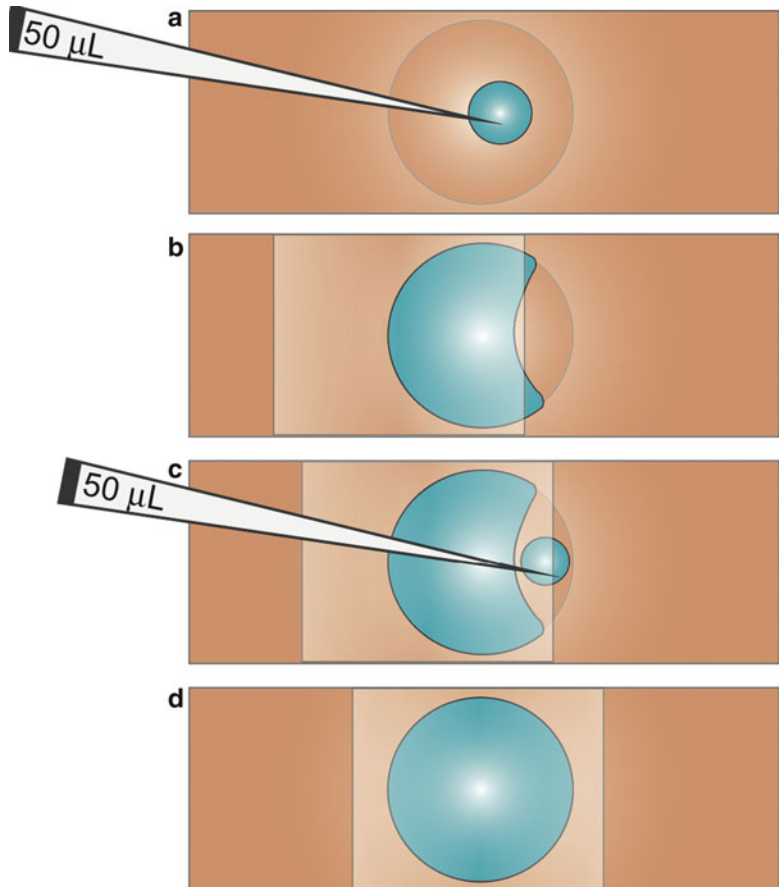
Fig. 5. Illustration of hanging drop sample cell preparation. (**a**) Initial introduction of 50 μL sample to the cell. (**b**) How the coverslip is being slid into position until an air bubble begins to form. Once this occurs a second aliquot of 50 μL of sample should be added to the cell (**c**) while continuing to slide the coverslip into a position centered on the hanging drop (**d**).

6. Allow the cell to sit for a few minutes on the bench top to allow the seal to set, Fig. 5d. The seal of the cell prevents the evaporation of water for about 7 days at room temperature.

While waiting for the sample cell seal to set, start the routine set up of the Raman microspectrometer, Subheading 4. The next section provides a deeper look deeper at the instrument and its alignment. The latter has to be performed whenever a new laser or laser line is selected. This type of adjustment should be performed by a well-trained individual, as even minor misalignments will deteriorate the quality of the Raman spectrum acquired.

3. Methods

3.1. Raman Instrument and Alignment

This section provides an overview of the routine operation of the Renishaw Raman imaging microscope with spectrometer RM1000 (Gloucestershire, UK) and the use of the GRAMS/WiRE software to record and analyze Raman spectra and imaging. The Renishaw Raman RM1000 consists of an optical microscope, spectrometer, laser, and personal computer. This section provides an overview of the technical process of setting up and acquiring a Raman spectrum.

3.2. Laser Operation

Figure 2 illustrates the intensity differences between the inelastically scattered light associated with the Raman effect (both, Stokes and anti-Stokes scattering) and the aforementioned elastic Rayleigh scattering. The intensity of the former is several orders of magnitude lower than that of the latter. The recording of Raman spectra therefore requires intense light to excite a sufficiently large number of molecules into a higher vibrational state. Nowadays, this is always accomplished by using a laser, since it can deliver highly intense, stable, and coherent monochromatic light. Our laboratory is equipped with a SpectraPhysics Stabalite 2018 Argon/Krypton (Irvine, CA) and K4601R-E, Kimmon Electric Helium/Cadmium (Tokyo, Japan) lasers. These continuous-wave lasers provide eleven emission lines between 400 and 700 nm. This tunability capability allows the selection of the optimal excitation. This is particularly useful for use in resonance Raman spectroscopy, which is not explained in this chapter.

Since all the wavelengths of our laser system are far off any electronic resonances of polypeptides and proteins, sample heating and photochemical effects do not interfere with the measurements. This allows us to use 514 nm as excitation wavelength, which is the most intense line of our two lasers. This high intensity is beneficial allowing for greater ease of detection and thus producing larger Raman intensities (see Note 1). However, one should keep in mind that the scattered intensity increases approximately with the fourth

power of the excitation frequency, v_{ex}^4 (7). Additionally, as indicated above, amide II and III are subject to a pre-resonance enhancement, which increases towards shorter wavelengths (10). Moreover, one has to consider a higher sensitivity of commonly used CCD detectors in the green region of the spectrum. Since Raman spectra are generally measured on the lower energy Stokes side of the scattering spectrum, an overlap of the recorded spectrum with the maximum of the CCD detector sensitivity generally requires excitation wavelengths in the blue region, i.e., below 500 nm. A more detailed account of the CCD described for our spectrometer will be discussed later in this chapter.

3.3. Plasma Line Rejection Optics

The optical setup for Raman measurements is shown in Fig. 6. Before the beam enters into the spectrometer (between mirrors 3 and 1) it has to pass through a plasma line laser band pass filter, which rejects plasma emissions from the laser. In the instrument setup this device is placed between the third and first beam steering mirror (Fig. 6, top view).
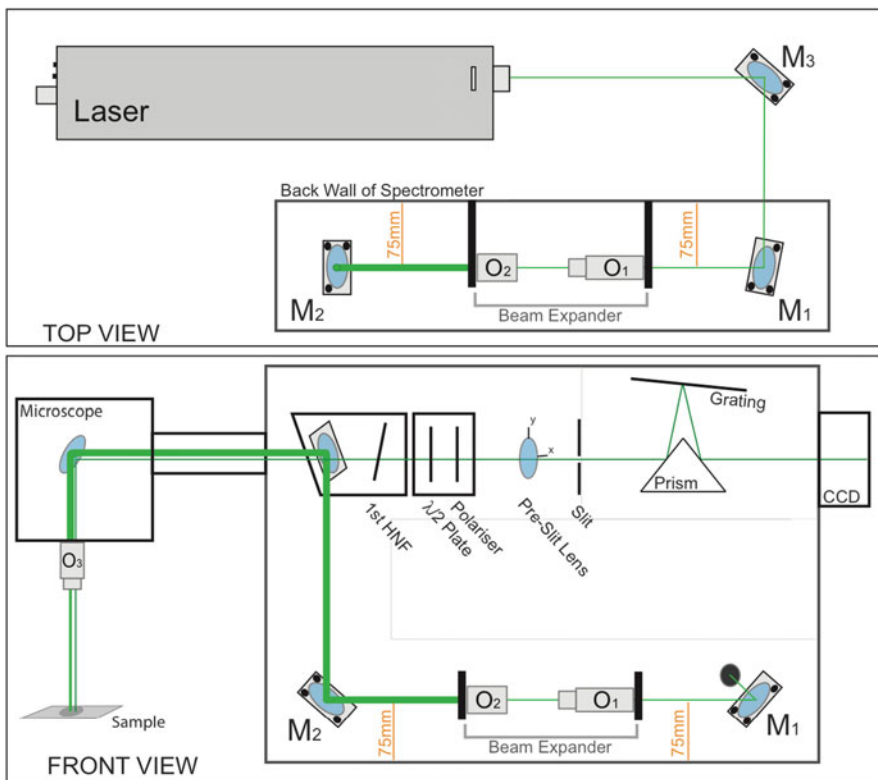
Fig. 6. Illustration of the optical layout of the Renishaw RM1000 Raman spectrometer. (**a**) Top view with the laser positioned behind the spectrometer and the laser beam steering mirror M₃ directing the laser beam into the micro-spectrometer. (**b**) Inside view of the spectrometer optics, which direct the laser beam (*thick line*) to the sample and the scattered light (*thin line*) through the spectrometer to the CCD camera.

Before aligning the optical pathway of the laser beam between laser output and sample, a holographic notch filter (HNF), which eliminates the Rayleigh component , must be placed into the path of the scattered light before it enters the spectrometer (front view of Fig. 6). This easily swappable filter uses kinematic direction to set it into position and is held firmly in place with a long screw that affixes it to the spectrometer.

The HNF serves a dual purpose in the Raman spectrometer in that it acts as a Rayleigh line rejection optic and a beam splitter while it allows the transmission of inelastic Raman scattering (11). It has to be adjusted for optimal function. This alignment step is pivotal for obtaining high-quality Raman spectra. It is the final step of the alignment procedure and should be carried out only by a properly trained user.

The tuning of the rejection angle of the HNF can be carried out over a few nanometers range by varying the incident angle of the first HNF in Fig. 6 (front view). Once properly adjusted the filter blocks the comparatively narrow band of the Rayleigh line and allows any scattered light outside of that bands rejected region to be backscattered into the spectrometer (see Note 2).

The HNF also acts as a beam splitter for the microscope thus directing the laser beam into the microscope and the scattered light into the spectrometer. This reduction in optical components allows for the highest efficiency of transmitted light into the spectrometer.

*3.5. Illumination
Path Alignment*

An optimal function of the Raman (micro) spectrometer requires the path of light to coincide with the optical axis of the instrument. Proper time should therefore be dedicated to the initial alignment of the optical setup.

The process of initially adjusting the optical axis begins with the beam steering mirror, M_3, located on the optical table just behind the Raman spectrometer, top view of Fig. 6. This mirror is set in front of the laser at a $45°$ angle to reflect the beam into an opening in the rear of the spectrometer as shown in Fig. 6. M_3 should be positioned so that the beam hits the center of the mirror. The mirror should be oriented by means of its flexure mounts so that the reflected beam is directed onto the center of the second mirror, M_1, the first mirror within the spectrometer housing. This can be accomplished by adjusting the roll, pitch and yaw of M_3's kinematic flexure mount $(\theta_X, \theta_Y, \theta_Z)$. The alignment step can be facilitated by using a $3''$ by $5''$ card with a line drawn at 75 mm to correct for smaller misalignments of the laser beam. The card should first be held up against the opening within the instrument to allow for the adjustment of the pitch θ_Y. In a second step adjust the roll and yaw together to direct the laser to the center of M_1. Once this is accomplished, one can proceed to the next step of the alignment process.

The laser beam should now be directed onto the center of mirror M_2 by properly adjusting mirror M_1. Next the card should
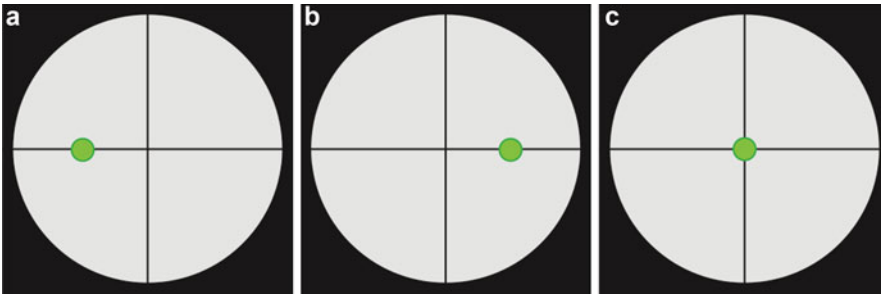
Fig. 7. Illustration of the "doubling the error" technique for correcting a misalignment. (**a**) Initial error as shown on the mirror. (**b**) Adjust the error in the previous optic to be "doubled" in the direction towards the other side of the defined axis. (**c**) Realign the mirror that is off axis to reestablish the defined optical axis. Repeat this process until there are no deviations as the beam travels to the next optical element.

be placed at the base and back wall close to M_1 and M_2 and the 75 mm guide should be used to ensure that the laser beam is at a distance of 75 mm from the back wall and the base of the instrument while traveling from M_1 to M_2. Since the two mirrors both exhibit an angle of ~45° with respect to the optical axis, the reference card will allow to visually verify that the laser is centered and coincident with the optical axis (see Note 3).

If the optical path does not stay on the "75 mm line" when the reference card is moved between M_1 and M_2, the optical axis is not properly aligned. To correct for this aberration one must use the "double the error" method. This entails the doubling of the error using M_3 and the subsequent correction by readjusting the orientation of M_1. This process is illustrated in Fig. 7. The success of the process should be checked with the reference card by verifying that the 75 mm distance is maintained from the wall and the base between M_1 and M_2. Once the optical axis is properly aligned, the ×40 objective, O_1, should be placed into the optical path by adding the objective to the beam expander, as shown in Fig. 6. This will expand the beam. Adjust the roll and the yaw of M_1 until the expanded beam is at the center of the circle and concentric circles are seen, this should only be a very slight adjustment. Add the ×4 objective into the optical path, O_2 as shown in Fig. 6. Adjust the pitch, roll and yaw on this objective's flexure mount so that the beam remains on axis, verifying 75 mm from the wall and base at M_1. With the 40 and ×4 objectives in place the beam should be uniformly bright; if it does not give this appearance then the optical axis is not perfectly aligned. Repeat this procedure until the axis is perfectly aligned (see Note 4).

Once the beam is on-axis at M_2, this procedure can be extended to align the optical path between M_2 and the pre-notch filter mirror. This mirror may have a different orientation depending on the holographic notch filter type being used. Following the beam path, after M_2 in Fig. 3, the pre-notch filter mirror is the next mirror

illuminated in the optical pathway. The pitch roll and yaw of M_2 should be adjusted until the expanded beam is centered on the pre-notch filter mirror. Place a piece of paper, or the reference card onto the microscope stage, verifying that the ×50 objective is in place. Remove the microscope-focusing lever, or lock, and adjust the stage until concentric circles appear illuminated on the stage. If the circles are not centered then one will have to adjust the x and y of the pre-notch filter mirror until the concentric circles are centered on the stage of the microscope (see Note 5).

To see if the laser beam coincides with the optical axis of the microscope, the beam should be focused onto a reflective surface, so that the spot can be inspected through the eyepiece of the microscope. If the spot is not perfectly centered in the crosshairs use the "double the error" method again to correct for the aberration, this time by using M_2 and then adjusting the pre-HNF mirror (Fig. 7). Once centered in the eyepiece crosshairs of the microscope use the fine adjustment knob to focus in and out, focusing about 5–8 μm above and below the reflective surface. If the reflection looks as though it changes from disperse to a well-defined point and back to disperse upon altering the focus, no further optimization is necessary (see Note 6).

3.6. Beam Expander Optics

The above-mentioned beam expander optics affects the size of the spot at the sample. For any objective setting the focus for 0 % will yield the smallest most strongly focused spot at the sample. Likewise, the 100 % setting produces the highest degree of defocusing, illuminating the largest amount of sample area without overfilling the microscope objective. Depending on the back aperture on the microscope objective this may add variance to the backscattering (see Note 2). Since the beam expander also expands the laser's Gaussian beam-shape, the beam expander objectives, O_1 and O_2, are removed for most peptide experiments so that the aperture of focus on the sample is not reduced (see Note 7). With the removal of O_1 and O_2 the symmetric illumination of the beam will return to a defined point. After removal of the beam expander, verify one last time that the confocal image on the stage is centered to guarantee that the optical axis in the microscope is fully aligned.

3.7. Optimization of Detection

For the next step, the reference material should be placed onto the XYZ stage of the microscope. For calibration purposes we use either a silicon wafer or sodium perchlorate, which gives rise to a very intense Raman lines at 520 and 934 cm^{-1}, respectively. $Silicon_{111}$ is generally recommended for this type of calibration due to the relative ease of the detection of an intense band assignable to Si-O phonons of the material.

First, the Si-wafer is placed onto the microscope stage so that it lies in the focus of the laser beam. The focus can be adjusted with the coarse knob until the scattered light goes through the viewing objectives. Subsequently, the stage should be locked into this
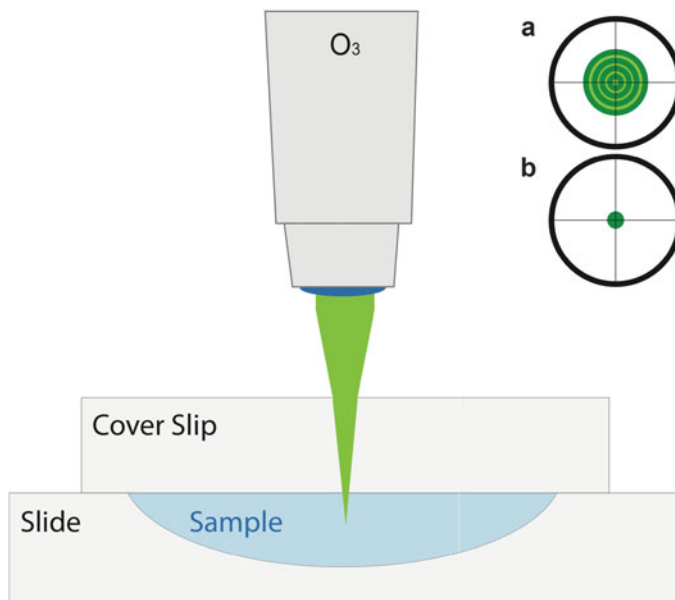
Fig. 8. Illustration of how the laser beam has to be focused onto a sample, in each media the beam will look like (**a**) series of *concentric circles*. This will turn into a well-defined point (**b**) at each interface, i.e., air–glass coverslip and glass coverslip–sample. This reflective surface facilitates the identification of the appropriate focus of the laser beam on the sample surface.

position (Fig. 8a). The fine adjustment knob can be used to move the focus just slightly beyond the reflective surface, until concentric circles are established in the viewing objectives again (Fig. 8b).

1. Now we discuss the data acquisition process. Click on "Collect/Experimental Setup" in the WiRE/GRAMS software package, and choose the following settings to record the silicon calibration spectra, as shown in Fig. 9:

    (a) "Static" mode.

    (b) Centered at: 520 cm^{-1}.

    (c) Time: 1 s.

    (d) Verify that the correct laser wavelength is selected (e.g., 514 nm) and verify that the laser power is at 100 %.

    (e) Verify that the correct objective is selected; for this measurement a ×50 confocal objective is used.

2. Click "Apply" followed by "Ok."

3. Click "collect" followed by "collect" to get a single spectra of the Silicon.

This should collect a spectrum as shown in Fig. 10. If the peak is not centered at 520 cm^{-1} an offset correction may have to be made if a deviation of larger than ± 1 cm^{-1} from the true position is found (see Note 8) (12).

Fig. 9. WIRE/GRAMS software settings for measuring the Silicon reference spectrum.
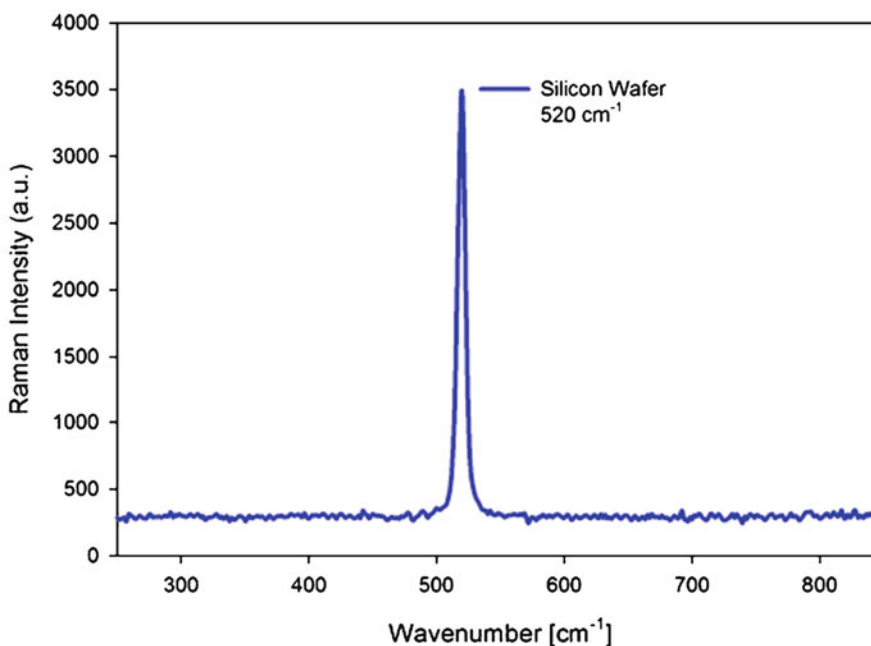
Fig. 10. Representative Silicon spectrum centered on 520 cm^{-1}, obtained using the 514 nm laser wavelength.

3.8. Pre-slit Lens and Slit

The optimization of the Si peak can be achieved through the adjustment of the pre-slit lens. With the Silicon in place, one begins to collect a continuous series of spectra while adjusting the pre-slit lens. This procedure should be performed under minimal light conditions in the lab.

1. To acquire a continuous spectra, use the previously mentioned setting, as found in Fig. 13.

2. Click "collect."

3. Followed by "cycle" to get continuous series of Si spectra.

4. Once the first spectrum of the Silicon appears, click on "History." This allows to monitor how the adjustments affect the silicon signal intensity in real time.

5. Minimally open the front door to the spectrometer, just enough so that an arm fits through the opening, adjust the perpendicular preslit lens position found just in front of the slit. This is the top knob of the pre-slit lens (see Note 9).

6. Adjust the position of this lens in either direction until the maximum signal level at 520 cm^{-1} is obtained, Fig. 10.

7. Click "abort" to stop the cycling of the spectra once the signal is maximized to stop the system from cycling the spectra.

3.9. Grating and CCD Optimization

The next step is to obtain a grating image to check whether Raman scattering hits the CCD surface. This becomes necessary after changing lasers or shifting the laser to different wavelengths can slightly change the path of the exciting laser beam as well as the focus position in the microscope. The grating image is taken at the CCD after the scattered light has passed the dispersive element of the spectrometer (Fig. 6, top view), before entering the detector housing.

The Raman microscope is equipped with a single plane diffraction grating, which needs to be centered on a wavenumber of 520 cm^{-1}. A charge coupled device (CCD) array detector is employed to detect the photons of the scattered light. To accomplish this:

1. Click on "collect" in the main menu.

2. Then click on "experimental setup."

3. The same window will open, Fig. 9, that was used in the previous section and choose the following settings to acquire the silicon grating spectrum:

   (a) "Static" mode.

   (b) Centered at: 520 cm^{-1}.

   (c) Time: 1 s.

4. Click "Image Area."

5. A new window will open showing the scattering image on the CCD; to acquire this image choose the following settings, as shown in Fig. 11.

   (a) Type: "Grating."

   (b) Centered at: 520 cm^{-1} (this may slightly shift as the image is taken as seen in Fig. 11).

   (c) Time: 1 s.

   (d) Click on the "use whole area" checkbox.

   (e) Click "get image."
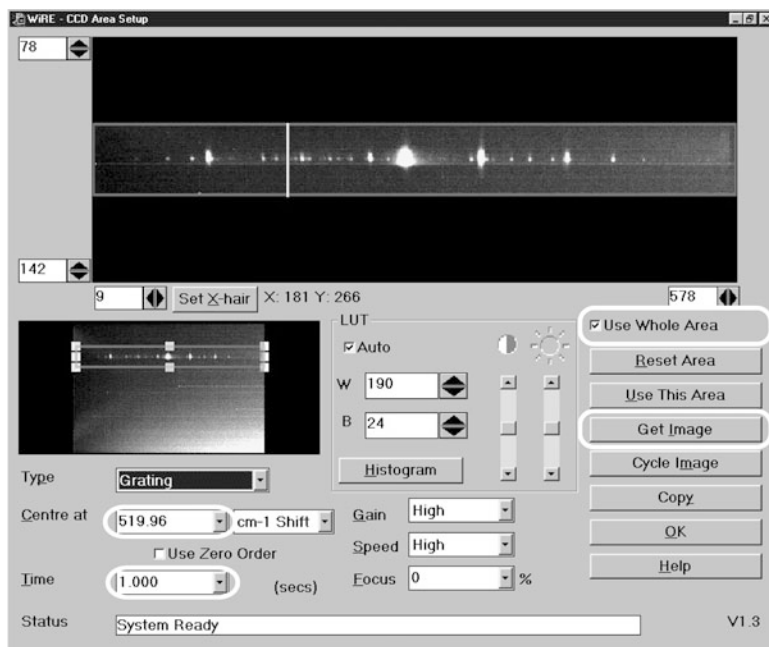
Fig. 11. WiRE/GRAMS software settings for acquisition of the Silicon scattering on the CCD.

This will expose the full area of the CCD for 1 s with the grating centering the spectrum at a 520 cm^{-1} shift. The first-order diffraction of the grating is used (see Note 10). Set the captured image area to have a height of around 2–20 pixels focusing in on the scattered light spots, as shown in Fig. 12. This is the area that is shown to have a shorter axis on the left of the CCD image. This area should be as small as possible to ensure a better signal-to-noise ratio. On the longer, spectral axis, the entire length of the CCD should be exposed, from pixel 2 to 557. Each pixel of the CCD consists of a thin conducting electrode and a thin insulating oxide layer located on top of a p type silicon substrate. This positive bias allows for the collection of electrons to form below the well, storing about 10^5–10^6 electrons (13).

6. Click "use this area."

7. Click "Ok."

By means of this protocol, one optimizes the signal to noise of the scattered light. If at this point the Rayleigh scattering (most intense) is not found, a much more involved alignment must be performed to find the Raman scattering. To find the weaker bands is a difficult process and should be performed by trained personnel only (see Note 11).

The CCD detects the optical signals, which are produced by the sample, temporarily storing the charge row-by-row as photons
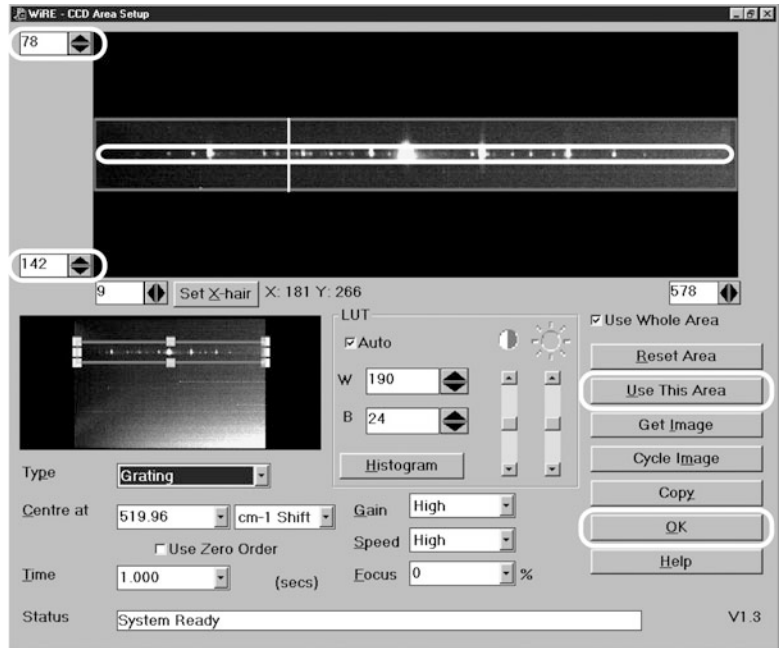
Fig. 12. CCD image of Silicon, the highlighted areas show where the adustments should be made to optimize the detection of the scattered light.

strike the pixel array. The accumulated charge is then transferred down to a preamplifier, which is housed within the instrument, and then sent to the computer. This high selectivity of minimally illuminating the pixel array allows for the CCD to be highly sensitive to the low-intensity levels of the scattered light. Upon optimization of the CCD the instrument is prepared to acquire a sample spectrum, if no major changes have been made to the instrument setup, the acquisition of a spectrum can begin from Subheading 4 see Note 3.

3.10. Software + Start of Experiment

As the full alignment is only performed when the laser line is changed or when a different laser is used, this simplified method can be used to acquiring sample spectra when the system has been properly aligned.

3.10.1. Proper Instrument Startup

1. Turn on Laser.
   (a) For the SpectraPhysics Stabilite 2018 Argon/Krypton Laser (Irvine, CA) begin by turning on the water.
   (b) Turn the key on the control box to power the amplifier.
   (c) Wait 20 min, to allow for the plasma tube to warm.
   (d) Open laser aperture.
2. Turn on the RM1000 Raman spectrometer system.

3. Turn on the PC and start the Renishaw WiRE/GRAMS software package.

4. Select "Check Motors" at startup, and allow the CCD to cool for approximately 20 min.

5. Verify that the correct objective and the correct holographic notch filter are in place for the laser wavelength selected, generally 514 nm is selected for peptides.

3.11. Optimization of the System

3.11.1. Running a Calibration on Silicon

1. Load the Silicon wafer onto the microscopes XYZ stage.

2. Positioning the wafer so that it is optimally oriented so that the laser beam can be focused onto its surface as described above.

3. The fine adjustment knob should be used to move the laser focus just slightly beyond the reflective surface (Fig. 8b), until one observes concentric circles in the viewing objectives (Fig. 8a).

4. Click on "Collect/Experimental Setup" in the WiRE/GRAMS software package, Fig. 9, and choose the following settings to acquire the silicon calibration spectrum:

   (a) Static mode.

   (b) Centered at: 520 cm^{-1}.

   (c) Time: 1 s.

   (d) Verify that the correct laser wavelength is selected, in this case the 514 nm laser will be used, and verify that the laser power is at 100 %.

   (e) Verify that the correct objective is selected, for this measurement a ×50 confocal objective is used.

5. Click "Apply" followed by "Ok."

6. Click "collect" followed by "cycle" to begin spectrum cycling.

7. To monitor this cycling process click "collect" followed by "monitor."

8. After exposure of the CCD for 1 s a spectrum will appear in the main WiRE window, Fig. 10. Click "History" this window to see all changes that are being made in real time during this optimization process.

9. Begin to adjust the pre-slit lens to maximize the silicon peak by adjusting the *x*-axis (top) knob, in small increments, and waiting the 1 s for the spectrum to reflect the changes made.

10. Once this is optimized, i.e., the Peak intensity is maximized, click "abort" to stop the system from collecting spectra.

11. By clicking on the spectrum peak in the spectral window verify that the center of the peak is centered at 520 ± 1 cm^{-1}. To ensure that the center of the peak is used, zoom into this region

by clicking on the plus (+) in the upper left corner of the GRAMS interface (see Note 8).

3.11.2. Measuring the Spectrum of a Sample

1. Place the cell, prepared as described in Subheading 2.2, onto the XYZ stage of the microscope and lock into position, as previously described.

2. Adjust the coarse know until the scattered light can be seen through the viewing objectives, Fig. 8a, and lock the stage into position.

3. With the fine adjustment knob, slowly focus into the sample, Fig. 8.

   (a) From the first scattered image, found in the viewing objectives, begin to focus until the image becomes a well defined point. This is the reflection off the coverslip at the air–glass interface, Fig. 8.

   (b) Once the focus has moved beyond this point the light will begin to scatter in concentric circles, Fig. 8a. This will continue until a point is reached where two sets of concentric circles are overlaid, this is the center of the coverslip.

   (c) Continue until a second spot is seen in the objective, this is the coverslip–sample interface, Fig. 8b.

   (d) Focus just slightly beyond this point, turning the fine adjustment knob a few micrometers thus ensuring that light is fully focused upon the sample and that the glass coverslip profile in the acquired spectrum is minimized. If the focus is moved too far an additional reflective point is observed, although with much weaker intensity. It originates from the sample–slide interface.

4. Using the WiRE/Grams software again click "collect" followed by experimental setup. Here again one will see the main interface, Fig. 13, for the instrument as described previously. For a diagnostic run of the sample choose the following settings:

   (a) Gradient: extended.

   (b) Range: 200–2,000 cm^{-1}.

   (c) Time: 10 s (the minimal exposure time for the CCD in this mode).

   (d) Verify that the correct laser wavelength is selected (514 nm) and that the power is at 100 %.

   (e) Finally verify that the correct objective is selected, for this measurement the ×50 confocal objective is used.

5. Click "Apply" followed by "Ok."
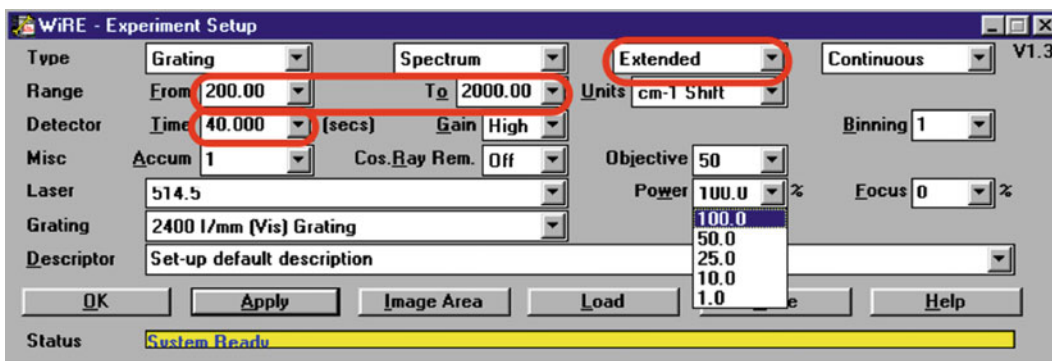
Fig. 13. WiRE/GRAMS software settings for the measurment of a spectrum.

6. Click "collect" followed again by "collect" to begin acquisition of the spectrum.

   To monitor this acquisition process click "collect" followed by "monitor."

7. Upon completion the spectrum acquisition the spectra will be displayed in the main WiRE/GRAMS window.

8. To maximize the quality of the spectrum, take the intensity of the major peak present in the spectrum and divide 500,000 by that number to get the optimized collection time for the spectra (see Note 12).

9. Return to step 4 and replace the minimum 10 s exposure time with the experimentally optimized exposure time (see Note 13).

10. Repeat steps 5–8 to obtain optimized x-polarized spectra.

11. Once the acquisition of the spectrum is complete, the spectrum must be saved (go to "file/save as"). For the file name one should be as specific as possible. One should also include the CCD exposure time duration in the name; this information will be necessary for analysis of the spectra.

12. In the Raman spectrometer flip down the polarizer to acquire the y-polarized spectra, Fig. 6.

13. Repeat this process, collect both the x- and y-polarized spectra a minimum of three times each for averaging. This will further clean up any minor signal-to-noise issues that may be present in the spectrum as well.

3.11.3. Spectral Manipulation and Baseline Correction

1. Once the acquisition of the spectrum is completed, export the data (File/Export) as an ASCII text file (.txt) for use in external analysis software packages such as MultiFit and Microsoft Excel™.

2. Open the spectra data files in MultiFit, by dragging the text file onto the main window for averaging. This averaging of the data can also be performed relatively easily in Microsoft Excel™. The following is the procedure to perform this averaging in MultiFit, a progam described in greater detail in the previous chapter.

   (a) To perform this function click on the file menu "Spektrum."

   (b) Then click on sugery, "Chirurgie."

   (c) In the modifier window "Modifizierer" select the next spectrum to be averages with the primarily selected spectra, the one shown in black in the main screen.

   (d) In operation select the summation operation, "Summe (HS + NS)."

   (e) Repeat this process until all of the spectra taken have been added together.

   (f) Once the spectra have been superimposed click on "Bearbeiten," the Edit menu.

   (g) Then click on "Transformation," to open the transformation window.

   (h) In the transformation window one can apply any mathematical function to the selected spectra.

   (i) Click "Neu," to add the new function to apply to the spectrum.

   (j) In the window on the left type, $Y = Y/(\#$ of spectra added), to finish the averaging of the spectra.

   (k) Click Ok, and the spectra are now averaged.

3. The next step is the process of careful baseline correction. Ideally, the settings defined above scans well beyond the area of interest. This allows for the trimming of the spectra to "trick" the fit in establishing a suitable baseline. This is highly dependent on what the spectrum looks like and the judgment of the individual performing the fit. This task is not trivial and with repetition a intuition for the choice of the best baseline model will develop. In consideration of this it is most important to be consistent across all spectra.

   (a) To perform this function click on the file menu, "Spektrum."

   (b) Then click on sugery, "Chirurgie."

   (c) Select the cut operation, "Ausschnitt (nurHS)."

   (d) The cursor will turn into a magnifying glass, select the area of interest by clicking and dragging a box around the area of interest.

   (e) Upon release of the mouse the spectra will be trimmed to the region selected.

(f) Click on the edit menu, "Bearbieten" to start the baseline correction.

(g) Select Baseline in the dropdown menu (see Note 14).

(h) Under refinement, "Feinheit," select the appropriate fitting type. Generally for this type of correction the roughest fit, "sehr grob," produces the most credible baseline (see Note 15).

(i) Under the function menu, "Funktion," choose Spline.

(j) Click "Ok."

(k) Multifit will then show the baseline that it fits to the spectrum and ask if you would like to subtract it, "Baseline vom Spektrum substrahieren?"

(l) Upon inspection:

- Click Yes, "Ja" if you believe the baseline fit to be good.

- If it is not a good baseline click no, "Nein," and then repeat the process, starting from grob, changing the fit of the baseline to be more fine and adding a hyperbolic function if present if your baseline.

4. This process should be carefully followed for all spectra taken so that the change in the baseline is consistent. Once this process is completed the spectra are ready for data analysis.

3.12. Data Analysis

Trialanine will be used to best illustrate our analysis process. This analysis begins with the baseline corrected x and y polarized Raman spectra. Figure 14 shows the representative polarized Raman spectra (x and y polarization) of trialanine in H_2O recorded between 1,200 and 1,800 cm^{-1}. The scattering angle is 180° (see Note 2). The program MultiFit, as described in the preceding chapter, can be used to decompose the two spectra into Voigtian bands (3). In order to achieve self-consistency, the spectral parameters, i.e., wavenumber positions as well as Gaussian and Voigtian halfwidths should be the same for the two Raman and the corresponding IR spectrum. Generally this works well with the FTIR spectrometer used for our experiments. The complexity of the region between 1,200 and 1,400 cm^{-1} is noteworthy. The region contains three distinct bands at 1,257, 1,291, and 1,340 cm^{-1}, which can be assigned to amide III modes due to the fact that the respective normal modes contain contributions from NH in-plane bending (ipb) and CN stretch (s) (10, 14). The multiplet structure of the amide III is caused by mixing between these classical amide III components and CH bending (in-plane and out-of-plane) (14). As shown by Asher and coworkers (14, 15) and confirmed in one
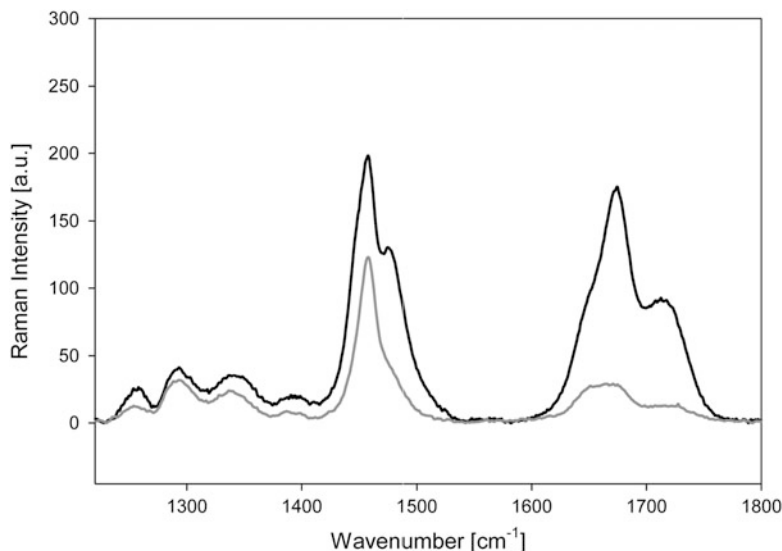
Fig. 14. Polarized Raman spectra of trialanine in D_2O between 1,200 and 1,800 cm^{-1}. The *x*-polarized spectrum is plotted as *solid black line*, the *y*-polarized spectrum is shown as *grey line*. The D_2O solvent spectrum has been subtracted as described above.

of our earlier papers (10) the amide III$_1$ (1,257 cm^{-1}) exhibits a dependency on the dihedral angle which can be described by:

$$\tilde{\nu} = \tilde{\nu}_0 + 30\,\text{cm}^{-1} \cdot \cos\left(\psi + \frac{\pi}{3}\right) \tag{15}$$

with $\tilde{\nu}_0 = 1,220\,\text{cm}^{-1}$. Using the position of AIII$_1$, we obtain $\psi = 165°$ (10). This is of course a representative value for the entire conformational distribution of alanine, but the value is consistent with the predominantly extended (PPII) conformation inferred from the IR and VCD data.

We now turn to the amide I' band of trialanine in D_2O, the isotropic and anisotropic part of which is shown in Fig. 14. These band profiles have to be analyzed with the same distribution function, which are used in the preceding chapter for the IR and VCD band profiles. Before this can be performed the Raman tensor of amide I' in the coordinate system of a given residue has to be formulated. This coordinate system has been introduced in Chapter 18 and is shown again here in Fig. 15 for the sake of convenience. Since the xy-plane is (nearly) colinear with the peptide group, the Raman tensor can be written as:

$$\hat{\beta} = \begin{pmatrix} a & c & 0 \\ c & b & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{16}$$

Of course, our coordinate system does not coincide with *part*. Since the latter is known for visible, nonresonant excitation from polarized Raman studies on a diglycine crystal (9), c and d can be
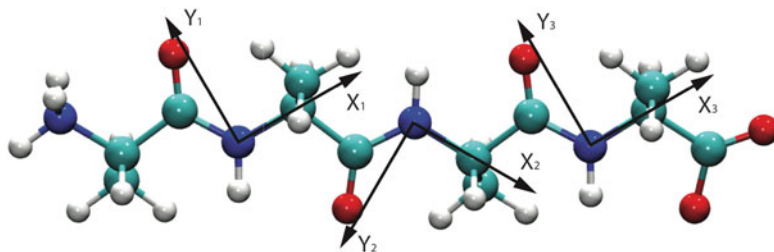
Fig. 15. Planar structure of tetra-alanine ($\phi = 180°$, $\psi = 180°$). Planar structure of tetra-alanine ($\phi = 180°$, $\psi = 180°$). The coordinate systems S1($x1$, $y1$, $z1$), S2($x2$, $y2$, $z2$), and S3($x3$, $y3$, $z3$) were used to express the Raman tensors of the individual, uncoupled amide I modes, and their transition dipole moments (the z-component for S2 has been omitted for the sake of clarity). The structure was obtained by using the program TITAN from Schrödinger, Inc. Taken from Ref. 19 with permission.

related based on an orthogonal transformation of Eq. 17 into the *part*-system. This yields (16):

$$b = -9.3 \cdot c + 1. \tag{17}$$

Since we do not determine absolute Raman cross sections, one of the tensor elements can be set to unity and the other two components can thus be expressed in fractions of this element. It makes sense to normalize a and automatically c on b also to yield $a^{\star}$ and $c^{\star}$, since the $\beta_{AI,yy}$ is the dominant contribution to the Raman tensor (16).

The theory for excitonic coupling between the two amide I′ modes is described in detail in Chapter 19. For trialanine, the Raman tensor of the two excitonic states can be written as:

$$\beta_- = \cos v \cdot \beta_{AI_1'} - \sin v \cdot \beta_{AI_2}$$
$$\beta_+ = \sin v \cdot \beta_{AI_1'} + \cos v \cdot \beta_{AI_2}, \tag{18}$$

where $\hat{\beta}_{A1_1}$ and β_{A1_2} are the Raman tensors of the N- and C-terminal amide I′ modes. The mixing parameter v has been introduced in Chapter 19. The prime notation for $\beta_{AI_1'}$ indicates that the Raman tensor has to be expressed in the coordinate system of the C-terminal residue, which requires that the original N-terminal Raman tensor is subjected to the rotations which we used also in Chapter 19 for the respective transition dipole moment of amide I:

$$\hat{\beta}'_{A1}(C) = R^{\mathrm{T}}(\tau) \cdot R^{\mathrm{T}}(-\eta) \cdot R^{T}(\psi) \cdot R^{\mathrm{T}}(-\xi) \cdot R^{\mathrm{T}}(\phi)$$
$$\cdot \hat{\beta}_{A1}(N) \cdot R(\phi) \cdot R(-\xi) \cdot R(\psi) \cdot R(-\eta) \cdot R(\tau) \tag{19}$$

Now, the amide I′ profiles of the isotropic and anisotropic Raman spectrum of tripeptides can be written as:

$$I_{\mathrm{iso}}(\tilde{v}) = A_{\mathrm{Raman}} \left[\frac{45\beta_{s+}^2}{\sigma_+ \sqrt{2\pi}} e^{-\frac{(\tilde{v}-\Omega_+)^2}{2\sigma_+^2}} + \frac{45\beta_{s-}^2}{\sigma_- \sqrt{2\pi}} e^{-\frac{(\tilde{v}-\Omega_-)^2}{2\sigma_-^2}} \right]$$

$$I_{\mathrm{aniso}}(\tilde{v}) = A_{\mathrm{Raman}} \left[\frac{7\gamma_+^2}{\sigma_+ \sqrt{2\pi}} e^{-\frac{(\tilde{v}-\Omega_+)^2}{2\sigma_+^2}} + \frac{7\gamma_-^2}{\sigma_- \sqrt{2\pi}} e^{-\frac{(\tilde{v}-\Omega_-)^2}{2\sigma_-^2}} \right], \tag{20}$$
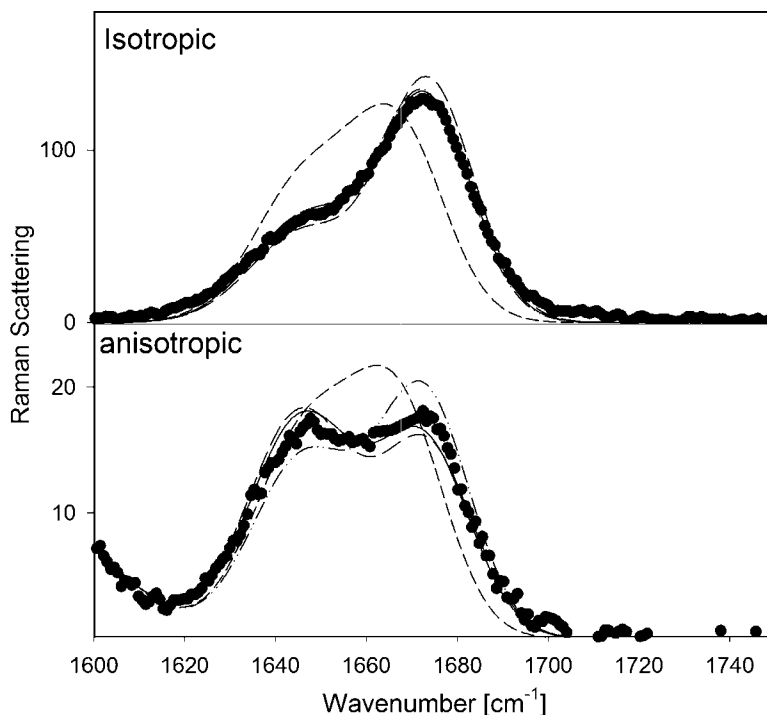
Fig. 16. Simulated isotropic and anisotropic amide I′ band profile of AAA in D₂O with (Isotropic) the conformational distribution visualized in Fig. 22 (3) (*solid line*) and distributions (Anisotropic) with 100 % PPII conformers (*dashed line*), 100 % β-strand (*short dash*) and 100 % right-handed helical (*dash-dot-dot*). The experimental data have been added for comparison. These data were taken from Eker et al. (17).

where the + and − subscripts indicates the respective tensor invariants relation to the excitonic states. For our simulations, we set the tensor element b to 1 and use the scaling factor A_{Raman} for fitting Eq. 21 to match the experimental profile. It should be remembered from Chapter 19, that the bandwidths in Eq. 21 can be estimated from a MultiFit decomposition of the amide I band profile (17).

The simulation visualized by the solid line in Fig. 16 reflects the distribution considered in Chapter 19 for reproducing the IR and VCD band profile of amide I′. Details of this distribution are reported by Hagarman et al. (18). It reflects the predominance of PPII like conformations in the conformational ensemble of the central alanine residues of trialanine. For comparison, we also calculated the profiles for more homogeneous conformations containing only right-handed helical and (antiparallel) β-strand conformations. For the former both profiles exhibit more intensity at the position of the higher wavenumber band, whereas a clear noncoincidence is obtained for the latter, in that the anisotropic Raman scattering is now more intense for the low wavenumber band (19).

As we did for IR and VCD, the approach described in this chapter can be extended for longer peptides. The Raman tensor $\beta_{\mathrm{ex},i}$ for transitions into the $i = $ th excitonic states is written as:

$$\left\langle \hat{\beta}_{\mathrm{ex},i} \right\rangle = \sum_{j=1}^{N} \alpha_{ij} \left\langle \hat{\beta}_j \right\rangle \tag{21}$$

In order to carry out this summation, all Raman tensors of the polypeptide have to be transferred into the same common coordinate system at the C-terminal by means of the consecutive application of the rotations described by Eq. 19. Once this is accomplished, the resulting Raman tensor can be decomposed into its isotropic and anisotropic parts as described above. The final band profiles can then be calculated by:

$$I_\kappa(\tilde{\nu}) = \frac{A_{\mathrm{Raman}}}{\sqrt{2\pi}} \sum_{i=1}^{N} \frac{I_{\kappa,i}}{\sigma_i} \exp\left(\frac{-(\tilde{\nu} - \Omega_i)}{2\sigma_i^2} \right). \tag{22}$$

where $\kappa = $ iso, aniso. The respective intensities $I_{\kappa,i}$ can be obtained by using Eqs. 9 and 10. Generally, we assume the same half-halfwidth of ca. 11 cm^{-1} for all transitions. This parameter can slightly be changed in order to fine-tune the fitting to best match the experimental data.

We demonstrate how this formalism can be extended to simulate the Raman profiles of the amide I mode of the hexapeptide A_2KAAW (Fig. 17). This peptide has five amide I oscillators. The calculation was carried out, as described in Chapter 19, showing the central three amide I$'$ modes were assumed to all exhibit a local wavenumber of 1652.5 cm^{-1}. End effects were considered for the terminal modes, which yields 1,648 cm^{-1} for the C-terminal and 1,675 cm^{-1} from the N-terminal mode. The transition dipole moment is 2.7×10^{-19} esu cm (used for the calculation of the respective IR and VCD profiles) (20), as indicated above. The best simulation was obtained with a distribution model which considered contributions from polyproline II, β-strand, right-handed helical and inverse χ-turn like conformations. To reduce the computer run time for individual simulations the respective Gaussian distributions were substituted by a set of representative values at the position of the distribution maximum and its halfwidths along the ϕ and ψ coordinate. Hence, each Gaussian was represented by five points located in the Rama-chandran space. The result of the best simulation is shown in Fig. 17. The respective coordinates and the statistical weights of the utilized sub-distributions are shown in table 3 of ref. 21.

Taken together this paragraph outlines an algorithm which can be used to calculate the isotropic and anisotropic Raman profiles of amide I for polypeptides. A program using the equations presented herein can easily been written in MATLAB. Our most current version is available free of charge, but at this stage it is not very user friendly. We are planning some upgrading in this regard for the near future.
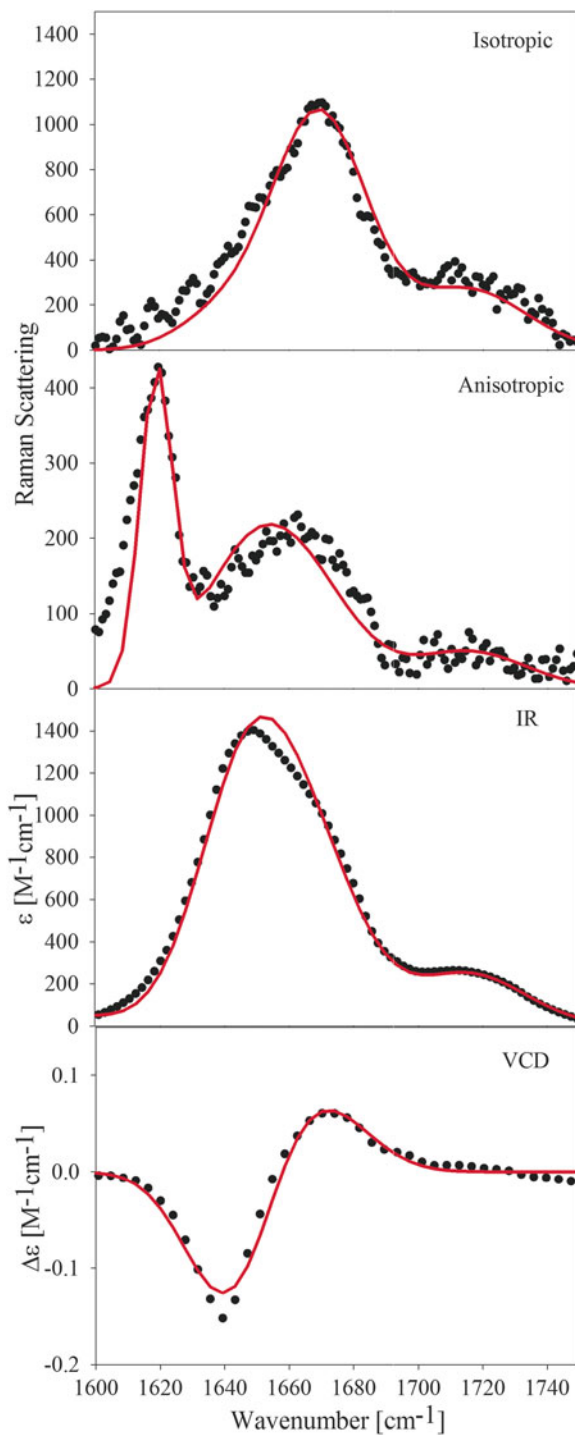
Fig. 17. Simulated isotropic and anisotropic amide I′ band profiles of A_2KAAW in D_2O based on a conformational distribution model explained in the text. A similar model has recently been introduced by Verbaro et al. (21). The simulations of the IR and VCD profiles, which are discussed, are shown for comparison (3).

4. Notes

1. A variety of laser wavelengths may be employed to provide excitation in a region that may not be absorbed by the sample or solvent, again unless resonance Raman is being performed. For peptide analysis we also commonly use the 488 nm and the 514 nm laser wavelengths.

2. 180° backscattering, Fig. 3, illustrates the collection geometry of the Olympus BH-2 microscope attached to the spectrometer. Allowing for the incident light to propagate to the sample (excitation) and the scattered light (collected) to propagate back up the microscope at an 180° orientation relative to the incident light.

3. It may be easier for visualization to draw two lines crossing at 75 mm to pinpoint where the defined axis is until more experience in the alignment is gained. Also beware of motors that are attached to the back wall. A cut may have to be made to ensure that the axis is 75 mm from the back wall.

4. Alternately, if difficulties persists with the alignment of the beam expander, it can be driven into and out of focus through the grams software. This will emphasize the deviation in the optical axis allowing for the fine adjustment of the flexure mounts to correct for this deviation and get the laser on axis through the beam expander.

   (a) Place the 75 mm reference card at M_2.

   (b) Click on "collect" followed by "experimental set-up"

   (c) Under "Focus" on the right-hand side of the window that opens drive the beam-expander in and out of focus (0–100 % Focus).

   - If the beam remains centered on the axis as it expands the axis is properly aligned.
   - If the beam expands at an angle relative to the optical axis, adjust the yaw pitch and roll on the O_2 flexure mount until the beam is properly realigned.

5. A change of the focus reflects a misalignment of the laser beam. In this case, the "double the error" method must be used between M_2 and the pre-notch filter to correct for this aberation. If this misalignment is significant, an adjustment may have to be made using the first holographic notch filter, depending on the filter type. The orientation of the pre-notch filter mirror may be different depending on the type of HNF used for the laser wavelength of interest. Do not adjust the first holographic Notch filter, first HNF as shown in Fig. 6, unless necessary. The adjustment of this filter affects the alignment of the

scattered light into the spectrometer, or microscope depending on the notch filter type. Only adjust this as a last resort if you are still unable to find the scattered light on the CCD, this will maintain manufacturers specifications.

6. If the image indicate that the laser beam does not hit the sample with propagation direction perpendicular to the sample surface, an adjustment has to be made at the pre-HNF mirror or the first HNF, depending on the HNF used. Perform this action with extreme caution, as it will alter the scattering of light into the microscope and the spectrometer. This correction should only be performed under the supervision of a trained individual.

7. If the system was not perfectly aligned when the 4 and ×40 objectives are removed a slight adjustment of M_2 has to be carried out. A small adjustment to the roll and yaw should quickly correct for such an abberation.

