Residual Structure in Disordered Peptides and Unfolded Proteins from Multivariate Analysis and Ab Initio Simulation of Raman Optical Activity Data

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2007

Impact Factor: 2.63 · DOI: 10.1002/prot.21593 · Source: PubMed

7 AUTHORS, INCLUDING:

Paul D A Pudney
Unilever
73 PUBLICATIONS **1,410** CITATIONS

WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Residual structure in disordered peptides and unfolded proteins from multivariate analysis and ab initio simulation of Raman optical activity data

Fujiang Zhu,[1] Josef Kapitan,[1] George E. Tranter,[2] Paul D. A. Pudney,[3] Neil W. Isaacs,[1] Lutz Hecht,[1] and Laurence D. Barron[1]*

[1] WestCHEM, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom

[2] Chiralabs Ltd., BCIE, Oxford University Begbroke Science Park, Yarnton, Oxfordshire OX5 1PF, United Kingdom

[3] Unilever Research and Development, Colworth Laboratory, Sharnbrook, Bedford MK44 1LQ, United Kingdom

ABSTRACT

Vibrational Raman optical activity (ROA), measured as a small difference in the intensity of Raman scattering from chiral molecules in right- and left-circularly polarized incident light, or as the intensity of a small circularly polarized component in the scattered light, is a powerful probe of the aqueous solution structure of proteins. The large number of structure-sensitive bands in protein ROA spectra makes multivariate analysis techniques such as nonlinear mapping (NLM) especially favorable for determining structural relationships between different proteins. We have previously used NLM to map a large dataset of peptide, protein, and virus ROA spectra into a readily visualizable two-dimensional space in which points close to or distant from each other, respectively, represent similar or dissimilar structures. As well as folded proteins, our dataset contains ROA spectra from many natively unfolded proteins, proteins containing both folded and unfolded domains, denatured partially structured molten globule and reduced protein states, together with folded proteins containing little or no α-helix or β-sheet. In this article, the relative positions of these systems in the NLM plot are used to obtain information about any residual structure that they may contain. The striking differences between the structural propensities of proteins that are unfolded in their native states and those that are unfolded due to denaturation may be responsible for their often very different behavior, especially with regard to aggregation. An ab initio simulation of the Raman and ROA spectra of an alanine oligopeptide in the poly(L-proline) II-helical conformation confirms previous suggestions that this conformation is a significant structural element in disordered peptides and natively unfolded proteins. The use of ROA to identify and characterize proteins containing significant amounts of unfolded structure will, inter alia, be valuable in structural genomics/proteomics since unfolded sequences often inhibit crystallization.

Proteins 2008; 70:823–833.

Key words: Raman optical activity; multivariate analysis; unfolded proteins; disordered structure; poly(L-proline) II helix; structural genomics/proteomics.

INTRODUCTION

The problem of the existence and nature of residual structure in unfolded proteins and in unfolded sequences in proteins comprising both folded and unfolded domains is of great current interest in the context of protein folding and function.[1] It is also central to understanding detailed mechanisms of protein misfunction in the protein misfolding diseases, especially why some sequences have a propensity to associate into ordered amyloid fibrils rather than the more usual amorphous aggregates. Furthermore, an increasing number of proteins are being discovered which, despite lacking a compact tertiary fold in their native states, exhibit a variety of functions.[2–6] The DisProt database (http://www.disprot.org) provides an archive of information about the structure and function of these rather extended flexible proteins, which are variously described as "natively unfolded," "intrinsically unfolded," "intrinsically unstructured," or "intrinsically disordered". They should be distinguished from those having a well-defined compact

Grant sponsor: BBSRC.
*Correspondence to: Professor L. D. Barron, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK. E-mail: laurence@chem.gla.ac.uk
Received 12 February 2007; Revised 18 April 2007; Accepted 1 May 2007
Published online 29 August 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21593

fold in their native states yet containing little or no α-helix or β-sheet as revealed by X-ray crystallography or solution NMR: the polypeptide sequences now comprise relatively fixed structure, described here as irregular, similar to that found in the longer loops of proteins with folds based on large amounts of α-helix or β-sheet.