8. If a Raman line position deviates from its expectation value by more than 1 cm$_,^{-1}$ an offset correction must be made. This can be carried out as follows:

   (a) Notate the deviation from the true position.

   (b) Click "collect" from the menu bar.

   (c) Click "Applications."

   (d) Click "Calibrate system."

   (e) Click "Offset correction."

   (f) Type in the value that the peak needs to be adjusted by in wavenumbers, to shift the band position to the correct band position to 520 cm−1.

   (g) Click "apply" (this will adjust the grating to the correct position).

   (h) Retake the silicon spectra until the error is negligible, i.e., within the acceptable error.

9. Never adjust the pre-slit lens that is parallel to the slit during this adjustment. This is the knob that is located on the front side of the lens mount.

10. The order is a label for the range of wavelengths that will satisfy the grating equation:

$$v = \frac{5\,gm}{\sin(\alpha)\cos((\gamma/2))},\qquad(23)$$

where v represents the absolute wavenumber, g is the number of grooves of the grating per mm, m is the order ($m = \pm 1$ for first order), α is the grating angle between the incoming and outgoing beam, γ is the full included angle between the incoming and out going beam.

11. Begin by adjustment of the pre-slit lens, if one goes to both extremes with the CCD image area opened wider. If one is still unable to find neither the scattering in the CCD image nor a Si peak for scans around 520 cm^{-1}, the holographic notch filter might be misalligned. To fix this one has to either cycle the CCD or the Silicon spectra while adjusting the first HNF until scattering is seen; this process can be very time consuming. Ideally this filter should never be touched as it changes the entrance angle of the laser beam into the spectrometer. This should only be performed by or under the supervision of properly trained personnel.

12. This optimization is obtained by maximizing the amount of time the scattered light is exposed to the CCD. The scan-time is maximized by using the detection limit of the CCD, 50,000, and the initial 10 s exposure time, divided by the maximum peak intensity of the spectrum, which was initially obtained.

$$\frac{50,000 \times scan\ time}{experimental\ max\ peak\ intensity}$$
$$= \frac{50,000 \times 10\ [\text{s}]}{experimental\ max\ peak\ intensity} \tag{24}$$

13. Longer CCD exposure times can bring about a variety of issues that can easily be troubleshot depending on the encountered issue. Data can be optimized by following the suggestions listed below:

    (a) To improve the signal-to-noise ratio one can either increase the exposure time or increase the number of accumulations.

    (b) For the elimination of an intense background due to auto-fluorescence one can either decrease the laser power, quench the fluorescence by exposing the sample to incident laser light for an extended period of time, or lastly choosing a different laser excitation wavelength.

    (c) If the signal appears saturated one should reduce the exposure time or reduce the laser power.

14. Alternately one can press "Control + B" to get to this same menu.

15. If baseline problems persist one can utilize a higher order fit using the peakfit menu, as thoroughly described in the previous chapter.

Acknowledgment

References

1. Small EW, Fanconi B, Peticolas W (1970) Raman spectra and the phonon dispersion of polyglycine. J Chem Phys 52:4369

2. Bandekar J (1992) Amide modes and protein conformation. Biochim Biophys Acta 1120:123–143

3. Schweitzer-Stenner R, Soffer JB, Verbaro D (2011) Structure analysis of unfolded peptides by vibrational circular dichroism spectroscopy. In: This book

4. Chen XG, Schweitzer-Stenner R, Asher SA, Mirkin NG, Krimm S (1995) Vibrational assignments of trans-N-methylacetamide and some of its deuterated isotopomers from band decomposition of IR, visible, and resonance Raman spectra. J Phys Chem 99:3074–3083

5. Chen XG, Schweitzer-Stenner R, Mirkin NG, Krimm S, Asher SA (1994) N-Methylacetamide and its hydrogen-bonded water molecules are vibrationally coupled. J Am Chem Soc 116:11141–11142

6. Sieler G, Schweitzer-Stenner R (1997) The amide I mode of peptides in aqueous solution involves vibrational coupling between the peptide group and water molecules of the hydration shell. J Am Chem Soc 119:1720

7. Long DA (2002) The Raman effect. A unified treatment of the theory of Raman scattering by molecules. Wiley & Sons, New York

8. Raman CV, Krishnan KS (1928) A new class of spectra due to secondary radiation, part I. Indian J Phys 2:399–419

9. Pajcini V, Chen XG, Bormett RW, Geib SJ, Li P, Asher SA, Lidiak EG (1996) Glycylglycine p→p* and charge transfer transition moment orientations: near resonance Raman single crystal measurements. J Am Chem Soc 118:9716–9726

10. Schweitzer-Stenner R, Eker F, Huang Q, Griebenow K, Mroz PA, Kozlowski PM (2002) Structure analysis of dipeptides in water by exploring and utilizing the structural sensitivity of amide III by polarized visible Raman, FTIR-spectroscopy and DFT based normal coordinate analysis. J Phys Chem B 106:4294–4304

11. Carrabba MM, Spencer KM, Rich C, Rauh D (1990) The utilization of a holographic Bragg diffreaction filter for Rayleigh line rejection in raman spectroscopy. Appl Spectrosc 44:1558–1561

12. Temple PA, Hathaway CE (1973) Multiphonon Raman spectrum of silicon. Phys Rev B 7:3695–3697

13. Ingle JD, Crouch SR (1998) Optical sources, transducers, and measurement systems. In: Spectrochemical analysis. Prentice Hall, Upper Saddle River, NJ, pp 87–134

14. Asher SA, Ianoul A, Mix G, Boyden MN, Karnoup A, Diem M, Schweitzer-Stenner R (2001) Dihedral ψ angle dependence of the amide III vibration: a unique sensitive UV resonance Raman secondary structural probe. J Am Chem Soc 123:11775–11781

15. Mikhonin AV, Ahmed Z, Ianoul A, Asher SA (2004) Assignments and conformational dependencies of the amide III peptide backbone UV resonance Raman bands. J Phys Chem B 108:19020–19028

16. Schweitzer-Stenner R (2002) Dihedral angles of tripeptides in solution determined by polarized Raman and FTIR spectroscopy. Biophys J 83:523–532

17. Eker F, Cao X, Nafie L, Schweitzer-Stenner R (2002) Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR and VCD spectroscopy study. J Am Chem Soc 124:14330–14341

18. Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R (2010) Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I band profiles and NMR scalar coupling constants. J Am Chem Soc 132:542

19. Schweitzer-Stenner R (2006) Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure. A critical review. Vib Spectrosc 42:98–117

20. Measey T, Hagarman A, Eker F, Griebenow K, Schweitzer-Stenner R (2005) Side chain dependence of intensity and wavenumber position of amide I′ in IR and visible Raman spectra of XA and AX dipeptides. J Phys Chem B 109:8195–8205

21. Verbaro D, Ghosh I, Nau WM, Schweitzer-Stenner R (2010) Discrepancies between conformational distributions of a polyalanine peptide in solution obtained from molecular dynamics force fields and amide I′ band profiles. J Phys Chem B 114:17201–17208

Chapter 20

Isotope-Edited Infrared Spectroscopy

Ginka S. Buchner and Jan Kubelka

Abstract

Isotope-edited infrared (IR) spectroscopy is a powerful tool for studying structural and dynamical properties of peptides and proteins with site-specific resolution. Labeling of selected amide carbonyls with ^{13}C results in detectable sidebands of amide I′ vibrations, which provide information about local conformation and/or solvent exposure without structural perturbation to the protein. Incorporation of isotopically labeled amino acids at specific positions is achieved by the chemical synthesis of the studied proteins. We describe the basic procedures for synthesis of ^{13}C isotopically edited protein samples, experimental IR spectroscopic measurements, and analysis of the site-specific structural changes from the thermal unfolding IR data.

Key words: Isotopic editing, Infrared spectroscopy, Amide I, Solid-phase peptide synthesis, Secondary structure, Tertiary structure

1. Introduction

Infrared (IR) spectroscopy is frequently used as an experimental probe for structural changes in peptides and proteins during folding or unfolding, ligand binding and catalysis, as well as other important biochemical processes ([1]). The sensitivity of the amide I band (predominantly amide backbone C=O stretch) to the polypeptide chain conformation and its changes, as well as to hydrogen bonding, has been well established. However, due to the inherently low resolution, only the average secondary structural information can be obtained. This fundamental limitation can be overcome by introducing isotopically labeled amino acids on the amide C=O at specific locations within the peptide or protein sequence ([2], [3]). Amide I vibrations of ^{13}C-labeled residues are decoupled from ^{12}C vibrations and result in a separate signal shifted to lower wavenumbers (Fig. [1]), which contains the information about the local structure within the labeled protein segment.
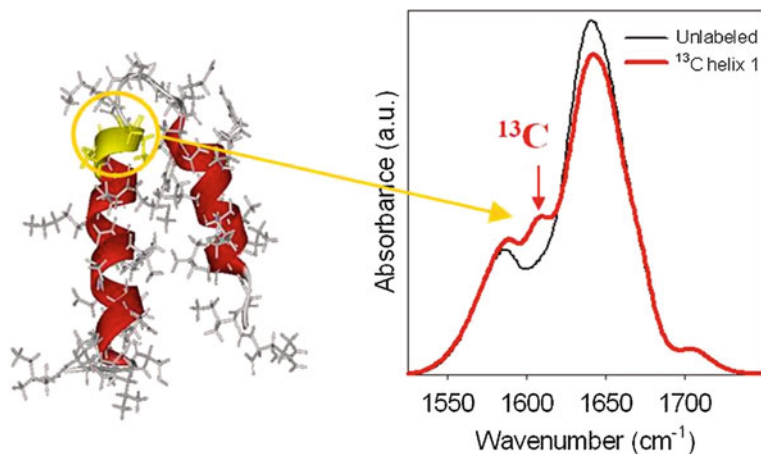
Fig. 1. Site-specific resolution of protein structure by means of ^{13}C isotopically edited IR spectroscopy. Amide I′ signal of the model protein selectively labeled in the highlighted region with three ^{13}C amino acids (*red*) overlaid with the amide I′ of the same, unlabeled protein (*black*).

Chemical synthesis of peptides and small proteins offers the possibility to insert ^{13}C=O-labeled amino acids into selected, known positions. Incorporation of two or three ^{13}C amino acids in neighboring or alternating positions provides enough spectral signal as well as specific coupling patterns to report on the backbone conformation (secondary structure) of the labeled region (4). Furthermore, changes in tertiary structure can be probed through the solvent exposure of labeled amides (5–8). In this case, labeling of a single buried amide is often sufficient (7), and the isotopic shift can be further enhanced (at higher cost) by ^{13}C=^{18}O substitution (9). The ^{13}C amide I′ signal is also a very sensitive probe of the β-sheet structure (10) and has been used in studies of polypeptide aggregation (11). Other isotopic labeling schemes are possible as well: for example deuteration of specific side-chain residues can be used as a probe of the burial of the hydrophobic core in proteins (12).

Isotopic editing therefore provides a versatile and minimally perturbing experimental probe of the local protein secondary and tertiary structure. In combination with the inherent fast timescale of the IR spectroscopy, it also allows direct measurements of protein conformational dynamics with site-specific resolution (13). However, thus far the applications have focused predominantly on model short peptides. Only recently has systematic incorporation of isotopically labeled amino acids at multiple positions within a protein been used to draw a detailed picture of its folding mechanism (14).

One of the challenges has been the analysis of site-specific structural transitions from the ^{13}C sidebands, which in proteins are more obscured by overlap with the ^{12}C (unlabeled) amide I′ and with the amino acid side-chain signals (Fig. 1). In peptides, intensity changes at specific ^{13}C frequencies, assigned to the particular secondary

structure (e.g., helix) and/or unfolded "coil", are most commonly used ([15], [16]). In a protein, we consider the frequency, intensity, and shape changes of the ^{13}C bands together, using singular value decomposition (SVD) of the data ([17]) and subsequent fitting to the empirically determined basis spectra ([14]). This analysis of ^{13}C IR protein thermal unfolding data is presented below as an example. Alternatively, the ^{13}C amide I' of interest can be resolved from the other band components by band fitting ([18]), with the aid of resolution enhancement techniques, such as Fourier self-deconvolution (FSD) ([19]). For analysis of the solvent exposure of buried amide groups, the shift of the ^{13}C peak maximum to the lower frequency is utilized ([7]).

2. Materials

2.1. Peptide Synthesis

Peptides and small proteins ($\lesssim$100 amino acids) can be relatively easily produced on an automated peptide synthesizer applying standard solid-phase peptide synthesis (SPPS) techniques ([20]). Amino acids with 9-fluorenylmethoxycarbonyl (Fmoc) protection on the amino group and acid labile side-chain protecting groups are commercially available from a number of suppliers (EMD Novabiochem, Anaspec, Sigma-Aldrich, etc.), as are preloaded resins and all reagents for the peptide synthesis. Fmoc-protected ^{13}C-labeled amino acids for synthesis are also commercially available (e.g., Cambridge Isotope Laboratories, Sigma-Aldrich/Isotec).

1. Resin, preferably preloaded with the C-terminal amino acid of the desired sequence (see Note 1). Protected amino acids: fourfold excess (e.g., 0.2 mmol for 50 μmol scale synthesis).

2. Activator (HBTU or HCTU): Fourfold excess.

3. Solvent: Dimethylformamide (DMF).

4. Deprotection of the amino function: 20 % v/v piperidine in DMF.

5. Activation of the carboxylic acid group: 0.4 M N-methylmorpholine (NMM) in DMF.

6. Peptide synthesizer (e.g., Tribute Peptide Synthesizer, Protein Technologies, Inc.).

7. Cleavage: 81.5 % v/v trifluoroacetic acid (TFA), 5 % v/v thioanisole, 5 % w/v phenol, 1 % v/v triisopropylsilane, 5 % v/v H$_2$O, 2.5 % v/v ethanedithiol (EDT) (see Note 2).

8. 13 mL filter tubes (see Note 3), 50 mL centrifuge tubes.

9. Precipitation: Methyl tert-butyl ether (MTBE) or diethyl ether.

10. Lyophilizer.

11. Low-temperature (0 °C) centrifuge.

12. Purification: HPLC, 0.1 % v/v TFA in H_2O (buffer A), 0.1 % v/v TFA in acetonitrile (buffer B), C18 column.

2.2. FTIR Spectroscopy

1. D_2O (e.g., Cambridge Isotope Laboratories), sodium phosphate salts (monobasic, dibasic).

2. FTIR spectrometer containing a constantly purged sample compartment with cell holder, equipped with a sample temperature control via a water recirculator bath and temperature jacket or a Peltier device, liquid sample cell, CaF_2 windows, gaskets, Teflon spacer (50 μm) (see Note 4).

3. Methods

3.1. Peptide Synthesis

1. Use standard procedures for the Fmoc SPPS (20).

2. Fill vials with amino acids and activator (HBTU, HCTU).

3. Place the resin into the reaction vessel.

4. Fill the solvent and reagent bottles and connect to the peptide synthesizer.

5. Load the vials into the synthesizer, starting with the C-terminus.

6. Start the synthesis.

7. When the synthesis is complete, wash the resin with the peptide into a filter tube using DMF.

3.2. Post-synthesis Workup and Cleavage

1. Remove DMF under reduced pressure and wash once again with DMF.

2. Wash three times with dichloromethane (DCM).

3. Wash twice with methanol and dry under reduced pressure for at least 30 min.

4. Add 10 mL of the cleavage mixture for every 100 mg resin/peptide, nutate the mixture for 2–3 h, and then let the liquid drop into prechilled (−20 °C) ether (30 mL ether/8 mL cleavage solution), where it precipitates (see Note 5).

5. Wash the resin twice with 1 mL cleavage solution or TFA, and add wash solutions to the ether.

6. Vortex the ether mixture well for 3 min, and then chill for 1 h at −20 °C.

7. Centrifuge the mixture for 5 min at ~4850 *g* and 0 °C, and discard the ether carefully.

8. Add 20 mL chilled ether to the peptide, vortex thoroughly, centrifuge again, and discard ether.

9. Repeat step 8 two more times.

10. Remove the remaining ether under vacuum.

11. Dissolve the dry peptide in a small amount of water, freeze, and lyophilize.

12. Purify by reversed-phase HPLC by dissolving the peptide in 100 % buffer A, and run a linear gradient over 50 min from 0 to 50 % buffer B (see Note 6).

13. Freeze the collected peaks in liquid N_2, lyophilize, and analyze by MS, if available (see Note 7).

3.3. FTIR Sample Preparation

1. H/D exchange of the peptide: Dissolve 3–4 mg of peptide in 2 mL D_2O, nutate for 1 h, heat to ~45 °C for 10 min, add 2 μL HCl (conc.) to remove TFA (see Note 8), nutate for another 45 min, freeze in liquid N_2, and lyophilize.

2. Buffer: Exchange the sodium phosphate buffer salts by dissolving in D_2O, nutating for 1 h, freezing, and lyophilizing. Mix to get ~0.1 M deuterated buffer with the desired pH.

3. Protein sample: Dissolve 1 mg of the exchanged peptide in 60 μL deuterated phosphate buffer while trying to minimize air contact with the sample.

3.4. FTIR Spectroscopy

1. Measure the FTIR spectrum of the water vapor with the empty sample compartment (see Note 9). Usually, quickly opening and closing the sample compartment cover before starting the scan will introduce enough water vapor.

2. Fill the (clean and dry) sample cell with ~30 μL of the deuterated phosphate buffer solution while trying to minimize air contact with the sample (see Note 10).

3. Measure the buffer spectrum or spectra (e.g., at preset temperatures).

4. Disassemble the sample cell, clean, and dry all components thoroughly, in particular the windows and the spacer.

5. Fill ~30 μL of the peptide solution into the IR sample cell (see Note 10).

6. Measure the protein solution FTIR spectrum or spectra.

3.5. FTIR Data Processing

The processing of raw FTIR data follows standard, established procedures (1, 21).

1. Subtract the FTIR spectra of the buffer and of the residual water vapor from the corresponding protein solution spectra (see Note 11).

2. Truncate the corrected protein spectra to contain the region of interest (e.g., amide I′ ~1,500–1,750 cm^{-1}, see Fig. 1).

3. If necessary, correct the residual spectral baseline, e.g., by a second-order (quadratic) polynomial.

3.6. Example Data Analysis: Thermal Unfolding

The analysis of ^{13}C data depends on the objective of the particular study, choice of isotopic labels, etc. We present an example of the analysis of thermal unfolding of a model protein (14).

1. Normalize the ^{13}C-labeled and unlabeled amide I′ FTIR spectra to the integral amide I′ intensity at the highest measured temperature.

2. Perform an SVD of each set of the processed temperature-dependent experimental data (all ^{13}C labeled, unlabeled FTIR, and additional experimental data, e.g., circular dichroism, fluorescence, etc.). Determine the significant number of SVD components (see Fig. 2, Note 12).

3. Based on the number of significant components (e.g., two, three), fit the temperature-dependent weights (V vectors from SVD, weighted by their corresponding singular values S) simultaneously to an appropriate thermodynamic model (two-state, three-state, etc., see Note 13). This will yield the global thermodynamic parameters as well as the basis spectra for each of the (two, three) thermodynamic states for the unlabeled protein and all the ^{13}C isotopically labeled variants (see Fig. 3).

4. Decompose the basis spectra corresponding to each of the (two, three) states (see Note 14) into the ^{12}C and ^{13}C components (see Fig. 4).
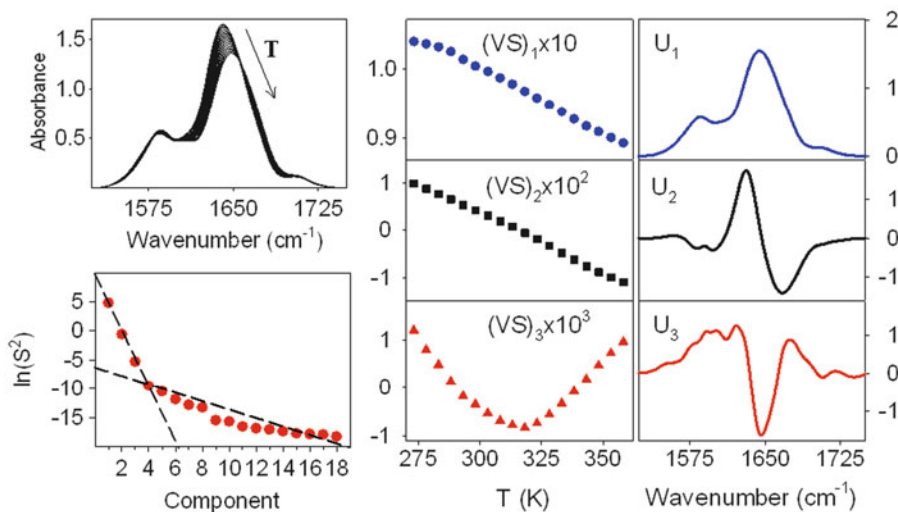
Fig. 2. Singular value decomposition (SVD) of the FTIR data. *Left*: Amide I′ experimental spectra measured at 18 temperatures (*top*) were decomposed by SVD. The singular values S (*bottom*) show a characteristic change in the slope of the $\ln(S^2)$ plot, which indicates that the first three components (containing 99 % of the data variance) are significant. The remaining components are discarded as noise. *Right*: The first three singular value-weighted V vectors and the corresponding basis spectra (U vectors).
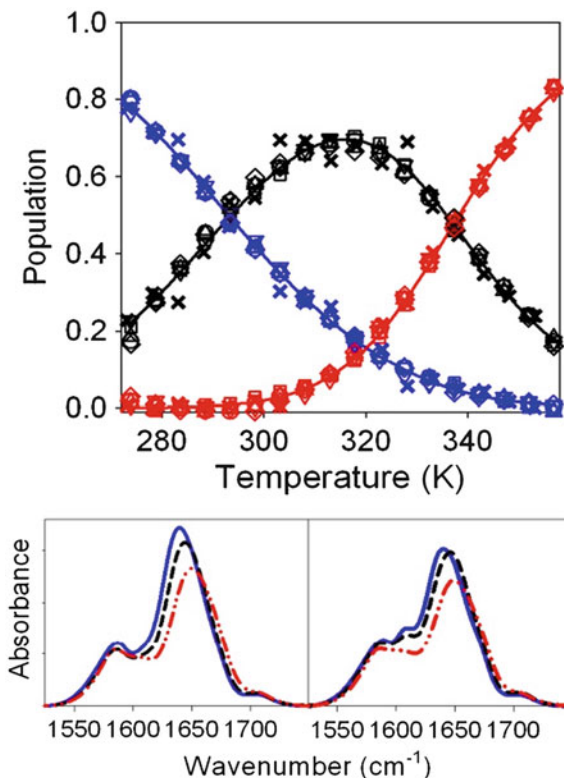
Fig. 3. Example analysis of the experimental thermal unfolding data with a three-state model. *Top panel* shows the fractional populations of the three thermodynamic states, obtained by fitting the three significant SVD components (*V* vectors weighted by their singular values). *Solid lines* are fitted populations (*F blue line*, *I black line*, *U red line*), symbols correspond to the experimental data, which include circular dichroism data (*crosses*) and amide I′ IR for an unlabeled (*circles*) and four ^{13}C-labeled variants of the model protein. The *bottom panels* show the corresponding basis spectra (*F solid blue*, *I dashed black*, *U dash-dot-dot red*) for the unlabeled (*left*) and one ^{13}C isotopically labeled protein (*right*).

5. Calculate the contribution of each of the ^{12}C and ^{13}C basis components to the temperature-dependent amide I′ IR spectra set for each ^{13}C-labeled protein (see Note 15).

6. Fit the resulting temperature-dependent contributions of the individual ^{13}C basis spectra to the experimental spectra with an appropriate thermodynamic model (see Fig. 5).

4. Notes

1. The first amino acid on the resin defines the success of the whole synthesis. Therefore, use of preloaded resins with a known loading density of the first amino acid is preferred.
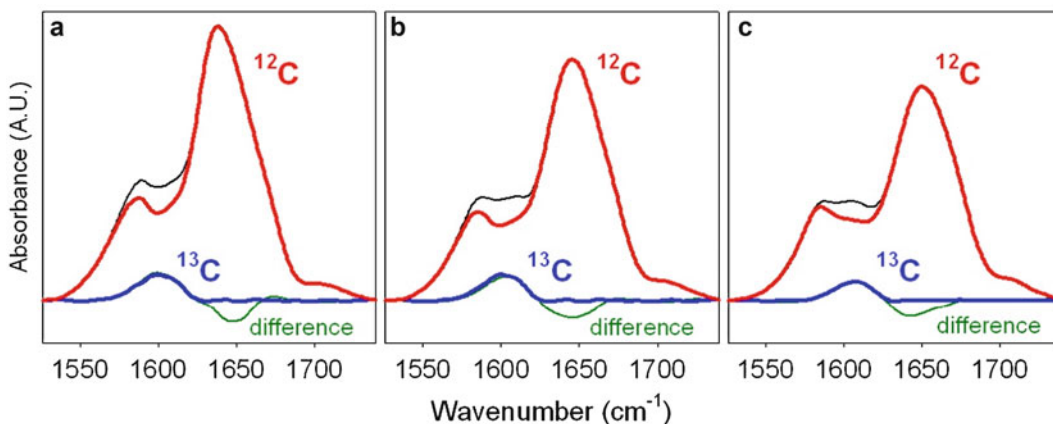
Fig. 4. Decomposition of the amide I′ IR into the ^{13}C and ^{12}C contributions. The positive sides of the difference spectra (*thin green lines*)—the best guess for the ^{13}C signal—represent the reference spectra for the subsequent analysis of the ^{13}C spectral signals. The complete IR spectrum (*thin black*) is decomposed into the ^{13}C signal (*blue lines*) and complementary ^{12}C signal (*thick red lines*). The decomposition is performed for the three basis spectra describing the unfolding thermodynamics: (**a**) *F*, (**b**) *I*, and (**c**) *U*.
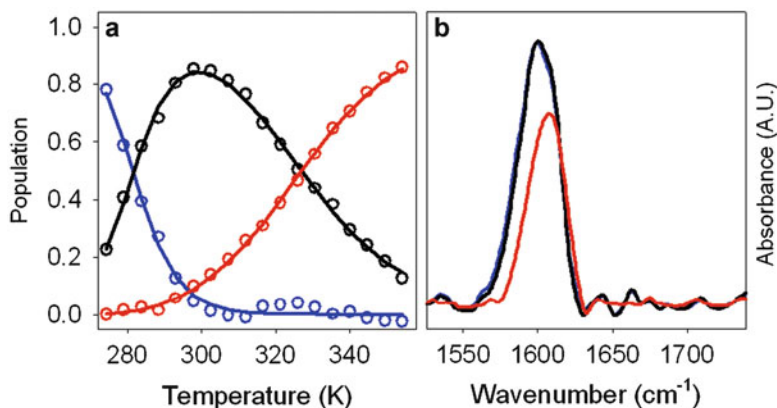
Fig. 5. Analysis of the ^{13}C amide I′ components. The decomposition of the amide I′ band (Fig. 4) yields three ^{13}C spectral components, whose corresponding temperature-dependent weights were fitted to a three-state model (**a**). The resulting basis spectra (**b**) for the first two states (*blue* and *black lines*) are virtually identical. As a consequence, the first transition does not correspond to any significant structural change and the local unfolding can be described as a two-state process with the *red trace* in (**a**) corresponding to the local unfolding transition.

Low loaded resins are beneficial for longer peptides or more difficult sequences, since high loading density facilitates aggregation during the synthesis. We usually synthesize unlabeled peptides on a scale of 50 μmol, while for isotopically labeled ones the scale is reduced to 25 μmol due to higher cost of the labeled amino acids. The latter still yields enough peptide for repeated IR spectroscopic measurements.

2. These ingredients include all recommended scavengers for different kinds of side chain protecting group removals. Due to the toxic and corrosive properties of these ingredients only work with the cleavage mixture in a fume hood wearing appropriate protection (gloves, safety glasses). EDT due to its particularly offensive odor is recommended to be added last.

3. We find that disposable polypropylene columns (Pierce, Thermo Scientific) serve best as filter tubes for that purpose.

4. Using a path length of more than 100 μm will result in an intense D_2O signal in the amide I′ region. Spectral resolution of 4 cm^{-1} is more than sufficient as the spectral bands in aqueous solutions are much broader.

5. If the peptide does not precipitate in the ether, evaporation of the TFA might be necessary.

6. If the peptide elutes at more than 30 % buffer B, it can also be dissolved in 10 % B and the gradient started from there.

7. Several peaks collected from HPLC may contain the desired peptide, but usually there is a single most intense peak with a purity of >95 % of the desired peptide.

8. TFA strongly absorbs at 1,673 cm^{-1} and therefore will interfere with the amide I′ band if not removed.

9. In order to obtain the protein amide I′ band, water vapor and buffer spectra have to be subtracted from that of the protein, which all have to be collected under conditions as close to identical as possible. Furthermore, since the analysis of the site-specific unfolding relies on the difference spectra of the ^{13}C-labeled and unlabeled protein, high degree of consistency is required between the experimental measurements of the isotopically edited and unlabeled samples. For thermal unfolding experiments, the spectra are collected at temperatures from 0 to 90 °C at steps of 3–5 °C. Preferably, the temperatures are changed automatically, through a software interfaced with the spectrometer control. Prior to the spectral scan, a delay of 3–5 min is applied after reaching each desired temperature to allow for full sample equilibration. The standard experimental parameters are 4 cm^{-1} spectral resolution and 256 or 512 repeated scans.

10. Make sure that the cell is completely filled and no air bubbles are visible after closing the cell. Use more solution if necessary.

11. The buffer spectrum is subtracted from that of the protein solution (protein plus buffer) by minimizing the absorbance signal in the spectral region with no contribution from the protein, e.g., 1,800–1,900 cm^{-1}. The water vapor is subtracted to eliminate the sharp features (either positive or negative) associated with the atmospheric H_2O spectrum.

12. The temperature-dependent set of the amide I′ spectra is modeled as a linear combination of the contribution of individual protein structural states:

$$D(\tilde{v}, T) = X_F(T)B_F(\tilde{v}) + X_{I_1}(T)B_{I_1}(\tilde{v}) + X_{I_2}(T)B_{I_2}(\tilde{v}) \\ + \ldots + X_U(T)B_U(\tilde{v}) \tag{1}$$

where $D(\tilde{v}, T)$ is the spectrum at temperature T, $X_I(T)$ are the populations of the protein "states" (F = folded, I_1, $I_2 \ldots$ intermediate, U unfolded), and $B_I(\tilde{v})$ are the corresponding spectra of each state. SVD is used to determine how many distinct "states" are necessary to describe the data (Fig. 2). SVD decomposes the experimental data as

$$\mathbf{D}(\tilde{v}, T) = \mathbf{U}(\tilde{v})\mathbf{S}\mathbf{V}^{\mathbf{T}}(T) = \mathbf{U}(\tilde{v})(\mathbf{V}(T)\mathbf{S})^{\mathbf{T}} = \mathbf{U}(\tilde{v})\mathbf{C}^{\mathbf{T}}(T) \tag{2}$$

where $\mathbf{U}(\tilde{v})$ is the matrix of the basis spectra, $\mathbf{S}$ is the diagonal matrix of singular values, and $\mathbf{V}(T)$ is the matrix of temperature-dependent coefficients (superscript T denotes matrix transpose). The number of significant components to be retained (the rest are discarded as "noise") can be determined based on the singular values $\mathbf{S}$ (Fig. 2) as well as by visual inspection of the individual $\mathbf{U}$ and $\mathbf{V}$ (14, 17).

13. Fitting of the thermodynamic models corresponds to finding the optimum parameters that determine the population matrix $\mathbf{X}(T)$ so that

$$\chi^2 = |(\mathbf{C}(T)) - (\mathbf{X}(T)\mathbf{A})|^2 \tag{3}$$

is minimized, where $\mathbf{C}(T) = \mathbf{V}(T)\mathbf{S}$ from the SVD and $\mathbf{A}$ is the transformation matrix which represents $\mathbf{C}(T)$ as a linear combination of the $\mathbf{X}(T)$ components. After parameter optimization, the $\mathbf{X}(T)$ and $\mathbf{A}$ are used to transform the SVD matrices as

$$\mathbf{D}(\tilde{v}, T) \approx \mathbf{U}(\tilde{v})(\mathbf{C}'(T))^{\mathbf{T}} = \left(\mathbf{U}(\tilde{v})\mathbf{A}^{\mathbf{T}}(T)\right)\mathbf{X}^{\mathbf{T}}(T)$$
$$= \mathbf{B}(\tilde{v})\mathbf{X}^{\mathbf{T}}(T) \tag{4}$$

which is eq. 1.

14. To extract the site-specific structural information it is necessary to decompose the spectra into ^{12}C and ^{13}C components (Fig. 4). As the ^{13}C signal is relatively weak, it is advantageous to use as basis spectra those obtained from the thermodynamic analysis of the complete amide I′, which also represent the underlying thermodynamic states common to all ^{13}C-labeled as well as the unlabeled protein.

15. In terms of the ^{12}C and ^{13}C basis functions ($\mathbf{B}_{12}(\tilde{v})$ and $\mathbf{B}_{13}(\tilde{v})$, respectively) the experimental spectra are

$$\mathbf{D}(\tilde{v}, T) = \mathbf{B}_{12}(\tilde{v}) \cdot \mathbf{C}_{12}^{\mathbf{T}}(T) + \mathbf{B}_{13}(\tilde{v}) \cdot \mathbf{C}_{13}^{\mathbf{T}}(T) \qquad (5)$$

where $\mathbf{C}_{12}(T)$ and $\mathbf{C}_{13}(T)$ are the corresponding temperature-dependent contributions. The $\mathbf{C}_{13}(T)$ is obtained from the matrix eq. 5 and fitted to a thermodynamic model in exactly the same manner as described above in eqs. 1–4 (Fig. 5).

Acknowledgements

References

1. Barth A, Zscherp C (2002) What vibrations tell us about proteins. Q Rev Biophys 35:369–430

2. Tadesse L, Nazarbaghi R, Walters L (1991) Isotopically enhanced infrared spectroscopy: a novel method for examining secondary structure at specific sites in conformationally heterogeneous peptides. J Am Chem Soc 113:7036–7037

3. Decatur SM (2006) Elucidation of residue-level structure and dynamics of polypeptides via isotope-edited infrared spectroscopy. Acc Chem Res 39:169–175

4. Huang R, Kubelka J, Barber-Armstrong W, Silva RA, Decatur SM, Keiderling TA (2004) Nature of vibrational coupling in helical peptides: an isotopic labeling study. J Am Chem Soc 126:2346–2354

5. Manas ES, Getahun Z, Wright WW, DeGrado WF, Vanderkooi JM (2000) Infrared spectra of amide groups in alpha-helical proteins: evidence for hydrogen bonding between helices and water. J Am Chem Soc 122:9883–9890

6. Walsh STR, Cheng RP, Wright WW, Alonso DOV, Daggett V, Vanderkooi JM, DeGrado WF (2003) The hydration of amides in helices; a comprehensive picture from molecular dynamics, IR, and NMR. Protein Sci 12:520–531

7. Brewer SH, Song BB, Raleigh DP, Dyer RB (2007) Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. Biochemistry 46:3279–3285

8. Fesinmeyer RM, Peterson ES, Dyer RB, Andersen NH (2005) Studies of helix fraying and solvation using C-13 isotopomers. Protein Sci 14:2324–2332

9. Marecek J, Song B, Brewer S, Belyea J, Dyer RB, Raleigh DP (2007) A simple and economical method for the production of C-13, O-18-labeled Fmoc-amino acids with high levels of enrichment: applications to isotope-edited IR studies of proteins. Org Lett 9:4935–4937

10. Kubelka J, Keiderling TA (2001) The anomalous infrared amide I intensity distribution in ^{13}C isotopically labeled peptide β-sheets comes from extended, multiple-stranded structures. An ab initio study. J Am Chem Soc 123: 6142–6150

11. Petty SA, Decatur SM (2005) Intersheet rearrangement of polypeptides during nucleation of beta-sheet aggregates. Proc Natl Acad Sci U S A 102:14272–14277

12. Sagle LB, Zimmermann J, Dawson PE, Romesberg FE (2004) A high-resolution probe of protein folding. J Am Chem Soc 126: 3384–3385

13. Huang CY, Getahun Z, Wang T, DeGrado WF, Gai F (2001) Time-resolved infrared study of the helix-coil transition using C-13-labeled helical peptides. J Am Chem Soc 123: 12111–12112

14. Amunson KE, Ackels L, Kubelka J (2008) Site-specific unfolding thermodynamics of a helix-turn-helix protein. J Am Chem Soc 130: 8146–8147

15. Huang CY, Getahun Z, Zhu YJ, Klemke JW, DeGrado WF, Gai F (2002) Helix formation

via conformation diffusion search. Proc Natl Acad Sci U S A 99:2788–2793

16. Silva RA, Kubelka J, Decatur SM, Bour P, Keiderling TA (2000) Site-specific conformational determination in thermal unfolding studies of helical peptides using vibrational circular dichroism with isotopic substitution. Proc Natl Acad Sci U S A 97:8318–8323

17. Henry ER, Hofricheter J (1992) Singular value decomposition: application to analysis of experimental data. Methods Enzymol 210:129–192

18. Lorenz-Fonfria VA, Padros E (2004) Curve-fitting overlapped bands: quantification and improvement of curve-fitting robustness in the presence of errors in the model and in the data. Analyst 129:1243–1250

19. Kauppinen JK, Moffatt DJ, Mantsch HH, Cameron DG (1981) Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. Appl Spectrosc 35: 271–276

20. Chan WC, White PD (2000) Fmoc solid phase peptide synthesis: a practical approach. Oxford University Press, Oxford

21. Goormaghtigh E, Cabiaux V, Ruysschaert JM (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. II. Experimental aspects, side chain structure, and H/D exchange. In: Hilderson HJ, Ralston GB (eds) Subcellular biochemistry. Plenum, New York, pp 363–403

Part IV

Other Spectroscopic Techniques

Chapter 21

Monitoring Structural Transitions in IDPs by Site-Directed Spin Labeling EPR Spectroscopy

Johnny Habchi, Marlène Martinho, Antoine Gruet, Bruno Guigliarelli, Sonia Longhi, and Valérie Belle

Abstract

Electron paramagnetic resonance (EPR) spectroscopy is a technique that specifically detects unpaired electrons. EPR sensitive reporter groups (spin labels or spin probes) can be introduced into biological systems via site-directed spin labeling (SDSL). This is usually accomplished by cysteine-substitution mutagenesis followed by covalent modification of the unique sulfhydryl group with a selective nitroxide reagent. SDSL EPR spectroscopy has been shown to be a sensitive and powerful method to study structural transitions within intrinsically disordered proteins (IDPs). In this chapter, we provide a detailed experimental protocol for this approach and present a few examples of EPR spectral shapes illustrative of various mobility regimes of the spin probe, reflecting different protein topologies.

Key words: Site-directed spin labeling, Electron paramagnetic resonance spectroscopy, Nitroxide radical, SDSL EPR, MTSL, Spin labels, Spin probes, Intrinsically disordered proteins, Induced folding, Structural transitions

1. Introduction

Electron paramagnetic resonance (EPR, also termed electron spin resonance, ESR) spectroscopy studies chemical species that have unpaired electrons. The application of EPR to the study of proteins requires the presence of either paramagnetic metal ions or organic free radicals. Thus, EPR is used either for studying metalloproteins (containing natural paramagnetic centers such as Mn^{2+}, Cu^{2+}, or Fe^{3+}), or proteins labeled by specific spin labels (spin probes). Site-directed spin labeling (SDSL) combined with EPR is a powerful technique for detecting structural changes in proteins. The strategy of SDSL involves the insertion of a paramagnetic label at a selected site of a protein and its observation by EPR spectroscopy (for reviews
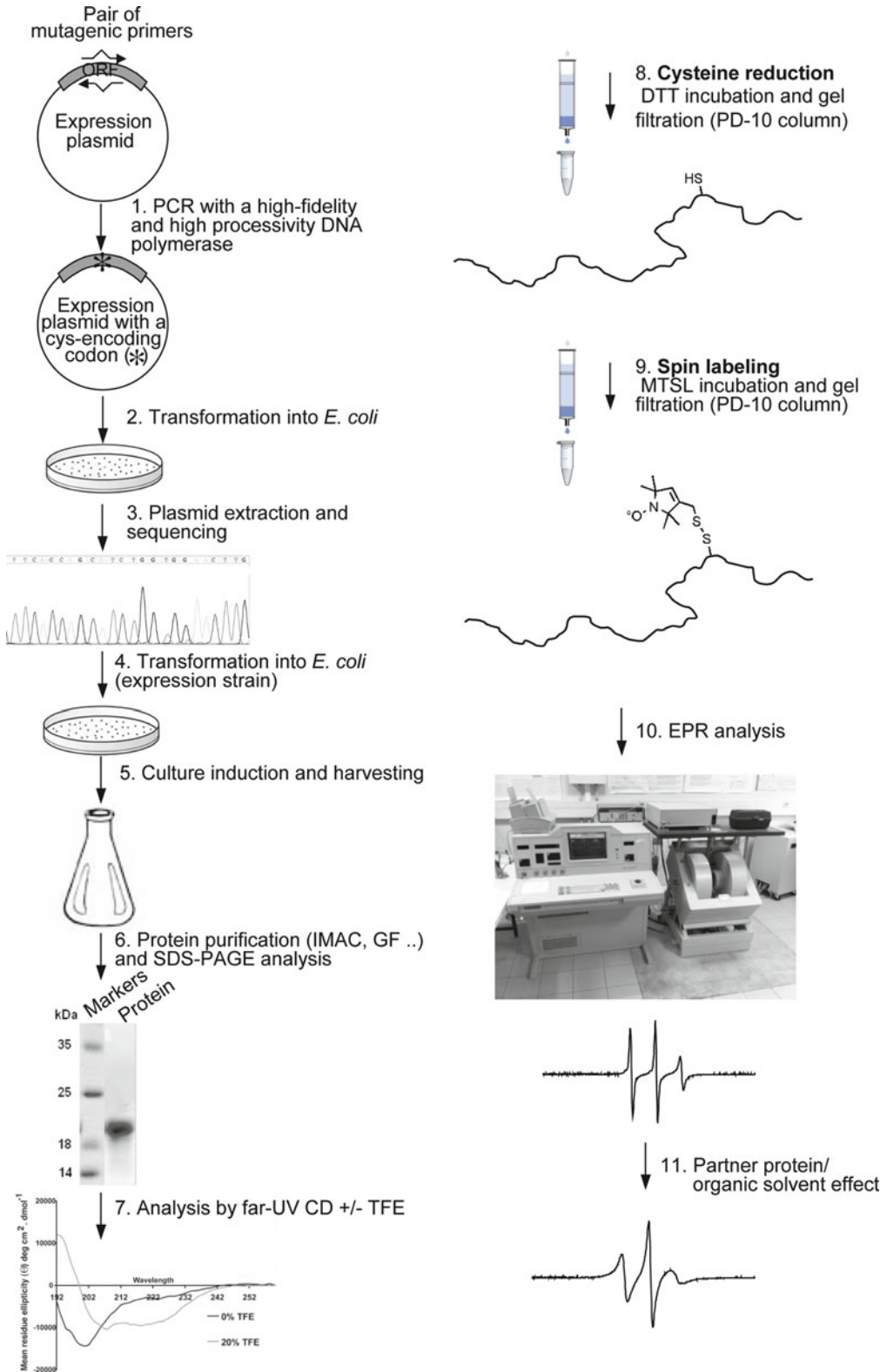
Fig. 1. Schematic outline of SDSL EPR spectroscopy.

see refs. 1–9). The paramagnetic label is usually a stable nitroxide reagent such as 1-oxyl-2,2,5,5-tetramethyl-δ3-pyrroline-3-methyl methane thiosulfonate (MTSL) that is covalently attached to a specific residue (e.g., cysteine). The latter is introduced at the desired position via site-directed mutagenesis. The presence of the label does not usually affect the protein properties thanks to its rather small size which is comparable to that of a tryptophane residue (7 Å).

Since the pioneering works of the group of Hubbell (10, 11), SDSL EPR spectroscopy has emerged as a valuable tool for studying structural transitions in a wide range of proteins, and has been shown to be very well suited to study induced folding events within intrinsically disordered proteins (IDPs) (for examples see refs. 12–14).

Depending on the system under study, a variety of EPR approaches could be used. The common approach consists in the analysis of the influence of the nitroxide motion on EPR spectral shapes. This approach is particularly well suited to monitor conformational changes induced by either protein–protein or protein–ligand interactions, since such interactions modify the mobility of the label, and consequently, the EPR spectrum of the spin-labeled protein (4, 12, 13). In addition, SDSL can also be used to measure distances in biological systems. The method requires the introduction of two spin probes in the protein of interest. Then, the distance between the two probes can be measured through spin–spin interaction analysis detected by double electron electron resonance (DEER) sequence requiring pulsed EPR technology (15–18).

This chapter provides a general guideline to SDSL EPR spectroscopy (see Fig. 1). The described procedure is applicable to many (if not all) types of protein samples and is intended to address the following questions: how to design a single-cysteine-substituted protein? How to graft the spin label on the cysteine residue? Does this substitution affect the protein secondary structure and its folding capacities? And finally, how to record and analyze EPR spectra?

2. Materials

Prepare all solutions using ultrapure MilliQ® water (prepared by purifying deionized water to attain a resistivity of 18 MΩ cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations when disposing waste materials.

2.1. In Vitro Site-Directed Cysteine-Substitution Mutagenesis

2.1.1. Polymerase Chain Reaction Mutagenesis for Cysteine Substitution and Transformation of Escherichia coli Competent Cells

1. Expression plasmid template encoding the protein of interest with a purification tag.
2. Synthetic, mutagenic oligonucleotide primers (sense and reverse) bearing a cysteine-encoding TGC/T codon (Operon).
3. *Pfu* turbo thermostable DNA polymerase (Stratagene) for polymerase chain reaction (PCR) amplification (or a similar, high fidelity and highly processive thermostable DNA polymerase).
4. dNTPs mixtures (2.5 mM each).
5. 10× reaction buffer.
6. *Dpn* I restriction enzyme (NE Biolabs).
7. Agarose gel electrophoresis equipment and TBE running buffer.
8. PCR thermocycler.
9. Ethidium bromide stock at 10 mg/mL.
10. DNA ladder (Invitrogen).
11. UV trans-illuminator for visualizing DNA in agarose gels stained with ethidium bromide.
12. Agarose gel loading buffer 6×.
13. *E. coli* TAM1 chemically competent cells (Active Motifs).
14. Petri agar plates with appropriate medium for bacterial growth (e.g., Luria Bertani or 2YT), sterile super optimal broth with catabolite repression (SOC), and appropriate antibiotics stock solutions.
15. Sterile 10 mL culture tubes and 1.5 mL Eppendorfs.
16. Incubator set at 45 °C.
17. Incubator set at 37 °C.

2.1.2. DNA Plasmid Extraction

1. Shaking incubator set at 37 °C, sterile culture tubes, culture medium, and antibiotics.
2. Macherey-Nagel (or equivalent) kit for plasmid DNA extraction.
3. Microfuge and 1.5 mL Eppendorfs.
4. Ethidium bromide stock at 10 mg/mL.
5. DNA ladder (Invitrogen).
6. Agarose gel electrophoresis equipment (Bio-Rad) and TBE running buffer.
7. UV trans-illuminator for visualizing DNA in agarose gels stained with ethidium bromide.

2.2. Protein Expression and Purification

2.2.1. Transformation of E. coli Expression Strain with the Mutated Plasmid Construct

1. Chemically competent *E. coli* Rosetta [DE3] pLysS strain (Novagen) for recombinant proteins expression.
2. Incubator set at 45 °C.
3. Sterile SOC medium and appropriate antibiotics stock solutions.
4. Sterile 10 mL culture tubes and 1.5 mL Eppendorfs.
5. LB (or 2YT) agar plates supplemented with appropriate antibiotics.
6. Shaking incubator set at 37 °C.

2.2.2. Cell Growth and Protein Expression

1. Rosetta [DE3] pLysS cells containing the protein expression vector that produces the cysteine-substituted protein.
2. Appropriate medium for cell growth and protein expression (e.g., Luria Bertani).
3. Sterile appropriate antibiotics stock solutions.
4. Sterile 250-mL and 3-L baffled-bottom flasks.
5. Inducer (such as Isopropyl-thio-β-D-galactopyranoside (IPTG) at 0.5 M).
6. Shaking incubator set at 37 °C.
7. Perkin UV–vis absorption spectrophotometer.
8. High-speed centrifuge (Sorvall).

2.2.3. 6× His-Tagged Protein Purification

1. Cellular pellets from induced cultures expressing the recombinant cysteine-substituted protein.
2. Buffers: 50 mM Tris–HCl pH 8.0, 300 mM NaCl, and 1 mM phenyl-methyl-sulphonyl-fluoride (PMSF) supplemented with either 10 mM imidazole for cell lysis, or 20 mM imidazole for resin wash or 250 mM for protein elution.
3. Stock solutions of lyzozyme (50 mg/mL), DNase (2 mg/mL), and $MgSO_4$ (2 M).
4. Complete EDTA-free protease inhibitor cocktail (Roche).
5. Sonicator, Sorvall centrifuge, microfuge, Perkin UV–vis absorption spectrophotometer.
6. Chelating Sepharose Fast Flow Resin (GE Healthcare) pre-loaded with Ni^{2+} ions.
7. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel, 2× Laemli SDS-PAGE sample buffer, running buffer, molecular mass markers (GE Healthcare), electrophoresis cell, and power supply (Bio-Rad).
8. Staining and destaining solutions for sodium SDS-PAGE gels.
9. ÄKTA Explorer Chromatography System (GE Healthcare).

10. 0.22-μm polyethersulfone filters (Millipore).

11. Prepacked Superdex 200 or 75 HR 16/60 columns (GE Healthcare).

12. Degassed and filtered 10 mM sodium phosphate pH 7 buffer, either supplemented or not with NaCl (50–500 mM) for protein elution from gel filtration columns.

13. ÄKTA collecting tubes, 1.5-mL Eppendorfs.

14. Centricon Plus-20 concentrators (Millipore) with appropriate molecular cutoff.

15. UV–vis absorption spectrophotometer and Bradford reagent (Bio-RAD).

2.3. Protein Characterization Using Far-UV Circular Dichroism

1. Stock solution of the protein of interest.

2. 2,2,2-trifluoroethanol (TFE) (Roth).

3. Jasco 810 dichrograph spectropolarimeter equipped with a Peltier thermoregulation system.

4. 1-mm thick quartz cells.

5. PD-10 desalting columns containing 8.3 mL of Sephadex™ G-25 Medium (GE Healthcare) for buffer exchange.

6. 10 mM sodium phosphate buffer at pH 7.

7. SpectraManager program for spectra smoothing and analysis.

2.4. Spin Labeling

1. Single-cysteine mutated protein.

2. 10 mM MES 150 mM NaCl buffer pH 6.5.

3. 10 mM sodium phosphate buffer pH 7.0.

4. 1.5 and 0.5 mL Eppendorfs.

5. Dithiotreitol solution (DTT), Sigma-Aldrich.

6. Spin label: 1-oxyl-2,2,5,5-tetramethyl-δ3-pyrroline-3-methyl-methane-thiosulfonate (MTSL) (Toronto Research Chemicals Inc.).

7. PD-10 Desalting Columns containing 8.3 mL of Sephadex™ G-25 Medium (GE Healthcare).

8. Argon gas.

9. Vivaspin 4-mL concentrators (Sartorius Stedim Biotech) with appropriate cutoff.

10. Perkin UV–vis absorption spectrometer.

2.5. EPR Spectroscopy

1. X-band (9.9 GHz) BRUKER ESP 300E EPR spectrometer.

2. ELEXSYS Super High Sensitivity resonator (height of detection of 40 mm).

3. Quartz capillary of 1.0–1.1 mm inner diameter corresponding to a sample volume of 30–40 μL.

4. 1.5 and 0.5 mL Eppendorfs.

5. 100 μL syringe (SGE or Hamilton) equipped with a 20 cm-long needle.

6. Purified partner protein known to interact with the labeled protein.

7. Sucrose (Sigma-Aldrich).

8. TFE (Roth).

3. Methods

SDSL EPR spectroscopy is a powerful approach to monitor and map disorder-to-order transitions that IDPs may undergo in the presence of their partner(s). Deciphering these folding events by SDSL EPR spectroscopy implies the generation and the study of a collection of spin-labeled variants of the IDP of interest, each of them bearing a single spin label grafted at a specific position. The higher the number of variants, the more detailed the study will be, as information about the mobility of spin labels grafted at numerous protein sites will be collected. Typically, the interaction site can be mapped and information about the precise boundaries of the region undergoing structural transitions can be obtained, as well as indications about possible preconfiguration in solution of molecular recognition elements (MoREs) (19–22), thereby allowing the molecular mechanism of complex formation to be inferred and plausible binding models to be proposed (for examples see refs. 12, 13, 23).

3.1. Site-Directed Spin Labeling

3.1.1. Construction of a Cysteine-Substituted Protein

The first step in view of protein labeling for EPR studies is the analysis of the amino acid sequence of the protein of interest. The latter should in fact contain a single reactive amino acid residue that can be selectively modified with an appropriate spin label. The sulfhydryl group of cysteines constitutes a suitable target for grafting nitroxide radicals such as the MTSL. Accordingly, site-directed mutagenesis is often required to introduce a cysteine residue at a specific protein site. On the other hand, if a cysteine residue occurs naturally within the protein, two situations can be distinguished: (1) the cysteine residue does not occur in the region of interest, and therefore an additional site-directed mutagenesis step aimed at replacing the sulfhydryl group by an hydroxyl group of a serine is required (unless the natural cysteine is buried and hence not accessible to the spin label) or, (2) a second cysteine residue is desirable in view of measuring interspin interactions under various

conditions so as to infer interspin distances (reflecting interresidue distances). In either case, a step of site-directed mutagenesis is required. Whenever possible, residues allowing isosteric substitutions (i.e., serines) should be targeted for cysteine substitution so as to introduce minimal structural perturbation. Putative MoREs within a given IDP represent the primary choice for the introduction of cysteine residues in view of mapping the site(s) involved in binding to a partner. Indications about putative MoREs can be obtained by the use of the PONDR VL-XT predictor (24), by the ANCHOR server (25), or by inspection of the HCA plot (26) (for examples see refs. 19, 27–31 and references therein cited).

Numerous mutagenesis methods have been developed based upon PCR. The simplest and most broadly applicable protocol is currently the QuickChange™ Site-Directed Mutagenesis System (QCM) developed by Stratagene. With this approach, the mutation is introduced in a single PCR with one pair of complementary primers containing the mutation of interest. However, this method is restricted to primer pairs of 25–45 bases in length with melting temperature (T_m) > 78 °C (see Note 1). Otherwise, primer dimer formation will become more favorable compared to the primer–template annealing, in particular for primer pairs with multiple mismatches, and the protocol has to be modified case by case, or more steps have to be carried out, e.g., by using an alternative two-stage PCR protocol.

3.1.2. PCR Mutagenesis Reaction and Transformation of E. coli Competent Cells with PCR Product

The following site-directed mutagenesis protocol is adapted to the use of *Pfu* turbo thermostable DNA polymerase (see http://www.stratagene.com). It allows for the introduction of a cysteine residue by modifying the DNA sequence using a single step of PCR. For optimal results, PCR primers should be designed so as to have similar T_m (see Note 1).

1. The PCR reaction mix is prepared as follows (see Note 2): 1 μL template DNA (≈10–20 ng/μL), 5 μL thermostable DNA polymerase 10× reaction buffer, 4 μL dNTP mix stock solution at 10 mM, 1.5 μL of each primer (39-mer each) at 10 μM (see Note 1), 1 μL *Pfu* turbo DNA polymerase (2.5 U/μL), and sterile H_2O up to 50 μL. If the thermocycler is devoid of a heated lid, overlay each reaction mixture with a few drops of mineral oil.

2. The reaction is placed in the PCR thermal cycler with the following program (see Note 3): an initial denaturation step of 1 min at 95 °C, followed by 18 cycles (see Note 4) consisting each of a denaturation step at 95 °C for 30 s, an annealing step at 50 °C for 1 min, and an elongation step at 68 °C for 10 min for a template of approximately 5 kb (the elongation rate of *Pfu* turbo DNA polymerase is of 0.5 kb/min).

3. The amplification can be checked by electrophoresis of 5 μL of the product on a 1 % agarose gel (see Note 5).

4. Add 1 μL of *Dpn* I restriction enzyme (10 U/μL) directly to the amplification reaction.

5. Gently and thoroughly mix the reaction mixture by pipetting the solution up and down several times, then immediately incubate the reaction at 37 °C for 1 h to digest the parental (i.e., the nonmutated) plasmid DNA.

6. An additional electrophoresis step for PCR yield estimation can be carried out (see Notes 5 and 6).

7. Transform TAM1 chemically competent cells using standard protocols (32).

8. Plate the cells onto agar plates containing the appropriate antibiotic and incubate at 37 °C for 18 h.

9. Peak the brightest colonies, culture them overnight in liquid medium at 37 °C with agitation (250 rpm), and extract plasmid DNA using the Macherey-Nagel kit for DNA extraction (according to manufacturer's instructions).

10. Check plasmids by agarose gel electrophoresis

11. Sequence candidate clones (GATC Biotech, Genome Express.) to ensure sequence integrity and proper reading frame.

3.1.3. Cell Growth and Protein Expression

The introduction of a foreign DNA into cells is a standard procedure that is routinely applied to produce recombinant proteins for biochemical studies. Different expression systems (i.e., bacteria, mammals, insects, yeasts) have been described and experimentally validated for the production of recombinant proteins. Although all available expression systems are used for roughly the same applications (i.e., structural analysis, functional assays, protein interactions), each one possesses its own advantages as well as its potential drawbacks. For instance, although expression in bacteria is commonly and widely used (scalable, low cost, compatible with different cloning systems), protein solubility can constitute a serious concern, especially when dealing with eukaryotic proteins or with proteins undergoing posttranslational modifications required for proper protein folding. Therefore, using the best expression system for your protein is the key to your success. In general, solubility, functionality, speed and yield are often the most important factors to be considered when choosing an expression system. Moreover, the growth conditions should be properly controlled. Culture conditions and the induction of expression have profound effects on the way the recombinant proteins are produced. It is therefore advisable to empirically establish optimal conditions with small-scale cultures before purification on a large-scale is attempted. Moreover, other problems with protein expression may occur, including cell toxicity

(see Note 7), protein instability (see Note 8), formation of inclusion bodies (see Note 9), improper processing or posttranslational modification, and inefficient translation.

As a general guideline, expressing the cysteine-substituted protein of interest using the same heterologous system used for the expression of the *wt* protein is advised.

The following protocol focuses on bacterial (*E. coli*) expression. Although the bacterial heterologous system is not applicable to all kinds of proteins, especially eukaryotic ones with intricate folds, it is still the most commonly used owing to its low cost, rapidity and easiness of implementation, as well illustrated by the fact that 70 % of structures deposited within the PDB correspond to proteins expressed in *E. coli*. With careful choice of host strains, vectors, and growth conditions, many recombinant proteins can expressed at high levels in *E. coli*, and tuning of these parameters often allows increasing expression and solubility levels (33).

1. Transform chemically competent Rosetta [DE3] cells, already containing or not the pLysS plasmid (see Note 10), with the expression vector encoding the protein of interest bearing a cysteine at the desired position, and then select transformants on LB agar plates containing the appropriate antibiotics (see Note 11). Incubate the plate overnight at 37 °C.

2. Inoculate 0.5 mL of LB medium containing the appropriate antibiotics in a culture tube with a single colony from the plate, and incubate for 3–4 h at 37 °C with shaking (250 rpm).

3. Withdraw an aliquot (250 μL) of the culture, add glycerol to a final concentration of 30 % and store the cell stock at −80 °C.

4. Inoculate 50 mL of the same medium in a 250-mL baffled-bottom flask with the rest of the liquid culture and incubate overnight at 37 °C with shaking until saturation is reached.

5. The next morning, inoculate 1 L culture of the same medium (see Note 12) with the appropriate antibiotics in a 3 L baffled-bottom flask with the saturated overnight culture.

6. Grow the cells at 37 °C with shaking to mid-log phase ($OD_{600nm} \approx 0.6$–0.8).

7. Add IPTG to a final concentration of 0.2–1 mM. Incubate for 3–4 additional hours (see Note 13).

8. Recover the cells by centrifugation ($5,000 \times g$). Freeze the cell pellets at −20 °C.

3.1.4. Cell Lysis and Protein Purification

This section describes a smooth path to protein purification. Very often, more than one purification step is necessary to reach the desired purity. Therefore, the key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements and combine them in a

logical way to maximize yield and minimize the number of steps required. Various purification tags are available (such as Trx, NusA, GST, Strep, 6×-His, MBP, GB1) and can be used to both improve protein solubility and allow purification by affinity chromatography of recombinant proteins (34). Unlikely larger tags (GST, MBP, Trx) that generally have to be cleaved off by specific proteases recognizing cleavage sites located between the tag and the protein of interest, the 6× His tag does not have necessarily to be removed, as its very small size in most cases does not affect the protein properties.

The purification procedure described in this section concerns 6×-His tagged proteins (see Note 14) and comprises two steps of chromatography. The protein is first bound to a resin (either Ni-NTA or Sepharose resin preloaded with Ni^{2+} ions) in an Immobilized Metal Affinity Chromatography (IMAC) step. Subsequently, the eluted protein is loaded onto a size exclusion chromatography (SEC) column.

IMAC purification may be achieved with either a batch or a column procedure. In the procedure described below, the IMAC step is performed in batch. This strategy enhances binding of 6×-His tagged proteins and is therefore recommended when the tag is not fully accessible or when the protein of interest is present at very low concentrations. Even when the tag is not completely accessible, the protein will bind as long as more than two histidine residues are available to interact with the nickel ions. A big advantage of the IMAC approach is that the affinity of histidine for the Ni resin is greater than that of antibody–antigen or enzyme–substrate pairs. Consequently, the addition of low concentrations of imidazole in the binding and washing buffers is very effective in preventing contaminating proteins from binding to the resin. The imidazole ring is part of the structure of histidine and is the moiety responsible for binding to the nickel ions. When the imidazole concentration is increased, the 6×-His tagged protein will dissociate from the resin because it can no longer compete for binding sites on the resin.

After IMAC, we recommend to carry out a gel filtration chromatography (or SEC) step so as to remove contaminants and aggregates, but also to estimate the monodispersity and oligomeric state of the protein. SEC separates molecules according to differences in their size (i.e., in their hydrodynamic radius). As such, this step also enables estimation of the impact of the cysteine substitution onto the compactness of the protein. Unlike affinity chromatography, molecules do not bind to the column matrix and consequently conditions can be varied to suit the type of sample or the requirements for further purification, analysis or storage.

Below is provided a general protocol for the purification of a 6× His tagged protein, where all steps, except gel filtration, are performed at 4 °C. Note that the cysteine substitution may result in

a protein form behaving slightly differently with respect to the *wt* protein, thereby rendering necessary to adjust and optimize the purification protocol on a case-by-case basis.

1. Thaw the cellular pellets on ice and suspend in 5 volumes (v/w) of ice-cold buffer A (50 mM Tris–HCl buffer pH 8, 300 mM NaCl, 10 mM imidazole, and 1 mM PMSF). Supplement buffer A with 0.1 mg/mL lysozyme, 10 μg/mL DNAse, 20 mM MgSO$_4$ and protease inhibitor cocktail (Roche) (one tablet for 50 mL of bacterial lysate) (see Note 15).

2. After a 20-min incubation with gentle agitation, disrupt the cells by sonication.

3. Clarify the crude cell extract by high-speed centrifugation (30,000 × *g*) followed by filtration (see Note 16).

4. Incubate the clarified supernatant for 1 h with gentle shaking with Chelating Sepharose Fast Flow resin preloaded with Ni^{2+} ions previously equilibrated in buffer A (2–4 mL resin per liter of bacterial culture).

5. Wash the resin with 10 volumes of buffer A supplemented with 20 mM imidazole (see Note 17) and then elute the protein with 3 volumes of buffer A supplemented with 250 mM imidazole.

6. The presence of the desired product may be checked by SDS-PAGE analysis.

7. Load the protein onto a gel filtration prepacked Superdex column (GE Healthcare) previously equilibrated in elution buffer (see Notes 18 and 19). A Superdex 200 or 75 HR 16/60 column can be used depending on whether the molecular mass of the protein of interest is higher or smaller than 70 kDa, respectively.

8. Monitor protein elution at 280 nm (see Note 20) and collect 1.5 mL fractions. Analyze by SDS-PAGE the fractions of interest and pool those exhibiting a sufficient purity level (see Note 21).

9. Concentrate the protein up to a concentration of 10 mg/mL (see Note 22), using Centricon Plus-20 concentrators of the appropriate cutoff.

10. Stock the protein at −20 °C (see Note 23).

3.1.5. Far-UV Circular Dichroism Analysis

A crucial point prior to further analysis is to ascertain that the cysteine substitution does not affect the overall secondary structure content of the protein of interest. It is therefore strongly recommended to record the far-UV circular dichroism (CD) spectrum of the cysteine variant and to compare it to the spectrum of the native protein (see Note 24). In addition, it is also suggested to record the CD spectrum of the variant in the presence of TFE so as to rule out

possible dramatic impact of the cysteine substitution on the folding abilities of the variant. The TFE solvent is a secondary structure stabilizer that mimics the hydrophobic environment experienced by proteins in protein–protein interactions, and that is therefore widely used as a probe to unveil disordered regions having a propensity to undergo an induced folding (35). Depending on the protein structural propensities, TFE concentrations as high as 40 % can be required to observe structural transitions (36). In most cases however, a significant disorder-to-order transition takes place with 20 % TFE. It is recommended to record the CD spectrum of the cysteine variant in the presence of the TFE concentration that triggers the most dramatic structural transitions in the native protein, and then to compare the estimated secondary structure content of the variant to that of the *wt* protein.

1. Prepare a sample of both the *wt* and cysteine-substituted protein at 0.1 mg/mL in 10 mM sodium phosphate pH 7.0 either in the presence or in the absence of TFE (see Note 25).

2. Record the CD spectra at 20 °C on a spectropolarimeter (Jasco 810 dichrograph), equipped with a Peltier thermoregulation system, using 1-mm thick quartz cells. Record the spectra in the 185–260 nm range, with a scanning speed of 20 nm/min, and a data pitch (wavelength increment) of 0.2 nm. Average each spectrum from three scans.

3. Record the CD spectra of buffer (containing or not TFE) and correct each protein spectrum for buffer signal. After smoothing the spectra by using a third-order least square polynomial fit, calculate mean ellipticity values per residue ($[\Theta]$) as follows; $[\Theta] = 3{,}300 \ m \ \Delta A/(lcn)$, where l (path length in cm), $n =$ number of residues, $m =$ molecular mass in daltons and $c =$ protein concentration expressed in mg/mL.

4. Analyze the spectra using the DICHROWEB Web site (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (37, 38). Use the CDSSTR deconvolution method and the reference protein set 7 to estimate the content in α-helical and disordered structure of each protein sample.

3.1.6. Spin Labeling

A critical point for achieving high labeling yields is the reduction of cysteine residues prior to incubation with the nitroxide reagent (MTSL). Incubation of the protein with an appropriate reducing agent, such as DTT, is thus carried out as the first step and a gel filtration onto a desalting column is then carried out to remove excess of reducing agent. Once DTT has been removed, it is recommended to immediately proceed to spin labeling to avoid reoxidation of cysteines. In our experience, incubating the protein with MTSL under an argon flow gives rise to higher labeling yields,

although successful spin labeling has also been reported in the absence of argon (14).

1. Add a 100 molar excess of DTT to 200 µL of purified cysteine-substituted protein typically in a concentration range of 200–500 µM and incubate for 30 min in an ice bath.

2. Equilibrate the PD-10 desalting column with ≈30 mL of 10 mM MES pH 6.5, 150 mM NaCl.

3. Add the mixture to the column and let it enter the packed bed. Add equilibration buffer to adjust the volume up to 1 mL and let it enter the packed bed completely. Discard the flow through.

4. Use Eppendorf (1.5 mL) tubes for fraction collection. Elute with 500 µL equilibration buffer and collect the eluate. Repeat four times. This constitutes the dead volume that should not contain any protein (check anyway that this is the case).

5. Use Eppendorf (0.5 mL) tubes for fraction collection. Elute with 250 µL equilibration buffer and collect the eluate. Repeat eight times. Eight fractions of 250 µL each are thus collected. Keep these fractions in an ice bath.

6. Each fraction is checked for protein concentration by measuring the absorbance at 280 nm (see Note 22). The fractions containing the protein are pooled (typical final volume of 1 mL).

7. The MTSL spin label is added to the pool of reduced protein in a molar excess of 10. The mixture is kept under a flow of Ar in an ice bath, in the obscurity and under gentle stirring for 1 h.

8. Equilibrate a second PD-10 desalting column with ≈30 mL of 10 mM sodium phosphate pH 7.0 buffer so as to remove the excess of MTSL.

9. Use Eppendorf (1.5 mL) tubes for fraction collection. Elute four fractions of 500 µL each using equilibration buffer.

10. Use Eppendorf (0.5 mL) tubes for fraction collection. Elute 12 fractions of 250 µL each (numbered from 1 to 12) using equilibration buffer and keep them in an ice bath.

11. Every fraction is checked (1) for protein concentration by measuring the absorbance at 280 nm (see Note 22) and (2) for spin concentration by recording an EPR spectrum at room temperature 296 K (±1 K) on approximately 30 µL of solution injected in a quartz capillary (see below for details) (for an example of a typical elution profile see Fig. 2).

12. Fractions containing the labeled protein and giving rise to a sufficient EPR signal are pooled (usually fractions 2–6, see Note 26) and concentrated up to a final concentration of 200–500 µM using a concentrator with an appropriate cutoff.
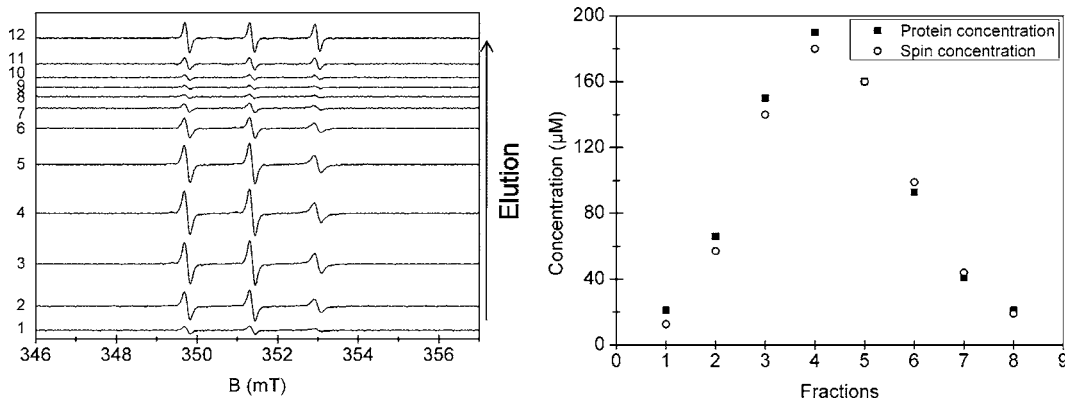
Fig. 2. *Left*: EPR spectra recorded at room temperature of fractions (numbered from 1 to 12) eluted from the PD-10 column following incubation with the spin label. *Right*: Protein and spin concentration analyses for each eluted fraction. Protein concentration is obtained by the measurement of the absorbance at 280 nm, and spin concentration by the double integration of the EPR signal (see text for details). Note that the labeled protein is eluted from fraction 1 to 8 and that the EPR signals of fractions 9–12 correspond to the elution of the free radical (typical line shape with three narrow lines of equivalent amplitude).

The concentrated sample is then checked for protein and spin concentration in order to determine the labeling yield (see below for details on EPR and spin concentration calculation). Concentrations of labeled proteins in the range of 20–100 μM are required to obtain EPR spectra with a good signal to noise ratio.

3.2. EPR Spectroscopy

3.2.1. Optimizing Acquisition Parameters

The analysis of the nitroxide spin probe mobility is based on the analysis of the EPR spectral shape recorded at room temperature (39). It is thus crucial to record spectra without distorting the EPR spectral shape and with an optimized signal to noise ratio (SNR). Two major parameters have to be carefully adjusted to avoid signal distortions: the microwave power (P) and the modulation amplitude of the magnetic field (MA).

1. Record EPR spectra with increasing values of P. The EPR signal increases linearly with $\sqrt{P}$ until saturation occurs, a condition leading to a distortion of the EPR line shape. Evaluate the optimal value of P by choosing a value slightly inferior to the beginning of saturation phenomenon.

2. Record EPR spectra with increasing values of MA. The intensity of the EPR signal increases linearly with MA until it becomes larger than the linewidth of the EPR signal. In that case, the detected EPR signal broadens and becomes distorted. Evaluate the optimal value of MA by choosing a value slightly inferior to the beginning of distortion.

3. Record the final spectrum of the labeled sample with the optimized values of both P and MA and accumulate if necessary to obtain a spectrum with a convenient SNR.

1. Record the EPR spectrum of a standard solution of known concentration (ideally a 100 μM stable nitroxide radical solution) and measure the double integration of the EPR signal I_{std}.

2. Record the EPR spectrum of the labeled protein of unknown concentration and measure the double integration of the EPR signal I.

3. Spin concentration C_{spin} is obtained by calculating:

$$C_{spin} = C_{std} \frac{I}{I_{std}} \frac{RG_{std}}{RG} \frac{MA_{std}}{MA} \sqrt{\frac{P_{std}}{P}} \frac{NA_{std}}{NA},$$

where I is the result of the double integration of the signal of the labeled protein, RG the receiver gain, MA the modulation amplitude of the magnetic field, P the microwave power, and NA the number of accumulations. The subscript std corresponds to the standard solution. Insure that the field resolution of the spectrum is taken into account in the calculation of I (check in the user's manual of the spectrometer).

4. Divide the spin concentration by the protein concentration (deduced from the absorbance at 280 nm, see Note 22) to obtain the labeling yield. Labeling yields for IDPs typically range from 50 to 80 %, reflecting the overall accessibility of cysteine residues.

3.2.3. Spectral
Shape Analyses

The EPR signal of the nitroxide radical is characterized by three lines as a result of the hyperfine interaction between the spin of the unpaired electron and the nuclear spin of the nitrogen atom ($I = 1$ for ^{14}N). When grafted onto a protein, the mobility of the spin probe is restricted and, as a consequence, the anisotropy of the hyperfine interaction is partially averaged, leading to shape modification of the EPR signal (for reviews see refs. 2–4, 8, 9) (see Note 27 and Fig. 3). Semiquantitative analyses can be used to describe this restricted mobility of the probe that arises from backbone fluctuations, internal dynamics of the spin-labeled lateral chain and possible tertiary contacts. Semiquantitative analysis relies on measuring in EPR spectra a parameter that directly depends on the mobility of the spin probe (40, 41) (see Fig. 3). In the fast motional regime of the spin label, the parameter of choice is the ratio between the peak-to-peak amplitude h of the lateral lines arising from the different spin states of the nitroxide ^{14}N nucleus $M_I = -1$ and $M_I = +1$, referred to as $h(+1)$ or $h(-1)$, and that of the central line ($M_I = 0$), referred to as $h(0)$. For a fast anisotropic motion of the label, the $h(+1)/h(0)$ ratio was found to be a more sensitive parameter than the $h(-1)/h(0)$ ratio (12) (for an example of such a motion see Fig. 3, spectrum b). On the other hand, in case of a probe being in the intermediate/slow motional regime, the outer line splitting ΔB (see Fig. 3) is a well-suited and sensitive parameter
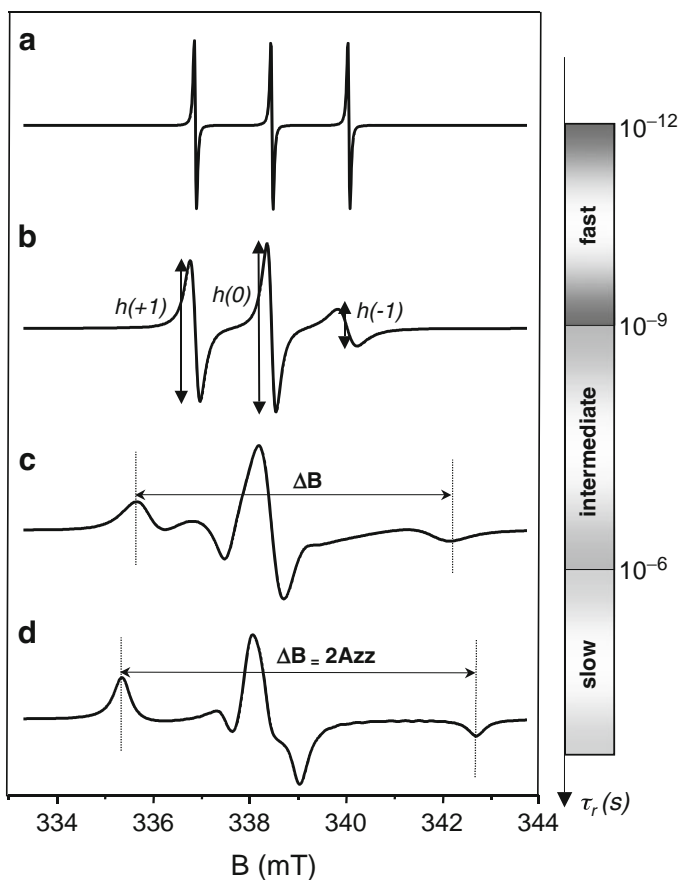
Fig. 3. Set of various simulated EPR spectra (EasySpin software, (44)) of a nitroxide radical corresponding to different isotropic motional regimes characterized by various rotational correlation times τ_r (1.10^{-12} s (a); 1.10^{-9} s (b); 1.10^{-8} s (c); 1.10^{-5} s (d)). The different semiquantitative parameters used to characterize the spectra, namely, the amplitude of the derivative lines (b) and the outer line splitting ΔB (c) are represented. When the probe is immobile (d), the outer line splitting corresponds to $2A_{zz}$, A_{zz} being the z-component of the hyperfine tensor (Modified from (23)).

to report the restricted mobility of the probe (for an example of such a restrained motion see Fig. 3, spectrum c) (42). Using these parameters, chosen as a function of the motional regime, the impact of a partner protein and/or of organic solvents on the spin label mobility can be evaluated.

Effect of Binding to a Partner Protein

Addition of a partner protein can profoundly affect the spectral shape of a spin-labeled protein. Indeed, as a function of the location of the spin label, complex formation may lead to a decrease in the spin label mobility arising from gain of regular structure, establishment of interactions at the binding interface or steric hindrance, or from a combination of these effects.
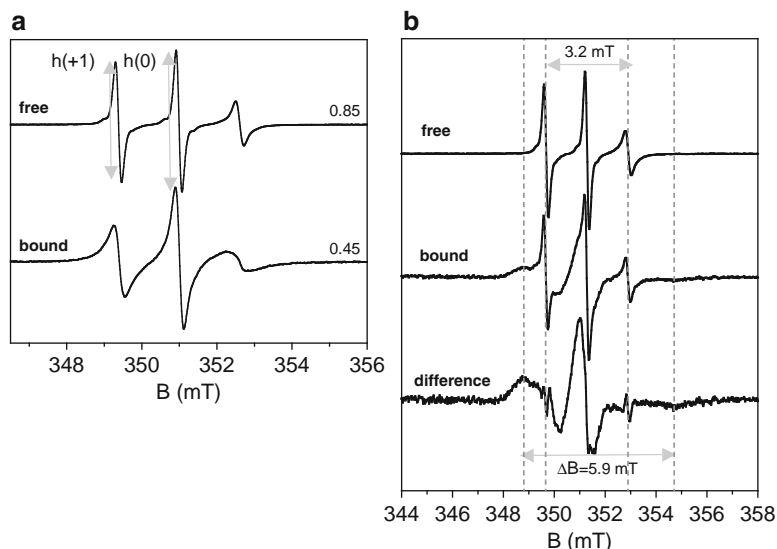
Fig. 4. Interaction of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL} aa 401–525) with the C-terminal X domain (XD, aa 459–507) of the phosphoprotein. Normalized EPR spectra of the labeled L496C N_{TAIL} variant (**a**) and S491C N_{TAIL} variant (**b**) recorded at room temperature in the absence and presence of a saturating amount of XD. The microwave power was 10 mW and the magnetic field modulation amplitude was 0.1 mT. For the S491C spin-labeled N_{TAIL} variant (**b**), the EPR spectrum at the *bottom* is obtained by subtracting the spectrum of unbound labeled N_{TAIL} from the spectrum of recorded in the presence of XD. The proportion of the unbound and bound form are 25 ($\pm$5) % and 75 ($\pm$5) %, respectively. Semiquantitative parameters used to describe the spectral shape modification are illustrated: $h(+1)/h(0)$ (**a**) and ΔB (**b**). Data are taken from (12, 13).

1. Record the spectrum of a labeled variant (40–100 μM) in the absence and in the presence of the partner protein. Gradually increase the molar excess of the partner protein to achieve saturation. In the presence of saturating amounts of the partner, the extent of reduction in the spin label mobility is maximal, and further increasing the molar excess of the partner does not result in any further EPR signal variation (see Note 28).

2. If the spin probe is in the high regime of mobility (see Note 27), measure the $h(+1)/h(0)$ ratio of the EPR spectrum in the presence and in the absence of the partner protein. A decrease of this parameter indicates a decrease of mobility of the probe giving information on the local gain of rigidity of the protein (for an illustrative example see Fig. 4a).

3. If in the presence of the partner the spectrum is composite (i.e., it consists of a mixture of the unbound and bound form), subtract the appropriate proportion of the unbound form from the composite spectrum. The proportion of each form can be determined by double integration of each spectrum (the composite and the one resulting from subtraction) (see Fig. 4b).

If, after addition of the partner protein, the spin probe is in the intermediate/slow regime of mobility (see Note 27 and Fig. 3), measure the outer line splitting ΔB. The more this value is far from twice the averaged value of the hyperfine interaction ($2\bar{A} \approx 3.2$ mT), the more the spin label has a restricted mobility, the maximum value being around 7.3 mT corresponding to an immobilized probe. The appearance of a broad spectral shape can reveal tertiary contacts and thus indicate a position located at the interface of the partner protein.

4. If no effect is observed upon addition of the partner, this can reflect two quite different situations, namely, lack of effect because the spin label is located far from the binding interface or loss of ability to interact with the partner by the spin-labeled protein because of steric hindrance. To discriminate between these two possibilities it is recommended to assess the binding abilities of the spin-labeled protein either by direct determination of the equilibrium dissociation constant (K_D) by isothermal titration calorimetry or plasmon surface resonance, or by EPR equilibrium displacement experiments (see Note 29).

Effect of Sucrose

For proteins of molecular mass smaller than 50 kDa, the observed spectral shape contains a contribution arising from the rotational mobility of the entire protein. This contribution can be reduced by increasing the viscosity of the solution by adding up to 30 % w/v sucrose (43). Hence the observed spectral shape in the presence of sucrose reflects the local mobility of the spin probe that can arise from local backbone fluctuations, internal dynamics of the spin-labeled lateral chain and possible tertiary contacts.

1. Check by CD that the global secondary structure content of the protein remains unchanged in the presence of 30 % sucrose.

2. Check that 30 % sucrose does not induce any spectral shape modification in the EPR spectrum of the free radical in solution.

3. Record the spectrum of a labeled variant in 30 % sucrose in the absence and in the presence of the partner protein and measure the variation of the semiquantitative parameters.

Effect of TFE

The solvent TFE is a secondary structure stabilizer that mimics the hydrophobic environment of protein–protein interactions and can therefore be used to unveil the inherent folding propensities of IDPs in the absence of a partner protein.

1. Check by CD that the spin-labeled variant undergoes folding in the presence of increasing amounts of TFE (10–40 %).

2. Record the EPR spectrum of the spin-labeled variant in the absence and in the presence of TFE (20 %) and measure the variation of the semiquantitative parameters.

4. Notes

1. The following considerations should be made for designing mutagenic primers:

   – Both mutagenic primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid. The cysteine-encoding codon (i.e., TGC or TGT) should be chosen so as to minimize mismatches.

   – Primers should be between 25 and 45 bases in length, with a melting temperature (T_m) of ≥78 °C. Primers longer than 45 bases may be used, but using longer primers increases the likelihood of secondary structure formation, which may affect the efficiency of the mutagenesis reaction.

   – The desired mutation should be in the middle of the primer with ~10–15 bases of sequence perfectly matching the template on both sides.

   – The primers optimally should have a minimum GC content of 40 % and should terminate in one or more C or G bases. The following formula is commonly used for estimating the T_m of primers:

   $$T_m = 81.5 \ + \ 0.41(\%GC) - 675/N - \% \text{ mismatch,}$$

   where N is the primer length in bases and values for %GC and % mismatch are whole numbers, N is the primer length in bases.

2. The PCR reaction can be modified in numerous ways to optimize results, depending on the nature of the template and primers and the DNA polymerase used for amplification.

3. PCR cycle conditions can also be varied. For example, the extension time should be increased for long DNA templates and that depending on the elongation rate of the DNA polymerase. The annealing temperature could be modified depending on the primers melting temperature (see Note 1) (i.e., the specificity of the annealing of the primers to the template increases with the temperature).

4. It is important to adhere to the 18-cycle limit when cycling the mutagenesis reactions. More than 18 cycles can have deleterious effects on the reaction efficiency and may introduce undesired mutations. A PCR with a control DNA plasmid template and a control couple of primers is advised.

5. A band may or may not be visualized at this stage. In either case, proceed with *Dpn* I digestion.

6. The DNA band intensity on the agarose gel electrophoresis should decrease after digestion with *Dpn* I owing to the

degradation of the methylated, nonmutated DNA template (note that *Dpn* I only digests methylated DNA).

7. High transcription rates lead to slow growth, and this in turn is compounded by metabolic demands imposed by translation of the recombinant protein. Gene products that affect the host cell's growth rate at low concentrations are considered to be toxic. To reduce effects of protein toxicity on cell growth prior to induction, the level of basal transcription that occurs in the absence of induction "leakiness" should be repressed as much as possible, and the number of generation before induction should be kept to a minimum.

8. Some proteins, particularly those that are smaller than 10 kDa, are not stable in *E. coli*, and may be degraded rapidly by proteases. This may be overcome by different means: reducing the growth temperature, inducing for a shorter period of time, adding low concentrations of PMSF (up to 1 mM) to the culture medium before induction...

9. In many cases, many gene products, when expressed at high levels in *E. coli*, accumulate as insoluble aggregates that lack functional activity. The formation of inclusion bodies is influenced by the nature of the protein, by the host cell, and by the level of expression resulting from the vector choice and the growth and induction conditions.

10. If the ORF encoding the target protein contains codons that are rarely used in *E. coli*, this can adversely affect the yield of the protein. In such cases, it is advisable to introduce an additional plasmid into the host cells that carries the cognate tRNA genes for rare codons. The pLysS plasmid contains such tRNA genes. It also encodes T7 lysozyme (a natural inhibitor of T7 RNA polymerase), which allows reducing the basal expression of genes under the control of the T7 promoter for better inhibition of expression under noninduced conditions. In addition, T7 lysozyme helps facilitating lysis of bacterial cells. The pLysS plasmid confers resistance to chloramphenicol and contains the p15A origin of replication, leading to compatibility with pUC- or pBR322-derived plasmids.

11. The working concentrations of mostly used antibiotics can be found in (32).

12. We recommend to perform a small-scale protein expression test to ensure that the cysteine substitution does not significantly impact the expression and/or solubility level of the protein of interest. For well-expressed proteins (i.e., at ≥ 5 mg/L), 1-L culture is sufficient to get protein amounts suitable to the ensuing spin labeling step. Otherwise, larger culture volumes are recommended.

13. Induction temperature and period can be modified based on a small-scale protein expression test. For example, it is known that decreasing the induction temperature in some cases increases the stability and the solubility of the protein without significantly decreasing the yield.

14. Primers are usually designed to introduce a 6× His tag encoding sequence at the 3′ end of the ORF (corresponding to the C terminus of the protein) so as to avoid purification of truncated forms arising from possible abortive translation. Note, however, that N-terminal 6× His tags can also be used.

15. Protease inhibitor cocktails should be completely EDTA-free. The latter is a strong chelating agent that could bind to nickel ions and strips them from the NTA/Sepharose matrices. Resins become white in the absence of nickel ions.

16. Centrifugation of the disrupted cells for at least 30 min at $30,000 \times g$ is recommended. Filtration through a 0.2-μm polyethersulfone is helpful to remove residual particulates prior to chromatography.

17. The washing step of the resin can be repeated several times and can be monitored by mixing 2 μL of washing buffer and 100 μL of Bradford reagent. In case a blue color (indicative of the presence of protein contaminants) appears, it is recommended to carry out additional washing steps.

18. Set an appropriate pressure and flow rate on the chromatography system so as to avoid damage to the column (adequate flow rates and pressure threshold are generally provided in the manufacturer's instructions).

19. For adequate resolution, the volume of the injected protein must not exceed 5 % of the total bed volume. Protein amounts as high as 30 mg can be injected, unless the protein is known to be aggregation-prone.

20. For proteins devoid of trp and tyr residues, the absorbance can be followed at 254 nm.

21. In most cases, the introduction of a cysteine causes the protein to at least partly dimerize. It is recommended to check by SDS-PAGE that the fractions of the first elution peak do correspond to a dimeric form of the protein of interest and not to a larger protein contaminant. The fractions of the monomer and of the dimer can be pooled in view of a DTT reduction step prior to spin labeling.

22. The protein concentration can be determined spectrophotometrically using either the Bradford reagent (Bio-Rad) or the absorbance at 280 nm. In case a protein assay is used, a calibration curve with a protein solution of known concentration has to be established. Protein concentrations can also be inferred

from the absorbance at 280 nm using the Beer-Lambert relation. The absorption coefficient ε (M^{-1}/cm) at 280 nm can be obtained either using the program ProtParam at the EXPASY server (http://www.expasy.ch/tools) or, preferentially, by direct analysis of the amino acid composition.

23. Some proteins require the addition of 10–20 % glycerol to prevent them from precipitating upon freezing.

24. It is strongly recommended to record the CD spectrum (either in the presence or in the absence of 20 % TFE) of the spin-labeled form either so as to rule out possible dramatic effects brought by the spin label itself.

25. In case the cysteine-substituted protein is dissolved at a rather low concentration in a NaCl containing buffer, it is suggested to exchange the buffer to get rid of salt (which otherwise will give rise to strong absorption in the far-UV region) using a PD-10 column.

26. It can happen that some free radical contaminates the fractions containing the labeled protein. The resulting EPR spectrum will then be a superimposition of the signals arising from the labeled protein (see fraction 2 in Fig. 2a for example) and the free radical. If this is the case, the contaminated fractions are discarded and not pooled with the other protein containing fractions.

27. Nitroxide radicals are anisotropic paramagnetic centers characterized by the interaction between an electronic spin $S = 1/2$ and a nuclear spin $I = 1$ arising from the magnetism of the ^{14}N nucleus located in the vicinity of the unpaired electron. The anisotropy of both g tensor (Zeeman interaction) and A tensor (hyperfine interaction) makes EPR spectra critically dependent on the mobility of the radical. When nitroxide spin labels are allowed to tumble rapidly, magnetic interactions (g and A-tensors) are completely averaged. As the motion becomes progressively slower, the magnetic anisotropy is no longer totally averaged and this results in a differential broadening of lines in the spectrum, while line positions remain constant. This is the so-called fast motional regime. By further reducing the mobility of the radical, the averaging of tensor components becomes less and less efficient, leading to shape distortions of the EPR spectrum corresponding to the intermediate motional regime. Finally, the slow motional regime is reached when the full effects of the anisotropy of g and A-tensors are observed, this situation corresponding to the spectrum observed on frozen solution of nitroxide radical (39) (see Fig. 3).

28. In our experience (12, 13), addition of stoichiometric amounts of the partner protein was not sufficient to lead to saturation, although the concentrations of both proteins were largely

above the estimated K_D of the binding reaction with the native (i.e., nonsubstituted) form of the protein. This requirement indicates that the spin-labeled variant has a reduced affinity towards the partner protein as compared to the *wt* protein. In addition, various spin-labeled variants can display significant differences in their apparent affinity as a function of the radical location. In particular, in our experience (12, 13), the grafting of the nitroxide radical at a position located at the interface between the two proteins impaired the ability of the protein to yield 100 % complex formation. This observation suggests that two protein populations can occur in solution, one of which is unable to bind to the partner protein even at oversaturating conditions (an illustration of partial complex formation is provided in Fig. 4b).

29. In case addition of a partner protein does not trigger any effect on the spin label mobility of a spin-labeled variant, EPR equilibrium displacement experiments are very useful to assess whether this behavior arises from inability of the variant to interact with the partner or rather from location of the spin label outside the region of interaction. These experiments rely on the addition of an "EPR silenced" form of the labeled variant to a complex between the partner protein and a distinct spin-labeled variant whose spin label mobility is dramatically affected by the partner (i.e., a spin-labeled variant giving rise to a different spectral signature according to its bound or free state). The "EPR silenced" labeled variant can be obtained by reduction of the nitroxide (R–NO•) into a hydroxylamine group (R–NOH) by ascorbic acid. If the addition of the EPR silenced labeled variant triggers a reduction in the percentage of bound form, then this will indicate its ability to sequester the partner protein from the complex and hence will attest of its ability to interact with the latter (for example see ref. (13)).

References

1. Feix JB, Klug CS (1998) Site-directed spin-labeling of membrane proteins and peptide-membrane interactions. In: Berliner L (ed) Biological magnetic resonance. Plenum Press, New York

2. Hubbell WL, McHaourab HS, Altenbach C, Lietzow MA (1996) Watching proteins move using site-directed spin labeling. Structure 4 (7):779–783

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

4. Hubbell WL, Cafiso DS, Altenbach C (2000) Identifying conformational changes with site-directed spin labeling. Nat Struct Biol 7 (9):735–739

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

6. Columbus L, Hubbell WL (2002) A new spin on protein dynamics. Trends Biochem Sci 27 (6):288–295

7. Hubbell WL, Altenbach C, Hubbell CM, Khorana HG (2003) Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. Adv Protein Chem 63:243–290

8. Fanucci GE, Cafiso DS (2006) Recent advances and applications of site-directed spin labeling. Curr Opin Struct Biol 16(5):644–653

9. Klare JP, Steinhoff HJ (2009) Spin labeling EPR. Photosynth Res 102(2–3):377–390

10. Altenbach C, Flitsch SL, Khorana HG, Hubbell WL (1989) Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. Biochemistry 28(19):7806–7812

11. Altenbach C, Marti T, Khorana HG, Hubbell WL (1990) Transmembrane protein-structure: spin labeling of bacteriorhodopsin mutants. Science 248(4959):1088–1092

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

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

14. Pirman NL, Milshteyn E, Galiano L, Hewlett JC, Fanucci GE (2011) Characterization of the disordered-to-alpha-helical transition of IA by SDSL-EPR spectroscopy. Protein Sci 20 (1):150–159

15. Pannier M, Veit S, Godt A, Jeschke G, Spiess HW (2000) Dead-time free measurement of dipole-dipole interactions between electron spins. J Magn Reson 142(2):331–340

16. Jeschke G, Godt A (2003) Co-conformational distribution of nanosized [2]catenanes determined by pulse EPR measurements. Chemphyschem 4(12):1328–1334

17. Jeschke G, Bender A, Paulsen H, Zimmermann H, Godt A (2004) Sensitivity enhancement in pulse EPR distance measurements. J Magn Reson 169(1):1–12

18. Jeschke G, Wegener C, Nietschke M, Jung H, Steinhoff HJ (2004) Interresidual distance determination by four-pulse double electron-electron resonance in an integral membrane protein: the Na$^+$/proline transporter PutP of Escherichia coli. Biophys J 86(4):2551–2557

19. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44(37):12454–12470

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

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

22. Fuxreiter M, Tompa P, Simon I (2007) Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23(8):950–956

23. Longhi S, Belle V, Fournel A, Guigliarelli B, Carrière F (2011) Probing structural transitions in both structured and disordered proteins by site-directed spin-labeling EPR spectroscopy. J Pept Sci 17(5):315–328

24. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered proteins. Proteins 42(1):38–48

25. Dosztanyi Z, Meszaros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25(20):2745–2746

26. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 53(8):621–645

27. John SP, Wang T, Steffen S, Longhi S, Schmaljohn CS, Jonsson CB (2007) Ebola virus VP30 is an RNA binding protein. J Virol 81 (17):8967–8976

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

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

30. Longhi S (2010) Conformational disorder. Methods Mol Biol 609:307–325

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

32. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratoty manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

33. Benoit I, Coutard B, Oubelaid R, Asther M, Bignon C (2007) Expression in Escherichia coli, refolding and crystallization of Aspergillus niger feruloyl esterase A using a serial factorial approach. Protein Expr Purif 55(1):166–174

34. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of

affinity tags for protein purification. Protein Expr Purif 41(1):98–105

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

36. Van Hoy M, Leuther KK, Kodadek T, Johnston SA (1993) The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell 72 (4):587–594

37. Whitmore L, Wallace BA (2004) DICHRO-WEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32(Web Server issue):W668–W673

38. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89(5):392–400

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

40. Qu K, Vaughn JL, Sienkiewicz A, Scholes CP, Fetrow JS (1997) Kinetics and motional dynamics of spin-labeled yeast iso-1-cytochrome c: 1. Stopped-flow electron paramagnetic resonance as a probe for protein folding/unfolding of the C-terminal helix spin-labeled at cysteine 102. Biochemistry 36 (10):2884–2897

41. Goldman A, Bruno GV, Polnaszek CF, Freed JH (1972) An ESR study of anisotropic rotational reorientation and slow tumbling in liquid and frozen media. J Chem Phys 56:716–735

42. Belle V, Fournel A, Woudstra M, Ranaldi S, Prieri F, Thome V, Currault J, Verger R, Guigliarelli B, Carriere F (2007) Probing the opening of the pancreatic lipase lid using site-directed spin labeling and EPR spectroscopy. Biochemistry 46(8):2205–2214

43. McHaourab HS, Lietzow MA, Hideg K, Hubbell WL (1996) Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. Biochemistry 35 (24):7692–7704

44. Stoll S, Schweiger A (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. J Magn Reson 178 (1):42–55

Chapter 22

Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains

Lucía B. Chemes, Leonardo G. Alonso, María G. Noval, and Gonzalo de Prat-Gay

Abstract

Circular dichroism (CD) spectroscopy is a simple and powerful technique, which allows for the assessment of the conformational properties of a protein or protein domain. Intrinsically disordered proteins (IDPs), as discussed throughout this series, differ from random coil polypeptides in that different regions present specific conformational preferences, exhibiting dynamic secondary structure content [1]. These dynamic secondary structure elements can be stabilized or perturbed by different chemical (solvent, ionic strength, pH) or physical (temperature) agents, by posttranslational modifications, and by ligands. This information is important for defining ID nature. As IDPs present dynamic conformations, circular dichroism measurements (and other approaches as well) should be carried out not as single spectra performed in unique conditions, but instead changing the chemical conditions and observing the behavior, as part of the determination of the ID nature.

In this chapter, we present the basic methodology for performing Far-UV CD measurements on a protein of interest and for identifying and characterizing intrinsically disordered regions, and several protocols for the analysis of residual secondary structure present in the protein under study. These techniques are straightforward to perform; they require minimal training and can be preliminary to more complex methodologies such as NMR.

Key words: Circular dichorism, Intrinsically disordered protein, Secondary structure, CD spectrum, Polyproline type II (PII) conformation, Alpha helix conformation

1. Introduction

Many substances display optical activity, which is reflected by two phenomena: optical rotatory dispersion (ORD) and circular dichroism (CD). While ORD involves changes in the plane of rotation of polarized light, CD is related to the differential absorption of left handed (lcp) and right handed (rcp) circularly polarized

light (2). Circular dichroism is defined as the difference in molar absorption coefficient between lcp and rcp at a given wavelength.

For a given chromophore, CD is a particular absorptive phenomenon that is observed in the same spectral regions where the absorption band of unpolarized light occurs. On the other hand, not all chromophores display CD bands, since the chromophore needs to be intrinsically asymmetric, or placed in an asymmetric environment, in order to produce a CD signal. Biological macromolecules can be considered as heteropolymers that contain many chromophores, such as the amide bonds and the aromatic side chains of proteins and the purine and pyrimidine bases of nucleic acids, all of which are intrinsically symmetric due to their planar geometry. However, these heteropolymers usually fold into complex structures, placing symmetric chromophores into asymmetric environments and giving rise to CD phenomena.

Proteins present CD bands in two different spectral regions, the far-UV or amide region (175–250 nm) and the near-UV or aromatic region (250–300 nm). Far-UV CD informs about the secondary structure content, mostly based on the asymmetric conformation attained by the main polypeptide backbone. Near-UV CD reflects changes in the environment of aromatic side chains. These side chains are sensitive to the chemical environment often related to differential exposure to either solvent or to the inner core of the protein, and therefore these changes report on the protein's tertiary structure. CD bands in the far-UV region are characteristic for different types of secondary structure. Alpha helix structure displays the most invariable band pattern: a characteristic spectrum with a positive band at 190 nm and two negative bands at 208 and 222 nm. Beta sheet elements, however, are more variable with a positive band at around 198 nm and a single negative band ranging from 214 to 218 nm, depending on the type of structure. CD is not restricted to highly regular secondary structure elements, since beta turns and polyproline type II structures also show characteristic CD bands in the far-UV region, although these are much more variable and of lower intensity. What is often referred to as "random coil" structure is characterized by a negative band below 200 nm; however, the current belief is that pure "random coil" polypeptides are unlikely to exist, and that the spectrum of "disordered" polypeptides includes different populations and degrees of canonical and noncanonical secondary structure.

The far-UV CD spectrum of proteins results from the combination of the spectral bands corresponding to the different secondary structure elements that compose it, yielding complex spectra with minima and maxima different from those observed in pure secondary structure elements. Nonetheless, CD is widely used to assess the secondary and tertiary structure content of a protein, and is particularly useful for detecting conformational changes that occur, for example, during protein unfolding or upon ligand binding; see ref. 3 for a textbook review.

Intrinsically disordered proteins (IDPs) lack many of the attributes present in globular proteins, such as stable canonical secondary structure and compact and cooperative folding (4–6). However, IDPs display particular conformational properties such as an extended and dynamic conformation, and the presence of significant secondary structure content in certain regions (7), including a high content of polyproline type II structure (8). CD has become a very useful and straightforward method to analyze IDPs because of its sensitivity for reporting different conformational transitions involving changes in secondary structure.

2. Materials

1. *Equipment.* CD measurements are performed in spectropolarimeters (Jasco J-810 or similar equipment from other manufacturers such as Aviv or Applied Photophysics). Temperature can be controlled by a water bath or, ideally, by a peltier thermostat (Model PTC-423S/L or similar) connected to the measurement cell. The apparatus must be purged with nitrogen gas, which increases the quality of the spectrum, avoids ozone production, and increases the lamp lifetime (see Note 1). This is achieved by using a nitrogen tank coupled to the equipment. Data is usually acquired by data collection software provided by the equipment manufacturer and installed on a computer connected to the spectropolarimeter.

2. *Cuvettes and holders.* The protein sample must be loaded into a quartz cuvette for measurement. These cuvettes display high optical transparency below 200 nm, which is required for CD measurements. Cuvettes of different path lengths and their corresponding adaptors for fitting into the measurement cell are available, ranging from 0.02 to 1 cm. The choice of the cuvette best suited for a particular measurement depends on the sample concentration and the desired signal to noise ratio (see Subheading 3.1 and Note 2).

3. *Protein samples.* The proteins, peptides or other macromolecules of natural, synthetic, or recombinant origin used for CD determinations must be highly purified, i.e., >90 %, because one cannot anticipate the contribution of impurities to the spectra. For example, and especially in the case of IDPs, even a small amount of a protein contaminant present in the sample with a high content of alpha helix structure (which yields a high intensity CD signal) can distort the CD spectrum of the protein under study significantly. Purity can be assessed by polyacrylamide gel electrophoresis or by mass spectrometry (see Note 3).

4. *Buffers and additives.* The type of buffers and additives present in the sample greatly affect the spectrum quality, since these reagents absorb in the low end of the far-UV spectrum (at and below 200 nm) and therefore increase measurement noise in this wavelength range. Therefore, the sample buffer must be carefully chosen in order to ensure a good quality spectrum. Best quality spectra are obtained at the lowest buffer and salt concentration possible. Furthermore, it is important to use chemical reagents of the highest quality grade available in order to avoid introducing contaminants into the sample under study. If the use of a buffer or additive component which increases measurement noise is unavoidable, protein concentration can be increased and/or the path length decreased in order to improve the signal to noise ratio. A guideline for the use of different buffers, salts and additives is provided in Note 4.

3. Methods

3.1. Measurement of a Far-UV CD Spectrum

3.1.1. Sample Preparation

The CD signal is proportional to protein concentration. A better signal to noise ratio is achieved at higher protein concentration, but care should be taken to keep sample absorbance below 1.0 at either of the spectral regions, considering the path length used in the experiment. For a 15-kDa polypeptide, using a 1-mm quartz path length cuvette, prepare a 300 µl solution of 10 µM protein or roughly, for a double sized protein, use half the concentration, i.e., 5 µM for a 30 KDa protein (also see Subheading 3.4).

Be careful to use purified samples, which do not contain unknown contaminants (see Subheading 2, item 3 and Note 3). Choose the cuvette path length according to protein concentration (see Note 2), and be careful to clean the cuvette prior to loading the protein sample (see Note 5). Scattering can be avoided by filtering or centrifuging the sample immediately prior to performing measurements.

3.1.2. Instrument and Parameter Setting

The spectropolarimeter should be properly calibrated prior to use, as indicated in the equipment manual or by the person in charge. A simple procedure for checking the calibration of the instrument is included in Note 6. The instrument must be purged with nitrogen for 5 min prior to turning on the instrument. After this time, the xenon lamp, thermostat and computer can be turned on.

The measurement parameters are usually set through the software installed in the computer after the cuvette containing the sample has been placed in the measurement cell. The first step is to establish the wavelength range that is experimentally attainable. This is determined by performing an exploratory scan starting at
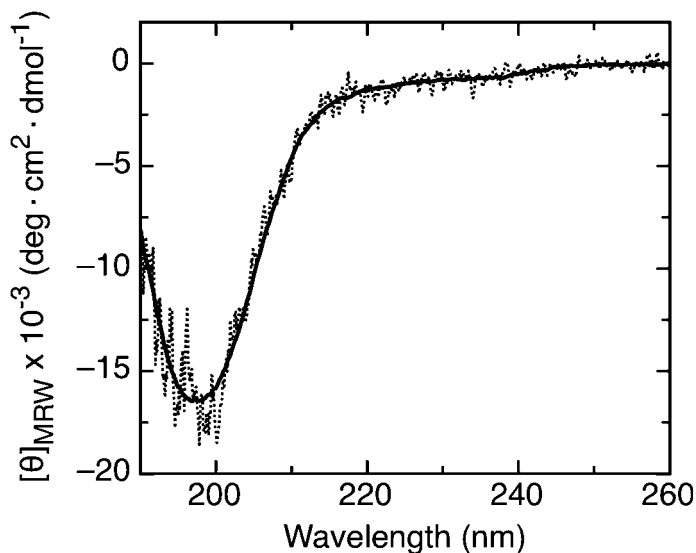
Fig. 1. Effect of acquisition parameters on the quality of a Far-UV CD spectrum. Single CD spectra of the intrinsically disordered HPV E7N domain (15) measured using a response time (DIT) of 8 s (*full line*) or 0.5 s (*dotted line*). Scans were performed at 50 nm/min scan speed, 0.1 nm data pitch and 2 nm bandwidth in 5 mM Tris–HCl buffer at pH 7.5 and at 20 °C in a 0.1-cm path length cuvette and at 50 μM protein concentration.

around 260 nm, and setting the lowest limit as the wavelength at which the photomultiplier (PMT) tube reaches a signal of 800 V. Extending measurements into the lower end of the far-UV region (below 200 nm) can be achieved by increasing the excitation bandwidth in the parameter window up to a value which does not produce distortion of the spectrum. We routinely use a bandwidth of 2 nm with good results. Also, the spectrum measurement range can be improved by increasing sample concentration, using shorter path length cuvettes (see Note 2), and by keeping buffer and additive absorbance low (see Notes 3 and 4).

Next, it is important to set those parameters that determine the spectrum quality. These are mainly the scan speed and the instrument integration time constant (also referred to as response or DIT), which are related to each other and to the data pitch (see Note 7). A further improvement of spectrum quality can be achieved by averaging several scans, usually five to ten. Figure 1 shows the effect of changing response time on the spectrum quality.

3.1.3. Spectrum Measurement

Once the sample has been equilibrated at the desired temperature, and the parameters have been set, perform five to ten scans in the wavelength range selected (usually 260 nm as an upper limit, and 180–200 nm as a lower limit). The instrument usually averages spectra while performing measurements. Check that the signal to noise level is acceptable or otherwise change parameters or increase repetitions accordingly in order to obtain a good quality spectrum.
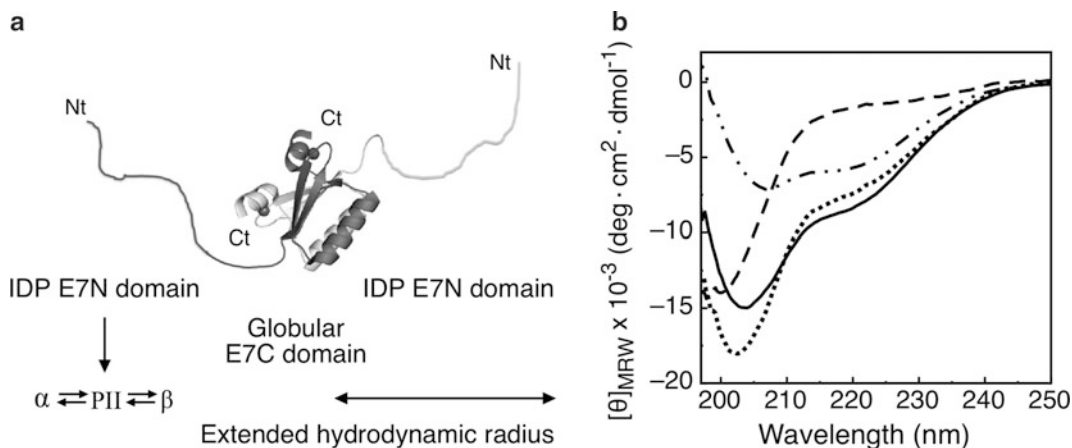
Fig. 2. Spectral features of IDP and globular protein domains. (**a**) Schematic representation of the HPV E7 protein as an extended dimer composed of two monomers depicted in *dark* and *light gray* respectively. Each monomer is composed of an intrinsically disordered (IDP) N-terminal domain (E7N), which exhibits dynamic secondary structure content (15) and a globular C-terminal domain (E7C) (28). (**b**) CD spectrum of the full-length HPV16 E7 protein (*full line*), the intrinsically disordered E7N domain (*broken line*) and the globular E7C domain (*broken and dotted line*). The sum of the spectra of both individual domains (E7N + E7C, *dotted line*) is similar to the spectrum of the full-length protein (*full line*). CD measurements were performed in 10 mM Tris–HCl buffer at pH 8.0, 50 mM NaCl, and 2 mM DTT at 20 °C in a 0.1-cm path length cuvette and at 10 μM protein concentration. Adapted from ref. 29.

Always remember to subtract a spectrum of the buffer alone from the spectrum of the sample, in order to ensure that one is analyzing signals originating from the sample and not from buffer components.

3.1.4. Data Processing   The CD scan is normally acquired by software provided by the manufacturer, installed on a computer connected to the equipment. The data collected is a CD signal, measured in millidegrees (mdeg), as a function of wavelength, measured in nanometers (nm). However, it is important to convert raw ellipticity to molar ellipticity, which is expressed as $[\theta]_{MRW}$ and can be calculated from a simple transformation of the raw data (see Note 8). Since molar ellipticity is normalized for protein concentration, path length, and number of peptide bonds, it allows for comparing across measurements performed in different experimental conditions.

3.2. Spectral Features of Intrinsically Disordered Proteins   IDPs present a CD spectrum resembling what was considered for a long time a purely disordered or "random coil" polypeptide, which is characterized by a minimum centered around 200 nm, and by the absence of strong negative signals above 205 nm, characteristic of alpha helix or beta sheet structure (compare Figs. 1 and 2b). Frequently, proteins contain a combination of globular and intrinsically disordered domains. In this case, if both domains are independent and do not interact, the CD spectrum will result from a combination of the different structural elements present in each domain.

In these cases, the presence of an intrinsically disordered domain can be detected by subtracting the CD spectrum of the globular domain from that of the full-length protein. As an example, the HPV16 E7 protein contains two modular and independent domains: an intrinsically disordered N-terminal domain (E7N) and a globular C-terminal domain (E7C) (see Fig. 2a). As shown in Fig. 2b, in this case the CD spectrum of the full length protein (full line) results from the sum of the spectra of the ID E7N domain (broken line) and the globular E7C domain (broken and dotted line). The negative band at 220 nm is indicative of alpha helix content present in the E7C domain, and the minimum centered at 205 nm results from a combination of the minimum at 200 nm due to the disordered nature of the E7N domain, and the minimum at 208 nm due to alpha helix structure present in the globular E7C domain (Fig. 2b).