X-ray crystallography is generally not useful for this problem since the presence of significant amounts of unfolded structure often inhibits crystallization; and if formed, the corresponding protein crystal structure may not be representative of that in solution. To date, NMR has been the most widely used technique for the study of unfolded proteins and has provided a great deal of valuable information.[7] Although NMR can easily recognize disordered structure associated with unfolded sequences, their intrinsic flexibility together with the relatively slow timescale of NMR makes the information less specific than that provided by X-ray (or indeed NMR) structures of folded proteins.[8] On the other hand, the much shorter timescale of optical spectroscopic techniques means that they provide a different perspective to that from NMR since they measure a superposition of "snapshot" spectra of all the interconverting conformers present in solution. Accordingly, techniques such as, inter alia, electronic circular dichroism measured in the ultraviolet (UVCD), vibrational circular dichroism (VCD), Raman optical activity (ROA), and fluorescence, have all been applied to this problem.[1] Here we focus on ROA which, like VCD, combines the additional sensitivity to three-dimensional structure of chiroptical methods such as UVCD with the advantages of vibrational spectroscopy and consequently reports on chirality associated with all the fundamental vibrational transitions of the molecule.

ROA measures a small difference in the intensity of vibrational Raman scattering from chiral molecules in right- and left-circularly polarized incident light or, equivalently, the intensity of a small circularly polarized component in the scattered light using incident light of fixed polarization.[9–12] The first and second experiments are called incident circular polarization (ICP) and scattered circular polarization (SCP) ROA, respectively. The ability to study aqueous solutions, with no restrictions on the size of the biomolecules, makes ROA ideal for the study of folded, unfolded and partially folded proteins, together with intact glycoproteins and viruses.[13–18]

We have recently shown that, on account of the large number of structure-sensitive bands in protein ROA spectra, multivariate analysis (pattern recognition) techniques such as principal component analysis (PCA)[13,17] and nonlinear mapping (NLM)[19] are especially favorable for determining structural relationships between different proteins. As well as folded proteins, our large set of ROA spectra contains data from many natively unfolded proteins, proteins containing both folded and unfolded domains, denatured partially structured molten globule and reduced protein states, together with folded proteins comprising mainly disordered sequences with little or no secondary structure. In this article we explore the application of NLM to obtain information about residual structure in these systems from their ROA spectra.

The peptides and proteins considered explicitly are listed in Table I.

MATERIALS AND METHODS

ROA spectroscopy

With the exception of that of the Type III antifreeze protein (supplied purified, M_W 6500, from ocean pout serum by A/F Protein), all the ROA spectra were acquired previously. The instrument used for the ICP Raman and ROA measurements was home built and has been described previously.[20] That used for the SCP ROA measurements was the Chiral*RAMAN* manufactured by BioTools and also described previously.[13,16] ICP and SCP ROA spectra are measured as $I^R - I^L$ and $I_R - I_L$ respectively, where I^R and I^L are the unpolarized Raman-scattered intensities in right- and left-circularly polarized incident light, and I_R and I_L are the Raman-scattered intensities with right- and left-circular polarization in unpolarized incident light. Details of typical protein sample handling, together with instrumental conditions, may be found in earlier publications.[13–18]

Multivariate analysis

The NLM analysis was performed using STATISTICA's multidimensional scaling (MS) procedures with Euclidean distances in the multidimensional space of standardized spectra.[21] Specifically, we used monotone nonmetric MS, with spectra standardized to put them all on an equal footing such that each spectrum is shifted and rescaled to have zero mean (so that the ROA band intensities sum to zero) and unit standard deviation.

Ab initio simulations

The ab initio simulation of the ROA spectrum of the 10-residue alanine oligomer formyl-Ala_{10}-amide in the poly(L-proline II) (PPII)-helical geometry was a preliminary calculation as part of an ongoing effort to simulate the ROA spectra of the major conformational elements of polypeptide and protein sequences, to be published later. Harmonic force fields and molecular polarizability and optical activity tensors were calculated for the model tripeptide fragment formyl-Ala_3-amide using the GAUSSIAN program,[22] and extended to longer poly(L-alanine) segments by transfer in Cartesian coordinates.[23] The standard PPII conformation ($\varphi = -75^\circ$, $\psi = +145^\circ$, $\omega = +180^\circ$)[24,25] was used as a starting geometry, and eight other possible conformations subsequently gener-

DOI 10.1002/prot

Table I
Peptides and Proteins Discussed in the Article (Numbers Correspond to Those in Fig. 1)