3.3. Analysis by CD Spectroscopy of Secondary Structure in Intrinsically Disordered Proteins

Although IDPs are devoid of consolidated secondary and tertiary structure, different regions present persistent or dynamic secondary structure that can be stabilized upon changes in the chemical environment or due to the interaction with other proteins, ligands, cofactors or lipid membranes. In this section we present CD techniques that aid in the identification of spectral features of different nature, which contribute to the overall spectrum of an IDP. The presence of alpha helix structure can be evidenced by the addition of trifluoroethanol (TFE) (9–11), and PII structure can be stabilized by guanidine chloride (Gdm.Cl) (12, 13) and by low temperatures (14). The presence of alpha helix or PII structure elements stabilized by these agents can be initially assessed by analyzing the difference spectrum resulting from subtracting the spectrum of the protein in buffer alone from that of the protein in the presence of the stabilizing agent. More detailed thermodynamic information on these structures can be extracted by performing titrations, as explained below.

3.3.1. Alpha Helix Structure

Trifluoroethanol (TFE) is widely used in order to stabilize regions with preexistent alpha helical structure (9–11). In order to save sample, TFE titrations can be performed on the same sample by measuring the CD spectrum of the protein sample at 0 % TFE and then adding increasing volumes of neat TFE to the cuvette containing the sample up to a final concentration of 65 % v/v. Samples must be allowed to equilibrate prior to each measurement (see Note 9). Since sample dilution by at least a factor of two occurs throughout the titration, the initial protein concentration must be at least double the minimal concentration required to yield a good signal to noise ratio in the CD spectrum. One should consider that the sample cuvette must allow for an increase in volume of at least a factor of two and for sample mixing, which can be performed using a Hamilton syringe.
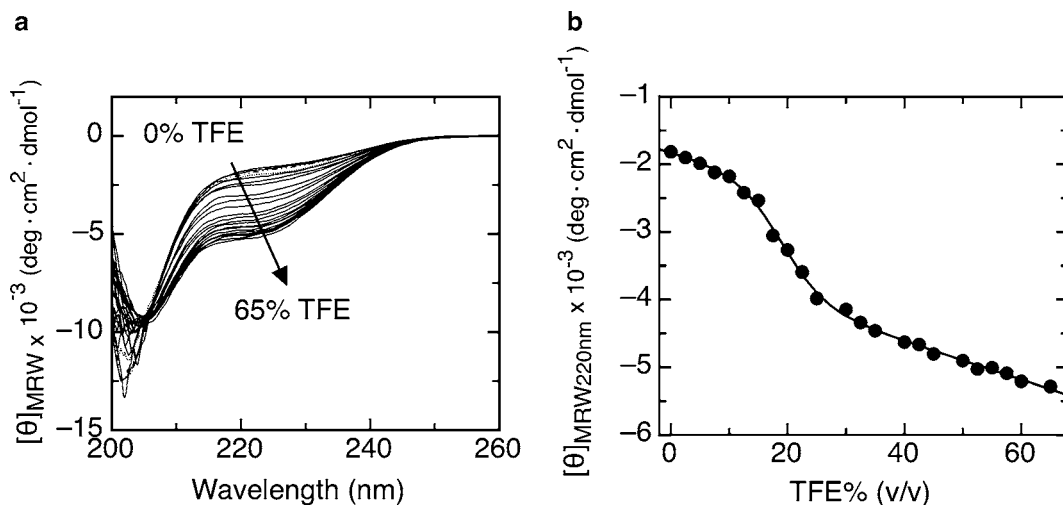
Fig. 3. Use of TFE titrations to estimate alpha helix content. (**a**) Far-UV CD scans of the HPV E7N domain performed at increasing TFE concentrations, expressed as % v/v. Induction of alpha helix structure is evidenced by an increase in the negative band at 222 nm and also as a decrease in the intensity and shift towards longer wavelengths of the minimum at 200 nm. (**b**) Fit of the TFE titration data to a two-state helix-coil model (see Note 10). The parameters from the fit were: $\Delta G^{H2O} = 1.5 \pm 0.1$ kcal/mol, $m = 0.5 \pm 0.05$ kcal/mol K. In this case, 5.6 residues were stabilized in alpha helix conformation, calculated as explained in Note 10 using eq. 22.3 and the value of ellipticity in high TFE ($[\theta]^{TFE}$ in eq. 22.2). Adapted from ref. 15. Measurements were performed at 20 μM peptide concentration at 20 °C in 20 mM Tris–HCl buffer pH 7.5 in a 0.5-cm path length cuvette.

Figure 3a shows a TFE titration of the HPV E7N domain, with data converted to molar ellipticity (see Note 8) taking into consideration sample dilution. As can be observed, TFE produces an increase in the absolute value of the band at 220 nm, and a shift in the position of the minimum from 200 to 208 nm, indicating the stabilization of alpha helical structure. The transition is cooperative, and the thermodynamic parameters (ΔG^{H2O} and m) for the formation of alpha helix structure can be calculated by considering a two-state coil-helix equilibrium model and by assuming a linear dependence of the thermodynamic stability of the alpha helix on TFE concentration. The change in molar ellipticity at 222 nm as a function of TFE% can be fit to this model to obtain the thermodynamic parameters for alpha helix formation (see Note 10 and Fig. 3b). The number of residues stabilized in alpha helical conformation can also be estimated from the molar ellipticity at 222 nm (see Note 10). A similar analysis has been performed on a number of proteins by us and by others with good results; for examples see refs. 12, 15–17.

3.3.2. Polyproline Type II (PII) Structure

Polyproline type II (PII) structure is a prevalent conformation in many intrinsically disordered regions and in unfolded polypeptides (8, 18). The CD spectrum of PII structure is characterized by a positive band centered around 218 nm, and by a minimum at around 198 nm (19). The positive band at 218 nm shows low

intensity, which makes the identification of this type of structure by CD difficult in the presence of other secondary structure elements. One way of identifying the presence of PII structure is by stabilizing it using low temperatures or high concentrations of Gdm.Cl, for examples see refs. 12, 15–17, 19, 20. Clearly, except in the case of short model peptides with pure PII conformation, a protein with different proportions of PII, coil, and other structures that contribute to the CD spectrum, will not present a unique distinctive band in its CD spectrum that identifies PII structure.

Whereas Gdm.Cl destabilizes folded proteins through solvation of hydrophobic side chains and screening of ionic interactions, this solvent has a stabilizing effect on polyproline type II (PII) structure (13), which also acts as a destabilizing force for the folded state. Therefore, one way of identifying PII structure in IDPs is through performing a Gdm.Cl titration (12) (see Note 11). As shown in Fig. 4a, increasing Gdm.Cl concentration produces an increase in the band at 218 nm, indicative of the stabilization of PII structure. If the transition reaches saturation within the Gdm. Cl range assayed (0–7.0 M), it is possible to extract the thermodynamic parameters (ΔG^{H2O} and m) performing a fit of the data to a two state PII-coil equilibrium model (see Note 10 and Fig. 4b).

It has been reported that PII structure is stabilized at low temperatures (14, 19). This can be observed by performing CD spectrum measurements at different temperatures, ranging from 5 to 85 °C (see recommendations in Note 12). As shown in Fig. 5a, as temperature is increased, there is a decrease in the absolute value of the intensity of both the positive band at 218 nm and of the minimum around 200 nm. Another characteristic of this transition
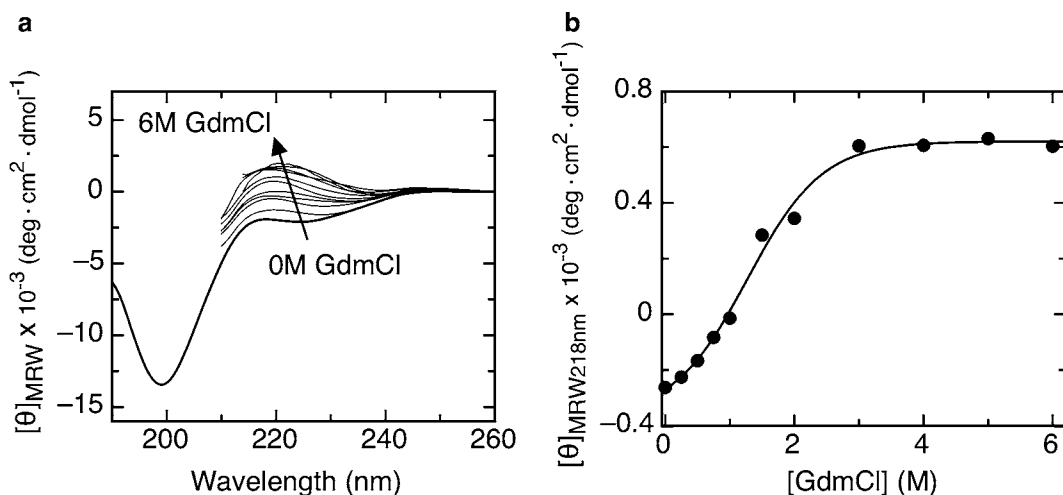
Fig. 4. Use of Gdm.Cl titrations to stabilize PII structure. (a) CD spectra of a peptide comprising the HPV E2 DNA binding helix. Increases in Gdm.Cl concentration cause a strong increase in the band at 218 nm, indicative of the stabilization of PII structure. (b) Fit of the titration data to a two state PII-coil transition (see Note 10). The parameters obtained from the fit were $\Delta G^{H2O} = 1.3 \pm 0.3$ kcal/mol, $m = 0.95 \pm 0.16$ kcal/mol K. Measurements were performed in 10 mM sodium phosphate buffer, pH 7.0 at 25 °C in a 0.1-cm path length cuvette and at 40 μM peptide concentration. Adapted from ref. 17.
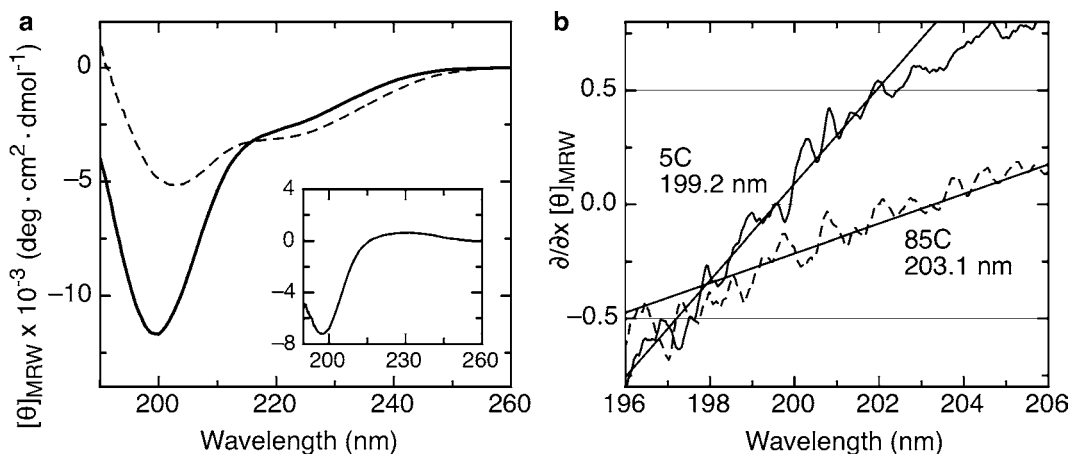
Fig. 5. Stabilization of PII structure by low temperature. (**a**) CD spectrum of the HPV E7N domain at 5 °C (*full line*) and 85 °C (*broken line*). Upon a temperature decrease, there is an increase in the intensity of the minimum at 198 nm and of the positive band at 218 nm, indicative of induction of PII conformation. These changes are also evidenced in the difference spectrum (5–85°C) (*inset*). (**b**) First derivative of both CD spectra. The position of the minimum corresponds to the zero point of the derivative. In this case, a shift of 3.9 nm towards longer wavelengths is evident upon a temperature increase. Measurements were performed in 5 mM Tris–HCl buffer, pH 7.5 at 20 °C in a 0.1-cm path length cuvette and at 40 μM protein concentration. Measurement parameters were 50 nm/min scan speed, 0.025 nm data pitch, 8 s response time, 2 nm bandwidth, 20 scans.

is a small but significant shift in the position of the minimum towards longer wavelengths. This shift can be better resolved by analyzing the first derivative of the CD spectrum, where the zero of the first derivative of the spectrum marks the position of the minimum (21) (see Note 13). An example of this analysis is shown in Fig. 5b, where the minimum is shifted 3.9 nm towards longer wavelengths upon increasing temperature from 5 to 85 °C.

It is important to underscore that the typical minimum at 195 nm in apparently disordered or "random-coil" polypeptides becomes an object of attention when examining noncanonical secondary structures such as PII, and in particular when analyzing the conformational properties of IDPs. Analyzing changes in the intensity of the minimum as well as small but detectable shifts in its position is essential for the investigation of conformational transitions of IDPs (e.g., see below the effect of phosphorylation on IDPs).

3.3.3. Beta Sheet Transitions   Some proteins present beta sheet transitions in the presence of submicellar concentrations of detergents such as sodium dodecyl sulfate (SDS) (22). In order to assess beta sheet transitions induced by SDS, take a spectrum of the protein or domain under study in buffer alone, and in buffer containing submicellar or supramicellar concentrations of SDS. For example, the critical micelle concentration (CMC) value for SDS is 4.8 in 20 mM Tris:Cl buffer at pH 7.5 and 5 mM in 20 mM formate buffer at pH 4.0 (15). For each buffer system the CMC value must be determined or obtained from tabulated values (see Note 14). An example for the HPV E7N
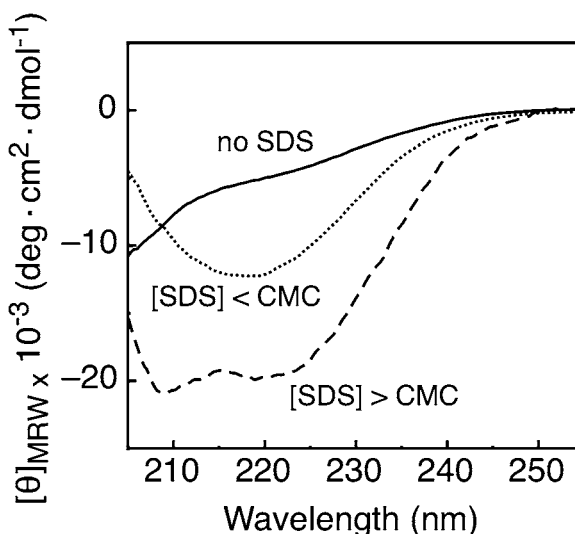
Fig. 6. Effect of SDS on secondary structure of an IDP domain. CD spectrum of the HPV E7N domain in the absence (*full line*) of SDS and in the presence of SDS at concentrations below ([SDS] = 1 mM, *dotted line*) and above ([SDS] = 10 mM, *broken line*) the critical micelle concentration (CMC). Below the [CMC], beta sheet structure is induced, whereas above the [CMC], alpha helix conformation is stabilized. Measurements were performed in 10 mM sodium formiate buffer, pH 4.0 at 25 °C and using 20 μM peptide concentration in a 0.1-cm path length cuvette. Adapted from ref. 15.

domain is shown in Fig. 6. In the presence of submicellar concentrations of SDS a spectrum characteristic of beta sheet structure can be observed, which presents a single minimum at 217 nm, as opposed to alpha helix structure induced above the CMC value.

3.3.4. Effects of Phosphorylation on IDP Secondary Structure

Posttranslational modifications can affect qualitatively and quantitatively the secondary structure content of IDPs. As an example, the E7 protein from human papillomavirus is regulated in vivo by casein kinase II (CKII), which phosphorylates serine residues 31 and 32, increasing the affinity for its main binding partner, the retinoblastoma (Rb) protein (16). Serine phosphorylation proximal to the Rb binding site leads to an increase in PII content, which can be observed both in the isolated HPV16 E7N IDP domain (E7$_{1-40}$) (15) and in a smaller peptide fragment (E7$_{16-40}$) containing the high affinity Rb binding site (16). The spectrum of the phosphorylated E7$_{16-40}$ peptide presents an increase in the intensity of the band at 218 nm and of the minimum at 198 nm, both indicative of an increase in PII content (Fig. 7a). The difference spectrum (5–85 °C) of the phosphorylated and unphosphorylated peptides shows a stronger band at 218 nm for the phosphorylated peptide (Fig. 7a, inset).

As a complementary example, an ID portion of the flexible PEST region of the bovine papillomavirus E2 protein (E2 PEST)
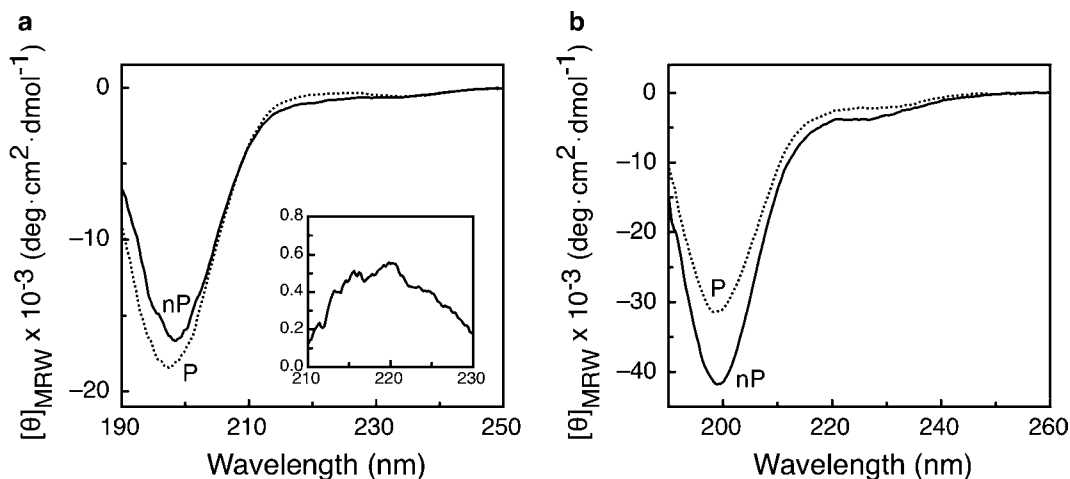
Fig. 7. Modulation of PII content by phosphorylation. (**a**) CD spectra of the unphosphorylated (*full line*, nP) and the phosphorylated (*dotted line*, P) E7$_{16-40}$ fragment at 5 °C. The phosphorylated fragment has a stronger positive band at 218 nm and a more negative minimum at 198 nm. *Inset*: Difference spectrum (P − nP) showing the increase in the band at 218 nm upon phosphorylation. Adapted from ref. 16. (**b**) CD spectra of a peptide corresponding to the unphosphorylated PEST region from the bovine papillomavirus E2 protein (*full line*, nP) and of its phosphorylated counterpart (*dotted line*, P). Phoshporylation destabilizes PII conformation, evidenced as a decrease in the intensity of the minimum at 198 nm. Adapted from ref. 12.

also displayed a substantial effect of phosphorylation on its structure. In this case, phosphoryation of a serine residue produced a decrease in PII content (12), and this structural change led in turn to an increase in the protein turnover in the cell (23), indicating that phosphorylation caused a thermodynamic local destabilization in this ID region. The CD spectra of the unphosphorylated (E2 PEST) and phosphorylated (E2 PEST 301p) peptides evidence a strong decrease in the intensity of the minimum at 198 nm upon phosphorylation (Fig. 7b). In both example cases discussed here, the change observed in the intensity of the minimum at 198 nm was too large to be ascribed to the single chemical difference: a phosphate moiety (Fig. 7b).

3.4. Quick-Start Guide

After reading the basics for obtaining Far-UV CD spectra of IDP proteins, domains, or of a protein combining globular and IDP domains, the reader may want to have a quick start guide with the sample under study. We assume that the equipment, irrespective of its age or manufacturer, is in optimal working conditions. We also assume that the protein under study is (or is suspected to be) at least in part IDP, which can be inferred initially from an anomalous behavior in a simple size exclusion chromatography experiment, or detected from the various powerful algorithms available (24, 25). Proceed as follows:

1. For a 15-kDa polypeptide using a 1-mm quartz path length cuvette, prepare a 300 μl solution of 10 μM protein, in the best

buffer possible, preferably phosphate, of the best quality possible (see Note 2). The size and the concentration of the protein keep roughly an inverse proportion (i.e., for two times the size, use half the concentration).

2. Spin down the sample for 5 min in a micro-centrifuge, and measure the protein concentration of the supernatant carefully by UV absorbance preferably, or by other alternative methods (i.e., Bradford or other), as accurately as possible.

3. Turn on the nitrogen flow to 5–7 l/min, wait 5 min and switch on the CD spectropolarimeter. Set the temperature (either circulating water bath or peltier), add the sample to the 1 mm cell, and cap it to prevent evaporation (this is critical when performing a thermal scanning experiment or a measurement at high temperature).

4. Set the following parameters: wavelength slit (or bandwidth): 2 nm; data interval (or data pitch): 0.2 nm; response (or DIT): 4 s; scan rate: 50 nm/min; number of scans: 10; wavelength range: 260–190 nm. Check that the PMT does not go higher than 800 V within the measurement range, and otherwise readjust the wavelength range (see Subheading 3.1). Run a buffer scan under the same conditions determined for the sample and subtract it from the sample scan. Export and convert the data in order to express ellipticity in normalized units (see Note 8).

5. You may either obtain a mainly IDP-like spectrum, or a spectrum with significant canonical secondary structure (i.e., α-helix, β-sheet, turns) if the protein presents globular as well as IDP domains (see Fig. 2 and Subheading 3.2).

6. If an IDP-like, apparently disordered spectrum is obtained, consider acquiring a spectrum at 5 °C and another at 80 °C (or whatever limit the equipment allows), under the same parameter conditions. The spectra will look different, and by subtracting the spectrum at 80 °C from that at 5 °C, you will obtain a difference spectrum that may show changes at 198 and 218 nm, indicative of the presence of PII-type conformation (see Subheading 3.3 and Fig. 5).

7. Select the wavelength at which the spectral change between the two temperatures is maximal and perform a temperature scan, which measures the ellipticity change with increasing temperature. A noncooperative (i.e., linear) transition, full reversibility, and an isodichroic point (a wavelength at which all the spectra intersect in a single point) are strong evidences of IDP nature.

8. You may want to consider registering similar spectra at 25 °C, adding suspected or known ligands, or modifying the solvent

conditions in order to evaluate the protein's conformational transitions (for example, using TFE, Gdm.Cl or SDS as explained in Methods Subheading 3.3). In each case, remember to subtract the CD spectrum of buffer with the ligand, as the ligand may also have a CD signal.

9. After performing temperature, ligand or pH studies on the IDP under study, one should consider centrifuging the sample for at least 10 min and re-measuring the spectrum and the protein concentration. This procedure will allow one to evaluate whether the changes observed were effectively due to conformational transitions in the IDP conformation as opposed to sample loss due to irreversible aggregation.

4. Notes

1. Nitrogen flow is required to remove oxygen from the equipment. Oxygen is prejudicial since it absorbs light, especially at low wavelengths, decreasing the energy of the light incident on the sample. Lighting the xenon lamp in air will also generate ozone gas, which is harmful to the lamp and to the human body. Nitrogen flow can be adjusted according to the lowest wavelength at which one wants to perform the measurement scans. As a practical guideline, set the nitrogen flow at 3–5 l/min for measurements up to 200 nm, and increase flow rate to 10–15 l/min for measurements up to 180 nm.

2. The cuvette path length should be chosen depending on the protein concentration and sample volume available for the measurements. For example, the signal measured from a 10 μM protein sample in 10 mM Buffer in a 1-mm path length cuvette (250 μl load volume) is equivalent to that of the same protein sample at 2 μM concentration measured in a 5-mm path length cuvette (1,000 μl load volume). Nevertheless, the measurement using a larger path length will have increased noise due to higher buffer absorption. Therefore, when possible it is advisable to use the lowest optical path length compatible with a measurable signal.

3. Ideally, CD measurements should be performed on protein samples obtained from recombinant or natural sources, which have been subjected to a thorough purification procedure, which usually goes beyond performing simple affinity column purification. The final purity can be evaluated by polyacrylamide gel electrophoresis (PAGE) or by chromatographic methods by using standard protocols, which will not be described in this chapter but are available from commonly used laboratory protocol manuals.

4. Buffers and additives: Most commonly used aqueous buffers are suitable for CD experiments, although they exhibit variable absorptivity in the far-UV range. For example, phosphate and acetate buffers have low absorption, Tris buffer has intermediate absorption and Hepes, MES, MOPS and Tricine buffer have the highest absorption. Use the best purity grade buffers available in order to avoid contaminants. Other additives commonly used in the molar concentration range in CD experiments are urea and guanidine chloride (Gdm.Cl). These additives have high absorption and will preclude measurements below 210 nm in a 1-mm path length cuvette. Other additives included in sample buffers, such as reducing agents (DTT, β-mercaptoethanol), detergents (Tween, SDS), glycerol, and salts will also introduce noise into the measurements, and should be kept at the lowest concentration possible. For measurements extending below 200 nm, we recommend keeping DTT or β-mercaptoethanol at concentrations at or below 1 mM, and buffer and salt concentration at 10–20 mM.

5. We recommend to carefully clean cuvettes before and after each use with acidic cleaning solutions. For this purpose, we commonly use 98 % grade sulfuric acid for 15 min, after which the cuvette must be rinsed repeatedly with distilled or MilliQ water.

6. Calibration of the CD scale can be checked by measuring a 0.06 % solution of ammonium d-10-camphor sulfonate in distilled water in a 0.1-cm path length cuvette, which should display a signal of 190.4 ± 1 mdeg at 190.1 nm. The spectropolarimeter's wavelength accuracy can be checked placing a neodymium glass, usually provided with the equipment, in the measurement chamber. When measuring a spectrum between 560 and 610 nm, the peak maximum should be centered at 586 ± 0.8 nm.

7. The scanning rate must be changed according to the response time. We routinely use scanning rates of 50–100 nm/min and response times of 2–8 s. As a general rule, when increasing the scan speed one should decrease the response time as stated in the manufacturer manual. The data pitch determines the spectral resolution of data acquisition. A data pitch of 0.2–1.0 nm works fine for most applications.

8. Mean residue molar ellipticity or $[\theta]_{\mathrm{MRW}}$ can be calculated applying the following formula to the raw data:

$$[\theta]_{\mathrm{MRW}}(\mathrm{deg\ cm^2/dmol}) = \frac{\mathrm{m\,deg}}{10 \times L \times [M] \times N} \qquad (1)$$

Where L is the path length (measured in cm), $[M]$ is the concentration (measured in molar units), and N is the number of peptide bonds (measured as: number of amino acids $-$ 1).

9. Since TFE titrations using a single sample will dilute the protein at least by a factor of two, this protocol is advised only if no change in oligomerization state occurs over the protein concentration range covered in the experiment. Otherwise, the titration must be performed in separate tubes keeping the protein concentration constant.

10. Data from TFE and Gdm.Cl titrations can be fit to a two state equilibrium (helix-coil or PII-coil) by considering no variation in the PII population for TFE titrations or in the alpha helix population for Gdm.Cl titrations. This model assumes the free energy for α-helix formation to depend linearly on [TFE]/[water] (10) and of PII structures to depend linearly on Gdm.Cl molar concentration (12). Molar ellipticity data can be fitted to the following equation to extract the values for ΔG and m:

$$[\theta] = \frac{[\theta]^{\text{TFE,Gdm.Cl}} + [\theta]^{\text{water}} \times \exp\left(-\Delta G^{\text{TFE,Gdm.Cl}}/RT\right)}{1 + \exp(-\Delta G^{\text{TFE,Gdm.Cl}}/RT)}, \quad (2)$$

Where $[\theta]^{\text{water}}$ and $[\theta]^{\text{TFE,Gdm.Cl}}$ are the mean residue ellipticities in water and at high cosolvent concentration, R is the gas constant, and T is the temperature.

The number of residues stabilized in alpha helix conformation before and after the TFE titration can be estimated by comparing the values of molar ellipticity in water or TFE obtained from the fit to equation [2] with the molar ellipticity value at 222 nm that corresponds to 100% helical content for the protein or peptide under study. This value can be obtained from the following empirical relationship, where n represents the number of residues in the peptide/protein under analysis: (26):

$$[\theta]_{222\text{nm}} = -39{,}500 \times \left(1 - \frac{2.57}{n}\right) \quad (3)$$

11. Gdm.Cl titrations are best performed using separate tubes for each Gdm.Cl concentration since the highest Gdm.Cl concentration attainable is 7.8 M. Prepare a series of tubes containing the same protein concentration, and increasing amounts of Gdm.Cl ranging from 0 to 7 M. It is advisable to measure at least 13 points covering the range of interest. Samples must be allowed to equilibrate at the measurement temperature (for best results, 5 °C) prior to measurement. Due to the high absorption of Gdm.Cl, the CD spectrum cannot be measured below 210 nm using 0.1-cm path length cuvettes.

12. When performing thermal scanning experiments, it is essential to use capped cuvettes, in order to avoid sample evaporation. Cuvettes with teflon caps are commonly available from quartz cuvette providers such as Hellma and others.

13. Calculation of the first derivative of the CD spectrum can be performed by applying standardized functions contained in most routinely used data analysis programs. In order to obtain a good quality derivative spectrum, it is important to use a small data pitch (ideally 0.025 nm) and extensive averaging which yields a high signal to noise ratio. It is useful to smooth the raw data in order to reduce noise prior to computing the spectrum derivative. Identification of the minimum is performed by fitting the data points contained at ±5 nm of the zero to a linear function (see Fig. 5b).

14. The critical micelle concentration (CMC) of SDS depends on the buffer used, on ionic strength, and on solution pH. Tabulated values can be found in ref. 27.

References

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

2. Woody RW (1996) Theory of circular dichorism of proteins. In: Fasman GD (ed) Circular dichroism and the conformational analysis of biomolecules. Plenum Press, New York, pp 25–67

3. Creighton TE (2010) Circular dichroism. In: Creighton, TE (ed) The physical and chemical basis of molecular biology. Helvetian Press, Eastbourne, pp 291-310

4. Fink AL, Creighton, TE (ed), (2005) Natively unfolded proteins. Curr Opin Struct Biol 15:35–41

5. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK (2005) Natively disordered proteins. In: Buchner J, Kiefhaber T (eds) Protein folding handbook. Wiley, Weinheim, pp 275–357

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

7. Uversky VN, Dunker AK (2010) Understanding protein non-folding. Biochim Biophys Acta 1804:1231–1264

8. Shi Z, Chen K, Liu Z, Kallenbach NR (2006) Conformation of the backbone in unfolded proteins. Chem Rev 106:1877–1897

9. Buck M (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. Q Rev Biophys 31:297–355

10. Jasanoff A, Fersht AR (1994) Quantitative determination of helical propensities from trifluoroethanol titration curves. Biochemistry 33:2129–2135

11. Nelson JW, Kallenbach NR (1986) Stabilization of the ribonuclease S-peptide alpha-helix by trifluoroethanol. Proteins 1:211–217

12. Garcia-Alai MM, Gallo M, Salame M, Wetzler DE, McBride AA, Paci M, Cicero DO, de Prat-Gay G (2006) Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. Structure 14:309–319

13. Tiffany ML, Krimm S (1973) Extended conformations of polypeptides and proteins in urea and guanidine hydrochloride. Biopolymers 12:575–587

14. Tiffany ML, Krimm S (1972) Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. Biopolymers 11:2309–2316

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

16. Chemes LB, Sanchez IE, Smal C, de Prat-Gay G (2010) Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. Structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7. FEBS J 277:973–988

17. Wetzler DE, Gallo M, Melis R, Eliseo T, Nadra AD, Ferreiro DU, Paci M, Sanchez IE, Cicero DO, de Prat Gay G (2009) A strained DNA binding helix is conserved for site recognition, folding nucleation, and conformational modulation. Biopolymers 91:432–443

18. Shi Z, Woody RW, Kallenbach NR (2002) Is polyproline II a major backbone conformation

in unfolded proteins? Adv Protein Chem 62:163–240

19. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR (2002) Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci U S A 99:9190–9195

20. Liu Z, Chen K, Ng A, Shi Z, Woody RW, Kallenbach NR (2004) Solvent dependence of PII conformation in model alanine peptides. J Am Chem Soc 126:15141–15150

21. de Prat Gay G (1996) Spectroscopic characterization of the growing polypeptide chain of the barley chymotrypsin inhibitor-2. Arch Biochem Biophys 335:1–7

22. Zhong L, Johnson WC Jr (1992) Environment affects amino acid preference for secondary structure. Proc Natl Acad Sci USA 89:4462–4465

23. Penrose KJ, Garcia-Alai M, de Prat-Gay G, McBride AA (2004) Casein Kinase II phosphorylation-induced conformational switch triggers degradation of the papillomavirus E2 protein. J Biol Chem 279:22430–22439

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

25. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. Cell Res 19:929–949

26. Chen YH, Yang JT, Chau KH (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. Biochemistry 13:3350–3359

27. Brito RM, Vaz WL (1986) Determination of the critical micelle concentration of surfactants using the fluorescent probe N-phenyl-1-naphthylamine. Anal Biochem 152:250–255

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

29. Smal C (2010) Oligomerización de la oncoproteína E7 del papilomavirus humano y su interacción con el regulador de la transcripción y replicación viral E2. In: Physics Department, PhD Thesis, University of Buenos Aires

Chapter 23

Deconstructing Time-Resolved Optical Rotatory Dispersion Kinetic Measurements of Cytochrome *c* Folding: From Molten Globule to the Native State

Eefei Chen and David S. Kliger

Abstract

The far-UV time-resolved optical rotatory dispersion (TRORD) technique has contributed significantly to our understanding of nanosecond secondary structure kinetics in protein folding and function reactions. For reduced cytochrome *c*, protein folding kinetics have been probed largely from the unfolded to the native state. Here we provide details about sample preparation and the TRORD apparatus and measurements for studying folding from a partly unfolded state to the native secondary structure conformation of reduced cytochrome *c*.

Key words: Time-resolved, Nanosecond, Optical rotatory dispersion, Circular dichroism, Cytochrome *c*, Molten globule, Protein folding

1. Introduction

Although circular dichroism (CD) is the most common polarization spectroscopy used today, it was developed several decades after the initial observation (early 1800s) of optical rotation (OR), which arises from circular birefringence (CB). OR was then the primary method of studying optical activity (1). Measurement of OR advanced from the earliest polarimeter (a sheet of glass and sunlight) to its detection at specific wavelengths of light (mercury lines, sodium D-line) to single wavelength determinations over a wide spectral range in the 1930s. About 20 years later, Keston and Lospalluto had developed a quasi-null optical rotatory dispersion (ORD) approach (2) that was later implemented as the Standard

Model D Keston polarimetric attachment in Beckman's DU spectrophotometer (3, 4). The quasi-null principle is the basis for the time-resolved optical rotatory dispersion (TRORD) design, which is a linear analog of the time-resolved circular dichroism (TRCD) technique, both of which are used in this laboratory today (5–7).

Nanosecond TRCD spectroscopy was developed in the 1980s. Initially designed to measure single wavelength traces, the TRCD method was extended to include multichannel detection, other time-resolved polarization methods (magnetic circular dichroism (MCD), linear dichroism, ORD, magnetic ORD), measurements from the visible into the far-UV spectral regions, faster time resolution, and most recently, it has been coupled with temperature-jump methods (8). The development of far-UV TRCD, and subsequently far-UV TRORD, has provided significant secondary structure insights into the kinetics of many protein function (phytochrome, myoglobin, PYP, LOV (9–13)) and folding (cytochrome c, azobenzene cross-linked peptides (14–20)) reactions. However, in recent years the far-UV TRORD method has been used in more of our studies because it allows for higher UV light intensity and lower sensitivity to birefringence artifacts.

In optical activity, CB and CD are Kramers–Kronig transform mates (21) as with unpolarized electronic transitions, where the refractive index and absorption are intrinsically related. Therefore, in theory, the two methods are expected to provide the same information. However, in practice each approach has advantages that make one preferable over the other depending on the experimental circumstances. CD is localized to the absorption band of the chromophore of interest and so the measured spectra can be more easily interpreted. In contrast, an OR spectrum comprises multiple Cotton contributions, making analysis of the data more difficult. In situations where the study is aimed at understanding protein structural changes, CD is generally chosen over ORD for its ability to distinguish between secondary structures such as the α-helix and β-sheet conformations. However, the fact that OR spectra are monitored away from the absorption band can lend a clear signal-to-noise advantage of the TRORD method over TRCD, especially in the case of protein folding kinetic studies.

The far-UV TRORD technique has been used extensively in this laboratory to follow the kinetics of cytochrome c (cyt c) folding. In 1998 and 1999, the first nanosecond polarization studies of folding in reduced cytochrome c (redcyt c) employed far-UV TRCD methods (14, 15). In those studies, redcyt c was bathed in guanidine hydrochloride (GuHCl), a denaturant that absorbs significantly in the far-UV region and obscures the 210 nm band, one of the spectral markers for the α-helix secondary structure.

Although the results of these TRCD studies successfully reported two major kinetic phases of α-helix formation, subsequent studies were executed with far-UV TRORD detection largely because of the signal-to-noise advantage mentioned above. Overall, these TRCD and TRORD studies suggested that before the slower, final folding of redcyt c into its native configuration there is an initial fast appearance (the time constant of formation is dependent on the concentration of GuHCl) of a partly unfolded, molten-globule-like intermediate (15–17). With increasing support for this deviation from two-state folding kinetics (15, 16, 22–30), the ability to manipulate the secondary structure of redcyt c using sodium dodecyl sulfate (SDS) provided the timely opportunity to test this proposed mechanism.

The molten globule (MG) state of cyt c can be observed under many different sample conditions (high temperatures, low or high pH, low to intermediate denaturant concentrations) and in the presence of alcohols, sugars, polyanions, anionic lysophospholipids, anionic lipid vesicles, and n-alkyl sulfates such as sodium octyl sulfate, SDS, and sodium decyl sulfate (31–38). In the presence of SDS micelles versus monomers, cyt c will respond differently. Below the critical micelle concentration (CMC) (2.2 mM in 45 mM sodium phosphate), interaction with monomers facilitates partial unfolding of cytochrome c, whereas above the CMC, recovery of structure results from interaction with SDS micelles (39–43). Using this information, titration of oxidized (oxcyt c) and redcyt c with SDS reveals differences in folding free energies for the two different states of cyt c that is critical for rapid initiation of the folding reaction (from the MG state) (20). These experimental conditions are particularly advantageous because one of the complications involved in understanding the role of a partly unfolded intermediate is that comparisons are difficult to make between the equilibrium and kinetic MG states. This is because the equilibrium MG state generally cannot be induced to fold without dramatic changes in the solution conditions. For example, to observe folding to the native state from the acid-induced MG species it is necessary to change the protein solution from low pH and high salt to more neutral buffer conditions (44). However, using the fast electron-transfer method, the equilibrium MG is triggered to fold to the native state without changing the initial solution conditions.

We will describe the steps involved in studying the kinetics of folding from the MG to the native state using the TRORD method, starting with the sample and apparatus materials required, in Subheading 2, followed by the SDS titration, molten globule (MG) characterization, and the TRORD setup and experiments in Subheading 3.

2. Materials

2.1. SDS Titration Samples

1. Fill and then empty a glove bag with argon gas. Repeat three times. Argon gas will be flowing to maintain an oxygen-free environment in the bag throughout the experiments on redcyt *c*.

2. Buffer solution: 50 mM sodium phosphate buffer, pH 7. Weigh 0.69 g monobasic sodium phosphate (MW 137.99 g/mol) and 0.71 g dibasic sodium phosphate (141.96 g/mol) and transfer into two separate beakers. Add 100 mL (see Note 1) deionized water to each beaker and stir until dissolved. Using all or some of the dibasic sodium phosphate, adjust the pH to 7 using the monobasic sodium phosphate solution. Deoxygenate 10 mL 50 mM sodium phosphate buffer with argon or nitrogen and move into the glove bag. Store unused buffers at 4 °C.

3. SDS solution: 100 mM SDS, pH 7 (see Note 2). In an argon- or nitrogen-purging glove bag, deoxygenate 0.029 g SDS (MW 288.38 g/mol) for about 15 min. Add 1 mL deoxygenated 50 mM sodium phosphate buffer, pH 7, to the solid SDS. Divide the solution into two aliquots, leaving one in the glove bag and removing the other from the glove bag.

4. Cytochrome *c* solution: 20 μM in 50 mM sodium phosphate buffer, pH 7. Weigh 0.0012 g oxcyt *c* (horse heart, MW 12,384 g/mol) and add 5 mL 50 mM sodium phosphate buffer, pH 7. Divide the oxcyt *c* solution into two aliquots. Deoxygenate one of the aliquots with argon or nitrogen and then place it into the purging glove bag.

5. In addition to the buffer, SDS, and cyt *c* solutions, add to the glove bag sodium hydrosulfite (more fondly known as sodium dithionite, MW 174.1 g/mol) to reduce oxcyt *c*, potassium ferricyanide (MW 329.3 g/mol) to oxidize redcyt *c*, a 2-mm quartz cuvette with screw cap, a 10 μL pipettor, appropriate pipette tips, and syringes and needles.

2.2. Circular Dichroism and Magnetic Circular Dichroism Measurements

1. Circular dichroism spectrophotometer.

2. 2-, 5-, and 10-mm path length quartz cuvettes for measurements in the far- and near-UV, Soret and visible spectral regions, respectively (see Note 3).

3. 0.64 T PM-2 permanent magnet (JASCO, Inc., Easton, MD).

2.3. TRORD Samples

1. Fill glove bag with argon gas and empty. Repeat three times. During the experiment, the glove bag will be under a constant argon environment.

2. Prepare 50 mM sodium phosphate buffer, pH 7, as described above in Subheading 2.1. Deoxygenate 15 mL 50 mM sodium phosphate buffer and place into the argon-purging glove bag.

3. 20 μM cytochrome c in 50 mM sodium phosphate buffer solution: Weigh 0.0025 g cyt c and place into the argon-purging glove bag. After about 15 min, add 9.94 mL deoxygenated 50 mM sodium phosphate buffer, pH 7 (see Note 4).

4. Weigh 0.029 g SDS for a final solution concentration of 0.65 mM. Place into glove bag.

5. Weigh 0.0035 g NADH (MW 709.4 g/mol) for a final solution concentration of 500 μM and place into glove bag.

6. Place a peristaltic pump, syringe and needles, a 100-μL pipettor, appropriate pipette tips, 2-mm quartz cuvette with screw cap, and sodium hydrosulfite into the glove bag, which now also contains deoxygenated 50 mM sodium phosphate buffer (pH 7), 20 μM cyt c, and solid SDS and NADH.

2.4. Time-Resolved Optical Rotatory Dispersion Apparatus

1. DCR-1 Nd:YAG pulsed laser.

2. Xenon flashlamp probe source.

3. Two MgF_2 prism polarizers.

4. Two UV-enhanced spherical mirrors.

5. Focusing lens.

6. Spectrograph with a 600 grooves/mm (200 nm blaze) grating.

7. CCD Detector.

8. Sample flow cell.

9. Computer.

3. Methods

3.1. SDS Titration (CD Detected) and Structure Characterization

1. Use the deoxygenated 100 mM SDS solution, stored in the glove bag, for the redcyt c samples and the 100 mM SDS solution, outside the glove bag, for the oxcyt c samples.

2. To the oxcyt c sample in the glove bag add sodium dithionite (see Note 9) to convert the oxcyt c to redcyt c. To the other oxcyt c aliquot (outside the glove bag) remove any trace redcyt c (usually <10 % from the bottle), either by stirring with the solution open to air or by addition of potassium ferricyanide (see Note 5).

3. Working with the oxcyt c outside the glove bag and redcyt c inside the glove bag, add small aliquots of 100 mM SDS with the pipettor to obtain protein samples with final SDS concentrations between 0 and 2 mM (see Note 6). SDS is cumulatively added to the sample cyt c solution.

4. Use a 2-mm path length quartz cuvette to make CD measurements in the far-UV region (190–300 nm). The data in this spectral region are collected with a step size of 1 nm, a bandwidth of 1 nm, and an integration time of 8 s. Several sets of data at each SDS concentration should be collected so that a standard deviation can be calculated at each titration point. The results of these titration experiments determine how much SDS to use in the TRORD experiments (see Note 7). A UV–visible measurement should be taken before data collection on the oxcyt *c* samples. For redcyt *c* samples, a UV–visible spectrum should be measured before and after data collection to verify the integrity of redcyt *c* before and during the CD measurement (see Note 9).

5. To characterize the structure of the protein at the concentration of SDS chosen for the TRORD experiments, measure near-UV CD and Soret and visible MCD spectra (see Note 8) for native oxcyt *c* and redcyt *c* and both proteins in the presence of 0.65 mM SDS. For near-UV CD measurements, use a 10-mm path length quartz cuvette with data collection specifications of 1 nm step size, 1 nm bandwidth, and 4 s integration time. Obtain a UV–visible measurement before data collection. For the reduced cyt *c* samples, a UV–visible spectrum should be measured before and after data collection to verify the integrity of the redcyt *c* (see Note 9).

6. Use a 5-mm path length cuvette for MCD measurements from 350 to 650 nm (Soret and visible regions). The 5-mm cuvette fits snugly between the two poles of the magnet. For each sample (oxcyt *c* and redcyt *c* with and without 0.65 mM SDS), collect data for the two different orientations of the magnetic field, antiparallel and parallel to the propagation direction of the light beam. Using an integration time of 2 s, data are collected every 1 nm with a 1 nm bandwidth. MCD spectra are calculated by taking the difference of the signals measured in the parallel and antiparallel magnetic field orientations divided by 2. Take a UV–visible measurement before data collection. For the reduced cyt *c* samples, a UV–visible spectrum should be measured before and after data collection to check that the sample is still in the reduced form (see Note 9).

3.2. TRORD Apparatus: For Small Sample Volumes

1. The xenon flashlamp (i in Fig. 3) is the probe source for the TRORD apparatus and is initially unpolarized.

2. On going from flashlamp to detector, the next optical component is the first MgF_2 prism polarizer (ii).

3. The third optic of this apparatus is a UV-enhanced spherical mirror (iii), which focuses the probe beam down from a

diameter of ~6 mm to ~300 μm, allowing for measurements of small sample volumes.

4. Because of the use of the spherical mirrors, optics ii, iii, iv, v, and vi are aligned in an off-axis configuration, mapping out the letter Z. Position the sample flow cell (iv) after the UV-enhanced mirror (iii) along the diagonal stroke of the letter Z.

5. With the focus at the sample cell, the probe beam is then collected by a second UV-enhanced spherical mirror (v) and directed to the second, analyzing MgF_2 polarizer (vi). This polarizer is oriented with its polarization axis perpendicular to the first MgF_2 polarizer.

6. After passing through the second polarizer, a lens (viii) focuses the probe beam onto the slit of the spectrograph/CCD detection system (ix).

7. The third harmonic of the Nd:YAG laser is used to initiate the photoreduction reaction. The 355 nm laser beam is directed to the sample with prisms and is then focused with a quartz lens down to a ~2-mm beam diameter.

8. Overlap the laser and probe beams (see Note 10).

9. To avoid sample photodegradation, synchronize the peristaltic pump that drives the flow of the sample to the cell to operate only between laser pulses.

10. Set the peristaltic pump at a flow rate that clears the entire volume of irradiated protein sample out of the laser beam footprint before the next laser pulse.

3.3. TRORD Measurements

1. All samples have been deoxygenated and are now handled in the glove bag.

2. Using the syringe, fill the 2-mm cuvette with oxcyt *c*. Measure a UV–vis spectrum to confirm the concentration of the oxcyt *c*, the formation of redcyt *c*, or the presence of SDS and NADH (see Notes 4 and 9). Place cuvette back into the glove bag, remove the oxcyt *c* sample and add back to the stock solution of cyt *c* (see Note 11).

3. Add 1 mL deoxygenated 50 mM sodium phosphate buffer, pH 7 to the solid SDS for a final concentration of 100 mM. To the 20 μM cyt *c* solution add 65 μL of this 100 mM SDS solution to obtain a final SDS concentration of 0.65 mM (see Note 2).

4. Repeat step 2.

5. Add the 20 μM cyt *c*, 50 mM sodium phosphate buffer, pH 7, and 0.65 mM SDS solution to the solid NADH (final 500 μM concentration) and mix well.

6. Repeat step 2.

7. Add a pinch of sodium dithionite to completely reduce the oxcyt *c* without overwhelming the region below 350 nm with absorbance from this reducing agent (see Note 9).

8. Use the peristaltic pump to circulate the deoxygenated buffer from the stock to the sample flow cell and then back.

9. Take equilibrium buffer ORD measurements by collecting data (see Note 12) with the first polarizer in one β orientation and then repeat with the first polarizer in the −β orientation (see Note 13).

10. Circulate the buffer out of the sample flow cell and back into the buffer vial.

11. Flow the protein solution into the sample flow cell. Circulate the solution for a few minutes to incorporate homogeneously any residual buffer with the protein sample (see Note 14).

12. For each time delay, collect data with the first polarizer in the β and then in the −β position (see Notes 13 and 15).

13. Calculate the ORD signal by normalizing the difference between the intensities measured in the +β and −β positions by their sum. That is, $S = (I_{+\beta} − I_{−\beta})/(I_{+\beta} + I_{−\beta})$, which is equal to CB/β.

3.4. Data Analysis

1. SDS titration experiments: Average the data sets for each concentration of SDS measured. Construct titration curves from the multiwavelength CD measurements by averaging the data from 217 to 227 nm to calculate the signal mean value at 222 nm for each concentration of SDS. To determine the error bar for each titration point, calculate standard deviations of the CD signal mean value at 222 nm.

2. Near-UV and MCD characterization: Average the respective data sets measured for the different species: oxcyt *c* and redcyt *c* with and without 0.65 mM SDS. Determine the error bars for the near-UV and MCD spectra by calculating standard deviations for the CD signal mean values at 264 and 552 nm. Obtain the mean values by averaging the respective sets of CD spectra from 262 to 266 nm and 547 to 557 nm.

3. TRORD experiments: Calculate difference TRORD data by subtracting the pre-trigger ORD signal from the time-resolved signals. If the OR spectra have significant baseline shifts they can be offset to zero in the 275–280 nm region (see Note 16). A kinetic trace is extracted from the multiwavelength data by averaging the time-dependent ORD signals from 228 to 236 nm. Fit the kinetic data to exponential processes (see Fig. 2b).

4. Notes

1. The final volumes of all solutions should be based on what is necessary for the experiment. The volumes reported here were designed for these experiments (on the horse heart cytochrome *c* MG) and for minimizing waste of protein and chemicals.

2. The stock SDS solution was prepared at 100 μM so that the volumes (0.5–5 μL) added to the cyt *c* solution to obtain the desired final SDS concentrations did not significantly change the 20 μM cyt *c* concentration.

3. The cuvettes used for all spectral regions probed were quartz. Glass cuvettes can be used for the MCD experiments in the Soret and visible regions.

4. The final concentration of the cyt *c* sample, after addition of SDS, will be 20 μM. However, at this stage the sample concentration is slightly higher, at 21.5 μM.

5. Sometimes it is possible to stir the oxcyt *c* solution in air to remove trace redcyt *c*, however, this method is not as effective as the use of potassium ferricyanide. The absence of redcyt *c* markers in a UV–visible spectrum of oxcyt *c* does not guarantee that all the redcyt *c* has been reoxidized. Rather, a more sensitive probe of redcyt *c* is necessary, such as MCD spectroscopy in the visible region (see Fig. 1). For the SDS titration, it is important to remove the trace redcyt *c* for the oxcyt *c* measurements. However, for the TRORD experiments, the trace redcyt *c* does not affect the results because it does not undergo further
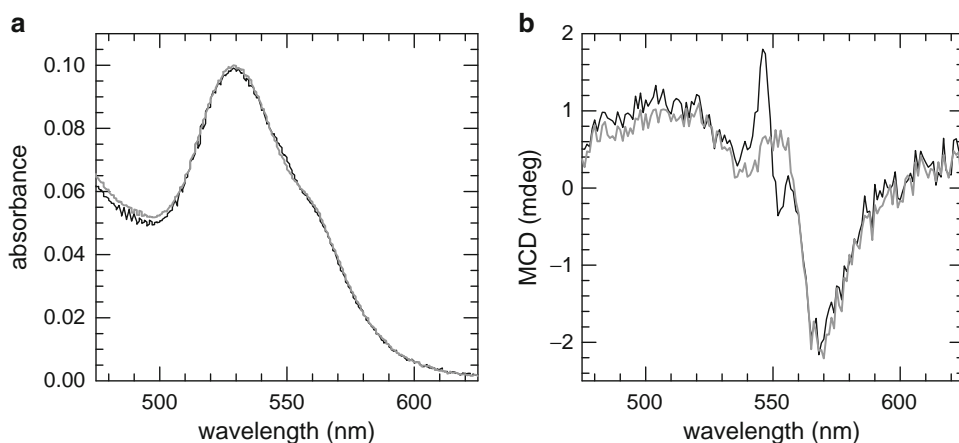
Fig. 1. UV–vis (**a**) and MCD (**b**) spectra of oxcyt *c* with (*gray*) and without (*black*) potassium ferricyanide. Upon addition of potassium ferricyanide, there is an almost imperceptible change in intensity near the 549 nm band, which is characteristic of redcyt *c*, suggesting that the initial sample was almost completely in the oxcyt *c* state. However, when the same samples are probed with MCD spectroscopy, it is clear that without the oxidant significant redcyt *c* is present.

reduction and is therefore subtracted out when difference spectra are calculated. Typically for TRORD experiments, potassium ferricyanide is not added because it absorbs below 350 nm, a region that is important for monitoring NADH changes and protein integrity and especially for signal to noise.

6. To probe the MG state, measurements were obtained below 2 mM because at concentrations above the CMC interactions with SDS micelles results in refolding of cyt c. Below the CMC, cyt c will interact with SDS monomers and undergo partial unfolding. Because very small volumes (0.5–5 μL) are used to change the SDS concentration, it is important to be cautious— wipe the pipet tips along the side so that you avoid adding too much or too little SDS. Cumulative addition of SDS is only one of many ways to obtain these data. For example, a larger volume of cyt c solution can be prepared so that a fresh sample can be used for measurement at each different SDS concentration.

7. For both oxcyt c and redcyt c, far-UV data were obtained at the following concentrations of SDS: 0, 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.2, 1.4, 1.6, and 1.8 mM. Using the CD intensity at 222 nm, the effect of SDS on the secondary structure of the two cyt c proteins is shown in Fig. 2a. Redcyt c is clearly more stable toward SDS than oxcyt c. The maximum decrease in the oxcyt c CD signal is about 30 % and this level is attained by 0.6 mM SDS. In contrast, the redcyt c CD signal only begins to decrease after about 0.7 mM SDS. Based on these observations, the SDS concentration chosen for the TRORD
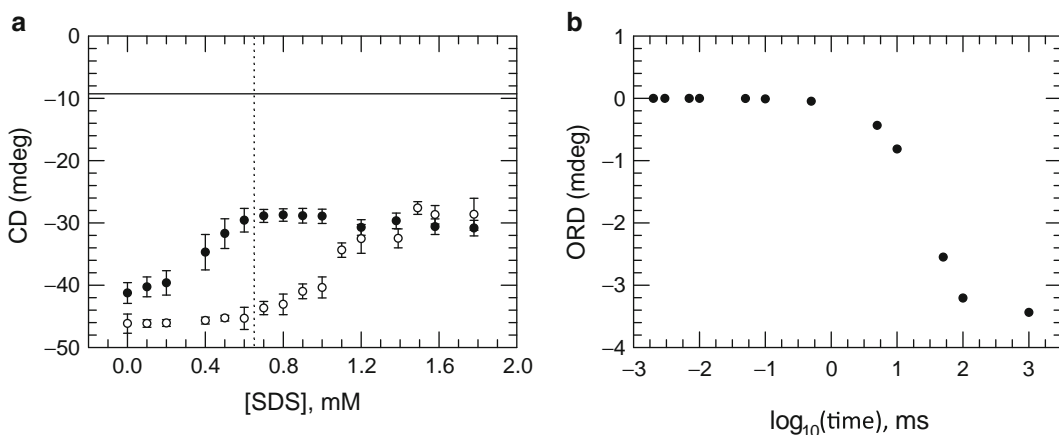
Fig. 2. (**a**) SDS titration curves for oxcyt c and redcyt c. The CD signal for oxcyt c (*black circles*) shows almost an immediate response to SDS, with a decrease in intensity detected at the lowest concentration of SDS (0.1 mM) measured. In contrast, the redcyt c (*white circles*) CD signal does not begin to change until after 0.7 mM SDS. From these results, which represent an average of 8 sets of data, it was determined that 0.65 mM SDS (at *dotted line*) would be used for the TRORD experiments. The *solid black line* represents the intensity of unfolded redcyt c. (**b**) TRORD kinetic results for folding of the redcyt c MG in 0.65 mM SDS. The kinetic trace represents the fit of the data (not shown) to a single exponential process with a time constant of 50 ± 16 ms. The increase in ORD signal magnitude is about 30 % upon folding from the redcyt c MG to the native state.

experiments was 0.65 mM. At this stage of the study, the initial oxcyt c sample (with 0.65 mM SDS) was considered a putative MG intermediate, with 30 % less secondary structure than the native protein, until further structural characterizations described next in Note 8. It was expected that upon photoreduction of the oxcyt c sample, the instantaneous photoproduct is redcyt c with the protein secondary structure of oxcyt c. Because at 0.65 mM SDS, redcyt c prefers to be in the folded conformation, the immediate redcyt c photoproduct will begin to fold. Before proceeding with the TRORD experiments, however, MCD spectra were measured to confirm that indeed oxcyt c assumes an MG conformation and redcyt c is in the native state at 0.65 mM SDS (see Note 8).

8. Different spectral regions are probed to fully understand the structural changes that the protein undergoes in the presence of SDS. In this study, the cyt c proteins were studied in three different spectral regions (far-UV, near-UV, and Soret and visible) with CD and MCD. The 282 and 290 nm bands in the near-UV region report on the integrity of the native tertiary interaction around Trp59 and on the SDS-dependent appearance of the MG state. With Soret and visible MCD measurements, the tertiary contact between the protein and the heme group can also be examined from the point of view of the heme axial coordination. The secondary structure of the proteins is monitored using far-UV CD. In summary, the following results confirmed that in 0.65 mM SDS, oxcyt c is indeed an MG species and redcyt c is in the native state. For oxcyt c, the secondary structure is 30 % less than the native state, there are nonnative tertiary fluctuations around Trp59, as indicated by the disappearance of the 282 and 290 nm Trp bands, and the MCD spectra show that the heme ligation in oxcyt c is largely His18-Fe-His (85 %) with about 15 % His18-Fe-Met. In 0.65 mM SDS, the redcyt c far-UV and near-UV CD are comparable to the corresponding native redcyt c spectra and a comparison of the respective MCD spectra shows only a slight difference in intensity for redcyt c with and without 0.65 mM SDS.

Different integration times and sets of averages are used because of differences in signal strength in the different spectral regions and in the wavelength range measured. For example, because the MCD data are measured over 300 nm, a shorter integration time (relative to the near- and far-UV measurements that step across about 100 nm) is used to avoid reversion of redcyt c back to oxcyt c during the collection time.

9. Use the Soret band ($\varepsilon_{409} = 106{,}000$ M^{-1} cm^{-1}) or the visible band ($\varepsilon_{530} = 10{,}370$ M^{-1} cm^{-1}) to determine the concentration of oxcyt c (45). The extinction coefficients used to calculated the concentration of redcyt c were: $\varepsilon_{519} = 15{,}600$ M^{-1} cm^{-1} or $\varepsilon_{414} = 134{,}770$ M^{-1} cm^{-1}. These values were calculated in

this lab and are slightly different from those reported by Marogliash (45).

In the presence of 0.65 mM SDS, the Soret band for oxcyt *c* blue-shifts to 407 nm (OD ~ 0.47 in a 2-mm path length cell). Upon addition of NADH, the absorbance at 340 nm is used to monitor its concentration—the solution OD_{340} is expected to be about 0.58, which includes the absorbance from both NADH and cyt *c*.

Sodium dithionite has an absorption maximum at 314 nm ($\varepsilon_{314} = 8,000$ M^{-1} cm^{-1} (46)). It is extremely sensitive to oxygen, making it difficult to make solutions of known concentrations by weighing the solid under aerobic conditions. Typically, a few crystals are added to avoid changing the concentration of the protein sample.

10. One approach to optimizing the overlap of the laser and probe beams is to maximize the difference absorption spectrum in the Soret region.

11. It is important to keep track of all your solution so that the concentrations are accurate, but this does not mean that every drop of solution will end up back in your original vial because it is difficult to retrieve the residual solution in the cuvette and the tip of the needle and syringe. To circumvent changes in concentration be sure to rinse the cuvette, syringe, and needle with the solution being prepared at each step. This maintains your sample concentration and is critical for the titration experiments. The final volume might be slightly less than 10 mL, but the concentration should not change if everything is rinsed diligently.

12. How much data to collect for the buffer depends on how many averages are planned for the protein sample. Typically, the latter ranges from 8 to 32 averages at each time point per day of data collection.

13. ORD measurements are based on the idea that when a small reference rotation, β, is introduced into the probe beam polarization axis, the CB from a chiral sample will add to or subtract from β. Thus, the first polarizer (iii, Fig. 3) is rotated ±β relative
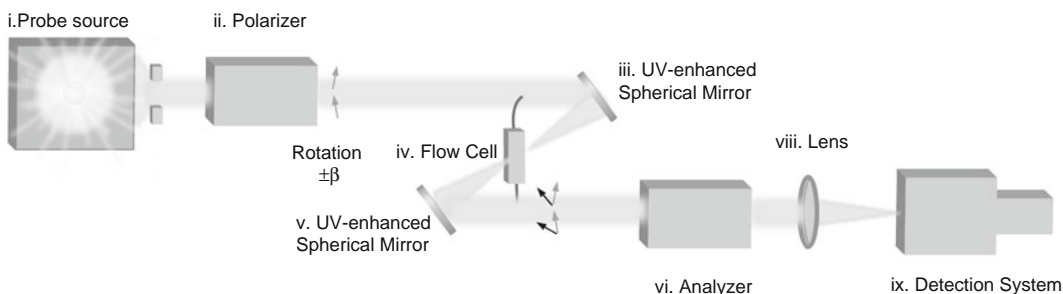
Fig. 3. Time-resolved optical rotatory dispersion apparatus for measurement of small sample volumes. The optics are numbered, i–ix, and described accordingly in Subheading 3.2.

to the cross position of the two polarizers. In general, for nanosecond TRORD studies in this lab, $|OR_{sample}| \ll \beta \ll 90°$, with β typically 0.5–1.5° for experiments in the visible region and 1–3° for experiments in the far-UV region.

14. The volume of buffer remaining in the flow cell after it is flowed out should not change the concentration of the protein sample significantly.

15. Typically TRORD results reflect an average of ≥ 15 experiments, with each experiment comprising 8–32 averages at each time point. The number of time delays depends on the kinetics of the folding reaction, a first determination of which is obtained through time-resolved absorption measurements. In this study, all TRORD/TRCD experiments on heme proteins were prefaced by TROD measurements in the Soret region (not reported). The results of those experiments provide kinetic information that guides the decision of the best time delays to measure for the TRORD studies (see Fig. 2b).

    During the course of the experiment there will be a buildup of redcyt c in the sample. A UV–vis spectrum of the sample will show the visible absorption bands that are hallmarks of redcyt c. The presence of redcyt c in the sample is not a problem because it is blind to additional laser pulses. Rather, the limiting factor will be the increasing absorption by NADH photoproducts at 260 nm (concomitantly with a decrease in absorption at the 340 nm band) during the course of the experiment.

16. At a concentration of 50 μM, oxcyt c in 0 M and 10 M urea have ORD signal intensities of about 0.0007 and 0.0003. At more than half the concentration, the OR spectrum in the 275–280 nm spectral region is effectively zero.

References

1. Lowry TM (1935) Optical rotatory power. Longmans, Green and Co., London
2. Keston A, Lospalluto J (1953) Simple ultrasensitive spectropolarimeters. Fed Proc 12:229
3. Poulsen KG (1960) Evaluation of standard model-D Keston polarimetric attachment for the Beckman Du spectrophotometer. Anal Chem 32:410–413
4. Gallop PM (1957) Simplified formula for the operation of the Keston-type polarimeter. Rev Sci Instrum 28:209
5. Lewis JW et al (1985) New technique for measuring circular dichroism changes on a nanosecond time scale—application to (carbonmonoxy)myoglobin and (carbonmonoxy)hemoglobin. J Phys Chem 89:289–294
6. Milder SJ et al (1990) Assignments of ground-state and excited-state spectra from time-resolved absorption and circular dichroism measurements of the ^2E state of (Δ)-Cr $(Bpy)_3{}^{3+}$. Inorg Chem 29:2506–2511
7. Chen E et al (2010) Nanosecond time-resolved polarization spectroscopies: tools for probing protein reaction mechanisms. Methods 52 (1):3–11
8. Chen EF et al (2005) Nanosecond laser temperature-jump optical rotatory dispersion: application to early events in protein folding/unfolding. Rev Sci Instrum 76:083120
9. Chen EF et al (1993) Time-resolved UV circular dichroism of phytochrome A: folding of the N-terminal region. J Am Chem Soc 115:9854–9855

10. Chen EF et al (1997) Dynamics of the N-terminal α-helix unfolding in the photoreversion reaction of phytochrome A. Biochemistry 36:4903–4908

11. Chen EF et al (2007) A LOV story: the signaling state of the Phot1 LOV2 photocycle involves chromophore-triggered protein structure relaxation, as probed by far-UV time-resolved optical rotatory dispersion spectroscopy. Biochemistry 46:4619–4624

12. Chen EF et al (2003) Dynamics of protein and chromophore structural changes in the photocycle of photoactive yellow protein monitored by time-resolved optical rotatory dispersion. Biochemistry 42:2062–2071

13. Chen EF, Kliger DS (1996) Time-resolved near UV circular dichroism and absorption studies of carbonmonoxymyoglobin photolysis intermediates. Inorg Chim Acta 242:149–158

14. Chen EF et al (1998) Time-resolved circular dichroism studies of protein folding intermediates of cytochrome *c*. Biochemistry 37:5589–5598

15. Chen EF et al (1999) Far-UV time-resolved circular dichroism detection of electron-transfer-triggered cytochrome *c* folding. J Am Chem Soc 121:3811–3817

16. Chen EF et al (2003) Earliest events in protein folding: submicrosecond secondary structure formation in reduced cytochrome *c*. J Phys Chem A 107:8149–8155

17. Chen EF et al (2004) The earliest events in protein folding: a structural requirement for ultrafast folding in cytochrome *c*. J Am Chem Soc 126:11175–11181

18. Chen EF et al (2003) The kinetics of helix unfolding of an azobenzene cross-linked peptide probed by nanosecond time-resolved optical rotatory dispersion. J Am Chem Soc 125:12443–12449

19. Chen E et al (2007) Non-native heme-histidine ligation promotes microsecond time scale secondary structure formation in reduced horse heart cytochrome *c*. Biochemistry 46:12463–12472

20. Chen E et al (2008) The folding kinetics of the SDS-induced molten globule form of reduced cytochrome *c*. Biochemistry 47:5450–5459

21. Moscowitz A (1962) Theoretical aspects of optical activity. 1. Small molecules. Adv Chem Phys 4:67–112

22. Latypov RF et al (2008) Folding mechanism of reduced cytochrome *c*: equilibrium and kinetic properties in the presence of carbon monoxide. J Mol Biol 383:437–453

23. Jones CM et al (1993) Fast events in protein folding initiated by nanosecond laser photolysis. Proc Natl Acad Sci U S A 90:11860–11864

24. Pascher T et al (1996) Protein folding triggered by electron transfer. Science 271:1558–1560

25. Mines GA et al (1996) Cytochrome *c* folding triggered by electron transfer. Chem Biol 3:491–497

26. Telford JR et al (1998) Protein folding triggered by electron transfer. Acc Chem Res 31:755–763

27. Pascher T (2001) Temperature and driving force dependence of the folding rate of reduced horse heart cytochrome *c*. Biochemistry 40:5812–5820

28. Hagen SJ et al (2002) Rapid intrachain binding of histidine-26 and histidine-33 to heme in unfolded ferrocytochrome *c*. Biochemistry 41:1372–1380

29. Pabit SA et al (2004) Internal friction controls the speed of protein folding from a compact configuration. Biochemistry 43:12532–12538

30. Colon W, Roder H (1996) Kinetic intermediates in the formation of the cytochrome *c* molten globule. Nat Struct Biol 3:1019–1025

31. Sanghera N, Pinheiro TJ (2000) Unfolding and refolding of cytochrome *c* driven by the interaction with lipid micelles. Protein Sci 9:1194–1202

32. Dolgikh DA et al (1985) Compact state of a protein molecule with pronounced small-scale mobility: bovine alpha-lactalbumin. Eur Biophys J 13:109–121

33. Moza B et al (2006) A unique molten globule state occurs during unfolding of cytochrome *c* by LiClO4 near physiological pH and temperature: structural and thermodynamic characterization. Biochemistry 45:4695–4702

34. Dolgikh DA et al (1984) "Molten-globule" state accumulates in carbonic anhydrase folding. FEBS Lett 165:88–92

35. Pinheiro TJ et al (1997) Structural and kinetic description of cytochrome *c* unfolding induced by the interaction with lipid vesicles. Biochemistry 36:13122–13132

36. Davis-Searles PR et al (1998) Sugar-induced molten-globule model. Biochemistry 37:17048–17053

37. Sedlak E, Antalik M (1999) Molten globule-like state of cytochrome *c* induced by polyanion poly(vinylsulfate) in slightly acidic pH. Biochim Biophys Acta 1434:347–355

38. Das TK et al (1998) Characterization of a partially unfolded structure of cytochrome *c*

induced by sodium dodecyl sulphate and the kinetics of its refolding. Eur J Biochem 254:662–670

39. Hiramatsu K, Yang JT (1983) Cooperative binding of hexadecyltrimethylammonium chloride and sodium dodecyl sulfate to cytochrome *c* and the resultant change in protein conformation. Biochim Biophys Acta 743:106–114

40. Takeda K et al (1985) Kinetic aspects of the interaction of horse heart cytochrome *c* with sodium dodecyl sulfate. Arch Biochem Biophys 236:411–417

41. Oellerich S et al (2003) Conformational equilibria and dynamics of cytochrome *c* induced by binding of sodium dodecyl sulfate monomers and micelles. Eur Biophys J 32:599–613

42. Moosavi-Movahedi AA et al (2003) Formation of the molten globule-like state of cytochrome *c* induced by n-alkyl sulfates at low concentrations. J Biochem 133:93–102

43. Xu Q, Keiderling TA (2004) Effect of sodium dodecyl sulfate on folding and thermal stability of acid-denatured cytochrome *c*: a spectroscopic approach. Protein Sci 13:2949–2959

44. Sosnick TR et al (1994) The barriers in protein folding. Nat Struct Biol 1:149–156

45. Margoliash E, Frohwirt N (1959) Spectrum of horse-heart cytochrome *c*. Biochem J 71:570–578

46. Di Iorio EE (1981) Preparation of derivatives of ferrous and ferric hemoglobin. Methods Enzymol 76:57–72

Chapter 24

The Use of UV–Vis Absorption Spectroscopy for Studies of Natively Disordered Proteins

Eugene A. Permyakov

Abstract

Absorption spectroscopy can be used to monitor structural changes upon transitions from ordered to disordered state in proteins. Changes in environment of tryptophan, tyrosine, and phenylalanine residues result in changes of their absorption spectra. In most cases the changes are small and can be measured only in a differential mode.

Key words: Absorption spectrum, Proteins, Tryptophan, Tyrosine, Phenylalanine, Environment, Natively disordered proteins

1. Introduction

Absorption of ultraviolet (UV) and visible (Vis) light by organic molecules causes electronic transitions. When a molecule absorbs UV–Vis radiation, the absorbed energy excites an electron into an empty, higher energy orbital. In ultraviolet and visible spectroscopy, the energy absorption causes transitions that occur between electronic energy levels of valence electrons, that is, orbitals of lower energy are excited to orbitals of higher energy. The absorbance of energy plotted against the light wavelength or frequency is called absorption spectrum. UV–Vis spectra are often called electronic spectra. Ultraviolet radiation has wavelengths of 200–400 nm. Visible light has wavelengths of 400–800 nm. UV–Vis spectroscopy has many uses including detection of eluting components in high performance liquid chromatography (HPLC), determination of the oxidation state of a metal center of a cofactor (such as a heme), determination of the maximum absorbance of proteins for measurement of their concentrations or monitoring of structural changes in proteins. Most organic compounds that absorb

UV–Vis radiation contain conjugated π-bonds. Both the shape of the peak(s) and the wavelength of maximum absorbance ($\lambda_{\max}$) give information about the structure of a compound and properties of its environment.

Let us consider a tresspassing of a monochromatic light beam through a cell with an absorbing solution. Intensity of the light beam before the cell is I_0, while its intensity after the cell is decreased down to I. The ratio of these intensities is called *transmission coefficient T*:

$$T = \frac{I}{I_0}. \tag{1}$$

In some cases it is useful to use *absorption coefficient* $(1 - T)$:

$$1 - T = \frac{I_0 - I}{I_0} = \frac{I_a}{I_0}, \tag{2}$$

where I_a is an intensity of the absorbed light.

An increase in concentration of the absorbing substance or thickness of the cell results in an increase in $(1 - T)$ and a decrease in T. It is of importance that these values are not directly proportional to the solution concentration and cell thickness.

One can easely find the dependence of T on absorbing substance concentration and cell thickness. Let us consider very thin layer dl of solution, which is perpendicular to the light beam (Fig. 1). Let concentration of the absorbing molecules is n (cm^{-3}). Each molecule is characterized by an effective cross section $s\,(cm^2)$, which is called *absorption cross section*. One can imagine it as an area absorbing photons hiting in it. The value of s depends upon light wavelength and is proportional to the probability of light absorption. Total area of the absorbing targets per 1 cm^2 of the
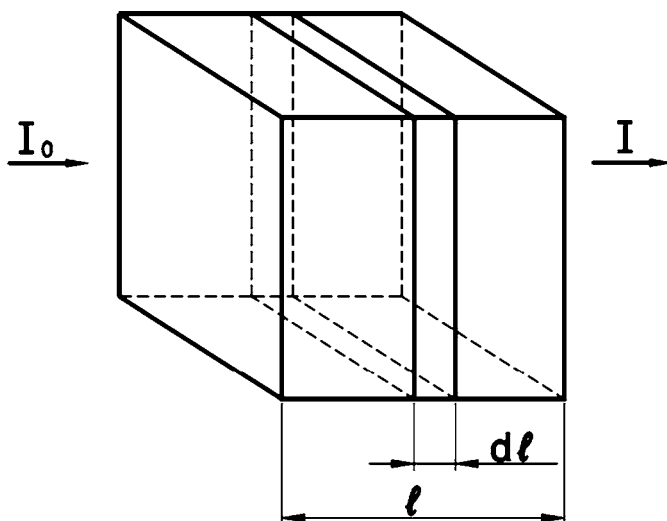
Fig. 1. Trespassing of a light beam through a cell with absorbing substance.

solution of the thickness dl is $snd l$. An attenuation of the light intensity by the layer dl is

$$\frac{\mathrm{d}I'}{I'} = -sn\mathrm{d}l. \tag{3}$$

After integration of Eq. 3 one can obtain:

$$\ln\left(\frac{I}{I_0}\right) = -snl, \tag{4}$$

where l is cell thickness. From Eq. 4 we can obtain an equation:

$$\frac{I}{I_0} = T = \mathrm{e}^{-snl}, \tag{5}$$

which is called the *Bouguer-Lambert-Beer law*. One can easily see that T and $(1 - T)$ depend exponentially on n and l. The exponetial dependence is not very convenient for determination of concentrations. It is more convenient to use decimal logarithm and molar concentration c (mol/L, M):

$$-\log\left(\frac{I}{I_0}\right) = -\log T = \varepsilon c l = A, \tag{6}$$

where A is *absorbance* and ε is *molar absorption coefficient* $(\mathrm{M}^{-1}/\mathrm{cm})$, which is connected with s by the equation:

$$\varepsilon = \mathrm{N_A} \times s \times \log e/1{,}000 = 2.62 \times 10^{20}\, s, \tag{7}$$

where $\mathrm{N_A}$ is Avogadro constant. Modern spectrophotometers measure a ratio of I to I_0, i.e., the ratio of light intensity after the cell with absorbing solution to the light intensity after a similar reference cell filled with solvent. Usually such instruments transform the ratio to absorbance A according to Eq. 6. In contrast to T and $(1 - T)$, A is additive value: absorbance of a mixture of two solutions is a sum of absorbances of these solutions. Interrelation between $(1 - T)$ and A is:

$$1 - T = 1 - 10^{-A}. \tag{8}$$

In the case of small A values Eq. 8 can be rewritten as

$$1 - T \approx A \ln 10 \approx 0.4343A. \tag{9}$$

The smaller is the A value, the more precise is Eq. 9.

The dependence of transmittance T on wavelength λ is called *transmittance spectrum*. Such spectra are usually used for characterization of optical filters but not substances, since their shape depends upon both substance concentration and cell width. The dependence of A on wavelength λ is called *absorption spectrum*. The absorption spectrum of a substance is usually normalized to a concentration unit and cell width, i.e., the absorption spectrum of a substance is the dependence of molar absorption coefficient ε on wavelength λ.

Absorption spectra of water solutions of biologically significant substances are relatively smooth curves with a single or multiple maxima. In contrast, atoms and simple molecules in a gas phase give line, discret spectra.

UV–Vis spectrometers come in four basic models: low-cost single-beam, dual-beam, array-based, and handheld. Single-beam techniques must apply a correction for the loss of light intensity as the beam passes through the solvent. Dual-beam spectrometers use a second solvent reference cell and perform the correction automatically. Single- and dual-beam benchtop instruments use a broad spectrum lamp as a light source, and most use a photomultiplier tube as a detector. Some instruments employ photodiode array detectors. Handheld instruments are usually dedicated to one wavelength and analyte, for example, for water testing in the field. Handhelds employ single wavelength light-emitting diode light sources and photodiode detectors. Array-based instruments use a very broad-spectrum tungsten lamp emitting between 200 and 1,000 nm, and use a photodiode array or charge-coupled device as a detector. Array instruments do not scan the UV–Vis spectrum in the traditional sense, but rather send and collect the entire spectrum at once. Array instruments are useful when spectra must be acquired quickly, for example, for fast kinetics experiments.

Figure 2a shows a functional scheme of a single-beam spectrophotometer. The light from a lamp L (most often xenon lamp
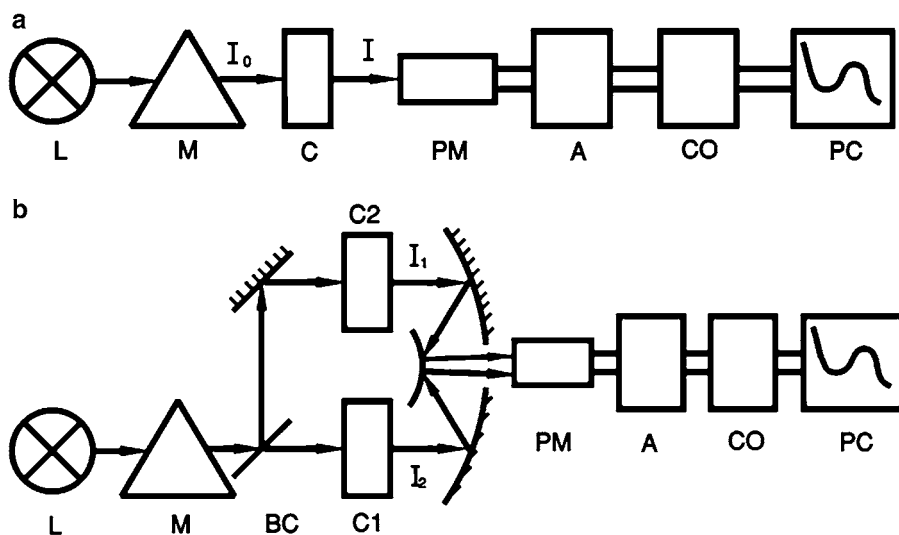
Fig. 2. (**a**) A scheme of a single-beam spectrophotometer. L is a lamp (most often a xenon lamp); M is a monochromator; C is a cell with a sample solution; PM is a photomultiplier. Photocurrent from PM is applied to an amplifier A and converter CO and then the signal goes to a computer. (**b**) A scheme of a dual-beam spectrophotometer. Monochromatic light from a lamp L is splitted into two beams by means of a beam chopper, BC. The two similar light beams pass through the measuring and reference cells C_1 and C_2. Photomultiplier PM measures light signals by turn coming from the two cells. After amplification by A the electrical signals are converted by CO to the ratio $I_1/I_2 = \Delta T$ and then to $\Delta A = -\log(\Delta T)$.

giving continuous emission spectrum in the UV and Vis regions) passes through a monochromator M. A beam of monochromatic light passes through a cell C with a sample solution and is collected by a photomultiplier PM. Photocurrent from PM is applied to an amplifier A and a converter CO and then the signal goes to a computer. In the single-beam spectrophotometer the spectrum of solvent is measured first. The spectrum is memorized in the converter and used as a zero level in the measument of absorption spectrum of a sample. The converter converts the signal into transmittance T or absorbance A according to Eqs. 1 and 6, respectively.

In dual-beam spectrophotometers (Fig. 2b) monochromatic light beam is splitted into two beams by means of a rotating mirror with sector cuts (beam chopper, BC). The two similar light beams pass through the measuring and reference cells C_1 and C_2. Photomultiplier PM measures light signals in turn coming from the two cells. After amplification the electrical signals are converted to the ratio $I_1/I_2 = \Delta T$ and then to $\Delta A = -\log(\Delta T)$. Thus, the dual-beam spectrophotometer is intended to measure difference absorption spectra. It can measure common absorption spectra as well if you use solvent as a reference. Dual-beam instruments are more sensitive compared with single-beam ones and allow measurement of a difference in absorbance up to 10^{-4}.

Proteins contain aromatic amino acid residues, which absorb light in the UV-range. These amino acids are tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe) (Fig. 3). All these molecules possess conjugated double bonds formed by π-electrons. Tryptophan has a nitrogen heteroatom.

Tryptophan has two absorption bands (see ref. 1 for a review): the first one with maxima at 280 nm ($\varepsilon_{280} = 5,500/M/cm$) and 288 nm ($\varepsilon_{288} = 4,500/M/cm$) and a shoulder at 265 nm ($\varepsilon_{265} = 4,500/M/nm$) and another one with maximum at 220 nm ($\varepsilon_{220} = 36,000/M/cm$) (Fig. 4a shows only long-wave band most important for absorption spectroscopy of proteins). The long-wave band demonstrates a weak vibrational structure: a shoulder at 271 nm, the main maximum at 279.5 nm and a
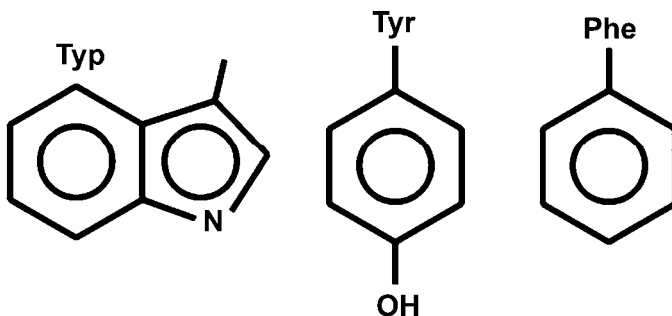
Fig. 3. Aromatic amino acid residues.
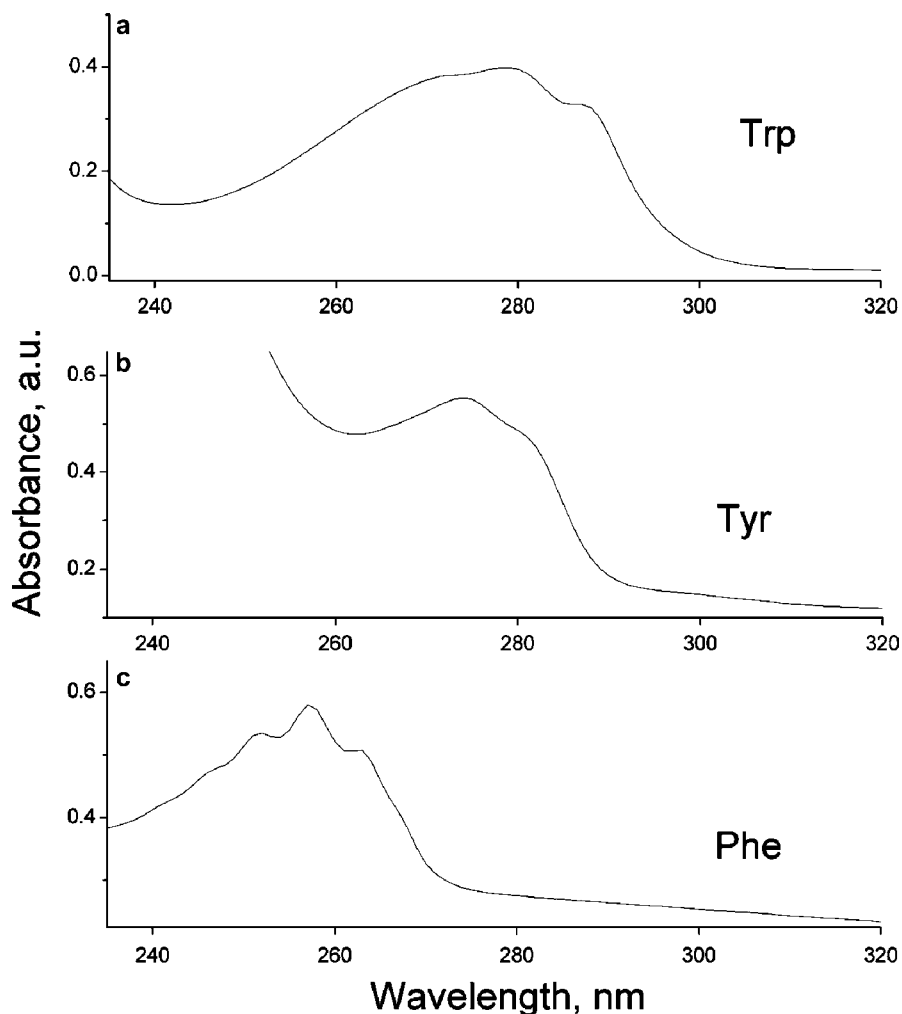
Fig. 4. Absorption spectra of tryptophan (**a**), tyrosine (**b**), and phenylalanine (**c**) in water.

sharp peak at 288 nm. The absorption bands are due to electron-vibrational $\pi \rightarrow \pi^*$ transitions in the indole ring, aromatic π-system of which is formed by ten π-electrons. Quantum mechanical calculations show uneven distribution of electron density among the indol ring atoms, which results in a large dipole moment of this molecule and a large change of the dipole moment upon the transition of the molecule to the excited state. This, in its turn, results in dipole–dipole and specific interactions between the chromophore and molecules of polar solvents. Calculations of the dipole moment of the ground state of indole gave the value 1.8–2.2 D (2).

The long-wave band of the indole chromophore in the region from 250 to 290 nm is assumed to consist of two electron transitions (see ref. 1 for review). An analysis of fluorescence polarization

spectra revealed two superimposed absorption bands with almost perpendicular transition vectors, which were ascribed to the $^1L_a \leftarrow A$ (transition vector along the short axis of the molecule, absorption maximum at 270 nm) and $^1L_b \leftarrow A$ (transition vector oriented along the long axis of the molecule, absorption maxima at 280 and 290 nm). The absorption band with the maximum at 218 nm contains one more electron transition—$^1B_b \leftarrow A$. It was found that the ^1L_a transition results in a change in charge distribution in the indole chromophore: electron density on the atoms 1 and 3 decreases, while electron density on the atoms 4, 7, and 9 increases (numbers start with the nitrogen atom and go counterclockwise). The ^1L_b transition does not cause such changes. It was revealed that the ^1L_a transition is mostly due to the transition from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) (3). The transition ^1L_b contains almost equal contributions of the HOMO → LUMO + 1 and HOMO − 1 → LUMO transitions.

Alkyl displacement in the position 3 of the indole ring results in a red shift of the ^1L_a and ^1L_b bands by 400 and 1,400/cm respectively. Semiempiric calculations show that orientations of the transition dipoles in tryptophan and indole differ no more then by 10°.

An increase in solvent polarity causes a red shift of the tryptophan absorption spectrum (see ref. 1 for review). The shift for ^1L_a is essentially more pronounced than for ^1L_b. The shift is caused by both universal and specific interactions because of formation of complexes with solvent molecules with participation of the NH-group of indole ring. It should be noted however that the spectral changes for indole chromophore induced by change in polarity are small and detectable only by means of differential spectrophotometry.

Figure 4b shows absorption spectrum of tyrosine in neutral water solution (only long-wave band is shown). The spectrum has maxima at 222 nm ($\varepsilon_{222} = 8,000/M/cm$) and 275–277 nm ($\varepsilon_{275} = 1,230/M/cm$) (see ref. 1 for review). The red band has a slightly resolved vibration structure with shoulders at 267 and 282 nm. The absorption is caused by the $\pi \rightarrow \pi^*$ electron transitions. The lowest singlet transition in tyrosine is the ^1L_b band, while a more intensive high energy transition is the ^1L_a band (3). Formation of hydrogen bonds causes red shift of tyrosine spectrum and increases its intensity.

Deprotonation of the hydroxylic group of tyrosine essentially changes its absorption spectrum: it shifts to longer wavelengths and its intensity increases (Fig. 4b). In this case the spectrum has maxima at 240 nm ($\varepsilon_{240} = 11,700/M/cm$) and 293 nm ($\varepsilon_{293} = 2,340/M/cm$). Ionization pK_a of tyrosine in water is 10.04 ± 0.03. The values of ionization pK_a for tyrosine residues in proteins vary in a wide range and depend upon their environment.

Figure 4c shows absorption spectrum of phenylalanine in neutral water solution. The spectrum is characterized by a rich vibration structure: maxima at 187.5, 205, 242, 252, 257 (main peak), 263, and 267 nm. Its molar absorption coefficient at 257 nm is about 200/M/cm. Changes in pH or solvent polarity cause very small changes in absorption spectrum of phenylalanine.

It should be noted that Trp, and especially Tyr, and Phe chromophores possess relatively small dipole moments in the ground state therefore their interactions with adjacent dipoles are weak and they hardly feel any changes in environment. Trp chromophore is the most sensitive to any changes in properties in its environment.

Absorption spectrum of a protein is a sum of absorption spectra of its aromatic amino acid residues Trp, Tyr, and Phe, and cystine residues. Most proteins contain all three absorbing aromatic amino acids. There exist few proteins which do not contain tryptophans and there are very few proteins which do not contain both tryptophans and tyrosines. Figure 5a shows absorption spectrum of pike parvalbumin pI 5.0, which does not have Trp and Tyr residues and its absorption is due to ten Phe residues. Figure 5b demonstrates absorption spectrum of pike parvalbumin pI 4.2, which does not contain Trp residues, but has ten Phe and one Tyr residue. The absorption spectrum of cod parvalbumin containing nine Phe, one Tyr, and one Trp is shown in Fig. 5c.

Many proteins lack rigid 3D structures under physiological conditions in vitro; they exist instead as dynamic ensembles of interconverting structures (see ref. 4 for review). These proteins are called intrinsically disordered (IDPs). IDRs have dynamic structures that interconvert on a number of timescales, with similarities to nonrigid states of normal globular proteins. The four different protein conformations are structured (folded, ordered), molten globular, premolten globular, and unstructured (unfolded). In other words, IDPs can be either collapsed (as a molten globule) or extended (as a coil or premolten globule). They are highly flexible and lacking stable secondary and tertiary structures. For this reason, environment of aromatic amino acid residues in IDPs is also highly flexible which can be reflected in their absorption spectra. The Trp, Tyr, and Phe absorption spectra of IDPs can be smoother and less structured compared with the spectra of ordered proteins. It should be noted however, that these differences are usually small and can be revealed in most cases only by means of differential spectrophotometry. Generally speaking, the absorption spectroscopy method is not very suitable for studies of IDPs. In this chapter we will try to show how the absorption spectroscopy can be applied for studies of IDPs.

Parvalbumin is a small (Mr 12 kDa), acidic (pI 3.9–6.6), cytosolic Ca^{2+}-binding protein of the EF-hand superfamily, found in lower and higher vertebrates, including humans (for review see ref. 5). It was detected in fast-twitch muscle cells, specific neurons of the central and
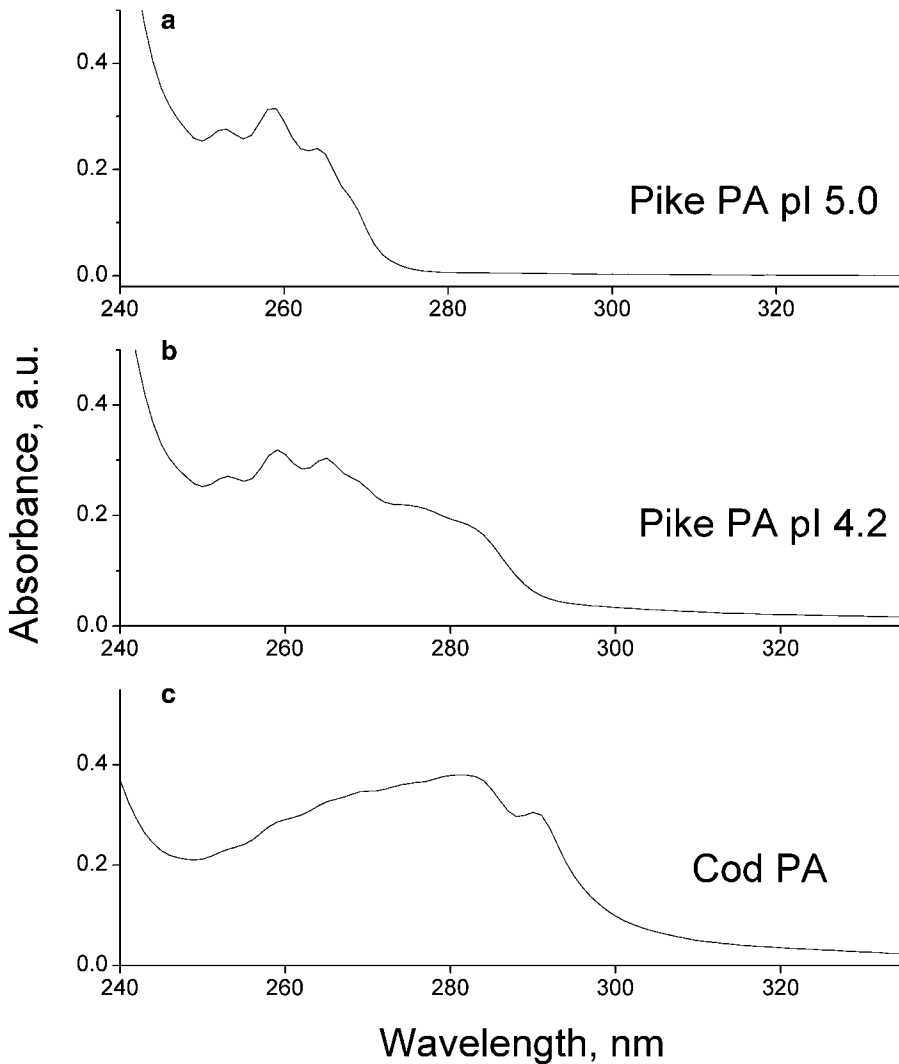
Fig. 5. Absorption spectra of pike parvalbumin pI 5.0 (0 Trp, 0 Tyr, 10 Phe) (**a**); pike parvalbumin pI 4.2 (0 Trp, 1 Tyr, 10 Phe) (**b**); and cod parvalbumin (1 Trp, 1 Tyr, 9 Phe) (**c**).

peripheral nervous system, certain cells of several endocrine glands, and sensory cells of the mammalian auditory organ, the organ of Corti, and some other cells. The highest concentration of parvalbumin (up to several millimoles per liter) was found in fast muscles (mainly in skeletal, but sometimes in cardiac). Parvalbumin was shown to serve as a soluble relaxing factor accelerating the Ca^{2+}-mediated relaxation phase in fast muscles, which was supported by direct gene transfer experiments and by studies of parvalbumin knockout mice. The exact functions of these proteins are still unknown, nevertheless their major role is assumed to be buffering, transport of Ca^{2+}, and regulation of various enzyme systems.

It has been found that metal-depleted forms of two pike parvalbumin isoforms lack first-order thermal transitions due to the absence of fixed tertiary structure, i.e., it seems to be an IDP in these conditions (6). Here we demonstrate how the transition of cod parvalbumin from the ordered Ca^{2+}-loaded form to disordered metal-free form can be monitored by means of absorption spectroscopy.

2. Materials

Prepare all solutions using ultrapure water and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise).

Prepare 50 mL 20 mM HEPES buffer pH 8.5 using deionized water. Unfortunately, tryptophan-containing cod parvalbumin is not produced for sale by any commercial company, nevertheless it can be relatively easily extracted from white muscles of cod or any other *Gadidae* fish (5). Since parvalbumin has two strong calcium binding sites per molecule, lyophilized parvalbumin samples usually almost totally saturated by calcium ions. Molar absorption coefficient for cod parvalbumin at 280 nm is 7,189/M/cm (7), its molecular mass is about 12 kDa.

Prepare 7 mL of 50 μM cod parvalbumin ($A_{280\,nm} = 0.36$) and add to it one equivalent of $CaCl_2$ to be sure that parvalbumin is saturated by calcium. It is better to use commercial stock solution of $CaCl_2$, since solid $CaCl_2$ is extremely hygroscopic and needs long drying before use.

Prepare 1 mL of 0.1 M EGTA, pH 8.5. Since parvalbumins possess relatively high affinity for sodium ions, do not use disodium salt of EGTA, use acid form of EGTA or its dipotassium salt instead. It is better to use commercial stock solution of EGTA.

3. Methods

Any type of dual-beam spectrophotometers can be used for these measurements, for example, Cary 100, Varian. Put 1 cm cell with 3 mL of Ca^{2+}-parvalbumin solution in 20 mM HEPES to the first cell-holder and put similar cell with 20 mM HEPES to the second cell-holder as a reference. Measure protein absorption spectrum in the region from 240 to 340 nm (Fig. 6). The spectrum contains contributions from the absorption of phenylalanine (several small peaks in the region from 250 to 280 nm), tyrosine, and tryptophan residues. The existence of a distinct tryptophan peak at 292 nm suggests a rigid and mostly unpolar environment of the single
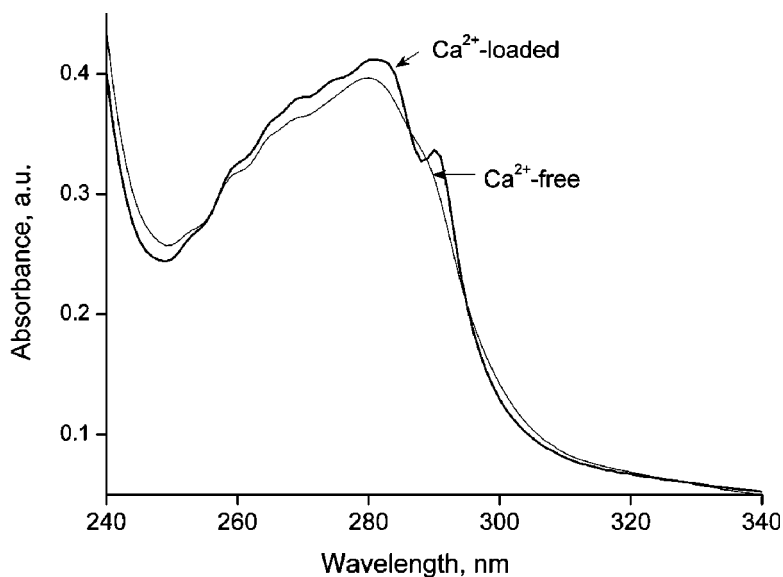
Fig. 6. Absorption spectra of Ca^{2+}-loaded and apo-states of cod parvalbumin in 20 mM HEPES, pH 8.5.
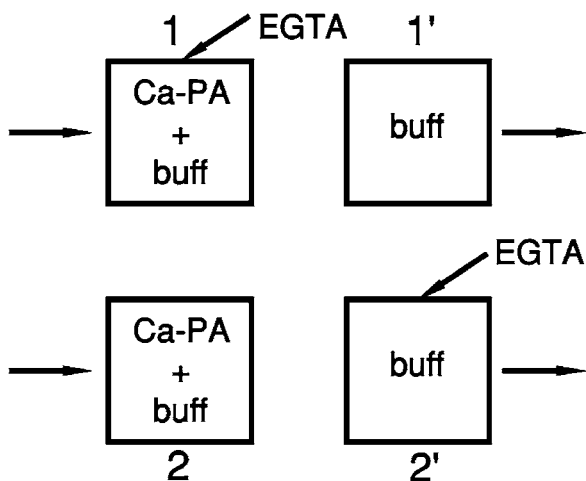
Fig. 7. Arrangement of cells for measuring of differential absorption spectrum (Ca^{2+}-loaded parvalbumin against apo-parvalbumin).

tryptophan residue in cod parvalbumin, which is typical for the ordered calcium-loaded state of the parvalbumin.

In order to study a transition of cod parvalbumin from the ordered calcium-loaded to intrinsically disordered apo-state one can measure a differential absorption spectrum. Put four 1-cm quartz cells to the spectrophotomer (Fig. 7). Two of them should contain 3 mL of 50 μM Ca-protein solution in 20 mM HEPES, pH 8.5, while two others should be filled with 3 mL of the buffer. Add 20 μL 0.1 M
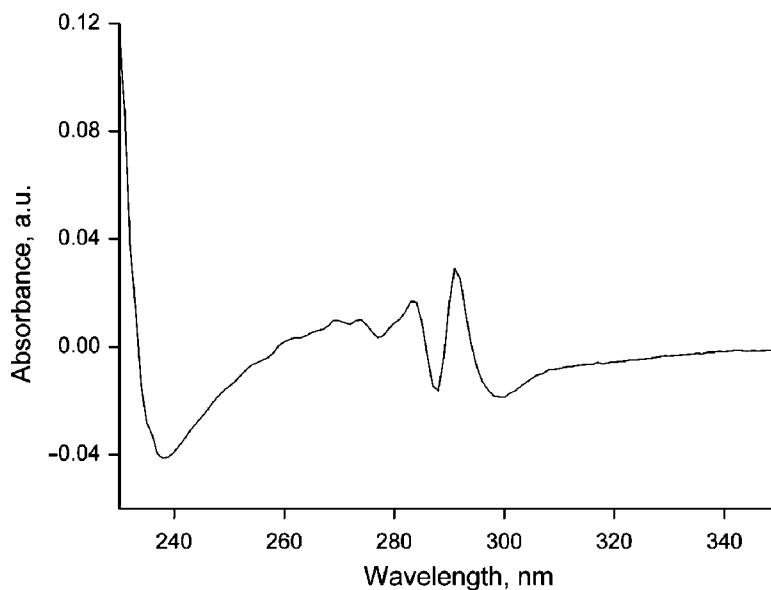
Fig. 8. Differential absorption spectrum of cod parvalbumin: Ca^{2+}-loaded against apo-state.

EGTA to the cells 1 and 2′ to remove calcium from the protein and to compensate an absorption of EGTA in the short vave region. Add also 20 μL of the buffer to the cells 2 and 1′ to compensate the dilution of the protein in the cell 1. Measure differential absorption spectrum for the sandwich system (Fig. 8). It is clearly seen in Fig. 8 that the removal of calcium changes absorption and hence environment of phenylalanine, tyrosine, and tryptophan residues.

One can obtain the absorption spectrum of apo-parvalbumin measuring absorption the cell 1 against the cell 2′. The resulting spectrum of cod apo-parvalbumin is shown in Fig. 7. The removal of calcium results in disappearance of the sharp peak in tryptophan absorption spectrum at 292 nm and slightly changes absorption of phenylalanine chromophores.

4. Note

It is worth to note that such pronounced changes in absorption spectra upon transition from ordered to disordered state do not happen very often. In most cases the spectral changes are very small and cannot be measured even in a differential mode.

Acknowledgments

References

1. Permyakov EA (1993) Luminescent spectroscopy of proteins. CRC, Boca Raton, FL
2. Slater LS, Callis PR (1995) Molecular orbital theory of the ^1L_a and ^1L_b states of indole. 2. An ab initio study. J Phys Chem 99:8572–8581
3. Callis PR (1997) ^1L_a and ^1L_b transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins. Methods Enzymol 278:113–150
4. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Curr Opin Struct Biol 18:756–764
5. Permyakov EA (2006) Parvalbumin. Nova Science Publishers, New York, NY
6. Permyakov SE, Bakunts AG, Denesyuk AI, Knyazeva EL, Uversky VN, Permyakov EA (2008) Apo-parvalbumin as an intrinsically disordered protein. Proteins 72:822–836
7. Closset JI, Gerday C (1976) Parvalbumins of white muscles of Gadidae. II. Properties and existence of two evolutionary lineages. Comp Biochem Physiol B 55:537–542

Chapter 25

Intrinsic Fluorescence of Intrinsically Disordered Proteins

Paolo Neyroz and Stefano Ciurli

Abstract

Resolution of the intrinsic emission properties of a protein by different fluorescence spectroscopy techniques is an invaluable tool to detect and characterize its structural architecture and conformational changes under different experimental conditions. Indeed, the multidimensional character of fluorescence can provide information on local chemical features, on solvent diffusional processes, and on rotational movements of peptide chains or whole proteins. Here, we describe the details of quenching fluorescence experiments and how to correlate the results to the peculiar structural information on the organization of intrinsically disordered proteins (IDPs).

   Key words: Fluorescence spectroscopy, Fluorescence quenching, Protein folding

1. Introduction

The emission of photons from an electronic excited state defines the fluorescence phenomenon (1). This light emission has a distinct energy distribution, a characteristic wavelength profile of its spectrum, as well as a distinct time-dependent distribution, that is, its fluorescence kinetics or lifetime (2). Since fluorescence emission takes place in the pico-to-nanosecond timescale, all the concomitant events that may interfere with fluorescence can be in principle investigated. In this respect, many biological reactions and processes take place during this time window and for this reason fluorescence techniques have gained a vast popularity among biology's scientific communities. In addition, the high sensitivity related to the emission fluorescence signal allows measurements with biological samples at very low concentrations, typically in the micromolar range. In protein studies (2, 3), the intense fluorescence of tryptophan residues dominates the contribution of the other aromatic residues showing weaker intrinsic emission (tyrosine and phenylalanine) and provides an ideal tool to perform fluorescence investigations using amounts of protein very

close to the biological concentration, under noninvasive experimental conditions. Albeit the apparently limited source of structural information offered by the unique environment of tryptophan residues in proteins, the variety of biophysical parameters that can be potentially investigated is rather wide because of the multidimensional character of fluorescence.

In particular, here we describe how to obtain information about tryptophan residues' solvent accessibility by steady-state fluorescence quenching experimental data (4). In quenching measurements, the diffusion rate by which a quencher molecule affects the intrinsic fluorescence of the emitting residue can be probed. If different conditions are compared (i.e., native protein vs. unfolded or partially unfolded forms), changes of the residue environment can be detected (i.e., buried inside the protein core vs. exposed to large interactions with the solvent).

2. Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ cm at 25 °C). Guanidine hydrochloride (GuHCl) and potassium iodide (KI) are analytical grade and stored at room temperature.

2.1. Preparation of Protein Sample

1. Buffer: 50 mM Tris–HCl at pH 8.0.
2. Protein solution: 20 μM.
3. Quenching stock solution: KI (4 M) freshly prepared in buffer (see Note 1).
4. Denaturing stock solution: GuHCl (6 M) and NaCl (4 M) (see Note 2).

3. Methods

3.1. Fluorescence Quenching Measurements

Steady-state fluorescence intensities and emission spectra of the protein solution are recorded using a steady-state spectrofluorometer operating in photon-counting mode, bandpasses of 2.5 nm, and an excitation wavelength of 295 nm (see Note 3). Use a quartz fluorescence cuvette of 1 cm path length and 4 mL volume capacity. In these experiments, the fluorescence of the protein solution is measured at increasing concentrations of potassium iodide (KI) used as quencher (see Note 4).

3.1.1. Sample Preparation in the Absence of Denaturant (Experiment 1)

Prepare 11 samples as follows:

$$sample_N = a_N + b + c + d.$$

where: a_N = volume of the protein sample (1,240 µL).

b = volume (µL) of KI stock solution (b = 0, 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200).

c = volume (200 − b µL) of NaCl stock solution.

d = volume (360 µL) of buffer.

3.1.2. Sample Preparation in the Presence of Denaturant (Experiment 2)

Prepare 11 samples as follows:

$$sample_U = a_U + b + c + d.$$

where: a_U = volume of the protein sample (1,240 µL).

b = volume (µL) of KI stock solution (b = 0, 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200).

c = volume (200 − b µL) of NaCl stock solution.

d = volume (360 µL) of GuHCl stock solution.

3.1.3. Fluorescence Intensity Measurements

The entire emission spectrum is recorded for each sample as a function of increasing KI concentration. The maximum fluorescence intensity of each sample is used as input for a spreadsheet software as shown in Fig. 1a (see Note 5).

3.1.4. Data Analysis

The data are analyzed according to the Stern–Volmer equation (2, 5):

$$\frac{F_0}{F} = 1 + K_{SV}[Q]$$

where F_0 and F are the fluorescence intensities measured in the absence and presence of the quencher, K_{SV} is the Stern–Volmer constant, and $[Q]$ is the quencher concentration (see Fig. 1a). A typical result of this analysis is reported in Fig. 1b. A linear fit of the data provides the Stern–Volmer constant, K_{SV}, as the slope of the straight line (see Note 6).

A typical example of the results is shown in Fig. 2. The Stern–Volmer constants $K_{SV}(N)$ of 1.2 M^{-1} and $K_{SV}(U)$ of 3.9 M^{-1} are recovered from a fit to the points obtained in experiments 1 and 2, respectively. The observed changes of the quenching constant K_{SV} indicate that the tryptophan residue moves from a native state, hydrophobic-buried environment to a more polar-exposed, partially unfolded state, upon addition of 1.2 M GuHCl.
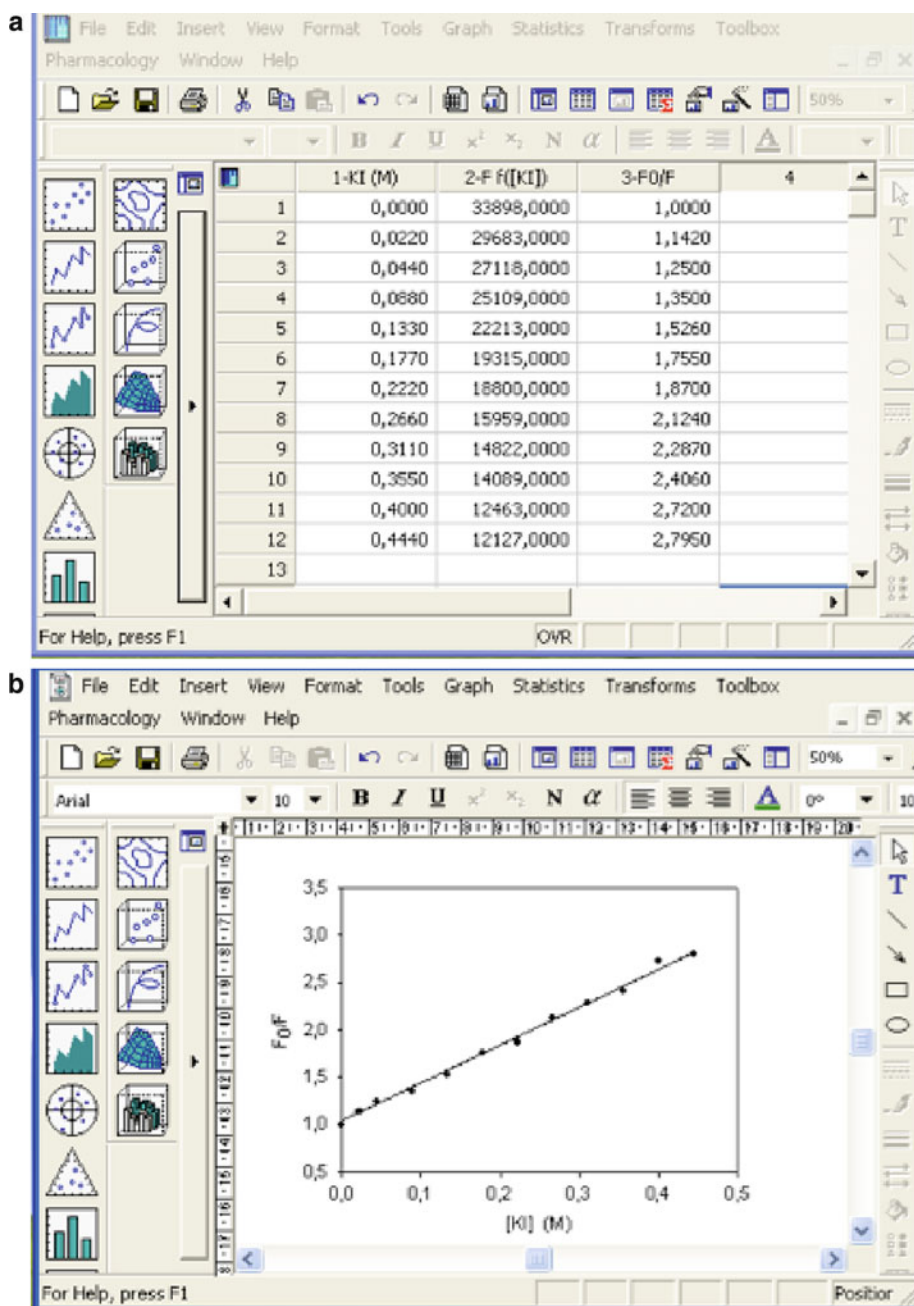
Fig. 1. Typical spreadsheet report of the experimental data and their display as Stern–Volmer plot. (**a**) Fluorescence intensity counts (column 2) recorded as a function of the quencher concentration (column 1). Their transformation as F_0/F ratios is listed in column 3. (**b**) The data presented in panel (**a**) are used to generate a graph according to the Stern–Volmer equation. The straight line is obtained by regression analysis of these data (*solid line*) to recover the slope representing the Stern–Volmer constant, K_{SV} (M^{-1}).
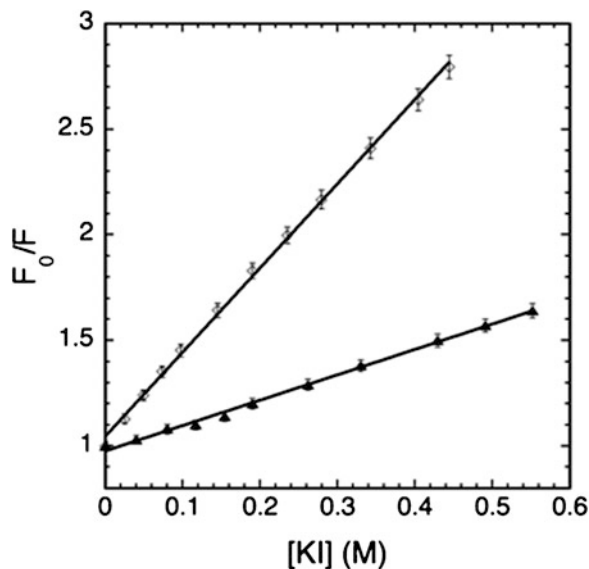
Fig. 2. Stern–Volmer plots of the intrinsic fluorescence of *Bacillus pasteurii* UreG (modified from ref. 4). Steady-state fluorescence intensity measurements were performed at 24 °C using an excitation wavelength of 295 nm and observing the emission at 335 nm and at 347 nm, with the protein in the absence (*filled triangle*) and presence (*open diamond*) of GuHCl 1.2 M, respectively.

4. Notes

1. The buffer should contain 1×10^{-4} M $Na_2S_2O_3$ to avoid I_3^- formation. It is a good practice to protect the glass vessel containing sodium thiosulfate with aluminum foil.

2. KI and GuHCl stock solutions are stored at 4 °C and are used for no more than 2–3 days. The NaCl stock solution is stored at 4 °C.

3. Excitation and emission bandpasses of 2.5 nm are set to obtain a good signal/noise ratio. The excitation wavelength of 295 nm is used to completely avoid the contribution of tyrosine and phenylalanine. The instrument is usually operates in photon-counting mode, instead of analogic mode, to observe very weak signals. In this way the operator can choose to accumulate counts for equal dwell times at each step of the monochromator, increasing the signal/noise ratio and improving the accuracy and precision of the measurement. Keep the absorbance of the sample at the excitation wavelength as low as possible (typically: $0.05 < OD_{295} < 0.20$) to record the correct fluorescence intensity and to avoid inner filter effects.

4. Keep the ionic strength constant using NaCl.

5. Reliable and less time-consuming data analysis can be attained using the counts at the maximum intensity fluorescence peak instead of performing full spectral integration.

6. A good linear distribution of the experimental data and the recovery of an intercept close to 1 from the fitting procedure are regarded as excellent markers of collisional quenching.

References

1. Jablonski A (1935) Uber den mechanismus der photolumineszenz von farbstoffphosphoren. Zeitschrift fur Physik 94:38–46

2. Lakowicz JR (1999) Principle of fluorescence spectroscopy. Plenum Publishing, New York

3. Beechem JM, Brand L (1985) Time-resolved fluorescence of proteins. Annu Rev Biochem 54:43–71

4. Neyroz P, Zambelli B, Ciurli SL (2006) Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy. Biochemistry 45:8918–8930

5. Lehrer SS (1971) Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. Biochemistry 10:3254–3263

Binding Stoichiometry and Affinity of Fluorescent Dyes to Proteins in Different Structural States

Anna I. Sulatskaya, Olga I. Povarova, Irina M. Kuznetsova, Vladimir N. Uversky, and Konstantin K. Turoverov

Abstract

Protocol of determination of binding stoichiometry and affinity of fluorescent dyes with proteins in different structural states is proposed. The proposed approach is based on the spectrophotometric determination of concentrations of dye bound to protein and free dye in solutions prepared by equilibrium microdialysis. This technique allows also determining spectral properties of the bound dyes. The restrictions of the use of dye fluorescence intensity for characterization of its interaction with the target protein are discussed. It is shown that the dependence of the dye fluorescence intensity on its optical density together with the data on its binding parameter can give information about the dye fluorescence quantum yield. All procedures are illustrated by interaction of 8-anilino-1-naphthalenesulfonate (ANS) with bovine serum albumin.

Key words: Fluorescent dyes, ANS, Binding stoichiometry, Binding affinity, Bovine serum albumin, Equilibrium microdialysis, Fluorescence intensity, Fluorescence quantum yield

Abbreviations

ANS   8-Anilino-1-naphthalenesulfonate
BSA   Bovine serum albumin
ThT   Thioflavin T
QS   Quinine sulfate

1. Introduction

Today, fluorescence methods penetrate all fields of biology. They are used for visualization of different biochemical processes in cells and tissues, for analysis of the structures of biomacromolecules and their complexes in solution, for pathogenic changes diagnostics in

medicine, and for environment monitoring. For this purpose, both intrinsic fluorescence of objects under study (e.g., tryptophan fluorescence of proteins) as well as extrinsic fluorescence of dyes covalently bound to target objects, or probes such as 8-anilino-1-naphthalenesulfonate (ANS) and thioflavin T (ThT) which bind to proteins by weak noncovalent coupling are used. Fluorescent probe ANS is used for testing the presence of hydrophobic clusters and hydrophobic "pockets" in the structure of target objects (1, 2). ThT is used for testing the appearance, kinetics of formation, and structural studies of amyloid fibrils (3–8). The very low fluorescence quantum yield of these dyes in free state in aqueous solutions and its significant increase on binding to proteins or their aggregates makes these probes very attractive for investigators. The use of fluorescent probes is always accompanied by the problem of determination of the parameters of dye binding to target proteins and the bound dye properties. This work offers a protocol of sequential procedures for determination of these characteristics by example of ANS and bovine serum albumin (BSA) interaction. BSA was chosen as a well studied protein which has hydrophobic clusters on the surface in the native state and due to this binds various ligands including ANS (2).

About 50 years ago Stryer published the work (1), in which, for the first time, it was shown that ANS interaction with apomyoglobin leads to significant increase in its fluorescence intensity and it was suggested that this hydrophobic probe binds to hydrophobic clusters of proteins. Later this suggestion was proven many times. In particular, it was shown that this probe can efficiently interact with serum albumins, which are known to have a variety of the so-called hydrophobic pockets. At present, fluorescent probe ANS is widely used as a test on hydrophobic clusters accessible for solvent molecules (reviewed in ref. 9).

Interest to ANS dramatically increased when it became clear that this dye can also be used as a probe for the presence of folding intermediate state known as the "molten globule" state. For several globular proteins which had no hydrophobic clusters on their surfaces, it was shown that ANS did not interact with their native/folded states or with completely unfolded states that realized at the high denaturant concentrations. It also did not interact with coil-like, α-helical, or β-structural conformations of several hydrophilic homopolypeptides. However, this dye formed highly fluorescent complex with protein at the intermediate denaturant concentration, i.e., under conditions in which equilibrium molten globule-like partially folded intermediate can be found (10–14). In the kinetic studies of protein folding, changes in the ANS fluorescence intensity typically followed the protein folding phases, reaching maximal values at the folding times corresponding to the maximal population of transiently formed partially folded molten globule-like intermediates. Based on these data, it was proposed that ANS interacts preferentially with the equilibrium and kinetic

intermediates in the globular protein folding pathway (10–14). This means that ANS can be used as a tool for simple and highly visual examination of globular protein folding. This observation opened a new era in the ANS use in equilibrium and kinetic studies on proteins' structural transformations, and the term "high affinity to ANS" became one of the main characteristics of a protein molecule in a molten globule state (15).

However the increase in ANS fluorescence intensity in the unfolding–refolding process of a protein cannot be unambiguously interpreted as an appearance of a molten globule-like partially folded intermediate (16). In fact, the peculiarities of GdnHCl as a denaturant need to be taken into account, since at low concentrations, it can cause stabilization and even aggregation of proteins. This effect, for example, was demonstrated for actin and carbonic anhydrase and it was shown that this effect (i.e., protein aggregation) can determine the increase in the ANS fluorescence intensity detected at low GdnHCl concentrations. In particular, it was shown that the increase in the ANS fluorescence intensity was not observed when these proteins were unfolded by urea due to the lack of the aggregate formation. The analysis of literature also showed that the increase in the ANS fluorescence intensity was typically observed when there were grounds to suggest protein aggregation. In particular, the increase in the ANS fluorescence intensity was accompanied by the increase in the polarization of intrinsic fluorescence of protein (13, 17, 18). Based on these data, it was concluded that hydrophobic dye ANS binds to the aggregates of proteins in the molten globule state rather than with the hydrophobic clusters on the surface of a protein in the molten globule state, as was commonly accepted (10, 15). Anyhow, this must be taken into account by the researchers who use ANS fluorescence intensity in their studies of protein structure.

Due to its characteristics ANS can be a useful tool for examining intrinsically disordered (ID) proteins. The degree of structural disorder in ID proteins can differ. In some cases, only a small portion of the polypeptide chain is unstructured; in the other cases, contrarily, only a small part of a protein polypeptide chain is structured. Quite often, the structure of ID proteins resembles the molten globule. The existence of unstructured regions might determine the ability of the ID proteins to interact with a great variety of partners and also can lead to aggregation. Therefore, ANS can be used to study the structure of the ID proteins, their complexes with partners, and their amorphous aggregates. Information on the spectral properties of ANS can be found in several studies (19–26). A few illustrative examples where ANS was utilized for the characterization of ID proteins are represented below.

UreG is an essential protein for the in vivo activation of urease. The exposure of protein hydrophobic sites, monitored using the fluorescent probe bis-ANS, indicated that the native dimeric state of BpUreG is disordered even though it maintains a significant amount of tertiary structure (27). ANS fluorescence also indicated that,

upon addition of a small amount of GuHCl, a transition to a molten globule state occurs, followed by formation of a pre-molten globule state at a higher denaturant concentration. The hydrodynamic parameters obtained by time-resolved fluorescence anisotropy at maximal denaturant concentrations (3 M GdmHCl) confirmed the existence of a disordered but stable dimeric protein core.

Spectrofluorimetric urea titration experiments with the use of ANS fluorescence could distinguish between the binding of ANS to the hydrophobic pocket of an ordered proteins and the binding of ANS to a proteins in molten globule state, including ID proteins (see below). This approach was utilized in the analysis of clusterin (28) and the N-terminal transactivation domain of the human androgen receptor (29).

Clusterin, also known as sulfated glycoprotein-2, TRPM-2, GP-80, SP 40,40, and ApoJ, has been found in many tissues including prostate, brain, kidney, liver, and plasma in many species including rat, human, ram, and bovine (28). Clusterin was proposed to have a wide range of biological functions including cell–cell interactions, sperm maturation, complement inhibition, and lipid transport. Furthermore, it was shown to have chaperone-like activity preventing the precipitation of denatured proteins in vitro (30). Finally, clusterin is able to interact specifically with a wide range of biological ligands including proteins such as complement components, peptides such as amyloid β_{1-40}, and lipids such as those found in high-density lipoproteins (31, 32). Clusterin is associated with cellular injury, lipid transport, and apoptosis, and it may be involved in the clearance of cellular debris caused by cell injury or death. A model was proposed where clusterin acted as a "biological detergent," binding to hydrophobic complexes and denatured proteins to aid in their clearance from ducts or lumen during tissue remodeling (33). To do so, clusterin was hypothesized to have a flexible or dynamic binding site or sites allowing numerous associations to take place. This hypothesis was tested by several methods, including prediction of disorder from amino acid sequence, limited protease digestion, far-UV CD, and ANS fluorescence (28). The comparison of the response of the clusterin–ANS complex to the increasing urea concentrations with those of proteins with structured binding pockets and molten globular forms of proteins revealed that clusterin likely contains a molten globule-like domain in its native state (28).

The androgen receptor is a ligand-activated transcription factor that mediates the actions of the steroid hormones testosterone and dihydrotestosterone. Its N-terminal domain (NTD) is intrinsically disordered and structurally flexible and participates in multiple protein–protein interactions (34). Using a set of computational and experimental approaches it has been shown that this domain, being intrinsically disordered, exists in a collapsed disordered conformation, distinct from extended disordered (random coil) and a stable globular fold (35). Particularly, the interaction of purified domain

and the hydrophobic fluorescence probe ANS was investigated. Incubation of the transactivation domain with ANS resulted in a significant increase in fluorescence intensity and a blue shift of the maximum emission to 465 nm (29). Furthermore, the urea-induced unfolding of this domain was compared with that of the structured protein BSA and α-lactalbumin in the molten globule conformation, since BSA was shown to unfold cooperatively, whereas the unfolding of the molten globular α-lactalbumin was noncooperative. The transactivation domain of the androgen receptor was similarly sensitive to urea and unfolded in a noncooperative manner. Based on these data it has been concluded that the domain in a native state is able to bind ANS and exhibited ANS binding characteristics similar to those of a well-characterized molten globule-state protein (29).

An ability to gain some ordered structure at extreme pH values is a characteristic property of extended ID proteins (see Chapter 20 for more details). pH-induced folding of a typical ID protein, α-synuclein, was analyzed using a variety of biophysical techniques, including changes in the ANS fluorescence (18). For example, a decrease in pH led to a noticeable increase in the ANS fluorescence intensity and detectable blue shift of the ANS fluorescence maximum (from ~515 to ~475 nm), reflecting the pH-induced transformation of α-synuclein from the natively unfolded state to the partially folded partially compact conformation. The transition from the natively unfolded to a partially folded conformation took place between pH 5.5 and 3.0, and was completely reversible. Furthermore, the pH-induced structural transitions observed by ANS fluorescence and far-UV CD ellipticity occurred simultaneously in a rather cooperative manner. This means that the α-synuclein protonation resulted in the transformation of this natively unfolded protein into a conformation with a significant amount of ordered secondary structure and with increased affinity to ANS. The position of the transition indicated that the protonation of one or more carboxylates was responsible for the detected structural changes (18).

The ability of ANS to interact with folded proteins possessing solvent-accessible hydrophobic regions and with highly flexible partially folded conformations brought the important question of how to discriminate these dye–protein two complexes. Three approaches developed to answer this important question are described below.

(a) *ANS fluorescence lifetimes.* The ANS interaction with proteins (both folded and partially folded) is accompanied by the characteristic changes in the fluorescence lifetime of the probe (35). The fluorescence decay of free ANS in aqueous or organic solvents is well described by the monoexponential law, whereas formation of complexes of this probe with proteins results in a

more complicated dependence (35). Analysis of the ANS fluorescence lifetimes in a number of proteins revealed that at least two types of ANS–protein interactions might exist. At interaction of the first type (characterized by fluorescence lifetime of about 1–5 ns) the probe molecules are bound to the surface hydrophobic clusters of the protein molecule and are in a relatively good contact with the solvent. At interaction of the second type (characterized by fluorescence lifetime of about 10–17 ns) the probe molecules are embedded into the protein molecule and are poorly accessible to the solvent and external quencher (35).

The changes in the long lifetime component correlate well with the overall conformational changes of the protein molecule observed upon its denaturation and unfolding (20, 35). In fact, the values of the longest lifetime component of fluorescence decay were measured for complexes of ANS with six different proteins: three of them, β-lactamase, lysozyme, and β-lactoglobulin, can bind this dye in the native state, whereas the others, bovine carbonic anhydrase, and human and bovine α-lactalbumins, acquire large affinity to ANS after the transformation to the molten globule state. It has been shown that the ANS interaction with the native proteins is characterized by the shorter fluorescence decay time compared to the ANS–molten globule complexes ($\tau_N \sim 10$–11 ns and $\tau_{MG} \sim 15$–17 ns, respectively) (20, 35). This sensitivity of ANS molecules to be inserted either into the folded protein or into the molten globule ($\tau_N \leq 12$ ns as compared with $\tau_{MG} \geq 15$ ns) was explained by taking into account the capability of the dye to self-associate (20). The penetration of the dye into the rigid hydrophobic pocket(s) of native proteins is not necessarily accompanied by its dissociation due to the considerable steric limitations, while the liquid-like core of the molten globules cannot prevent such dissociation. As a result, ANS molecules, being embedded into the native proteins, exist as self-oligomers and exhibit fluorescence with a shorter lifetime. Embedding into the molten globule proteins leads to the dissociation of ANS oligomers and, as a consequence, to the increase in characteristic times of fluorescence decay (20). Therefore, the interaction of ANS with both molten globules and pre-molten globules of different proteins results in fluorescence lifetimes characteristic of the second type, with molten globules reacting more strongly than pre-molten globules. In other words, there are some "magic numbers"—the values of ANS fluorescence lifetime $\tau_N \leq 12$ ns and $\tau_{MG} \geq 15$ ns—that show the conformational state (the folded or the molten globule) of a protein molecule one deals with (20).

(b) *Urea titration of ANS fluorescence*. This method is based on the important observation that ordered proteins unfold cooperatively, whereas unfolding of molten globular forms typically is

much less cooperative (19). It has been found that urea titration of ANS fluorescence therefore could be used to distinguish between the ANS binding to the hydrophobic pocket of an ordered protein and the binding of ANS to a molten globular form. In fact, this analysis revealed that the urea-induced unfolding of such ordered proteins as BSA, apomyoglobin, and hexokinase is characterized by typical sigmoidal curves, whereas unfolding of molten globular forms of apomyoglobin and α-lactalbumin is much less cooperative.

(c) *Comparison of the Stern–Volmer quenching by acrylamide and trichloroethanol (TCE)*. Folded and molten globular ANS binders might be potentially discriminated by comparing the Stern–Volmer quenching curves for a polar quencher, acrylamide, with the quenching curves for a nonpolar quencher, TCE. The essence of this method is based on the following. If the hydrophobic groups surrounding a fluorophore (e.g., ANS) are rigidly packed, then both acrylamide and TCE are excluded and so both show little quenching. On the other hand, if the hydrophobic groups surrounding the fluorophore are loosely packed and dynamic, then the hydrophilic quencher, acrylamide, is still excluded and so continues to show little quenching. However, the hydrophobic quencher, TCE, actually partitions into the hydrophobic region surrounding the fluorophore. This leads to quenching that is much stronger than if the fluorophore were completely exposed on the protein surface. These concepts were proven in a model system in which a tryptamine fluorophore was complexed with SDS micelles (36). This approach was applied for the characterization of three different forms of fd phage (37). The fluorescence emission maxima and intensities for the tryptophans in all three forms are nearly identical. Furthermore, there is very little difference in the acrylamide quenching, suggesting that the indole rings are in tightly packed environments. On the other hand, for the two contracted forms (I- and S-forms), quenching by TCE is enormously stronger than the quenching by acrylamide. Moreover, the data show that the quenching by TCE is even stronger than the quenching for a naked indole ring in water. How can this be? The data for tryptamine in SDS show a similar behavior. For SDS, the original interpretation was that the internal, dynamic micelle could actually dissolve the TCE, so its local concentration around the fluorophore is higher than in the surrounding solution. Thus, these data suggest a similar interpretation for the contracted forms of fd phage: the residues surrounding the tryptophan are likely dynamic similar to the inside of an SDS micelle, thus leading to high local accumulation of TCE and very high quenching values (37).

The existence of unstructured regions in ID proteins increases the danger of the formation of ordered aggregates—amyloid fibrils, which can be detected by the benzothiazole dye ThT. The first works where ThT fluorescence was used as a test for amyloid fibrils were done 50 years ago. The extremely low fluorescence quantum yield of this dye in the solutions with low viscosity was explained by the freedom of rotation of the benzothiazole and aminobenzoyl rings that leads to appearance of non-radiative state with φ angle between benzothiazole and aminobenzoyl rings about to 90°. Due to this phenomenon, ThT is typically referred to as a member of the class of molecules known as molecular rotors. The restriction of intramolecular mobility of ThT molecule caused by its interaction with fibrils leads to the significant (by several orders of magnitude) increase in its fluorescence intensity (8, 38).

The main characteristics of the probe–acceptor interaction are binding affinity and stoichiometry. Since fluorescence of ThT and ANS increases dramatically on interaction with proteins, it seems natural to use fluorescence for characterization binding parameters. These studies were based on the assumption that the fluorescence intensity as a function of dye concentration reaches a plateau when all of the binding sites are occupied. We show that the concentration dependence for any fluorophore is the curve with saturation, and that fluorescence, in principle, cannot be used for determining of the binding parameters of dye interaction with acceptor.

The protocol for the determination of binding stoichiometry and affinity of fluorescence probe with proteins is based on absorption spectrophotometry of solutions prepared by equilibrium microdialysis. It is shown that fluorescence intensity can be used for determination of fluorescence quantum yield (or quantum yields in the case of several binding modes) if binding parameters and molar extinction coefficient (or extinction coefficients) are already determined. The proposed procedure is illustrated by the results obtained for interaction of ANS with BSA. However, this method is universal. It has already been used for determination of ThT – amyloid fibrils binding parameters and spectral properties and fluorescence quantum yield of this dye in the bound state (39–41).

2. Materials

The samples of ANS from Serva (Germany), quinine sulfate from AnaSpec (USA), and BSA from Koch-Light Lab Ltd (England) were used without additional purification. Absorption spectra were recorded by spectrophotometer U-3900H (Hitachi, Japan). Fluorescence measurements were recorded by homemade spectrofluorimeter (42) and spectrofluorometer Cary Eclipse (Varian,

Australia). Equilibrium microdialysis was done with Harvard Apparatus/Amika (USA) device. It consists of two chambers (each of 500 μL) separated by membrane (MWCO 10,000) impermeable for particles larger than 10,000 Da.

3. Methods

3.1. Determination of the Binding Stoichiometry and Affinity of the Dye–Protein Interaction

1. Prepare solutions for microdialysis. Prepare the buffer solutions with the required pH values [Note 1]. Prepare the solution of protein with concentration C_p and solutions of ANS with optical density in the range from 0.2 to 3 (5–7 concentrations). Perform procedures described in steps 2–7 for each solution of ANS.

2. Determine the optical density of the stock solution of the dye, $D_0(\lambda)$.

3. Determine the dye concentration in the stock solution:

$$C_0 = \frac{D_0(\lambda_{max})}{\varepsilon_f(\lambda_{max}) \times l}, \tag{1}$$

where $\varepsilon_f(\lambda_{max})$ is molar extinction coefficient of free dye in solution at the wavelength corresponding to the maximum of absorption spectrum.

4. Put the dye solution of concentration C_0 in the chamber #1, and the protein solution of concentration C_p in the chamber #2 [Note 2]. Perform the dialysis at constant temperature (23 °C) and agitation overnight. To meet these requirements, put the microdialysis chambers on shaker in thermostat. If it is necessary to determine thermodynamic characteristics of dye binding to protein, the experiments are performed at several temperatures.

5. Record the absorption spectrum of ANS in chamber #1 after equilibration:

$$D(\lambda)_{\#1} = D_f(\lambda). \tag{2}$$

6. Determine the dye concentration in the chamber #1 (C_f), which, under the equilibrium conditions, is equal to the free dye concentration value in the chamber #2:

$$C_f = \frac{D_f(\lambda_{max})}{\varepsilon_f(\lambda_{max}) \times l}. \tag{3}$$

7. Determine the concentration of the dye bound to protein:

$$C_b = C_0 - 2C_f. \tag{4}$$

8. Plot the dependence $C_b = f(C_f)$ or $C_b = f(C_0)$. On the basis of the test experiment one has to decide on the principal possibility

of spectrophotometric determination of the values C_b и C_f: optical densities of these solutions (D_b and D_f) must correspond to the working range of the spectrophotometer. In some cases, the range of the C_b and C_f values can be expanded by recording optical density in the cells with longer optical pathway.

For reliable determination of the number of binding sites and binding constants, the procedures described in steps 1 through 8 have to be repeated for the large number of the D_0 (C_0) values. The strategy of choosing of the C_0 and C_p values for the following series of solutions is mainly determined on the basis of the result of the first series of experiment.

9. Determine the binding constants, the number of binding modes, and the number of binding sites in the dye–protein interaction. If all dye binding sites in protein are identical and independent from each other, then the binding constant of the dye to protein (K_b) is determined as the ratio of the ligand–receptor complex concentration (C_b) to the product of free receptor ($nC_p - C_b$) and free ligand (C_f) concentrations:

$$K_b = \frac{C_b}{(nC_p - C_b) \times C_f}. \tag{5}$$

Here n is the number of binding sites.

Using Eqs. 4 and 5 it is easy to obtain the relation between the bound dye concentration and the initial dye concentration:

$$C_b = \frac{2 + K_b nC_p + K_b C_0 - \sqrt{(2 + K_b nC_p + K_b C_0)^2 - 4K_b^2 nC_p C_0}}{2K_b}. \tag{6}$$

More frequently, binding parameters n and K_b (or dissociation constant $K_d = 1/K_b$) are determined from three equivalent-to-each-other equations:

$$C_b = \frac{nC_p C_f}{K_d + C_f}. \tag{7}$$

Klots plot:

$$\frac{1}{C_b} = \frac{1}{nC_p} + \frac{K_d}{nC_p}\frac{1}{C_f}, \tag{8}$$

or Scatchard plot:

$$\frac{C_b/C_p}{C_f} = nK_b - K_b\left(\frac{C_b}{C_p}\right). \tag{9}$$

Klots plots and Scatchard plots give visual presentation of the number of dye–protein binding modes in assumption that all
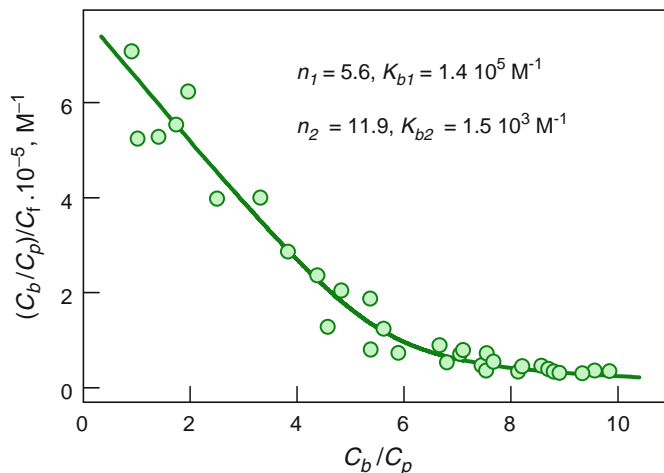
Fig. 1. Scatchard plots for ANS interaction with BSA. Experimental data (*circles*), best fit curve and the volumes of binding constants (K_{bi}) and number of binding sites (n_i) are presented.

binding centers are independent. Linearity of these dependencies suggests identity of all binding sites, and therefore supports the existence of one binding mode. Nonlinear character of these plots points on the existence of at least two different binding modes of dye–protein interaction. In this case, $C_b = \sum_i C_{bi}$, while C_{bi} is characterized by equations similar to those described in Eqs. 6–8. Thus, in the case of i modes, we have

$$C_b = \sum_i \frac{n_i C_p C_f}{K_{di} + C_f}. \tag{10}$$

The values K_{d1}, K_{d2}, n_1, and n_2 can be determined by multiple nonlinear regression (GraphPad), or by global analysis if alongside with C_0 the value of C_p was varied.

Figure 1 shows the results obtained for the ANS interaction with BSA, which is known to have hydrophobic clusters on the surface in the native state and therefore is known to bind various ligands including ANS. In this case, if the values of binding constants of modes 1 and 2 significantly differ, the dye interaction with low constant binding mode may have nonspecific character (43).

3.2. Determination of Molar Extinction Coefficient of Dye Bound to Protein

1. The case of one binding mode: In this case, the measured absorption spectrum can easily be presented in the units of the molar extinction coefficient:

$$\varepsilon_b(\lambda) = \frac{D_b(\lambda)}{C_b \times l}. \tag{11}$$
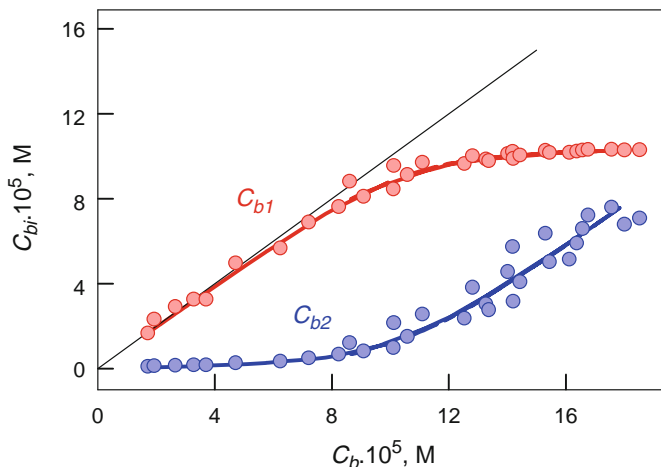
Fig. 2. Concentration of ANS bound to BSA as superposition of concentrations of the dye bound to mode 1 (C_{b1}) and mode 2 (C_{b2}).

2. The case of two binding mode: If the microdialysis results show the existence of two binding modes, then on the basis of the K_{d1}, K_{d2}, n_1, and n_2 values, the concentrations of dye bound to each mode can be calculated as

$$C_{b1} = \frac{n_1 C_p C_f}{K_{d1} + C_f} \qquad (12)$$

and

$$C_{b2} = \frac{n_2 C_p C_f}{K_{d2} + C_f}.$$

Figure 2 shows the decomposition of C_b into two components, C_{b1} and C_{b2}. Taking into account that

$$D_b(\lambda) = D_{b1} + D_{b2} = \varepsilon_{b1}(\lambda) C_{b1} l + \varepsilon_{b2}(\lambda) C_{b2} l, \qquad (13)$$

the values of $\varepsilon_{b1}(\lambda)$ and $\varepsilon_{b2}(\lambda)$ can be determined using the known values of $D_b(\lambda)$, C_{b1}, and C_{b2} by multiple linear regression (e.g., using SigmaPlot). Figure 3 shows the relation between D_b at 365 nm and C_{b1} and C_{b2} values that result in values of $\varepsilon_{b1}(365)$ and $\varepsilon_{b2}(365)$. Similarly, the values of ε_{b1} and ε_{b2} can be determined at the other wavelengths. Figure 4 shows the absorption spectra of the dye bound to each of the two modes in units of the molar extinction coefficient. These data illustrate that the molar extinction coefficient of ANS bound to BSA depends on the binding mode being greater than that of the free dye in solution (Table 1).
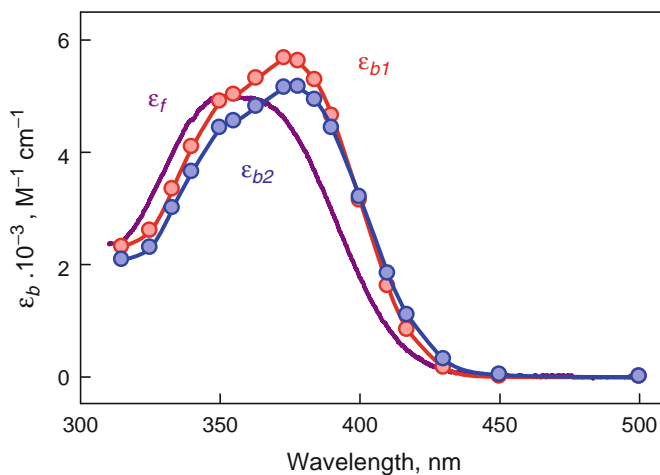
Fig. 3. Absorption spectra of ANS, bound to mode 1 and mode 2 in the units of the molar extinction coefficient.
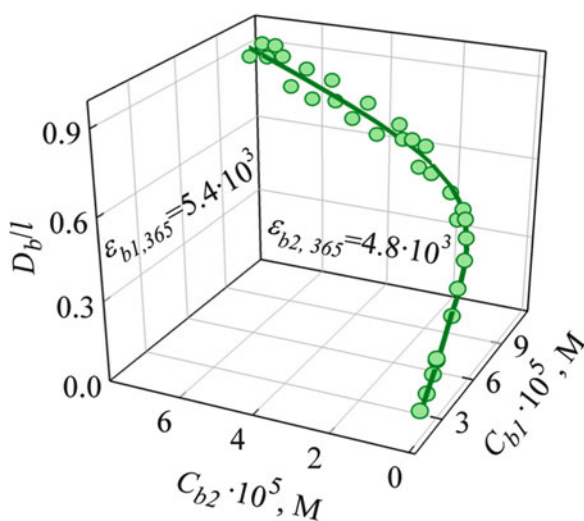
Fig. 4. The dependence $D_b = D_{b1} + D_{b2} = \varepsilon_{b1} C_{b1} l + \varepsilon_{b2} C_{b2} l$ for ANS bound to BSA. Best fit curve and the values of molar extinction coefficients ε_{b1} and ε_{b2} obtained by multiple nonlinear regression are presented.

Table 1

Binding parameters of ANS–BSA interaction and characteristics of bound dye

	Mode	$\varepsilon_{i,365} \times 10^{-3}$, M^{-1} cm^{-1}	$K_{bi} \times 10^{-3}$, M^{-1}	n_i	q
ANS + BSA	1	5.4	140	5.6	0.45
	2	4.8	1.5	11.9	0.08

3.3. Dependence of the Fluorescence Intensity on Optical Density of Fluorescence Substance and Total Optical Density of Solution

The solution of ANS in the presence of BSA is a two-component system, in which one component (free ANS unbound to protein) absorbs the excitation light (optical density, D_f) but does not fluoresce, while the other component (ANS bound to protein) absorbs the excitation light (optical density, D_b) and fluoresces with quantum yield q_b. Therefore, ANS fluorescence intensity can be presented as follows:

$$I_{ANS} = kI_0 \left(1 - 10^{-(D_b+D_f)}\right) \frac{D_b}{D_b + D_f} \times q_b. \qquad (14)$$

Here, I_0 is the intensity of the excitation light and k is the proportionality coefficient. Fluorescence intensity can be normalized $(kI_0 = 1)$ so that $I_{ANS} = [D_b/(D_b + D_f)] \times q_b$ at $(D_b + D_f) \to \infty$ and consequently, $I_{ANS} = q_b$ at $D_b \to \infty$ and $I_{ANS} = 0$ at $D_f \to \infty$. In Eq. 14, two factors can be selected. One factor is the function of total optical density of solution only and does not depend on the contribution of the optical density of the fluorescent component:

$$W_{calc,D_b+D_f} = \frac{1 - 10^{-(D_b+D_f)}}{D_b + D_f}. \qquad (15)$$

The other factor is a product of optical density and the quantum yield of the fluorescent component $(D_b q_b)$. Then the Eq. 14 can be written as follows:

$$I_{ANS} = W_{calc,D_b+D_f} D_b q_b. \qquad (16)$$

Similarly, the intensity of a single-component solution can also be presented. In particular, for fluorescent dye quinine sulfate (QS), which in this work was used as a standard substance with known quantum yield, fluorescence intensity is

$$I_{QS} = \frac{(1 - 10^{-D_{QS}})}{D_{QS}} D_{QS} \times q_{QS} = W_{calc,QS} D_{QS} q_{QS}. \qquad (17)$$

It is evident that the dependence $I_{QS} = f(D_{QS})$ is a curve with saturation $(I_{QS} = q_{QS}$ at $D_{QS} \to \infty)$, while the dependence $I_{QS}/W_{calc,QS} = f(D_{QS})$ is a straight line with a slope equal to the fluorescent quantum yield of the dye (Fig. 5).

The experimentally measured fluorescence intensity can decrease with the increase in the solution optical density, since the increase in the total optical density of the solution results in the increased absorption of the excitation light by the solution layers adjacent to the front wall of the spectrofluorometer cell, while the detecting system of spectrofluorometer "sees" the central part of the cell, which is reached by the smaller fraction of the excitation light. Due to this effect, after the total optical density reaches some definite value, the recorded fluorescence intensity begins to decrease with the further increase in the solution optical density (Fig. 5). This effect detected both for the dye solution in
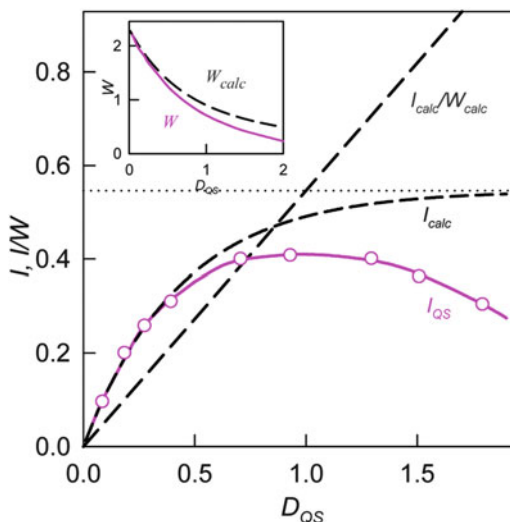
Fig. 5. The dependences of fluorescence intensity QS on its optical density (D_{QS}) calculated as $I_{calc} = [(1 - 10^{-D_{QS}})/D_{QS}]D_{QS} \times q_{QS}$ and experimentally recorded I_{QS}. The dependences $W_{calc} = (1 - 10^{-D_{QS}})/D_{QS}$ and $W = I_{QS}/D_{QS}q_{QS}$ on total optical density are given in the Insert. Strait line is the dependence of $D_{QS}q_{QS}$ on D_{QS} calculated as $D_{QS}q_{QS} = I_{calc}/W_{calc} = I_{QS}/W$.

the presence of protein and for the free dye solution is often interpreted even by experienced researchers as dye association and consequently fluorescence self-quenching associated with the increase in the dye concentration (44).

The discussed effects depend on the instrument used in the experiment and must be taken into account. This can be done by replacing W_{calc} with experimentally determined value W. As the value W_{calc}, the value of W is determined only by the total optical density of solution and does not depend on the contribution of the optical density of the fluorescent substance. That is why in Eqs. 16 and 17 the value of W_{calc} must be replaced with W[Note 3]. All this is also true in the case of the two-component solution. The fluorescence intensity corrected by factor W dependent on the total optical density of solution is the product of the optical density and fluorescence quantum yield of the bound dye: $I_{ANS}/W = q_b D_b = q_b \varepsilon_b C_b l$. In this case Eq. 6 will be as follows [Note 4]:

$$\frac{I_{ANS}}{W} = q_b \varepsilon_b \frac{2 + K_b n C_p + K_b C_0 - \sqrt{(2 + K_b n C_p + K_b C_0)^2 - 4 K_b^2 n C_p C_0}}{2 K_b} l. \tag{18}$$

Thus, experimental dependence of fluorescence intensity on the total concentration of the dye (C_0) gives information on the binding constant only if some special assumptions are made about the values of q_b and ε_b, which, as we showed, can differ from these values measured for the free dye.
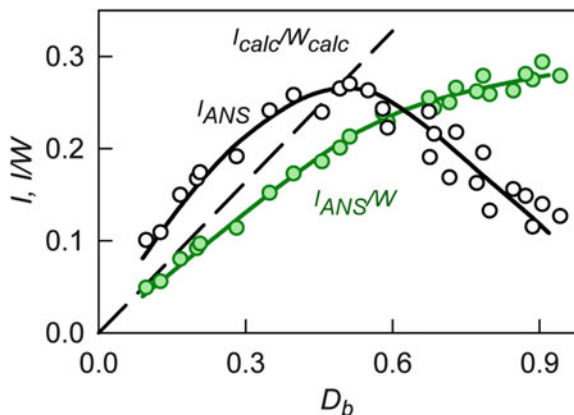
Fig. 6. The dependencies of experimentally recorded fluorescence intensity (*I*) and reduced fluorescence intensity (*I/W*) of ANS bound to BSA on the dye optical density (D_b). *Dashed line* is the dependence of $D_{QS}q_{QS}$ on D_{QS} for fluorescence dye QS.

Figure 6 shows the dependencies of I and I/W of ANS bound to BSA on the dye optical density (D_b). It should be noted that in the case of two binding modes the dependence of I/W on D_b is not linear. In the case of two binding modes, the use of fluorescence intensity for the determination of the binding stoichiometry and the binding constants is problematic, since in this case $I/W = \sum\limits_i q_{bi}D_{bi} = \sum\limits_i q_{bi}\varepsilon_{bi}C_{bi}l$. But these dependencies allow determination of the fluorescence quantum yield of ANS bound to different binding modes.

3.4. Determination of the Fluorescence Quantum Yield of Dye Bound to Protein

On the basis of Eqs. 16 and 17, in which W_{calc} is replaced by W, one can write

$$D_b q_b = \frac{I_{\text{ANS}}}{I_{\text{AQS}}} \frac{W_{D_{QS}}}{W_{D_b+D_f}} D_{QS} q_{QS}. \qquad (19)$$

To exclude the influence of the factor dependence on the total optical density it is necessary to record fluorescence intensity of etalon solution, whose optical density is equal to the summarized optical densities of free and bound ANS. In this case

$$D_b q_b = \frac{I_{\text{ANS}}}{I_{QS,D_b+D_f}} (D_b + D_f) \times q_{QS}. \qquad (20)$$

The q_b value is determined as an average of several values independently measured for samples with different dye concentrations after microdialysis.
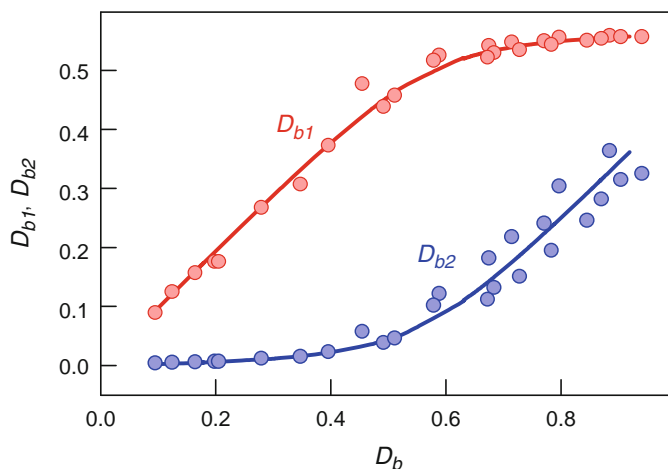
Fig. 7. Optical density of ANS bound to BSA as superposition of optical densities of the dye bound to mode 1 (D_{b1}) and mode 2 (D_{b2}).

In the case of two binding modes, a relation similar to Eq. 20 is valid:

$$D_{b1}\, q_{b1} + D_{b2}\, q_{b2} = \frac{I_{ANS}}{I_{QS,D_b+D_f}}(D_b + D_f) \times q_{QS}. \tag{21}$$

To determine the values of q_{b1} and q_{b2}, it is necessary to know the set of three related values D_{b1}, D_{b2}, and $D_{b1}\, q_{b1} + D_{b2}\, q_{b2}$ that correspond to one microdialysis experiment.

The dependencies of the $D_{b1} = \varepsilon_{b1} C_{b1} l$ and $D_{b2} = \varepsilon_{b2} C_{b2} l$ values on D_b can be determined using the values of the molar extinction coefficients and concentrations of ANS molecules bound to two modes and determined for the samples obtained after microdialysis (Fig. 7). The dependence of the value $(D_{b1}\, q_{b1} + D_{b2}\, q_{b2})$ on D_b can be determined from Eq. 21. On the basis of these data, the fluorescence quantum yields of the dye bound to each mode can be determined using multiple nonlinear regression (Fig. 8, Table 1).

4. Notes

1. Before starting the experiment it is necessary to perform examination of the dye binding to membrane (MWCO 10,000) and chamber material.

    Solution with the dye (ANS) in concentration C_0 is put in chamber #1 and solvent is put in chamber #2. The time needed for equilibration depends on the type of membrane between the chambers. For membrane MWCO 10,000 the equilibrium is reached overnight.
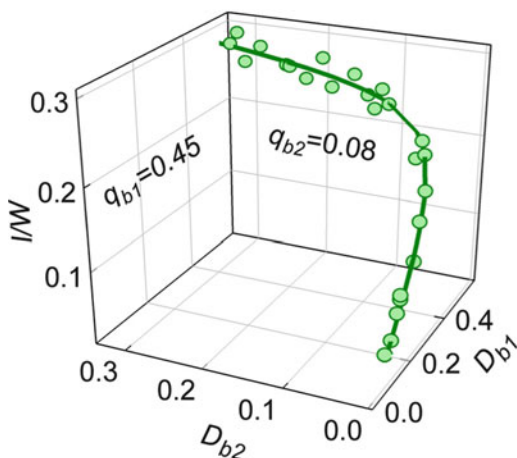
Fig. 8. Determination of the fluorescence quantum yield of ANS bound BSA. 3D dependence of $D_{b1}q_{b1} + D_{b2}q_{b2}$ from D_{b1} and D_{b2}. Experimental data, best fit curve and the values of q_{b1} and q_{b2} determined by multiple nonlinear regression are presented.

After equilibration it is checked whether:

$$D(\lambda)_{\#1} = D(\lambda)_{\#2}. \tag{22}$$

$$D(\lambda)_{\#1} = \frac{D_0(\lambda)}{2}. \tag{23}$$

If Eq. 23 is not true then the value of C_b determined by Eq. 4 must be corrected.

2. When incorporation of the dye to the receptor which solution has noticeable light scattering is studied this light scattering must be taken into account. In this case to determine true value of optical density it is necessary to use a well-known procedure (7).

3. It is necessary to take into account that the dependence of W on the total optical density of solution is determined by the type of the used spectrofluorimeter. The difference between W and W_{cal} is minimal for the spectrofluorimeter Cary Eclipse (Varian, Australia) which has horizontally orientated slits.

4. The Eq. 18 is true if the solutions for determination of the fluorescence intensity are prepared by microdialysis. If the dye is inserted directly in the solution containing protein ($C_b = C_0 - C_f$) this equation must be as follows:

$$\frac{I}{W} = q_b \varepsilon_b \frac{1 + K_b n C_p + K_b C_0 - \sqrt{(1 + K_b n C_p + K_b C_0)^2 - 4K_b^2 n C_p C_0}}{2 K_b} l. \tag{24}$$

Acknowledgement

References

1. Stryer L (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. J Mol Biol 13:482–495

2. Peters T Jr (1996) All about albumin: biochemistry, genetics, and medical application. Academic Press, New York

3. LeVine H III (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. Protein Sci 2:404–410

4. Levine H III (1995) Soluble multimeric Alzheimer beta(1–40) pre-amyloid complexes in dilute solution. Neurobiol Aging 16:755–764

5. Groenning M (2009) Binding mode of thioflavin T and other molecular probes in the context of amyloid fibrils-current status. J Chem Biol 3:1–18

6. Voropay ES, Samtsov MP, Kaplevsky KN, Maskevich AA, Stepuro VI, Povarova OI, Kuznetsova IM, Turoverov KK, Fink AL, Uversky VN (2003) Spectral properties of thioflavin T and its complexes with amyloid fibrils. J Appl Spectrosc 70:868–874

7. Maskevich AA, Stsiapura VI, Kuzmitsky VA, Kuznetsova IM, Povarova OI, Uversky VN, Turoverov KK (2007) Spectral properties of thioflavin T in solvents with different dielectric properties and in a fibril-incorporated form. J Proteome Res 6:1392–1401

8. Sulatskaya AI, Kuznetsova IM, Maskevich AA, Uversky VN, Turoverov KK (2010) Non-radiative deactivation of the excited state of thioflavin T: dependence on solvent viscosity and temperature. PLoS One 5:e15385

9. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK (2004) Natively disordered proteins, vol I. Wiley-VCH Verlag GmbH & Co, Weinheim, pp 271–353

10. Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers 31:119–128

11. Semisotnov GV, Rodionova NA, Kutyshenko VP, Ebert B, Blanck J, Ptitsyn OB (1987) Sequential mechanism of refolding of carbonic anhydrase B. FEBS Lett 224:9–13

12. Goto Y, Fink AL (1989) Conformational states of beta-lactamase: molten-globule states at acidic and alkaline pH with high salt. Biochemistry 28:945–952

13. Rodionova NA, Semisotnov GV, Kutyshenko VP, Uverskii VN, Bolotina IA (1989) Staged equilibrium of carbonic anhydrase unfolding in strong denaturants. Mol Biol (Mosk) 23:683–692

14. Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI (1990) Evidence for a molten globule state as a general intermediate in protein folding. FEBS Lett 262:20–24

15. Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem 47:83–229

16. Povarova OI, Kuznetsova IM, Turoverov KK (2010) Differences in the pathways of proteins unfolding induced by urea and guanidine hydrochloride: molten globule state and aggregates. PLoS One 5:e15035

17. Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov S, Ptitsyn OB (1981) Alpha-lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett 136:311–315

18. Uversky VN, Li, J, Fink AL (2001) Evidence for a partially-folded intermediate in α-synuclein fibrillation. J Biol Chem 276(14):10737–10744

19. Uversky VN, Ptitsyn OB (1996) All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. Fold Des 1:117–122

20. Uversky VN, Winter S, Lober G (1998) Self-association of 8-anilino-1-naphthalene-sulfonate molecules: spectroscopic characterization and application to the investigation of protein folding. Biochim Biophys Acta 1388:133–142

21. Klimtchuk E, Venyaminov S, Kurian E, Wessels W, Kirk W, Prendergast FG (2007) Photophysics of ANS. I. Protein–ANS complexes: intestinal fatty acid binding protein and single-trp mutants. Biophys Chem 125:1–12

22. Kirk W (2007) Photophysics of ANS. II: charge transfer character of near-UV

absorption and consequences for ANS spectroscopy. Biophys Chem 125:13–23

23. Kirk W, Klimtchuk E (2007) Photophysics of ANS. III: circular dichroism of ANS and anilinonaphthalene in I-FABP. Biophys Chem 125:24–31

24. Kirk W, Wessels W (2007) Photophysics of ANS. IV. Electron transfer quenching of ANS in alcoholic solvents and mixtures. Biophys Chem 125:32–49

25. Kirk W, Kurian E, Wessels W (2007) Photophysics of ANS. V. Decay modes of ANS in proteins: the IFABP–ANS complex. Biophys Chem 125:50–58

26. Hawe A, Sutter M, Jiskoot W (2008) Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res 25:1487–1499

27. Neyroz P, Zambelli B, Ciurli S (2006) Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy. Biochemistry 45:8918–8930

28. Bailey RW, Dunker AK, Brown CJ, Garner EC, Griswold MD (2001) Clusterin, a binding protein with a molten globule-like region. Biochemistry 40:11828–11840

29. Lavery DN, McEwan IJ (2008) Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry 47:3360–3369

30. Humphreys DT, Carver JA, Easterbrook-Smith SB, Wilson MR (1999) Clusterin has chaperone-like activity similar to that of small heat shock proteins. J Biol Chem 274:6875–6881

31. Wilson MR, Easterbrook-Smith SB (2000) Clusterin is a secreted mammalian chaperone. Trends Biochem Sci 25:95–98

32. Calero M, Tokuda T, Rostagno A, Kumar A, Zlokovic B, Frangione B, Ghiso J (1999) Functional and structural properties of lipid-associated apolipoprotein J (clusterin). Biochem J 344(Pt 2):375–383

33. Bailey R, Griswold MD (1999) Clusterin in the male reproductive system: localization and possible function. Mol Cell Endocrinol 151:17–23

34. Lavery DN, McEwan IJ (2005) Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. Biochem J 391:449–464

35. Uversky VN, Winter S, Lober G (1996) Use of fluorescence decay times of 8-ANS-protein complexes to study the conformational transitions in proteins which unfold through the molten globule state. Biophys Chem 60:79–88

36. Eftink MR, Ghiron CA (1976) Fluorescence quenching of indole and model micelle systems. J Phys Chem 80:486–493

37. Roberts LM, Dunker AK (1993) Structural changes accompanying chloroform-induced contraction of the filamentous phage fd. Biochemistry 32:10479–10488

38. Turoverov KK, Kuznetsova IM, Maskevich AA, Stepuro VI, Kuzmitsky VA, Uversky VN (2007) Tht as an instrument for testing and investigation of amyloid and amyloid-like fibrils. Proc SPIE 6733:1–7

39. Sulatskaya AI, Kuznetsova IM, Turoverov KK (2011) Interaction of thioflavin T with amyloid fibrils: stoichiometry and affinity of dye binding, absorption spectra of bound dye. J Phys Chem B 115:11519–11524

40. Sulatskaya AI, Kuznetsova IM, Turoverov KK (2012) Interaction of thioflavin T with amyloid fibrils: fluorescence quantum yield of bound dye. J Phys Chem B 116(8):2538–2544

41. Kuznetsova IM, Sulatskaya AI, Uversky VN and Turoverov KK (2012) Analyzing Thioflavin T binding to amyloid fibrils by an equilibrium microdialysis-based technique. Plos One 7(2): e30724

42. Turoverov KK, Biktashev AG, Dorofeiuk AV, Kuznetsova IM(1998) A complex of apparatus and programs for the measurement of spectral, polarization and kinetic characteristics of fluorescence in solution. Tsitologiia 40:806–817

43. Kurganov BI (2000) Analysis of negative cooperativity for glutamate dehydrogenase. Biophys Chem 87:185–199

44. Dzwolak W, Pecul M (2005) Chiral bias of amyloid fibrils revealed by the twisted confor-mation of thioflavin T: an induced circular dichroism/DFT study. FEBS Lett 579:6601–6603

Chapter 27

Fluorescence Lifetime Measurements of Intrinsically Unstructured Proteins: Application to α-Synuclein

Sarah Schreurs, Malgorzata Kluba, Jessika Meuvis, and Yves Engelborghs

Abstract

Lifetimes of fluorescent states and their fluorescence intensities are strictly coupled and very sensitive to the environment of the fluorophores. The advantage of measuring lifetimes, next to intensities, comes from the fact that it can reveal heterogeneity and dynamic properties of this environment. In this way lifetime analysis can be used to characterize static and dynamic conformational properties and heterogeneity of fluorescent groups in different areas of a protein and as a function of time for an evolving protein. The phenomena that determine the lifetime of a label are its intrinsic properties, dynamic quenching by neighboring groups, exposure to the solvent, as well as Förster resonance energy transfer (FRET) between different groups. The basic principles of these fluorescence phenomena can be found extensively described in the excellent book of Lakowicz (Principles of fluorescence spectroscopy, 3rd edn. Springer, New York, 2006). The fluorescent groups involved are either natural amino acid side chains like tryptophan (Trp) or tyrosine (Tyr), or fluorescent labels covalently engineered into the protein. Even a single fluorescent group can show indications of heterogeneity in the local environment. If several natural fluorescent groups are present, the properties of the individual groups can be separated using site-directed mutagenesis, and additivity of their contributions can be analyzed (Engelborghs, Spectrochim Acta A Mol Biomol Spectrosc 57 (11):2255–2270, 2001). If no fluorescent group is naturally present, site-directed mutagenesis can be used to introduce either a fluorescent amino acid or a cysteine allowing chemical labeling.

Key words: α-Synuclein, Fluorescence lifetimes, Tryptophan, Tyrosine, Nitro-tyrosine, Anisotropy, Förster resonance energy transfer, Alexa Fluor 488, Maleimide, Alexa Fluor 594, Oxazine, IAEDANS

1. Introduction

The excited state lifetime (τ) of a fluorescent molecule is determined by its radiative (k_r) and nonradiative rate (k_{nr}) constant: $\tau = (k_r + k_{nr})^{-1}$ where the radiative rate constant is principally an intrinsic property of the probe, and the non-radiative rate constant is the sum of all the rate constants of the different parallel pathways

Table 1
Distance dependent phenomena that influence the lifetime of the fluorescence source

Fluorescence source (donor)	Perturbant (acceptor)	Phenomenon	Distance of action [3] (R_0)
Tyr	Trp	FRET from Tyr to Trp	0.9–1.8 nm
Trp	Nitro-Tyr	FRET	2.6 nm
Trp	Cys-IAEDANS	FRET from Trp to IAEDANS	2.2 nm
Trp	Cys-Oxazine	Oxazine quenching by Trp	Contact
Trp	HisH+	Quenching	Contact
Trp	-SS-	Quenching	Direct or indirect contact
Cys-Alexa Fluor 488	Cys-Alexa Fluor 594	FRET from Alexa Fluor 488 to Alexa Fluor 594	5.4 nm

which return to the ground state without emission [1]. Therefore the lifetime is generally very sensitive to the static and dynamic properties of the probe's environment. Consequently lifetime analysis of fluorescent probes engineered into proteins can reveal kinetic information on the conformational state—heterogeneity, and—dynamics of the protein [2]. This is also true for intrinsically unstructured proteins. Diverse phenomena, such as Förster resonance energy transfer (FRET) or dynamic quenching, can be used to collect information on local or more remote interactions [3]. To simplify the interpretation a unique set of interacting pairs has to be present or introduced by mutagenesis and/or by orthogonal labeling. The easiest and cleanest system is to use a natural or an engineered Trp residue and a second group introduced by chemical coupling onto an engineered cysteine. The following pairs are orthogonal and easy to introduce, and can lead to lifetime changes based on the indicated phenomena (see Table 1).

Introducing two labels via thiol labeling always leads to complex mixtures that need extensive purification [4]. FRET calculated on the basis of reduced intensities of donor fluorescence has to be corrected for the percentage of labeling, while lifetime analysis can reveal the lifetimes of the donor-alone and the donor–acceptor pairs separately, showing again the advantage of lifetime analysis. However, it is difficult to resolve more than three separate lifetimes. Lifetime analysis is necessary to obtain the anisotropy decay. In this case the fluorescence decay has to be measured with the polarizer in the vertical ($I_V(t)$) and in the horizontal position ($I_H(t)$) assuming vertical excitation:

$$r(t) = \frac{I_V(t)/I_H(t) - 1}{I_V(t)/I_H(t) + 2}.$$

2. Materials

1. WT α-synuclein does not contain tryptophan (Trp) residues but has four tyrosine (Tyr) residues at positions: 39 (N-end) 125, 133, and 136 (C-end). It should be realized that mutating them alters the conformation of the protein (5). Many Trp mutants have been constructed at positions (F4W, Y39W, A69W, A90W, A124W, A140W) by Van Rooijen et al. (6) and Trp fluorescence spectra have been characterized in terms of local conformational properties. FRET between Trp and nitro-Tyr has been studied, based on lifetime measurements, for the pairs: 4–19, 19–39, 4–39, 74–94, 94–136, 4–136 by Lee and coworkers (7, 8). Of course again the effect of the modification itself has to be taken into account. We have studied the mutant Y133W to probe the conformation of the C terminus. FRET was also observed between Tyr39 and engineered 125W with distances evolving from 2.49 to 1.88 nm in going from the monomer to late oligomers (12).

2. His-tag labeled α-synuclein His-Y133W—was purified according to the protocol described by Gerard et al. (9). The purification of His-A107C/Y133W—α-synuclein was performed similarly with the addition of 1 mM DTT in all buffers. A HiPrep 26/10 desalting column (GE healthcare) equilibrated in 20 mM NH_4HCO_3 pH = 7.4 was used to remove salt and DTT after which the pure fractions were pooled and lyophilized.

3. Fluorescence lifetime instrumentation.
Fluorescence lifetime analysis requires specialized instrumentation, e.g., a pulsed light source and a fast, sensitive detection system which are available usually only in specialized laboratories. Nowadays several companies produce relatively cheap pulsed diode lasers, which achieve only subnanoseconds wide excitation pulses. More expensive laser systems are needed for obtaining shorter pulses. Our laboratory setup contains the mode-locked Titanium-Sapphire laser (Tsunami 3950C) powered by a MilleniaPro pumped laser and followed by a Picosecond Pulse Selector (Model 3980-2) used for reducing the repetition frequency from 80 to 8 or 0.8 MHz and subsequently by the Flexible Harmonic Generator (GWU-23PL) consisted of a compact frequency doubler and frequency tripler in one housing, used for changing the wavelength from 885 to 295 nm with ~10 ps pulse width. All instruments were purchased from Spectra Physics. In the emission path usually narrow bandpass filters or monochromators are used. In our setup we use Oriel cut-off filters in combination with a Glan—Thompson polariser (PTS15, Oriel) to select either the vertically or the horizontally polarized component, or the full intensity at the magic angle. A fast and sensitive

detector is also needed, such as a microchannel plate or a photomultiplier. Here the detection unit consists of a Lens-PhotoMultiplier Tube (Model H5023, Hamamatsu Photonics) with a wide band amplifier (1 GHz). Our instrument response function (IRF) is determined by its response function. Photon counting is usually done in the so-called reverse mode—the time between an observed fluorescence photon (in our setup registered by the ORTEC Model 425A Nanosecond Delay) and the following excitation pulse (collected by the Trigger Diode Assembly—TDA200, PicoQuant) is determined, which allows much more efficient memory usage than determination of the time between the excitation pulse and the following observed fluorescence photon. Several companies provide TCSPC cards for single photon counting (e.g., Becker and Hickl, Picoquant) and software for deconvoluting the decay curves from the width of the excitation pulse and the IRF (e.g., PicoQuant).

3. Methods

3.1. Labeling with IAEDANS (5-({2-[(Iodoacetyl)amino] ethyl}amino) Naphthalene-1-Sulfonic Acid)

Chemical labeling of (intrinsically unstructured) proteins by thiol modification is described for α-synuclein in the chapter on FCS in this book and in different articles in the reference list (4, 10). The concentration of labeled—α-synuclein was determined by correcting its spectrum for the contribution of IAEDANS to the absorbance at 278 nm using the following equation:

$$A_{278nm(protein)} = A_{278nm(protein)} - A_{336nm(IAEDANS)} \frac{\varepsilon_{278nm(IAEDANS)}}{\varepsilon_{336nm(IAEDANS)}}.$$

The following extinction coefficients were used: $\varepsilon_{278nm\ (Y133W\ \alpha - SYN)} = 9{,}970\ M^{-1}\ cm^{-1}$, $\varepsilon_{278nm\ (IAEDANS)} = 745\ cm^{-1}\ M^{-1}$ and $\varepsilon_{336nm\ (IAEDANS)} = 4{,}500\ cm^{-1}\ M^{-1}$ (see Note 1).

3.2. Fluorescence Lifetime Measurements

1. All fluorescence lifetime measurements of α-synuclein solutions were performed using the described MilleniaPro-pumped Tsunami mode-locked Ti:Saphire laser setup with the final excitation wavelength of 295 nm which is ideally suited to selectively excite the Trp residue in the presence of Tyr residues.

2. Each measurement consist of acquiring fluorescence emission photons using a polarizer at the so-called magic-angle ($\Theta_{mag} = 54.7°$) with respect to the direction of polarization of the excitation light. The acquisition was controlled by Timeharp (PicoQuant) software and continued until 10,000 counts were accumulated in the first channel. For emission wavelength determination, a set of Oriel small-band-pass filters is used, covering the wavelength region from 320 to 380 nm (width at half maximum absorbance is about 9 nm).

References

Fig. 1. Fluorescence decay measurements on α-synuclein with the purpose to probe the conformation of the C terminus and the influence of the proline residues. The instrument response function (IRF) is given and the fluorescence decay of His-tagged α-synuclein with mutation Y133W and His-Y133W-PA$_5$ in which all the proline residues are mutated to Ala. Furthermore the two proteins are labeled with IAEDANS on position A107C (for about 50 %). The measurements show the nice reproducibility of the data, and the surprising absence of an effect of all the five proline mutations as well as the absence of FRET between 133W and AEDANS at position 107.

References

1. Lakowicz JR (2006) Principles of fluorescence spectroscopy, 3rd edn. Springer, New York

2. Engelborghs Y (2001) The analysis of time resolved protein fluorescence in multi-tryptophan proteins. Spectrochim Acta A Mol Biomol Spectrosc 57(11):2255–2270

3. Wu PG, Brand L (1994) Resonance energy transfer: methods and applications [review]. Anal Biochem 218:1–13

4. Ferreon AC, Gambin Y, Lemke EA, Deniz AA (2009) Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. Proc Natl Acad Sci USA 106(14):5645–5650

5. Ulrih NP, Barry CH, Fink AL (2008) Impact of Tyr to Ala mutations on alpha-synuclein fibrillation and structural properties. Biochim Biophys Acta 1782(10):581–585

6. Van Rooijen BD, van Leijenhorst-Groener KA, Claessens MM, Subramaniam V (2009) Tryptophan fluorescence reveals structural features of alpha-synuclein oligomers. J Mol Biol 394(5):826–833

7. Lee JC, Langen R, Hummel PA, Gray HB, Winkler JR (2004) Alpha-synuclein structures from fluorescence energy-transfer kinetics: implications for the role of the protein in Parkinson's disease. Proc Natl Acad Sci USA 101(47):16466–16471

8. Winkler GR, Harkins SB, Lee JC, Gray HB (2006) Alpha-synuclein structures probed by 5-fluorotryptophan fluorescence and 19F NMR spectroscopy. J Phys Chem B 110(13):7058–7061

9. Gerard M, Debyser Z, Desender L, Kahle PJ, Baert J, Baekelandt V, Engelborghs Y (2006) The aggregation of alpha-synuclein is stimulated by FK506 binding proteins as shown by fluorescence correlation spectroscopy. FASEB J 20:524–526

10. Neuweiler H, Banachewicz W, Fersht AR (2010) Kinetics of chain motions within a protein-folding intermediate. Proc Natl Acad Sci USA 107(51):22106–22110

11. Boens N et al (2007) Fluorescence lifetime standards for time and frequency domain fluorescence spectroscopy. Anal Chem 79(5):2137–2149

12. Kaylor J, Bodner N, Edridge S, Yamin G, Hong D-P, Fink A (2005) Characterisation of oligomeric intermediates in α-synuclein fibrillation: FRET studies of Y125W/Y133F/Y136Fα-synuclein. J Mol Biol 353:357–372

Chapter 28

Ensemble FRET Methods in Studies of Intrinsically Disordered Proteins

Elisha Haas

Abstract

The main structural characteristic of intrinsically disordered proteins (IDPs) or intrinsically disordered regions of globular proteins is that they exist as ensembles of multiple conformers which can continuously interconvert, and at times, form ensembles of a more restricted number of conformers. Characterization of the disordered state and transitions to partially or fully ordered states of such ensembles must be expressed in statistical terms, i.e., determination of probability distributions of the various conformers. This can be achieved by measurements of time-resolved dynamic non-radiative excitation energy transfer within ensembles of site-specifically labeled IDP molecules. Distributions of intramolecular segmental end-to-end distances and their fast fluctuations can be determined and fast and slow conformational transitions within selected sections of the molecule can be monitored and analyzed.

Key words: IDP, Time resolved FRET, Site-specific labeling, Distance distributions, Intramolecular diffusion coefficient, Fast fluctuations and conformational transitions

1. Introduction

Measurements of dynamic non-radiative excitation energy transfer (FRET) (1–7), which is based on the distance-dependent interactions between excited-state dipoles, can be applied for the determination of long-range distances between amino acyl residues in disordered or partially folded proteins. This method has several advantages which make it uniquely suited to the study of intrinsically disordered protein (IDP) structure: close to ideal time resolution; very high sensitivity, up to single molecule detection; accuracy at distances representing molecular dimensions (10–100 Å); and the ability to recover distributions of intramolecular distances of fast fluctuating disordered protein molecules.

A series of parameters can be chosen to most fully characterize the disordered states of protein molecules. These include the mean,

the width, the number of subpopulations, and the shape of the distributions of the intramolecular distances. Ensemble time-resolved FRET (trFRET) and single molecule detected FRET measurements of double-labeled protein samples can provide all these parameters. Both ensemble and single molecule FRET experiments have been used in studies of IDPs. Comparison of intramolecular distance distributions of an IDP with those of model polypeptides which are known to be in a true fully disordered state (i.e., free of any attractive or repulsive constraints other than the steric hindrance) can be a sensitive indicator of the disordered state. A small bias relative to such model distributions of intramolecular distances can serve as a sensitive measure of sub-domain structures in otherwise disordered protein molecules.

In their disordered states, protein molecules undergo rapid conformational fluctuations. The lifetime of the excited states of donor probes in trFRET experiments (nanoseconds) defines time windows that enable the detection of rapid fluctuations of intramolecular distances both in the ensemble and in the single molecule modes. Slower fluctuations can be detected by single molecule FRET spectroscopy or autocorrelation analysis of intensity fluctuations of double-labeled protein samples (8, 9).

FRET measurements are also useful for detection and analysis of intermolecular interactions. A common characteristic of many IDPs is the tendency to form amyloid aggregates. These protein–protein interactions can be monitored by various modes of FRET measurements.

Below, we review methods and their applications based on the Förster type distance-dependent interactions, which were developed for detection of inter- and intramolecular distance distributions in IDPs. We shall first briefly review the theoretical basis of the FRET-based methods and then describe protocols for preparation of labeled samples of IDPs suitable for FRET experiments. This will be followed by sections describing the various analysis methods that were developed. The present chapter is limited to the ensemble methods.

1.1. Theoretical Background

1.1.1. Non-radiative Excitation Energy Transfer

Non-radiative transfer of excitation energy requires some interaction between a donor molecule and an acceptor molecule (Fig. 1). This transfer can occur if at least some vibronic levels in the donor have the same energy as the corresponding transitions in the acceptor. Energy transfer can result from different interaction mechanisms. The interactions may be Columbic and/or due to intermolecular orbital overlap. The columbic interactions consist of long-range dipole–dipole interactions (Förster's mechanism) and short-range multipolar interactions. In the energy transfer process, the initially excited electron on the donor, D, returns to its ground state orbital on D, while simultaneously, an electron on the acceptor, A, is promoted to an excited state. For permitted transitions on D and A, the
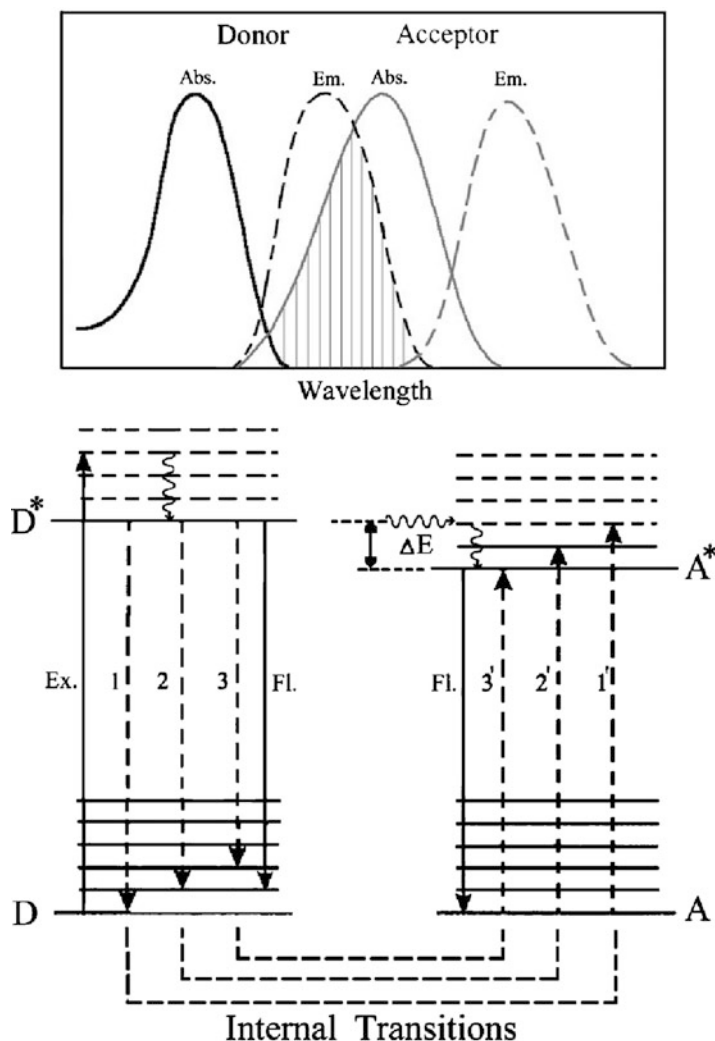
Fig. 1. Energy-level diagram demonstrating the mechanism of resonance coupling of the non-radiative transitions in the donor and acceptor probes under conditions where the vibrational relaxation is faster than the energy transfer (very weak coupling). The non-radiative transitions are shown in *dashed lines* and the radiative transitions are shown as *continuous vertical lines*. The *upper inset* shows a scheme of the corresponding absorption and emission spectra of the probes and the overlap integral.

columbic interaction is predominant, even at short distances. This is the case of the singlet–singlet transfer

$$^1D^* + {}^1A \rightarrow {}^1D + {}^1A^*$$
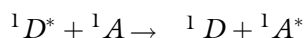
which is effective over a long interaction range (up to 80–100 Å).

Two main mechanisms are relevant in the context of the application of energy transfer for IDP research: the Förster mechanism for long-range interactions, and the triplet–triplet transfer for the detection of shorter range interactions.

The term fluorescence resonance energy transfer (FRET) is commonly used to describe singlet–singlet energy transfer via a mechanism based on long-range dipole–dipole resonance coupling[1](1).

1.1.2. Rate of Non-radiative Excitation Energy Transfer Within a Donor–Acceptor Pair

A molecule in an electronically excited state (donor) can transfer its excitation energy to another molecule (acceptor) (1) provided that the pair fulfills several conditions: (a) the energy donor must be luminescent; (b) the emission spectrum of the donor should have some overlap with the absorption spectrum of the acceptor; and (c) the distance between the two probes must not exceed an upper limit (usually within the range of up to 100 Å). The transfer is readily observed when two different probes are involved, but transfer within a homogeneous population of chromophores can occur, as well. Such transfer was first detected in 1927 by Perrin, who detected the loss of polarization of the emitted light (10).

The dipole–dipole interaction which leads to the transfer of excitation energy is very weak, usually of the order of ~2–4 cm^{-1}, while the spectroscopic energies that are transferred are much higher, ~15,000–40,000 cm^{-1}.

Förster's theory describes the rate of energy transfer for an isolated pair of chromophores which fulfill the requirements for energy transfer by the dipole–dipole interaction to be

$$k_T = \frac{9(\ln 10)\kappa^2 \Phi_D^o}{128\pi^5 N_A' r^6 \tau_D^o} \int_0^\infty f(\lambda)\varepsilon(\lambda)\lambda^4 \mathrm{d}\lambda \tag{1}$$

$$k_T = \frac{1}{\tau_D^o}\left(\frac{R_o}{r}\right)^6 \tag{1a}$$

where $\kappa = \cos\theta_{DA} - 3\cos\theta_D\cos\theta_A = \sin\theta_D\sin\theta_A\cos\phi - 2\cos\theta_D\cos\theta_A$ (θ_{DA} is the angle between the donor and the acceptor dipoles; θ_D and θ_A are the angles between the donor and the acceptor dipoles and the line joining their centers, and ϕ is the angle between the projections of the transition moments on the plane perpendicular to the line through the centers). κ^2 can in principle assume values from 0 (perpendicular transition moments) to 4 (collinear transition moments), or 1 (when the transition moments are parallel); Φ_D^0 and τ_D^0 are the quantum yield and the fluorescence lifetime of the donor in the absence of an acceptor; r is the distance between the centers of the two dipoles (donor and acceptor); N_A' is Avogadro's number per millimole; $f(\lambda)$ is the fluorescence intensity of the donor in the range of

---

[1]The acronym FRET denoting fluorescence resonance energy transfer is inaccurate, because the transfer does not involve any fluorescence but non-radiative transfer of the electronic excitation energy of the donor. This mechanism does not depend on the fluorescence properties of the acceptor. The correct term should be EET representing *electronic energy transfer* or *excitation energy transfer* or RET for *resonance energy transfer*. However, since the term FRET has been accepted in the literature, we continue to use it here.
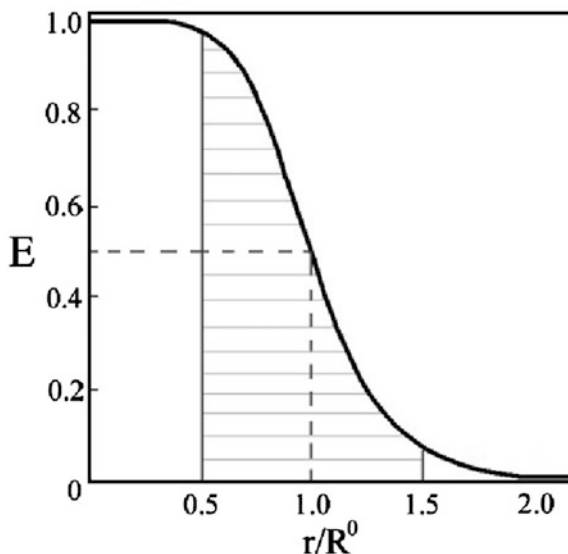
Fig. 2. Variation of the transfer efficiency, E, as a function of the distance between the donor and the acceptor probes. The *shadowed range* represents the limits of the distance range, where reliable measurements of distances are possible.

λ to $\lambda + d\lambda$ normalized so that $\int_0^\infty f(\lambda)d\lambda = 1$; and $\varepsilon(\lambda)$ is the absorption coefficient of the acceptor at the wavelength λ. In Eq. 1, r and λ are in cm units, $\varepsilon_A(\lambda)$ is in cm^2/mol, and $J(\lambda)$ in units of cm^6/mol. When those units are used R_o is defined by Eq. 2 and is given by

$$R_o^6 = 8.8 \times 10^{-28} \Phi_D^o \kappa^2 n^{-4} J \qquad (2)$$

where J is the overlap integral in Eq. 1. The energy transfer process competes with the spontaneous decay of the excited state of the donor, characterized by the rate constant $k_D^o = 1/\tau_D^o$. If $\varepsilon_\varsigma(\lambda)$ is given in cm^{-1} mol^{-1} units, then $R_o^6 = 8.8 \times 10^{-25} \Phi_D^o \kappa^2 n^{-4} J$. Thus the probability ρ for the donor to retain its excitation energy at the time t after excitation is given by

$$-\left(\frac{1}{\rho}\right)\frac{d\rho}{dt} = \frac{1}{\tau_D^o} + \frac{1}{\tau_D^o}\left(\frac{R_o}{r}\right)^6 \qquad (3)$$

And the efficiency of E of energy transfer is expressed by

$$E = \frac{R_o^6}{R_o^6 + r^6} \qquad (4)$$

R_o is thus the inter-dipole distance at which the transfer efficiency (and the donor lifetime) is reduced to 50 % and has the strongest dependence on changes of that distance (Fig. 2).

There is a strong experimental support for Förster's Eqs. 1–4 (reviewed by Steinberg (4). Weber and Teal (11) and Latt et al. (12) showed the dependence on the overlap integral. The R^{-6} dependence was demonstrated by Stryer and Haugland (13).

Equation 4 shows that the distance between a donor and an acceptor can be determined by measuring the efficiency of transfer, provided that r is not too different from R_o and that all molecules in the sampled ensemble share the same intramolecular distance between the labeled residues. If this is not the case, an average distance r_{av} that would be extracted from the measured transfer efficiency does not correspond to any simple average distance.

The efficiency of energy transfer is independent of the value of the lifetime of the excited state of the donor. Efficient energy transfer can occur even for "long-life" excited states (e.g., for phosphorescence emission) provided that the quantum yield of emission is reasonably high. Such transfer phenomena have been observed. Equations 3 and 4 thus define the characteristic time and distance ranges ("windows") in which the transfer efficiency, E, is most sensitive to conformational transitions and their rates (Fig. 2). Typical values available for probes that are used in protein chemistry are in the range of 10–80 Å and down to picosecond time intervals. Many conformational changes and processes in IDPs occur on these timescales. Hence, the researcher can design experiments with the time and distance resolutions most suitable for each molecular question.

1.1.3. The Orientation Factor   Distance determination by measurements of transfer efficiencies is complicated by the strong dependence of the probability for energy transfer on the orientation of the interacting dipoles. Therefore knowledge of the orientation factor, κ^2, is essential for applications of FRET in studies of protein conformations. When the two dipoles undergo rapid orientational averaging during the lifetime of the excited state of the donor, the orientation factor can be averaged to yield a numerical value of 2/3. The large span of possible κ^2 values, between 0 and 4, makes it the primary source of uncertainty in distance determination by FRET measurements (see Notes 1 and 2).

2. Materials

Prepare all solutions using ultrapure water and analytical grade reagents. Use spectrograde solvents for fluorescence studies. All materials used should be checked for fluorescence in the relevant spectral range. For an example of the materials used for preparation of labeled α-synuclein (αS) molecules see Note 7.

3. Methods

3.1. Determination of Intramolecular Distances in Protein Molecules Using FRET Measurements

Clear presentation of theoretical background for the methods described below may be found in the monographs by B. Valeur (14), J. Lakowicz (15), W. Van Der Mear (6), and numerous reviews such as (4, 16–20) published over the past 40 years.

3.1.1. Single Distance Between Donor and Acceptor

Determination of distances by determination of FRET efficiency, E, is possible by measuring the decrease of donor emission or an enhancement of the acceptor emission by steady-state and time-resolved methods.

$$r = \left(\frac{1}{E} - 1\right)^{1/6} R_o \tag{5}$$

Due to the comparative nature of this mode of determination of E, the concentrations of the probes (and hence the labeled protein molecules) and their microenvironments must be the same for each set of measurements (see Note 1).

3.2. Steady-State Detection Methods

The competition of the FRET process results in decreased quantum yield of donor emission:

3.2.1. Method 1: Determination of Decrease of Donor Emission

$$E = 1 - \frac{\Phi_D}{\Phi_D^o} \tag{6}$$

Since only relative quantum yields are to be determined, Eq. 13 can be written directly in terms of single wavelength donor emission intensities at wavelengths where the acceptor emission intensity is negligible:

$$E = 1 - \frac{A(\lambda_D)}{A_D(\lambda_D)} \frac{I_D\left(\lambda_D, \lambda_D^{em}\right)}{I_D^o\left(\lambda_D, \lambda_D^{em}\right)} \tag{7}$$

The factor A/A_D corrects for the absorption by the acceptor. This method can be readily applied at the single molecule level (see Note 2).

3.2.2. Method 2: Acceptor Excitation Spectroscopy

The most direct measurement of E, independent of correction factors or differences in concentrations, can be achieved via measurements of acceptor excitation spectra. In this procedure, the acceptor excitation spectra of the following three samples should be measured under uniform solution conditions. These spectra are (a) the acceptor excitation spectrum of the double-labeled protein (labeled by both the donor and the acceptor), $I_A(\lambda_D, \lambda_A^{em})$; (b) the excitation spectrum of the acceptor in the absence of a donor, using protein sample labeled by the acceptor alone, $I_{ref0}(\lambda_D, \lambda_A^{em})$

(a reference for $E = 0$); and (c) the excitation spectrum of the acceptor attached to a double-labeled (donor and acceptor) reference compound for which E is well known under the conditions of the measurements $I_{ref1}(\lambda_D, \lambda_A^{em})$. The acceptor excitation spectra of the three samples monitored at an acceptor emission wavelength (with negligible donor emission contribution) are then normalized at the acceptor excitation wavelength. E can then be obtained with high accuracy by

$$E = \frac{I_A\left(\lambda_D, \lambda_A^{em}\right) - I_{ref0}\left(\lambda_D, \lambda_A^{em}\right)}{I_{ref1}\left(\lambda_D, \lambda_A^{em}\right) - I_{ref0}\left(\lambda_D, \lambda_A^{em}\right)} \tag{8}$$

(Reduced noise can be achieved by using the ratio of the area under the excitation spectra in the donor absorption range.) This is the preferred method for the determination of E by steady-state methods. Non-FRET mechanisms of quenching of the donor might affect the determination of E by Method 1. Control experiments should be performed to confirm that the missing donor emission intensity is observed in the acceptor emission. Such non-FRET change of the donor emission does not affect the determination of E by method 2. Both steady-state methods might be affected by the "inner filter effect" and care must be taken to account for changes in emission intensity due to this phenomenon.

3.3. Time-Resolved Detection Methods

Time-resolved emission of the donor or the acceptor fluorescence provides direct information regarding the transfer rates independent of the concentrations, as this information is provided by the shape of the fluorescence decay curves, and not the amplitudes. The donor and the acceptor fluorescence decay curves contain additional information not available from the steady-state measurements. This information, which can be extracted directly by proper analytical procedures, is of particular interest for studies of IDPs. Analysis of trFRET experiments can resolve conformational subpopulations, distributions of distances, and fast conformational fluctuations on nanosecond timescales.

3.3.1. Determination of Donor Fluorescence Decay Rates

When an acceptor is located close enough to the donor, two competing processes affect τ_D, the lifetime of the excited state of the donor, and the transfer rate,

$$1/\tau_D = 1/\tau_D^o + k_{ret} \tag{9}$$

The transfer efficiency can be determined using the fluorescence lifetimes of the donor,

$$E = 1 - \frac{\tau_D}{\tau_D^o} \tag{10}$$

and the apparent mean distance is calculated by

$$r = \frac{R_o}{\left(\tau_D/\tau_D^o - 1\right)^{1/6}} \tag{11}$$

When a non-exponential decay of the donor alone (τ_D^o) is observed, a common situation due to heterogeneity of microenvironments, average lifetime values can be used to calculate E, provided that the deviation from monoexponentiality is moderate:

$$E = 1 - \frac{\langle \tau_D \rangle}{\langle \tau_D^o \rangle} \tag{12}$$

where

$$\langle \tau \rangle = \sum_i \alpha_i \tau_i \bigg/ \sum_i \alpha_i$$

When the heterogeneity of the decay rates is enhanced by the FRET effect, distribution analysis accompanied by appropriate control experiments should be applied.

3.3.2. Determination of Intramolecular Distance Distributions

The distance between two selected residues in an IDP is by definition nonuniform, and hence Eqs. 6 or 10 should be replaced by distance analysis, assuming a probability distribution of the intramolecular distance between the labeled sites, $N_o(r)$, with finite full width at half maximum (FWHM) determined. In that case, the FRET efficiency determined by steady-state measurements is averaged over the efficiencies of all fractions in the distribution,

$$<E> = \int_0^\infty N_o(r) \frac{R_o^6}{R_o^6 + r^6} \, dr \tag{13}$$

where $N_o(r)dr$ is the probability of the fraction of molecules with D–A pairs at distance r to $r + dr$. $N_o(r)$ can be resolved when the experiment is repeated with several pairs of probes of different R_o values (21). A single time-resolved experiment contains all the information needed for the determination of $N_o(r)$.

When the fast conformational exchange within the ensemble of labeled IDP molecules is slow relative to the lifetime of the excited state of the donor, $N_o(r)$ can be recovered by modeling $i(t)$ using the expression

$$i(t) = k \int_0^\infty N_o(r) \exp\left[-\frac{t}{\tau_D^o}\left(1 + \left(\frac{R_o}{r^6}\right)^6\right)\right] dr \tag{14}$$

where k is a proportionality factor. Knowledge of R_o and τ_D^o enables the extraction of $N_o(r)$ from the multiexponential decay curve, $i(t)$.

3.3.3. Evaluation
of the Effect of Fast
Conformational
Fluctuations and
Determination
of Intramolecular Diffusion
Coefficients

When exchange of conformers occurs between fractions of an IDP ensemble under equilibrium conditions (where the equilibrium probability distribution of the distances is unchanged) the transfer efficiency is enhanced. This enhancement is modeled as an intramolecular diffusion of the labeled sites on the chain, which acts toward restoration of the equilibrium probability distribution. This phenomenon can be viewed as a process of diffusion under the conformational force field which governs the equilibrium probability distribution of the intramolecular distance, $N_o(r)$ (22–24). Equation 15 considers both the energy transfer reaction term, $k(r)$, and the restoration force by the diffusion term. Equation 15 is the basis for the experimental data analysis (22, 25) for determination of the parameters of the distribution, $N_o(r)$:

$$\frac{\partial \overline{N}_i(r, t)}{\partial t} = \frac{D}{N_0(r)} \times \frac{\partial}{\partial r}\left[N_0(r)\frac{\partial}{\partial r}\overline{N}_i(r, t)\right] - k_i(r)\overline{N}_i(r, t) \quad (15)$$

where i denotes the ith species of the donor fluorescence decay (in the absence of an acceptor), $N_i^*(r,o) = N_0(r)$ (the equilibrium distance distribution); $N_i^*(r,t)\mathrm{d}r$ is the probability for molecules with an excited-state species i of the donor to have an intramolecular distance in the range r to $r + \mathrm{d}r$ and $\overline{N}_i(r,t) = N_i^*(r,t)/N_0(r)$. D is defined as the intramolecular diffusion coefficient of the segments carrying the two probes; τ_i and α_i are the lifetime and the normalized pre-exponential factor, respectively, of the excited state of the ith species of the donor in the absence of an acceptor. $k_i(r)$, the reaction term, includes the spontaneous emission rate and the Förster energy transfer rate:

$$k_i(r) = \frac{1}{\tau_i} + \frac{1}{\tau_r}\left(\frac{8.79 \times 10^{-25}\ n^{-4}\ \kappa^2\ J}{r^6}\right) \quad (16)$$

τ_r is the radiative lifetime of the donor, and the other terms were defined in Eqs. 1–1a. Equation 16 considers the variation of R_o due to variations of the donor quantum yield Φ_D^0. This is represented by the ratio of the donor lifetime components and the radiative lifetime, which is assumed to be independent of local environmental effects.

The solution of Eq. 15 is obtained by numerical methods. Since the boundaries are chosen in such a way that the distribution function is very small at both the upper and lower extreme distances, reflective boundary conditions are arbitrarily used (26):

$$[\delta \overline{N}(r, t)/\delta r]_{r=r\min, r\max} = 0 \quad (17)$$

$i_c^D(t)$, the calculated fluorescence decay curve of the donor, is readily obtained from $\overline{N}(r,t)$ by the relationship (22)

$$i_c^D = m\int_{r_{\min}}^{r_{\max}} N_0(r)\left[\sum_i \alpha_i \overline{N}_i(r, t)\right]\mathrm{d}r \quad (18)$$

where m is a constant, and α_i is the pre-exponential factor of the ith excited-state species of the donor in the absence of an acceptor.

The ideal case is such that the fluorescence decay of the donor in the absence of an acceptor is mono-exponential. In such a case, all molecules share a single value of R_o, the sum in Eq. 18 reduces to one term, and the reactive term, $k_i(r)$ (Eq. 16), can be replaced by a simpler expression:

$$k(r) = -1/\tau_D^o[1 + (R_o/r)^6] \qquad (19)$$

In a case where the diffusion rate is known to be negligible on the timescale of the lifetime of the excited state of the donor, the Fick term in Eq. 15 is zero and $N_o(r)$ can be derived directly from an analysis of the donor decay curve. In such a case, Eq. 15 can be replaced by (23) (see Note 3):

$$i_c^D(t) = k \int_0^\infty N_o(r)(\exp -k(r))\mathrm{d}r \qquad (20)$$

3.3.4. Global Analysis

The routine global analysis procedure includes joint analysis of at least one set containing at least two or four fluorescence decay curves of the probes attached to the protein: (a) *The "DD-experiment,"* determination of τ_D the fluorescence decay of the donor in the presence of FRET (excitation at the donor excitation wavelength and detection at the donor emission wavelength); (b) *The "DA-experiment,"* determination of τ_A, the acceptor emission, in the presence of FRET, with excitation at the donor excitation range, and detection of the acceptor emission at the long wavelength range. Two internal reference measurements are included for determination of the probes' lifetime when attached to the same corresponding sites as in the double-labeled mutant but in the absence of FRET; (c) The *"D-experiment,"* a measurement of τ_D^o, the emission of the donor in the absence of an acceptor, using protein molecule labeled by the donor alone; and (d) *The "A-experiment,"* measurement of τ_A^o, the decay of the acceptor emission in the double-labeled protein, excited directly at a wavelength where the donor absorption is negligible or using a protein preparation labeled by the acceptor only. When collected under the same laboratory conditions (optical geometry, state of the instrument, time calibration, linearity of the response, and solution parameters) used for the first two experiments, the *D*- and A-experiments serve as "internal standards" for the time calibration of this experiment and the dependence of the spectroscopic characteristics of the probes.

3.3.5. Rigorous Error Analysis

The statistical significance of the value of any one of the free parameters that is determined by any curve fitting method can be evaluated by a rigorous analysis procedure. This procedure, if repeated for every free parameter, actually generates projections of

the χ^2 hypersurface on specific axes in the parameter space. Error limits at any required significance level can then be determined using an *F*-test (24). This procedure considers all correlations between the parameters. The evaluation of each analysis and of the significance of the parameters is based on four indicators: (1) the global χ^2 values; (2) the distributions of the residuals; (3) the autocorrelation of the residuals; and (4) the error intervals of the calculated parameters obtained by the rigorous analysis procedure. In addition, each "DD-experiment" is analyzed by a multiexponential (three or four exponentials) decay model. The χ^2 values obtained in these analyses serve for evaluation of the significance of the χ^2 values obtained for each set of experiments in the global analysis, based on the solution of Eqs. 15–18.

3.4. Strategies for Site-Specific Double Labeling of Proteins

Site-specific labeling can be achieved by seven basic strategies: (a) Use of natural probes including the tyrosine or tryptophan residues and a few prosthetic groups. Site-directed mutagenesis can be applied for the preparation of single-tryptophan and single-tyrosine mutants; (b) Nonselective chemical modification of reactive side chains and chain terminal reactive groups (amine, carboxyl, and mainly sulfhydryl groups) combined with high-resolution or affinity chromatography-based separation methods; (c) Solid-phase synthesis of protein fragments and ligation of synthetic peptide fragments, or of a genetically truncated protein fragment and a synthetic, labeled complementing fragment. Total synthesis of labeled protein molecules is applicable for medium-size proteins, e.g., the total synthesis of Aβ fragments(27, 28); (d) Production of edited mutants with no more than one pair of exposed reactive side chains, preferably cysteine residues (29, 30), or one cysteine and one tryptophan (17, 31–33) at the desired sites, followed by chemical modification and chromatographic separation. Nitration of tyrosine residues has also been successfully applied (34); (e) Cell-free protein synthesis and the use of edited t-RNAs and edited genetic code for the incorporation of nonnatural amino acids into selected sites in the chain (35); (f) Enzyme-catalyzed insertion of fluorescent substrates, taking advantage of the specificity and mild conditions of the enzymatic reaction (36); and (g) Use of a protein genetically fused with one or two fluorescent proteins (37) (see Note 3).

A major consideration involved in the preparation of site-specifically labeled natively disordered proteins is avoiding aggregation. Procedures should be rapid with short incubations, and avoid extreme conditions. Therefore the simpler labeling strategies are preferred. Strategy (a) offers minimal posttranslational processing steps and avoids the potential artifacts and uncertainties involved in chemical labeling. On the other hand, the main limitations of using the natural probes are the limited arsenal of available probes; their less than ideal spectral characteristics; the limited range of R_o values

that can be achieved using them; and the fact that most proteins contain several Trp and Tyr residues. Applications of strategy (b) are limited to small proteins or proteins where a limited number of the reactive side chains can be found (38). Strategy (c) allows a high degree of flexibility and freedom for the design of labeled proteins (35, 39). The limited scope of application of this approach at the present time is mainly due to the multistep operations involved and the low yields. In most FRET experiments, the sensitivity of the fluorometric detection systems makes this method satisfactory and promising. Strategy (f) is not of general use due to the limited number of known suitable enzymatic reactions and the multiplicity of reactive side chains. The well-documented catalysis of peptide bond synthesis by proteolytic enzymes was described (for C-terminal labeling of RNase A) but it is not easy to apply as a general method. Transglutaminase catalysis of insertion of primary amino groups attached to fluorescent probes was also attempted (40–42), but the development of strategy (d) offers simpler route. Strategy (g) is good for in vivo experiments, but not so useful for structural studies IDPs in vitro due to the very large size of the "probes." Strategy (e) is theoretically the ideal strategy—using such techniques, any protein with any number of Cys or Trp residues can be labeled. But, even though the technique was developed more than 20 years ago, it is still not widely used due to its sophistication and low yields.

Strategy (d) (production of edited mutants with single labeling sites) is currently the most general and practical approach for several reasons: first, in general, natively disordered proteins do not have disulfide cross-links and cysteine residues are rare. In many proteins these residues can be replaced by Ser or Ala residues without major perturbations of structure and function. Thus, most natively disordered proteins can be engineered to have two exposed sulfhydryl side chains. Second, the arsenal of sulfhydryl reactive pairs of probes covering a wide range of spectroscopic characteristics and R_o values is abundant, and the experimental design can be optimized to meet a wide variety of distances and labeling conditions. That is also the main strategy that can be applied for single molecule FRET experiments (43) (see Note 4).

3.4.1. Preparation of Double-Labeled Mutants Using Engineered Cysteine Residues (Strategy d)

Preparation of protein mutants labeled by a single Trp and a single chemically modified Cys (or Tyr) residue is the simplest strategy. When this pair cannot satisfy the spectral requirements needed for a specific experiment, the challenge is then to develop methods for selective modification of each one of the two reactive sulfhydryl groups in the same protein molecule. This method was used by Grupi and Haas (33, 44) in a study of the molecule α-synuclein. This is a 140 residue molecule which is implied in the onset of the Parkinson disease. The labeling strategy is shown in Fig. 3 (see Note 5).
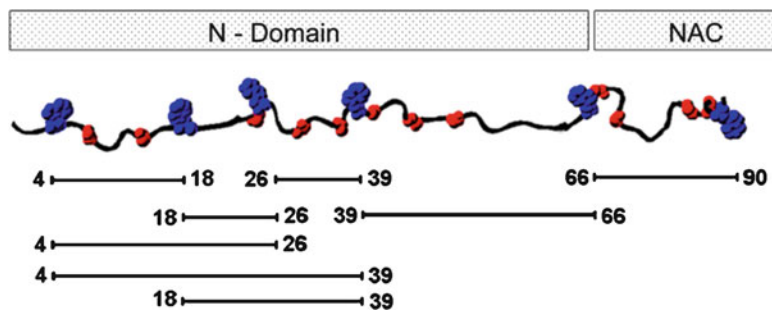
Fig. 3. The design of a labeling plan of overlapping backbone segments in the N terminal domain of the human α-Synuclein molecule. α-Synuclein is a 140-residue IDP. Strategy (d) (see text) was used by Grupi and Haas with engineered Trp residues as the donor and a coumarin derivative attached to engineered Cys residues as acceptor. Labeled segments are marked ▬. The residues labeled with either Tryptophan or coumarin are marked as large cloud-like entities indicating tryptophan side chains. Glycine residues are marked small knots to emphasize their expected flexibility and contribution to the dynamics of the segment. The possible disorder of the backbone is represented by *bends* and *squiggles* (reproduced from ref. 33, with permission).

Sub-milligram amounts of labeled proteins can suffice for a full set of FRET experiments under equilibrium conditions. Such minute amounts of site-specifically double-labeled protein samples can be prepared, based on high-resolution chromatography which can resolve a mixture of the possible products of a nonselective reaction of two-cysteine mutants with a fluorescent reagent. Homogeneous double-labeled products can be obtained in a second cycle of reaction and separation. A limited degree of selectivity of the first labeling reaction can sometimes be achieved by altering reactivity of selected side chains by means of changes in solvent composition (organic solvents, salt components, and pH) and temperature (45). More drastic effects can be obtained by addition of ligand (substrates, inhibitors, or protein–protein interactions) (46).

Chromatographic resolution of labeled disordered protein mutants based on the site of labeling is difficult. Separation between single- and double-labeled fractions following the first round of labeling of a two-cysteine mutant is feasible. In principle, a mixture of double-labeled mutants, where the donor and the acceptor can be attached to either one of the sites, can be used, provided that the spectral characteristics of the probes are the same at both sites.

It is also possible to use a mixed disulfide resin for separation of fractions with free cysteine residues in the process of labeling two-cysteine mutants. The single-labeled mutants which are used for the reference measurements can be prepared by straightforward labeling of single-cysteine mutants. Caution must be taken, since in several cases it was shown that the spectroscopic properties of a probe attached at one site on a protein are significantly modified by the addition of a second cysteine for the double labeling. In this case, it is desirable to use the corresponding single-labeled protein

sample of the two-cysteine mutant, in which the second cysteine residue is blocked by a nonfluorescent alkylating reagent such as iodoacetamide (see Note 6).

3.5. Experimental Aspects of FRET Applications in Characterization of Monomer Structure and Intermolecular Interactions

The transfer efficiency in an ensemble of disordered protein molecules cannot be used for calculation of mean inter-probe distances due to its nonlinear distance dependence and the intrinsic heterogeneity of the conformations. Only quasi-mean distances can be determined, and changes of FRET efficiency can be related to changes in the distances. This limitation highlights the strength of ensemble trFRET methods and the single molecule detection modes.

3.5.1. Steady-State Determination of Transfer Efficiency

Donor Emission

The most commonly used method for steady-state determination of transfer efficiency is based on the determination of the extent of quenching of donor emission in the double-labeled samples (Eqs. 6–7). Emission of the donor is recorded at a wavelength range free of acceptor emission; then, $E = 1 - I_{DA}/I_D$ where I_{DA} and I_D are the donor emission intensities in the presence and absence of an acceptor, respectively. Both intensities should be normalized to the same donor concentration. It may be difficult to determine the concentration of the donor in the presence of an acceptor. This difficulty (which is eliminated in the time-resolved experiment) can be circumvented by enzymatic digestion of each sample after determination of I_{DA} (47). A second emission intensity measurement immediately after the digestion thus provides I_D and an internal reference (48). The FRET interaction does not affect the shape of the spectra of the probes. Therefore, nonlinear least squares fitting of the absorption and emission spectra of the double- and single-labeled samples can be used to overcome the concentration normalization problem (49). The nonlinear least squares fitting of the entire emission spectral range can improve the accuracy of transfer efficiency calculations when the donor and acceptor emissions overlap (18, 50).

Acceptor Excitation Spectroscopy

The most direct and error-free determination of FRET efficiency by steady-state experiments is obtained when the excitation spectra of the acceptor are recorded for the following three samples and then normalized (at the acceptor excitation range) and superimposed (Fig. 4). The three samples (and corresponding spectra) are (a) the acceptor single-labeled mutant $\left(I_A^{\text{ex}}\right)$, (b) the double-labeled mutant $\left(I_{DA}^{\text{ex}}\right)$, and (c) a special double-labeled reference of well-documented transfer efficiency, E_{ref} (preferably a compound in which the two probes are very close, $E_{\text{ref}} = 1.0$), $\left(I_{\text{ref}}^{\text{ex}}\right)$. The integrated areas under the three curves in the donor excitation wavelength range provide the transfer efficiency free of instrumental artifacts and concentration effects:

$$E = \frac{I_{DA}^{\text{ex}} - I_A^{\text{ex}}}{E_{\text{ref}} I_{\text{ref}}^{\text{ex}} - I_A^{\text{ex}}} \tag{21}$$
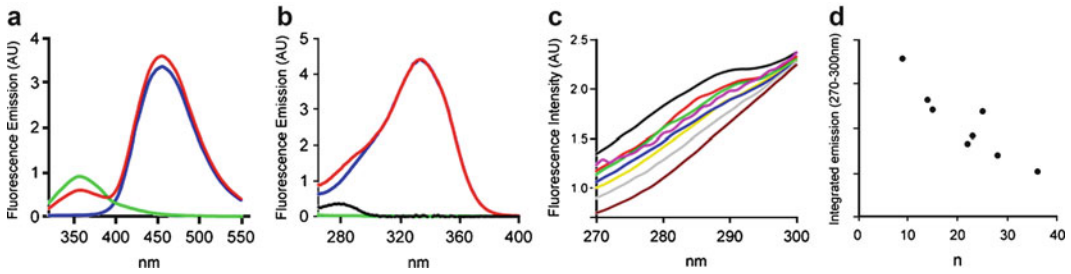
Fig. 4. Steady-state determination of relative FRET efficiencies of the labeled segments in the N terminal domain of the α-Synuclein molecule. (**a**) Emission spectra of the species labeled by donor in the absence of an acceptor (DO) (*green*), the species labeled by both donor and acceptor (DA) (*red*), and a species labeled by the acceptor without donor (AO) (*blue*)-labeled mutants. (**b**) Excitation spectra of DO (*green*), DA (*red*), and AO (*blue*)-labeled mutants; emission wavelength was 332 nm. FRET is observed as increased emission of the acceptor in the DA mutants relative to the AO mutants when both spectra are normalized at the acceptor excitation wavelength. Subtraction of the DA curve from the AO curve gives the absorption spectra of the protein (*black*). (**c**) Excitation spectra of eight DA-labeled mutants and a reference AO excitation spectrum. The labeled segments are AO (*brown*), 4–39 (*gray*), 39–66 (*light blue*), 66–90 (*purple*), 4–26 (*blue*), 18–39 (*yellow*), 4–18 (*green*), 26–39 (*red*), and 18–26 (*black*). (**d**) The integrated emission intensity between 270 and 300 nm of the excitation spectra reveals the transfer efficiency dependence on segment length (reproduced from ref. 33 with permission).

Time-Resolved Measurements

Equations 10–12 are used for the determination of the average FRET efficiency by time-resolved measurements. The main technical advantage of this measurement is that there is no need to know the concentrations of the proteins, and the accuracy of the lifetime measurements surpasses that of the steady-state intensity measurements.

The fluorescence decay curves of the probes in trFRET experiments contain detailed information on heterogeneity of the ensembles of proteins (Fig. 5).

The technology of time-resolved fluorescence measurements matured during the last quarter of the twentieth century. Fluorescence decay can be detected either by the single-photon counting method or by the phase modulation method, and in principle, both methods are equivalent. While phase experiments are relatively easy to perform, single-photon counting techniques offer more straightforward methods for incorporation of acceptor decay curves in the analysis, and easier control of background subtraction and other corrections.

Data Analysis

The basic procedure used for analysis of trFRET experiments is based on curve fitting methods. In this procedure, the parameters of the distance distribution function, the diffusion coefficient, the fluorescence lifetimes, and their associated pre-exponents are searched (24, 51).

The calculated fluorescence decay profile, $F_{\text{calc}}(t, \lambda_{\text{ex}}, \lambda_{\text{em}})$, is composed of both the donor and the acceptor decay curves,
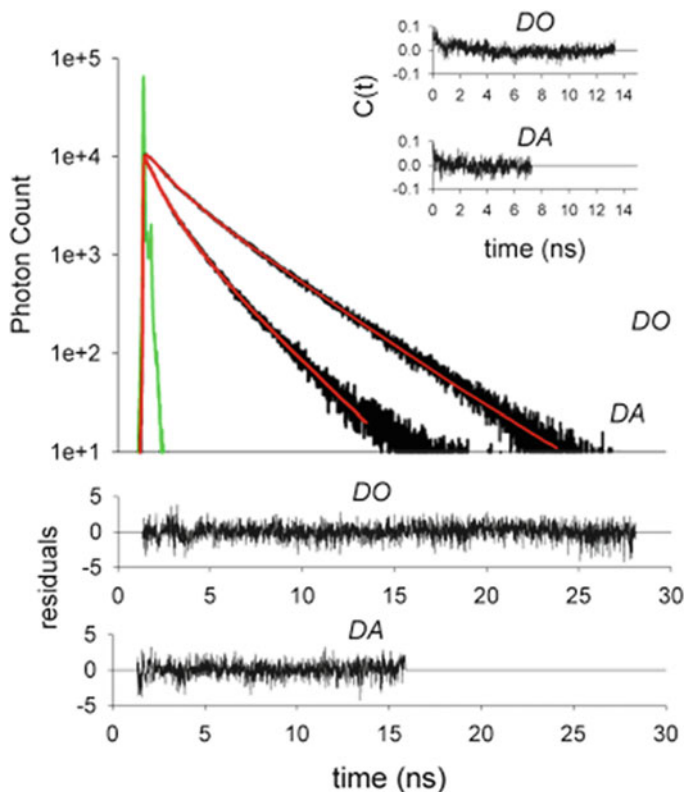
Fig. 5. Fluorescence decay curves of tryptophan in DO and DA mutants of α-Synuclein, the associated residuals, and the autocorrelation function. The decay rate of the tryptophan residue in the DA mutant is accelerated due to the FRET effect. The distribution of the residuals (*bottom*) and the autocorrelation function (*upper right*) served as indicators of the quality of fit of the experimental data to a multiexponential function. The average lifetime of tryptophan for all DO mutants was $2.2 < \tau_{\text{average}} < 2.7$. The lifetime components were $\tau_1 \approx 4$ ns (30–55 %), $\tau_2 \approx 1.8$–2.3 ns (30–45 %), and $\tau_3 \approx 0.8$ ns (15–30 %) (reproduced from ref. 33 with permission).

$F_D(t)$ and $F_A(t)$, respectively, which are calculated using the Förster theory applied to the relevant model distribution,

$$F_{\text{calc}}(t, \lambda_{\text{ex}}, \lambda_{\text{em}}) = F_A(t) + F_D(t) \qquad (22)$$

Several computational methods were developed to search for the set of free parameters. The maximum entropy method (MEM) is also applied to the analysis of time-resolved fluorescence (52) and FRET experiments (53). These yield the theoretical decay curve best fitted to the experimentally recorded fluorescence decay curves $F_{\text{exp}}(t, \lambda_{\text{ex}}, \lambda_{\text{em}})$ (54, 55). This procedure is inherently difficult since the analysis of any multiexponential decay curve is a mathematically "ill-defined" problem (56, 57). The quality of fit of the experimental and calculated curves is judged by the minimization to the χ^2 values and the tests for randomness of the deviations (51, 57).
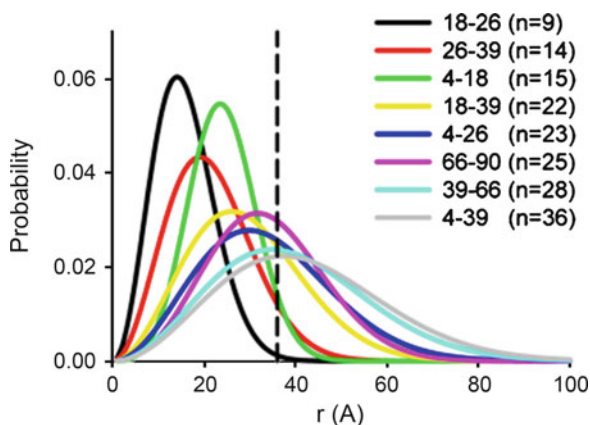
Fig. 6. End-to-end distance (EED) distributions of eight labeled segments in the N terminal domain of the α-Synuclein molecule at 20 °C. *EED* distributions were calculated by fitting the fluorescence decay data to a distance distribution model with a diffusion parameter. The *broken line* at 36 Å indicates the distance beyond which the uncertainty in the distance determination is increased due to the specific Förster critical distance of the pair used here (reproduced from ref. 33 with permission).

The differential Eq. 15 does not have a simple general analytical solution (except for specific cases). Consequently, a numerical solution based on the finite difference (FA) method (58, 59) was applied. Therefore the most common current practice is the use of model distribution functions for $N_o(r)$ with a small number of free parameters (Fig. 6). The model used in the experiments described below is a Gaussian distribution ($N_o(r) = C(f4\pi r^2 \exp - a_1(r - \mu_1)^2 (+ 1 - f)4\pi r^2 \exp - a_2(r - \mu_2)^2)$ where a_i and μ_i are free parameters which determine the width and the mean of the distributions, respectively, f is the fraction of subpopulation 1, and C is a normalization factor). A model including the contribution of intramolecular diffusion can also be used in this context (Fig. 7).

Uncertainties in the values of these parameters (due to correlations) are routinely reduced by global analysis of multiple independent measurements. In particular, coupled analysis of the fluorescence decay of both the donor and the acceptor probes contributes to over-determination of the free parameters (24). Whatever methods are used for data analysis, the best fit must be tested by the statistical tests and fine details of the distribution of the residuals.

3.6. Applications of trFRET for Characterization of Natively Disordered Proteins

The first pioneering application of FRET investigation of a natural flexible polypeptide was reported by Edelhoch, Brand, and Wilchek in 1963 (60, 61). Edelhoch and his colleagues applied strategy (a), using a pair of intrinsic probes (Tyr to Trp transfer). Since then, the technology of both site-specific labeling and steady-state and
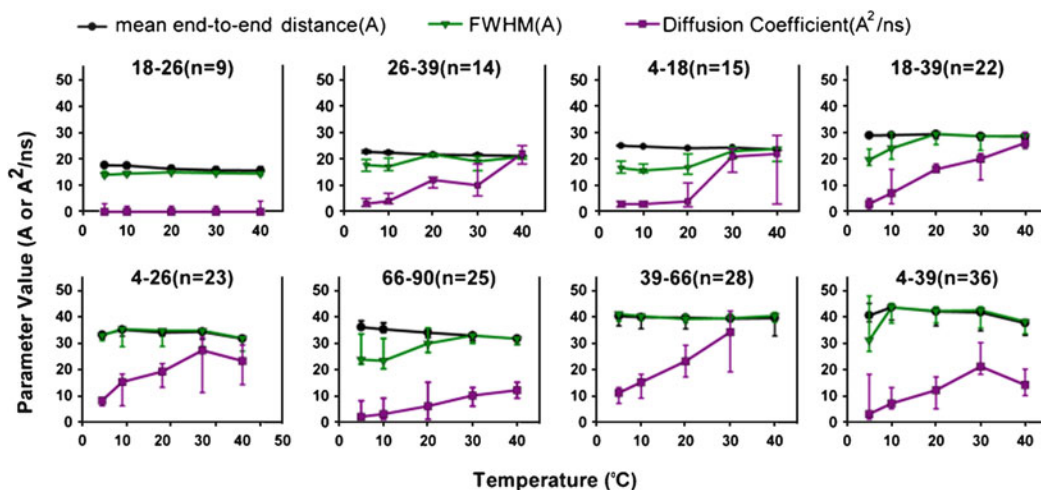
Fig. 7. Temperature dependence of the mean and width of the αS segmental end-to-end distance distributions as well as the intramolecular diffusion coefficient. The mean of the segmental end-to-end distance distribution (Å); the corresponding FWHM of each distribution (Å); and the intramolecular diffusion coefficient (Å2/ns). The parameters are plotted as a function of temperature over the range of 5–40 °C. Confidence intervals for the mean distance are within two std (97.7 %), and for the FWHM and diffusion coefficient are within one std (67.7 %) (reproduced from ref. 44 with permission).

time-resolved fluorescence measurements has advanced; together with the methodological developments, the number and power of the applications of FRET measurements in characterization of disordered peptides and proteins proliferated.

In this section, we describe several examples of the application of the methods presented above, which demonstrate the unique advantages of FRET-based methods in the field of IDP research. Two major types of FRET applications were reported: (a) characterization of the ensemble of conformations of the monomer IDP by means of intramolecular FRET measurements and (b) characterization of the structure and kinetics of formation of supramolecular structures, particularly of amyloid aggregates.

3.6.1. Characterization of Ensembles of Disordered Polypeptides or Disordered Sections Within Proteins by Intramolecular Ensemble FRET Methods

Ensemble Properties of Long IDPs

Several groups applied ensemble FRET-based methods in studies of full-length IDPs such as α-synuclein and tau protein. αS is a 140 residue IDP which is implicated in the etiology of Parkinson disease. Extensive mapping of the degree of disorder in segments of the α-synuclein backbone in the monomer state was reported by Lee et al. (34) and Grupi and Haas (33). Labeling strategy (d) was combined with ensemble trFRET experiments, and a series of intramolecular segmental end-to-end distance probability distributions were determined by analyses of the Trp fluorescence decay curves. Gray's group used nitration of Tyr residues to create a nonfluorescent acceptor, while the Haas group inserted Cys residues which were alkylated by Iodo-acetamidocoumarin-4-carboxylic. Subtle but significant variations in the extent of disorder along the αS backbone were reported. The determination of the temperature

dependence of the segmental end-to-end intramolecular diffusion coefficient yielded further indications of very subtle conformational constraints which might be relevant to the onset of aggregation (33). Fink et al. (62) used labeling strategy (a) and monitored the steady-state detected transfer between Tyr (39) and Trp (125) in αS at the initiation of oligomerization and found a two-step decrease of the distance. Extensive studies of the ensemble properties of the αS molecule using multiple methods were reported by the Jovin and Subramaniam groups (63–66) who obtained a general view of the properties of this ensemble under various conditions by combination of FRET experiments and very sophisticated NMR experiments. Weak but significant conformational constraints were detected in particular possible interaction between the C terminal domain and the NAC and the N terminal domains. Such interactions might be important in the control of the stability of the soluble disordered monomer conformation.

The tau protein was studied intensively by Von Bergen and Mandelkow who used both labeling strategies (a) and (d) to monitor intermolecular transfer (67) during initiation of aggregation. A series of labeled FRET pairs spanning the tau molecule were created by inserting tryptophan (donor) and cysteine (labeled with IAEDANS as an acceptor). The observed distances do not fit the random coil model and both ends of the molecule were shown to approach one another. The results indicate that tau protein retains some global folding even in its "natively disordered" state, combined with high flexibility of the chain. Similar to the study of Grupi and Haas (33), this report emphasizes the importance of the FRET-based methods in detection of subtle deviations from full disorder which might be key to the understanding of the mechanism of amyloid transition (68).

3.6.2. Characterization of the Ensemble of Aβ Fragments

The ensemble of the amyloid β (Aβ) conformers (a protein fragment which is involved in the etiology of Alzheimer's disease) was studied by several groups using labeling strategy (c), since even the full-length molecule is manageable by current peptide synthesis methods.

Kim and Lee (27) applied labeling strategy (c) and intramolecular trFRET in a study of end-to-end distance distributions in a synthetic $A\beta_{11-25}$ fragment in the course of amyloid aggregation. The distribution of donor fluorescence lifetimes in a DA species showed that during the course of amyloid aggregation induced by a stepwise increase of unlabeled peptide concentration, the amyloid peptides undergo multistep conformational changes. The peptides begin in a random coil (monomers) conformation; when aggregation is induced, they first form a collapsed coil configuration (multimers), followed by a micellar structure, and finally an extended β-sheet in fibrils. The authors demonstrate that the end-to-end distance of micellar $A\beta_{11-25}$ becomes much shorter than that of

the collapsed coil or fibril structure. These results indicate that the monomeric Aβ fragment undergoes specific conformational transitions in the formation of micelles.

3.6.3. Characterization of Supramolecular Structures by Intermolecular FRET Measurements

The ordered structures of globular proteins are quite stable, unaffected by moderate changes of the solution conditions, and undergo cooperative transitions upon strong perturbations of the native conditions. This is not the case with IDPs and intrinsically disordered regions (IDRs), which exist as ensembles of multiple conformations with very similar energy; hence they are prone to major conformational shifts under mild changes of the solution conditions. Therefore, amyloid transitions and aggregation of IDPs under physiological conditions are very common. Intermolecular FRET-based methods can be very useful in investigation of this transition, and several methods were developed in order to address the task of deciphering the mechanism of formation of amyloid oligomers and fibrils.

Glabe et al. (30) used strategy (d) for preparation of single-labeled Aβ peptides, and applied steady-state detected intermolecular FRET measurements. Intermolecular FRET was found immediately after transition to folding conditions, suggesting that soluble Aβ(1–40) is also dimeric at more physiological conditions.

Chakrabartty and colleagues (69) applied labeling strategy (c) in a study of the mechanism of fibrillogenesis of Aβ(9–25) at physiological concentrations of peptides. Using the FRET-based fibrillogenesis assay, they found that fibril formation by Aβ(9–25) is reversible and cooperative. Two imidazole-carboxylate salt bridges per monomer were identified, and the kinetics of exchange of monomers with the fibrils was measured.

Bu and Callaway (70) applied labeling strategy (c) with Trp and Cys residues as sites of labeling of synthetic Aβ peptide and detected steady-state transfer efficiency in search for hairpin structures within the monomer. They concluded that beta sheets involving the essential Aβ(14–23) fragment strongly prefer the antiparallel topology. Gustiananda et al. (28) also used a similar approach (strategy (c) and steady-state detected intramolecular FRET) to study the disordered domain of the prion protein (PrP). An interesting approach was used in which the FRET experiments were simulated by molecular dynamic (MD) methods. Conformational distributions and average dansyl-Trp distances were obtained directly from MD simulation with excellent agreement with the experimental data. This study showed that the repeat sections of the PrP molecules are partially disordered but collapsed so that the end-to-end distances are shorter than expected for random coil peptides of the corresponding length. This study demonstrates the possibility of incorporating MD, emission intensities, and anisotropy measurements in the analysis of steady-state detected FRET measurements. It also demonstrates the limitations of the

steady-state FRET detection mode in studies of partially disordered polypeptides, where the mean transfer efficiency does not relate to any simple mean distance.

3.6.4. Studies of the Ensemble Properties of IDRs

Several reports of studies of applications of FRET methods in studies of disordered domains in otherwise ordered proteins demonstrate the advantages of FRET methods, when the common structure determination approaches cannot be applied.

Tapley et al. used intramolecular FRET to show that the partially unfolded character of active small chaperone molecule, HdeA, allows the chaperone to adopt different conformations as required for the recognition and high-affinity binding of different substrate proteins. Tapley, Banta, and colleagues used labeling strategy (f) and determination of transfer efficiency (using the acceptor emission) between two fluorescent proteins (CFP and YFP) attached to a disordered section (RTX domain) of the *Bordetella pertussis* adenylate cyclase. The apparent average end-to-end distance changes across the nine segment repeats of the RTX domain were monitored. This is a particularly low resolution application of the FRET method, but the main advantage of this labeling strategy is the potential of in vivo studies of the conformational states and transitions. An example of this direction was reported by Xu et al. who studied the intermolecular interactions of p53 in the cell (71). In these studies the importance of the partial disorder in control and multiplicity of functions was demonstrated.

A major challenge of deciphering the conformations and dynamics in the p53 molecule is being met by the Fersht group. p53 exists in solution as a large tetrameric complex with partly IDRs, an intrinsically disordered N-terminal domain (NTD, residues 1–94), and a C-terminal negative regulatory domain (360–393) (72). The structure of the two ordered domains was determined by the standard structure determination methods. The IDRs were studied by NMR, SAXS, and SM-detected FRET and ensemble trFRET-based experiments. Labeling strategy (c) combined with trFRET was applied in a study of four synthetic peptide fragments of the NTD. Huang et al. (73) took advantage of solid-phase synthesis and introduced a nonnatural residue, naphthyl-alanine, as a donor. This probe has a long fluorescence lifetime and almost monoexponential decay (in the absence of an acceptor) which enhances the accuracy of the determination of the parameters of the distance distribution and the intramolecular diffusion coefficient. All the peptides gave wide end-to-end distance distributions, supporting the hypothesis that the NTD is intrinsically disordered (73). In addition, all four peptides gave very small intramolecular diffusion coefficient values, which are much smaller than the reported values for very flexible peptides (22, 33). The reduced intramolecular diffusion coefficient, which is a measure of the rates of fast conformational fluctuations, shows

that intersegmental interactions probably impose mild structural constraints on these ensembles of multiple conformers. This work is a good example of the strength of single molecule and ensemble trFRET-based methods in dealing with the challenge of characterizing IDRs within otherwise ordered proteins.

Future Challenges   FRET-based methods have capabilities which are particularly suitable for IDP research. With the further development of this field, these techniques will acquire even more utility. A major future challenge is a study of how protein conformations within the cell modulate various functions. This would provide the missing link between protein dynamics and function. The main factor that limits the application of FRET experiments in the study of IDPs is the difficulty in preparing double-labeled protein samples. Every protein is a different case and a protocol that is successful with one protein might be unsuccessful with the next one (see Note 6).

4. Notes

1. The main uncertainties that should be eliminated in FRET-based experiments are (a) the possible bias of the Förster critical distance used in the calculation of a distance by incomplete averaging of the orientations of the probes and (b) multiexponential decay of the donor emission that can result from the distribution of either the distances or the orientations in trFRET experiments. Various approaches were proposed for the reduction of these uncertainties and the determination of the range of possible errors in distance determination. Dale et al. (74–76) reported the most extensive studies and developed methods for estimation of the range of the orientation factor, $<\kappa^2>_{min}$, and $<\kappa^2>_{max}$, using anisotropy measurements. Rapid reorientation during the lifetime of the excited state of the probes is reflected in the anisotropy measurements and therefore can be estimated experimentally for each preparation (for details of all the angles taken into account see Dale et al. (75, 76) and Van der Meer (6) pp. 55–83).

   An ideal case, where full averaging is an intrinsic characteristic of the system, is achieved with single-atom probes, such as lanthanides and chelated transition metal ions (5, 77–79). An alternative approach to this problem is based on the fact that many chromophores show mixed polarizations in their spectral transitions, i.e., their absorption and emission across the relevant spectral range of overlap involve a combination of two or more incoherent dipole moments (80). Haas et al. (79) showed that the occurrence of mixed polarizations

in the energy donor and acceptor may markedly limit the range of possible values that κ^2 can assume, and thus, under favorable conditions may alleviate the problem of the orientation factor in FRET experiments. This is the case for many common probes which have high levels of mixed polarization, such as aromatic probes and derivatives of naphthalene. This effect, when considered in planning FRET experiments by using probes that show a high degree of internal depolarization, can reduce the κ^2-related uncertainties to a range in which they do not exceed the other common experimental uncertainties. An improved approximation of the value of κ^2 is obtained when the contribution of the rotation of the whole molecule to the emission anisotropy of the probes is first subtracted from the observed anisotropy. This procedure can be performed when the anisotropy decay is determined and the analysis shows the relative amplitude associated with the decay component corresponding to the molecular rotational diffusion.

The uncertainty in r can be further reduced by employing two independent FRET experiments. In these experiments, different pairs of probes are used for FRET determination of the distance between a single pair of sites in a protein molecule.

2. The distance range for the statistically significant determination of distances for each pair is defined by the characteristics of the probes and is given by the R_o value. Equation 4 shows that the limits of that range are $r = R_o \pm 0.5\ R_o$ (Fig. 2). Maximal sensitivity of the FRET effect to small changes in r is observed when $r = R_o$. Therefore optimal design of FRET experiments includes the selection of a pair of probes that have an R_o value close to the expected distance between the labeled sites.

Modulation of the value of R_o of any specific pair in order to adjust it to the expected r value is limited to a narrow range due to the sixth power relationship between R_o and the spectroscopic constants of the probes. Equation 1 shows that the donor quantum yield Φ_D^o and the overlap integral parameters are the most readily available for manipulation to modulate the R_o value.

The Donor Quantum Yield Φ_D^o. Φ_D^o should be determined for each probe when attached to the protein molecule at the same site used for the FRET experiment. Many experiments showed that the Φ_D^o is affected by the site of attachment of the probes, by modifications at other sites in the protein molecule, by attachment of an acceptor at second sites, by changes of solvent conditions, and by ligand binding or conformational changes. Therefore R_o should be determined repeatedly after any such changes in the conditions to avoid misinterpretation of changes of R_o as changes of r. R_o depends on the sixth root of Φ_D^o and is

therefore not very sensitive to uncertainties in Φ_D^o and its manipulation is quite limited. For instance, a twofold decrease of Φ_D^o would result in ~9 % decrease of R_o. Increasing Φ_D^o while maintaining the same probes usually requires change of solvent and therefore is not very useful in IDP studies; however, Φ_D^o can be decreased by adding quenching solute without a change of solvent.

Bound donor probes may have several modes of interactions with the macromolecular environment in the ground or in the electronically excited state. In such a case, each subpopulation which has a different lifetime also has a different R_o, as indicated in Eqs. 1–2.

The Spectral Overlap Integral, J. The most effective variation of R_o is available by selecting probes according to the overlap of their spectra. The value of J depends on the local environment of the probes. As a rule, absorption spectra vary to only a limited extent with a change of solvent, temperature, or local environment, but emission spectra may be very sensitive to such changes of the local environment of the probes. An example of this dependence is the fluorescence of indole derivatives (81). The selection of probes with narrow emission and absorption spectra and with high extinction coefficients of the acceptor can yield high R_o values.

3. Practical implementation of the FRET approach in characterization of IDP ensembles either in the ensemble or the single molecule modes involves three very different steps:

   1. Design and preparation of pure, homogenous, structurally and functionally intact double-labeled protein samples

   2. Spectroscopic determination of transfer efficiencies and fluorescence decay profiles of the probes and control experiments

   3. Data analysis
      A minimum of two labeled protein species should be prepared for each well-controlled trFRET experiment. These are (a) a site-specifically double-labeled (by both donor and acceptor) species and (b) a single donor-labeled species (at the same site as in (a)). The best results can be obtained when the following requirements are satisfied: (a) the mean distance between the two selected sites should be close to the R_o distance of the pair of probes used; (b) the ensemble characteristics of the labeled protein derivatives should be minimally perturbed; (c) the probes should be of small size and with minimal potential interactions with the backbone or the side chains; and (d) the probes should have ideal spectroscopic characteristics to enable detection and data analysis free of correction factors. Proteins are rarely totally unaffected by chemical or mutational modifications

and the spectroscopically ideal probes are usually also very large. Therefore, the design process is a matter of optimization, not maximization.

The extent of perturbation is an important factor in the evaluation of the FRET results. Occasionally, modified protein molecules are found to be unsuitable for planned experiments due to very strong perturbations. Consequently, an alternative modification design should be attempted. The concern for purity cannot be overemphasized, since background fluorescence or fluorescence of impurities can drastically distort the results of distance determinations.

Incomplete labeling of the reference single-labeled mutant in each set should not prevent the use of the preparation in FRET experiments, provided that the extent of labeling is known. Even incomplete labeling of the double-labeled mutant can be tolerated in ensemble steady-state or time-resolved detected experiments, provided that the extent of labeling is known, and (in some cases) the partially labeled fractions do not exceed ~25 % of the total population of labeled molecules. The strict requirements for complete labeling and accurate knowledge of the extent of labeling are relaxed in single molecule FRET experiments. In these measurements, it is possible to discard records of single molecule emission intensity that do not show both donor and acceptor emission during data processing.

4. For optimal accuracy of distance determination in disordered proteins by steady-state or time-resolved FRET, the following spectral characteristics of the probes are desired: (a) R_o of the selected pair should be as close as possible to the expected mean distance between the two labeled sites (including the linker of the probes); (b) the orientation of the probes should be averaged during the lifetime of the excited state of the donor; (c) for time-resolved ensemble experiments, the ratio should be $\tau_D/\tau_A \geq 3$; (d) photostability; (e) minimal environmental sensitivity of the spectral characteristics of the probes; (f) large Stokes shifts of the acceptor (to achieve minimal overlap of the emission spectra of the donor and the acceptor); and (g) a high extinction coefficient to enable measurements of dilute solutions (where aggregation problems are minimized).

5. Possible pitfalls associated with the preparation of labeled protein samples for fret folding experiments

   (a) Incomplete specificity of the labeling might prohibit meaningful interpretation of FRET data, particularly for flexible systems. Distributions of labeling reactions (e.g., mixtures of species with different distances between the labeled sites) might, when used in FRET experiment, be thought to represent distributions of a specific distance. In principle, under

conditions where the extent of mislabeling is known, it should be possible to adjust the model in the analysis algorithm and account for subpopulations of a distance distribution and relate them to the labeled species.

(b) Impurities: Two types of impurities are of concern: the first is the presence of fluorescent impurities and background emission, and the second is incomplete labeling. As long as the background emission is not too high (ca. <20 % of the signal), and the emission can be determined independently (control experiments), reliable distance determination is possible by careful subtraction of the background emission. Incomplete labeling of the single-labeled samples should not interfere with analysis of trFRET data, provided that the extent of the partially labeled fraction is independently known ([82]) and does not exceed ca. 20 % of the population.

(c) Conformational bias due to the labeling procedure: The ensembles of conformers of disordered proteins are particularly sensitive to effects of solvent ([83]) and chemical or mutational modifications.

Structural perturbations can be minimized by the proper choice of labeling sites and probes. Rigorous control experiments for determination of the extent of perturbation due to the modifications are mandatory. Control experiments can include the determination of biological activities, aggregation curves, and detection of structural properties using CD, IR, and other spectral characteristics.

6. *Expression and purification of αS mutants.* αS was expressed from a pT7–7 plasmid encoding the WT protein. All tryptophan- and cysteine-containing single and double mutants were introduced using the QuikChange site-directed mutagenesis kit (Stratagene), and were confirmed by DNA sequencing. BL21(DE3)-competent cells (Stratagene) were grown in LB in the presence of ampicillin (0.1 mg/ml). Cells were induced with IPTG when cultures reached an OD_{600} of 0.6–0.7, cultured at 37 °C for 4 h and harvested by centrifugation at $6,000 \times g$. The cell pellet was resuspended in 10 mM Tris–HCl, 2 mM EDTA, pH 8.0, stirred for 20 min on ice, and sonicated. DNA was sedimented by adding Streptomycin sulfate (Sigma) to a final concentration of 10 mg/ml, and the mixture was stirred for 10 min on ice. After centrifugation at $13,000 \times g$, the supernatant was collected and proteins were precipitated by ammonium sulfate at 300 mg/ml. The solution was stirred for 30 min on ice and centrifuged again at $13,000 \times g$. The pellet was resuspended in 10 mM Tris–HCl, pH 7.4, 2 mM DTT and desalted using Centricon columns (Millipore, 10 kDa MWCO). The solution was loaded on a

200 ml DEAE-Cellulose column (Sigma) and αS was eluted at ~300 mM NaCl with a salt gradient from 0 to 500 mM NaCl. αS concentration was determined by UV absorption at 280 nm using an extinction coefficient of 1,490 L/mol cm for each Tyrosine and 5,500 L/mol cm for Tryptophan. The molecular weight of the recombinant αS was verified by electrospray ionization mass spectroscopy (ESI-MS). Protein samples were stored at $-20\,^{\circ}$C.

Thawed protein samples were incubated with DTT for 15 min. Buffer was exchanged on a Sephadex G-75 column equilibrated with 50 mM Hepes buffer pH 7.3 containing 2 mM EDTA, or by dialysis. The Cys mutants (~2 mg/ml) were labeled with I-Cou or IAA in polystyrene Falcon tubes at a 5:1 molar ratio of label to protein, in the presence of 2 M Guanidine-HCl, and 50 μM TCEP. Typically, the reaction solution was incubated for 2 h at room temperature in the dark, under mild agitation, at the end of which, DTT was added in order to block excess dye molecules. Nonspecific labeling of positively charged residues was tested by a control labeling reaction of a WT protein. The labeled samples were dialyzed against 20 mM Tris–HCl, 2 mM EDTA, pH 8.

Samples were purified on an RP-HPLC c18 column (Vydac). The pH of the collected samples was immediately neutralized by addition of a concentrated solution of Tris–HCl buffer (pH 8) followed by dialysis against 50 mM Tris–HCl, 2 mM EDTA, pH 8 at $40\,^{\circ}$C. Sample purity was analyzed by RP-HPLC c18 column and by SEC-HPLC. Size exclusion was preformed with a G2000swXL column (Tosohaas), calibrated with globular protein standards (Aldolase, 161 kDa; Bovine serum albumin (BSA), 66.4 kDa; Albumin, 42.7 kDa; Chymotrypsin, 25.6 kDa; Trypsin inhibitor, 20.5 kDa; Lysozyme, 14.4 kDa; and pABA, 0.37 kDa). The running buffer was 50 mM Tris–HCl, 300 mM NaCl, pH 7 at a flow rate of 1 ml/min.

In order to accurately determine the extinction coefficient of αS mutants labeled with I-Cou, we prepared a sample of L-cysteine (Sigma 30129) labeled with I-Cou. The $\varepsilon 280$ of the dye was calculated at 280 nm from the known extinction coefficient at 332 nm ($\varepsilon 332nm = 18,000$ L/mol cm): $\varepsilon 280 = Abs332/Abs280 \times \varepsilon 332nm = 6,475$ L/mol cm. From this value, the total $\varepsilon 280$ of DA and AO protein samples labeled with I-Cou was estimated at 17,525 L/mol cm and 12,070 L/mol cm, respectively. The Abs332/Abs280nm ratio of a single-labeled DA and an AO-labeled protein was 1 and 1.5, respectively, and served as an additional confirmation for labeling at a 1:1 protein:dye ratio.

References

1. Forster TH (1948) Zwischen molekulare energie wanderung und fluoreszenz. Ann Phys (Leipzig) 2:55–75

2. Forster TH (1959) Transfer mechanisms of electroinc excitation. Discuss Fraday Soc 27:7–17

3. Forster TH (1965) Delocalized excitation and excitation transfer. In: Sinaonglu O (ed) Modern quantum chemistry, Istnabul lectures part III: action of light and organic crystals. Academic, New York, pp 93–137

4. Steinberg IZ (1971) Long-range nonradiative transfer of electronic excitation energy in proteins and polypeptides. Annu Rev Biochem 40:83–114

5. Stryer L, Thomas DD, Meares CF (1982) Diffusion-enhanced fluorescence energy transfer. Annu Rev Biophys Bioeng 11:203–222

6. Van Der Meer WB, Coker G III, Chen SYS (1994) Resonance energy transfer theory and data. VCH, New York

7. Haas E (2004) Fluorescence resonance energy transfer (FRET) and single molecule fluorescence detection studies of the mechanism of protein folding and unfolding. In: Kiefhaber T, Buchner J (eds) Protein folding handbook. Part I. Wiley-VCH Verlag GmbH & Co.KGaA, Weinheim, pp 573–633

8. Haas E, Steinberg IZ (1984) Intramolecular dynamics of chain molecules monitored by fluctuations in efficiency of excitation energy transfer. A theoretical study. Biophys J 46:429–437

9. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc Natl Acad Sci U S A 104:2655–2660

10. Perrin J (1927) Fluorescence et induction moleculaire par sesonance. C R Hebd Seances Acad Sci 184:1097–1100

11. Weber G, Teale FJW (1959) Electronic energy transfer in haem proteins. Discuss Faraday Soc 27:134–151

12. Latt SA, Cheung HT, Blout ER (1965) Energy transfer. A system with relatively fixed donor-acceptor separation. J Am Chem Soc 87:995–1003

13. Stryer L, Haugland RP (1967) Energy transfer: a spectroscopic ruler. Proc Natl Acad Sci USA 58:719–726

14. Valeur B (2002) Molecular fluorescence principles and applications. Wiley-VCH Verlag GmbH, Weinheim

15. Lakowicz JR (1999) Principles of fluorescence spectroscopy, 2nd edn. Kluwer Academic/Plenum, New York

16. Stryer L, Thomas DD, Carlsen WF (1982) Fluorescence energy transfer measurements of distances in rhodopsin and the purple membrane protein. Methods Enzymol 81:668–678

17. Wu P, Brand L (1994) Resonance energy transfer: methods and applications. Anal Biochem 218:1–13

18. Clegg RM (1992) Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol 211:353–388

19. Clegg RM (1996) Fluorescence resonance energy transfer. In: Wang XF, Brian H (eds) Fluorescence imaging spectroscopy and microscopy. Wiley, New York, pp 179–252

20. Edidin M (2003) Fluorescence resonance energy transfer: techniques for measuring molecular conformation and molecular proximity. Curr Protoc Immunol Chapter 18:Unit 18.10.

21. Cantor CR, Pechukas P (1971) Determination of distance distribution functions by singlet-singlet energy transfer. Proc Natl Acad Sci U S A 68:2099–2101

22. Haas E, Katchalsky-Katzir E, Steinberg IZ (1978) Brownian motion of the ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. Biopolymers 17:11–31

23. Steinberg IZ, Haas E, Katchalsky-Katzir E (1983) Long-range non-radiative transfer of electronic excitation energy. In: Cundall RB, Dale RE (eds) Time resolved spectroscopy in biochemistry. Plenum Publishing Corporation, New York, pp 411–451

24. Beechem JM, Haas E (1989) Simultaneous determination of intramolecular distance distributions and conformational dynamics by global analysis of energy transfer measurements. Biophys J 55:1225–1236

25. Lakowicz JR, Kusba J, Wiczk W, Gryczynski I, Szmacinski H, Johnson ML (1991) Resolution of the conformational distribution and dynamics of a flexible molecule using frequency-domain fluorometry. Biophys Chem 39:79–84

26. Chandrasekhar S (1943) Stochastic problems in physics and astronomy. Rev Mod Phys 15:1–89

27. Kim J, Lee M (2004) Observation of multi-step conformation switching in beta-amyloid peptide aggregation by fluorescence resonance energy transfer. Biochem Biophys Res Commun 316:393–397

28. Gustiananda M, Liggins JR, Cummins PL, Gready JE (2004) Conformation of prion protein repeat peptides probed by FRET measurements and molecular dynamics simulations. Biophys J 86:2467–2483

29. Kim Y, Ho SO, Gassman NR, Korlann Y, Landorf EV, Collart FR, Weiss S (2008) Efficient site-specific labeling of proteins via cysteines. Bioconjug Chem 19:786–791

30. Garzon-Rodriguez W, Sepulveda-Becerra M, Milton S, Glabe CG (1997) Soluble amyloid Abeta-(1–40) exists as a stable dimer at low concentrations. J Biol Chem 272:21037–21044

31. Navon A, Ittah V, Landsman P, Scheraga HA, Haas E (2001) Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. Biochemistry 40:105–118

32. Lakowicz JR, Kusba J, Szmacinski H, Gryczynski I, Eis PS, Wiczk W, Johnson ML (1991) Resolution of end-to-end diffusion coefficients and distance distributions of flexible molecules using fluorescent donor-acceptor and donor-quencher pairs. Biopolymers 31:1363–1378

33. Grupi A, Haas E (2011) Segmental conformational disorder and dynamics in the intrinsically disordered protein alpha-synuclein and its chain length dependence. J Mol Biol 405:1267–1283

34. Lee JC, Langen R, Hummel PA, Gray HB, Winkler JR (2004) Alpha-synuclein structures from fluorescence energy-transfer kinetics: implications for the role of the protein in Parkinson disease. Proc Natl Acad Sci USA 101:16466–16471

35. Liu DR, Magliery TJ, Pastrnak M, Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc Natl Acad Sci USA 94:10092–10097

36. Buckler DR, Haas E, Scheraga HA (1993) C-terminal labeling of ribonuclease A with an extrinsic fluorescent probe by carboxypeptidase Y-catalyzed transpeptidation in the presence of urea. Anal Biochem 209:20–31

37. Szilvay GR, Blenner MA, Shur O, Cropek DM, Banta S (2009) A FRET-based method for probing the conformational behavior of an intrinsically disordered repeat domain from Bordetella pertussis adenylate cyclase. Biochemistry 48:11273–11282

38. Amir D, Haas E (1987) Estimation of intramolecular distance distributions in bovine pancreatic trypsin inhibitor by site-specific labeling and nonradiative excitation energy-transfer measurements. Biochemistry 26:2162–2175

39. Santoro SW, Anderson JC, Lakshman V, Schultz PG (2003) An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. Nucleic Acids Res 31:6700–6709

40. Takashi R (1988) A novel actin label: a fluorescent probe at glutamine-41 and its consequences. Biochemistry 27:938–943

41. Folk JE, Chung SI (1985) Transglutaminases. Methods Enzymol 113:358–375

42. Fink ML, Chung SI, Folk JE (1980) Gamma-glutamylamine cyclotransferase: specificity toward epsilon-(L-gamma-glutamyl)-L-lysine and related compounds. Proc Natl Acad Sci U S A 77:4564–4568

43. Lymperopoulos K, Kiel A, Seefeld A, Stohr K, Herten DP (2010) Fluorescent probes and delivery methods for single-molecule experiments. Chemphyschem 11:43–53

44. Grupi A, Haas E (2011) Time-resolved FRET detection of subtle temperature-induced conformational biases in ensembles of alpha-synuclein molecules. J Mol Biol 411 (1):234–47

45. Jacob MH, Amir D, Ratner V, Gussakowsky E, Haas E (2005) Predicting reactivities of protein surface cysteines as part of a strategy for selective multiple labeling. Biochemistry 44:13664–13672

46. Jager M, Michalet X, Weiss S (2005) Protein-protein interactions as a tool for site-specific labeling of proteins. Protein Sci 14:2059–2068

47. Erbse AH, Berlinberg AJ, Cheung CY, Leung WY, Falke JJ (2011) OS-FRET: a new one-sample method for improved FRET measurements. Biochemistry 50:451–457

48. Epe B, Woolley P, Steinhauser KG, Littlechild J (1982) Distance measurement by energy transfer: the 3′ end of 16-S RNA and proteins S4 and S17 of the ribosome of Escherichia coli. Eur J Biochem 129:211–219

49. Flamion PJ, Cachia C, Schreiber JP (1992) Non-linear least-squares methods applied to the analysis of fluorescence energy transfer measurements. J Biochem Biophys Methods 24:1–13

50. Brand L (1992) Numerical computer methods. In: Brand L, Johnson ML (eds) Methods in enzymology. Academic, San Diego, pp 636–675

51. Grinvald A, Steinberg IZ (1974) On the analysis of fluorescence decay kinetics by the method of least-squares. Anal Biochem 59:583–598

52. Brochon JC (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. Methods Enzymol 240:262–311

53. Lakshmikanth GS, Sridevi K, Krishnamoorthy G, Udgaonkar JB (2001) Structure is lost incrementally during the unfolding of barstar. Nat Struct Biol 8:799–804

54. Beechem JM, Brand L (1985) Time-resolved fluorescence of proteins. Annu Rev Biochem 54:43–71

55. Ameloot M, Beechem JM, Brand L (1986) Simultaneous analysis of multiple fluorescence decay curves by Laplace transforms. Deconvolution with reference or excitation profiles. Biophys Chem 23:155–171

56. Lanczos C (1956) Applied analysis. Prentice Hall Inc., Engelwood Cliffs, pp 272–304

57. Bevington PR (1969) Data reduction and error analysis for the physical sciences. McGraw-Hill, New York

58. Ames WF (1977) Numerical methods for partial differential equations, 2nd edn. Academic, New York

59. Press WH, Flannery BP, Teukolski SA, Vetterling WF (1989) Numerical recipes: the Art of scientific computing. Cambridge University Press, Cambridge

60. Edelhoch H, Brand L, Wilcheck M (1963) Fluorescence studies of tryptophyl peptides. Isr J Chem 1:1–2

61. Edelhoch H, Brand L, Wilchek M (1967) Fluorescence studies with tryptophyl peptides. Biochemistry 6:547–559

62. Kaylor J, Bodner N, Edridge S, Yamin G, Hong DP, Fink AL (2005) Characterization of oligomeric intermediates in alpha-synuclein fibrillation: FRET studies of Y125W/Y133F/Y136F alpha-synuclein. J Mol Biol 353:357–372

63. Rasia RM, Bertoncini CW, Marsh D, Hoyer W, Cherny D, Zweckstetter M, Griesinger C, Jovin TM, Fernandez CO (2005) Structural characterization of copper(II) binding to alpha-synuclein: insights into the bioinorganic chemistry of Parkinson disease. Proc Natl Acad Sci U S A 102:4294–4299

64. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M (2005) Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci U S A 102:1430–1435

65. Celej MS, Caarls W, Demchenko AP, Jovin TM (2009) A triple-emission fluorescent probe reveals distinctive amyloid fibrillar polymorphism of wild-type alpha-synuclein and its familial Parkinson disease mutants. Biochemistry 48:7465–7472

66. Veldhuis G, Segers-Nolten I, Ferlemann E, Subramaniam V (2009) Single-molecule FRET reveals structural heterogeneity of SDS-bound alpha-synuclein. Chembiochem 10:436–439

67. von Bergen M, Li L, Mandelkow E (2005) Intrinsic fluorescent detection of tau conformation and aggregation. Methods Mol Biol 299:175–184

68. Jeganathan S, von Bergen M, Brutlach H, Steinhoff HJ, Mandelkow E (2006) Global hairpin folding of tau in solution. Biochemistry 45:2283–2293

69. Huang TH, Fraser PE, Chakrabartty A (1997) Fibrillogenesis of Alzheimer Abeta peptides studied by fluorescence energy transfer. J Mol Biol 269:214–224

70. Shi Y, Stouten PF, Pillalamarri N, Barile L, Rosal RV, Teichberg S, Bu Z, Callaway DJ (2006) Quantitative determination of the topological propensities of amyloidogenic peptides. Biophys Chem 120:55–61

71. Xu H, Tai J, Ye H, Kang CB, Yoon HS (2006) The N-terminal domain of tumor suppressor p53 is involved in the molecular interaction with the anti-apoptotic protein Bcl-Xl. Biochem Biophys Res Commun 341:938–944

72. Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J (2003) The N-terminal domain of p53 is natively unfolded. J Mol Biol 332:1131–1141

73. Huang F, Rajagopalan S, Settanni G, Marsh RJ, Armoogum DA, Nicolaou N, Bain AJ, Lerner E, Haas E, Ying L, Fersht AR (2009) Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. Proc Natl Acad Sci USA 106:20758–20763

74. Eisinger J, Dale RE (1974) Letter: interpretation of intramolecular energy transfer experiments. J Mol Biol 84:643–647

75. Dale RE, Eisinger J (1976) Intramolecular energy transfer and molecular conformation. Proc Natl Acad Sci U S A 73:271–273

76. Dale RE, Eisinger J, Blumberg WE (1979) The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. Biophys J 26:161–193

77. Selvin PR, Hearst JE (1994) Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer. Proc Natl Acad Sci U S A 91:10024–10028

78. Mersol JV, Wang H, Gafni A, Steel DG (1992) Consideration of dipole orientation angles yields

accurate rate equations for energy transfer in the rapid diffusion limit. Biophys J 61:1647–1655

79. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Effect of the orientation of donor and acceptor on the probability of energy transfer involving electronic transitions of mixed polarization. Biochemistry 17:5064–5070

80. Albrecht AC (1960) Forbidden characterizing allowed electronic transitions. J Chem Phys 33:156

81. Cortijo M, Steinberg IZ, Shaltiel S (1971) Fluorescence of glycogen phosphorylase b.

Structural transitions and energy transfer. J Biol Chem 246:933–938

82. Mansoor SE, Farrens DL (2004) High-throughput protein structural analysis using site-directed fluorescence labeling and the bimane derivative (2-pyridyl)dithiobimane. Biochemistry 43:9426–9438

83. McNulty BC, Tripathy A, Young GB, Charlton LM, Orans J, Pielak GJ (2006) Temperature-induced reversible conformational change in the first 100 residues of alpha-synuclein. Protein Sci 15:602–608

Chapter 29

Fluorescence Correlation Spectroscopy to Determine the Diffusion Coefficient of α-Synuclein and Follow Early Oligomer Formation

Sangeeta Nath, Manli Deng, and Yves Engelborghs

Abstract

Fluorescence correlation spectroscopy (FCS) can be used to determine the diffusion coefficient of fluorescently labeled α-synuclein. It is a technique based on the use of a confocal microscope. By applying FCS in a combination of short sampling times and repeated measurements, the disappearance of individual α-synuclein molecules (called monomers) and the formation of oligomers can be characterized during the early aggregation process.

Key words: α-Synuclein, Fluorescence correlation spectroscopy, Early intermediates, Fluorescence fluctuations, Heterogeneity in solution, Labeling, Alexa488, Maleimide, Heterogeneity

1. Introduction

Confocal microscopy not only reveals heterogeneity and structure, when scanning biological samples, but can also be used to study dynamics in the living cell, by analyzing the intensity fluctuations that appear in time, when fluorescent molecules are diffusing in and out of the confocal volume. This application of confocal microscopy is called fluorescence correlation spectroscopy (FCS). The method also reveals heterogeneity in solution by a distribution of diffusion coefficients and the formation of spikes when bright objects are formed. This analysis of heterogeneity in time can be quantified to study protein–DNA interaction (1) or complex formation (2). Here we study the diffusion of α-synuclein and the formation of early oligomers. The idea is to use trace amounts (nM) of labeled protein in the presence of a large excess of unlabeled protein and follow the aggregation process by measuring the reduction in time of the diffusion coefficient of the fluorescent species.

α-Synuclein with an engineered cysteine (A140C) was labeled with Alexa488 and was used as a fluorescent probe in trace amounts (3–4 nM) in the presence of 100 μM unlabeled α-synuclein. The combination of short sampling times and repeated measurements produce a size distribution of the oligomers. Initially, a very sharp distribution is obtained (diffusion coefficient 114 ± 15 μm²/s) probably corresponding to monomers (3). Subsequently a distinct transient population appears, followed by the gradual formation of broader sized distributions of higher oligomers. At longer time scales, further aggregation leads to the formation of big aggregates that are moving too slow to contribute to the short-term fluctuations. The overall aggregation process can then be followed by the decrease of the number concentration of fast diffusing fluorescent species.

2. Materials

1. Wild-type α-synuclein (140aa) was purified according to the protocol described in Hoyer et al. (4). The A140C mutant of α-synuclein was purified similarly to wild type, but in the presence of 1 mM DTT in the buffers. Salt and DTT was removed from the purified α-synuclein using a HiPrep™ 26/10 desalting column equilibrated with 20 mM Hepes buffer of pH 7.4. After desalting, samples were aliquoted (usually 300 μL with concentration between 100 and 200 μM) and flash frozen with liquid nitrogen and stored immediately at −80 °C.

2. Buffers required. 20 mM HEPES buffer pH 7.4; 20 mM Hepes buffer, 150 mM NaCl, pH 7.4. Prepare all buffers and aqueous solvents using ultrapure deionized water. Buffers and solvents should be filtered (pore-size 0.2 μm) and degassed directly before use.

3. Alexa488 C5 maleimide. Alexa488 is a well-standardized fluorescent probe with excitation maximum at 493 nm and emission maximum at 519 nm. It is quite photo stable and independent on pH between pH 4 and 10 (Molecular Probes, Invitrogen). Dry Alexa488 C5 maleimide (1 mg) was dissolved in 70 μL ultra-dry DMSO to obtain a 20 mM stock solution. Its concentration can be determined on the basis of its molar extinction coefficient of 72,000 at 493 nm (Invitrogen).

4. TCEP (Tris-(2-carboxyethyl)phosphine). Dry TCEP is dissolved in HEPES buffer to obtain a 20 mM solution. The solution is stored at −20 °C.

5. FCS measurements were performed on a commercial laser scanning microscope (LSM 510) equipped with an FCS unit,

i.e., the ConfoCor II combination (Carl Zeiss, Jena, Germany). A confocal microscope has to be equipped with a fast and sensitive detector, e.g., a photomultiplier or an avalanche photodiode. The fluctuating signal has to be stored with a short binning time, e.g., 50 ns. Autocorrelation analysis can be done by hardware correlators, or by software correlation. The latter is preferred because in this way the original data are stored and available for rebinning.

3. Methods

3.1. Labeling of Purified Protein Using the Fluorescent Probe

1. A sample of α-synuclein from the $-80\,°C$ store was left to thaw at room temperature for 20 min.

2. A140C-synuclein (100 μM, 100 μL) was first treated with TCEP by adding 5 μL from the TCEP stock solution. It was then labeled with a two- to fivefold molar excess of Alexa488 C5 maleimide (1–2.5 μL of stock solution) in 20 mM HEPES buffer of pH 7.4. If TCEP is used to regenerate the thiol of A140C-synuclein, it is not necessary to remove it before labeling (see Note 1).

3. Upon the addition of Alexa488 C5 maleimide stock solution, the protein solution was immediately vortexed to avoid local effects of the hydrophobic solvent. The mixture was incubated for 1 h at 20 $°C$ and vortexed (1,800 rpm, 5–10 s) twice during the incubation (see Note 2).

4. Removal of excess label. PD-10 columns were pre-equilibrated with HEPES pH 7.4 and labeled protein was eluted with the same buffer and separated from excess-free dye and TCEP. For samples up to 500 μL, two subsequent PD Minitrap G-25 are used, and for samples up to 100 μL two PD spintrap G-25, with column volumes of 2.1 and 0.5 mL, respectively (see Note 3).

5. The concentration of labeled-synuclein was calculated from its spectrum by correcting for the contribution of the dyes to the absorbance at A_{277}.

$$A_{277(\text{protein})} = A_{277} - A_{277(\text{Alexa})}.$$

$$A_{277(\text{Alexa})} = A_{493(\text{Alexa})} \frac{\varepsilon_{277(\text{Alexa})}}{\varepsilon_{493(\text{Alexa})}}.$$

The following molar extinction coefficients were used: α-synuclein $\varepsilon_{277(\text{synuclein})} = 5{,}800$ M^{-1} cm^{-1} and Alexa488 $\varepsilon_{493} = 72{,}000$ $M^{-1}cm^{-1}$. The ratio $\varepsilon_{277}/\varepsilon_{493}$ of Alexa $= 0.125$ as obtained from the spectrum of pure Alexa488 C5 maleimide (it is assumed that the spectrum at 277 nm does not change upon reaction). (The stoichiometry obtained in this way may be overestimated because there may be some non-covalently bound dye present, see further.)

3.2. Fluorescence Correlation Spectroscopy

1. FCS measurements were performed on the ConfoCor II. The instrument was calibrated with Alexa488-carboxylic acid–succinimidyl ester for which the diffusion coefficient ($435 \; \mu m^2/s$ at 22.5 °C) was determined with high accuracy by scanning FCS (5) (see Note 4). This calibration allows to calculate the dimensions of the confocal volume ($\omega_x = \omega_y$ and ω_z and calculate S, see further).

2. FCS measurements on α-synuclein were done at a concentration of 3–4 nM labeled protein mixed with different concentration of unlabeled protein, to ensure that the number of fluorescent molecules in the confocal volume (0.312 fL) is limited to 1–4.

3. The autocorrelation curves can be fitted with the following relation, using the general fitting programme Origin:

$$G(\tau) = 1 + G_T(\tau) \times G_D(\tau).$$

$$G_T(\tau) = \left(1 + \frac{T_R e^{-\tau/\tau_x}}{1 - T_R}\right).$$

$$G_D(\tau) = \left(\frac{1}{N}\right)\left\{\left(\frac{F_1}{1 + \tau/\tau_{D1}}\right)\left(\frac{1}{1 + S^2(\tau/\tau_{D1})}\right)^{1/2}\right.$$
$$\left. + \left(\frac{1 - F_1}{1 + \tau/\tau_{D2}}\right)\left(\frac{1}{1 + S^2(\tau/\tau_{D2})}\right)^{1/2}\right\}.$$

$$S = \frac{\omega_x}{\omega_z} \text{ and } D = \frac{\omega_x^2}{4\tau_D}.$$

$G(\tau)$ is the global autocorrelation curve, τ is the correlation time, $G_T(\tau)$ is the part of the autocorrelation curve at the fast timescale, representing different photo-physical processes, e.g., triplet formation, T_R and τ_x are the amplitude and the relaxation time of the photo-physical processes, $G_D(\tau)$ is the part representing diffusion, N is the average number of fluorescent particles in the confocal volume, τ_{D1} (F_1) and $\tau_{D2}(1 - F_1)$ are, respectively, the diffusion time (fraction) of free Alexa dye and dye bound to α-synuclein, S is the structure parameter, ω_x and ω_z are the radial and axial radii of the confocal volume, and D is the diffusion coefficient of the labeled species.

The FCS measurements often reveal the presence of a small fraction (<20 %) of free dye (see Note 3). A typical autocorrelation curve of labeled α-synuclein with its fitting is seen in Fig. 1.

3.3. Oligomers Formation of α-Synuclein

1. Oligomers are formed when nM concentrations of labeled protein are mixed with unlabeled α-synuclein at micromolar concentrations, at 45 °C in 20 mM Hepes buffer, pH 7.4 and 150 mM NaCl with constant magnetic stirring (500 rpm) or shaking.

2. Prior to the experiments, the unlabeled α-synuclein, after taken out from −80 °C stock and thawing, was centrifuged for 30 min
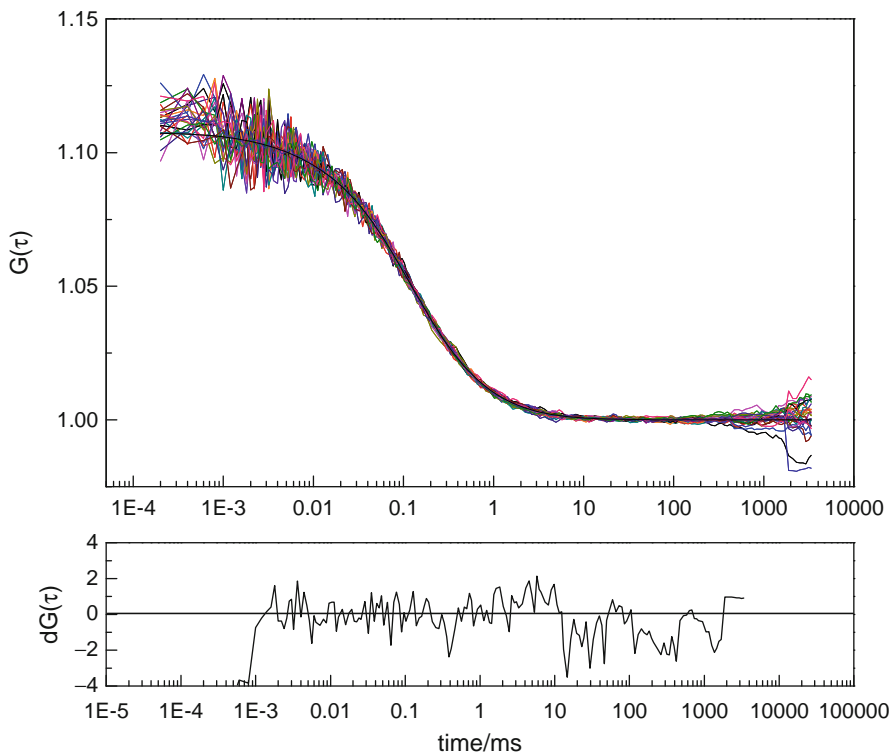
Fig. 1. Typical autocorrelation curves of labeled α-synuclein at nM concentrations. Twenty-five curves are shown and the fit (*dark line*) to their average. The fit determines two components: 5 % of free dye and 95 % protein with $D = 90$ μm^2/s. (The information about the triplet state, i.e., 3.8 % amplitude and a triplet lifetime of 4 μs is obtained from the autocorrelation curve of the free label and used as such here.) Residuals of the fit are shown below (1,000× enlarged). The obtained $\chi^2 = 1.012$.

at 13,400 rpm (18,500 × *g*) in a Galaxy 14D centrifuge (Sorvall) to remove any existing oligomers and fibrils.

3. Turbidity was measured at 350 nm at different time point to follow the kinetics of the oligomer formation.

4. The process of oligomer formation can also be followed by FCS. The FCS measurements can be done during a short sampling time (e.g., 10 s) and repeatedly to produce a distribution of diffusion coefficients. When applied to a solution of labeled A140C-Alexa488-Synuclein mixed with unlabeled protein, initially, a sharp peak is obtained (diffusion coefficient 114 ± 15 μm^2/s) probably corresponding to monomers. With increasing time, a distinct transient population appears, followed by the gradual formation of broader sized distributions of higher oligomers (see Fig. 2). It should be noted that this distribution reflects the heterogeneity of the solution, but cannot be assumed to reflect correctly the distribution of individual species.

5. At a later stage the large aggregates started to precipitate on the glass slide. Initially however, when the diffusion coefficients are

Aggregation of a-SYN at 100μM conc and 45°C
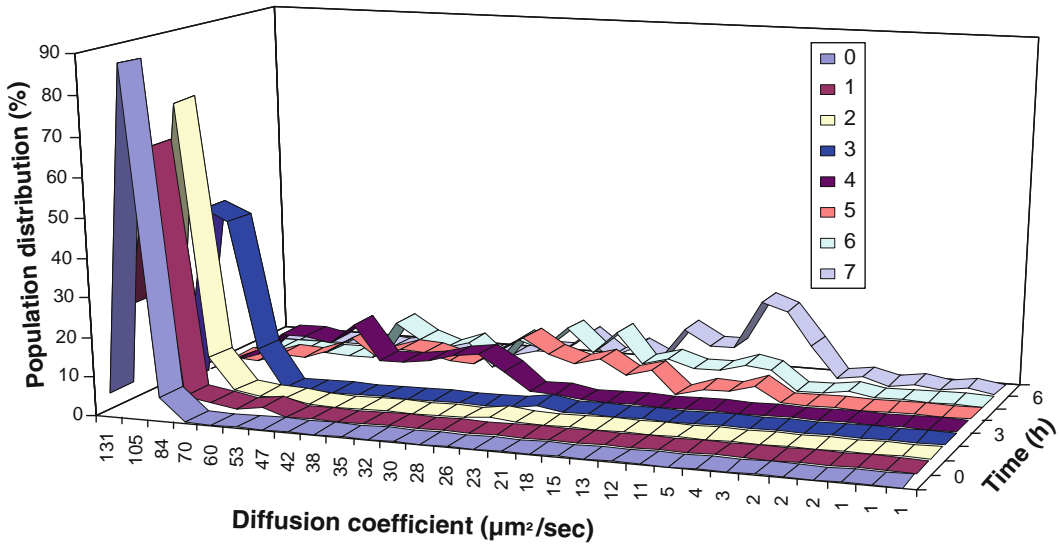
Fig. 2. Evolution of the measured diffusion coefficients as a function of aggregation time. Unlabeled α-synuclein (at 100 μM) was mixed with nM labeled synuclein and incubated at 45 °C with continuous stirring. Repetitive FCS measurements were performed and the distribution of the measured diffusion coefficients is plotted vs. the incubation time from 0 to 6 h. The distribution is initially very sharp but broadens considerably with time.

corresponding to small oligomers, no big aggregates are visible on the glass slide, nor in solution.

3.4. Monomer Disappearance Kinetics from the Reduction of the Number of Molecules in the Confocal Volume

1. The amplitude of the autocorrelation curves reflects the inverse of the average number of fluorescent molecules in the focal volume $(1/N)$.

2. The analysis of the data shows that there is an increase of the *fraction* of free dye and a concomitant decrease of labeled α-synuclein fraction with aggregation time.

3. In absolute terms the number of free dye remains constant and can be used as a kind of internal standard. There is a significant decrease in the number of labeled α-synuclein molecules and this decrease is concentration dependent. It reflects the disappearance of labeled molecules in big aggregates that are excluded from the measurements.

4. The number concentration of labeled α-synuclein molecules correlates nicely with the level of the fluctuating signal between the bright pulses that appear in the signal (3).

5. The reduction of the number concentration is not mainly due to the incorporation of multiple fluorescent molecules in big oligomers. FCS measurements were done after dilution (10^4–10^5 times) of labeled protein with unlabeled monomer α-synuclein. Therefore, the chance of having more than

one labeled molecule in an oligomer is insignificant, when the aggregates are not too big.

6. Experimentally this is confirmed by the fact that the brightness of the oligomers does not increase. Of course the spikes are due to big oligomers with multiple labeling.

7. Confocal images of the bottom of the glass slide and of the solution (100 μm from the glass bottom) can be made using laser scanning microscope (LSM 510) and show the gradual formation of larger aggregates.

4. Notes

1. In the past, we labeled proteins also at pH 8.8 to avoid oligomer formation during labeling, but the maleimide is less stable at this pH and the reaction is less specific. Nevertheless we could obtain nicely labeled A140C-synuclein with a high diffusion coefficient.

2. It is very important to get pure monomeric α-synuclein as starting material. Here, we use brief vortexing instead of continued stirring to prevent the aggregation of α-synuclein at neutral pH.

3. The separation of free dye and labeled protein is always a compromise between efficient chromatography and dilution of the sample. Insufficient removal of the free dye may be due to ineffective gel chromatography, but it could also be due to the presence of nonspecifically bound dye molecules that are not fully removed by the gel chromatography at 100 μM, but are released upon dilution to nanomolar concentrations.

1. It is best to calibrate the confocal microscope with a dye that can be excited at the same wavelength of the label used. In the past, Rhodamine 6G was often used to calibrate FCS instruments (6). In this way we obtained fair agreement with D obtained by dynamic light scattering for a set of different proteins (7). The more precise technique of scanning FCS has allowed to determine an absolute D of a number of labels which can be found in ref. 5.

References

1. Vercammen J, Maertens G, Gerard M, De Clercq E, Debyser Z, Engelborghs Y (2002) DNA-induced polymerization of HIV-1 integrase analyzed with fluorescence fluctuation spectroscopy. J Biol Chem 277:38045–38052

2. Buyens K, Lucas B, Raemdonck K, Braeckmans K, Vercammen J, Hendrix J, Engelborghs Y, De Smedt SC, Sanders NN (2008) A fast and sensitive method for measuring the integrity of siRNA-carrier complexes in full human serum. J Control Release 26:67–76

3. Nath S, Meuvis J, Hendrix J, Carl SA, Engelborghs Y (2010) Early aggregation steps in alpha-synuclein as measured by FCS and

FRET: evidence for a contagious conformational change. Biophys J 98:1302–1311

4. Hoyer W, Antony T, Cherny D, Heim G, Jovin TM, Subramaniam V (2002) Dependence of α-synuclein aggregate morphology on solution conditions. J Mol Biol 322:383–393

5. Petrasek Z, Schwille P (2008) Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. Biophys J 94:1437–1448

6. Magde D, Elson EL, Webb W (1974) Fluorescence correlation spectroscopy. II. An experimental realization. Biopolymers 13:29–61

7. Krouglova T, Vercammen J, Engelborghs Y (2004) Correct diffusion coefficients of proteins in fluorescence correlation spectrosocpy. Application to tubulin oligomers induced by Mg2+ and Paclitaxel. Biophys J 87:2635–2646

INDEX