Peptide/protein		Structure class	Protein type/state
1	Acetyl-OOAla$_7$OO-amide	All disordered/irregular	
2	Disordered poly(L-glutamic acid)	All disordered/irregular	
3	Disordered poly(L-lysine)	All disordered/irregular	
4	Phosvitin (hen)	All disordered/irregular	Natively unfolded
5	Metallothionein (rabbit)	All disordered/irregular	Folded
6a	Ala$_3$	All disordered/irregular	
6b	Ala$_4$	All disordered/irregular	
6c	Ala$_5$	All disordered/irregular	
6d	Ala$_6$	All disordered/irregular	
7	β-casein (bovine)	Mainly disordered/irregular	Natively unfolded
8	κ-casein (bovine)	Mainly disordered/irregular	Natively unfolded
9a	α-synuclein (human)	Mainly disordered/irregular	Natively unfolded
9b	α-synuclein (human A30P mutant)	Mainly disordered/irregular	Natively unfolded
9c	α-synuclein (human A53T mutant)	Mainly disordered/irregular	Natively unfolded
10	β-synuclein (human)	Mainly disordered/irregular	Natively unfolded
11	γ-synuclein (human)	Mainly disordered/irregular	Natively unfolded
12a	tau46 (human)	Mainly disordered/irregular	Natively unfolded
12b	tau46 (human P301L mutant)	Mainly disordered/irregular	Natively unfolded
13	ω-gliadin (wheat)	Mainly disordered/irregular	Natively unfolded
14	T-A-1 peptide (wheat glutenin subunit)	Mainly disordered/irregular	Natively unfolded
15	Type III antifreeze protein (ocean pout)	Mainly disordered/irregular	Folded
16	Bowman–Birk inhibitor (soybean)	Mainly disordered/irregular	Folded
17a	A-gliadin (wheat)	α disordered	Natively unfolded
17b	A-gliadin (wheat) in MeOH	Unassigned (all α)	Natively unfolded
18a	Prion protein (ovine, 25–233)	Mainly α	Folded and unfolded domains
18b	Prion protein (ovine, 94–233)	Mainly α	Folded
18c	Prion protein (ovine, 94–233), reduced	Unassigned (all β)	Denatured
19a	Lysozyme (hen)	Mainly α	Folded
19b	Lysozyme (hen), reduced	Unassigned (mainly β)	Denatured
20a	Ribonuclease A (bovine)	αβ	Folded
20b	Ribonuclease A (bovine), reduced	Unassigned (mainly β)	Denatured
21a	α-lactalbumin (bovine)	Mainly α	Folded
21b	α-lactalbumin (bovine), A-state	Unassigned (αβ)	Denatured
22a	Lysozyme (equine)	Mainly α	Folded
22b	Lysozyme (equine), A-state	Unassigned (mainly α)	Denatured
23a	Lysozyme (human)	Mainly α	Folded
23b	Lysozyme (human), prefibillar intermediate	Unassigned (mainly α)	Denatured

Where assigned, the structure class has been determined using physical methods; where unassigned, the structure class suggested by the position in the NLM plot of Figure 1 is given in brackets.

ated by changing the φ,ψ angles by ±20° (changing all three φ,ψ angles in each fragment at the same time). These structures were subjected to a restricted normal mode optimization,[26] so that only the vibrations important for simulating the experimental spectrum could be relaxed with a minimal change of the main chain conformation. The QGRAD program[26] was used for the normal mode optimization with modes fixed in the range −100 to +200 cm^{-1}. The optimizations and the force-field computations were performed at the B3LYP/COSMO[27]/6-31G** level, while the polarizability and optical activity tensor computations were performed at the HF/vacuum/6-31G level. During the transfer to the longer poly(L-alanine) segments, the source tripeptides in the proper conformations were propagated along the regular polymer with the backbone torsion angles equal to the averages of the corresponding angles within the tripeptide fragment. The Raman and ROA spectra were then generated within the harmonic approximation using the usual procedure.[28] Spectra can be simulated for alanine oligomers containing an arbitrary number of residues, but there is not a strong dependence of the simulated spectra on chain length longer than 10 residues; thus formyl-Ala$_{10}$-amide is appropriate for this study. All nine optimized tripeptide fragments were only slightly distorted during optimization, in most cases by less than 5° for φ,ψ, and ω (e.g., the average angles of formyl-Ala$_{10}$-amide in a conformation close to the regular PPII structure were φ = −74°, ψ = + 147°, ω = + 175° after optimization) and these average torsion angles were used to construct the final 10-residue oligomer.

RESULTS AND DISCUSSION

Protein ROA

We are currently building up a set of ROA spectra of polypeptides and proteins in aqueous solution that covers

as many different structure and fold types as possible. Some recent examples of typical protein ROA spectra may be found in Refs. 13–19. One advantage of ROA for protein structure analysis is that resolved signatures of loops and turns appear in addition to those of secondary structure, thereby also providing information about tertiary structure. One consequence is that, although some individual band assignments may be uncertain, or not even valid due to the extensive vibrational coupling often involved in the generation of large ROA signals, overall ROA band patterns can be characteristic of certain motifs and sometimes even folds. ROA is also unique among spectroscopic methods in having the ability to distinguish between hydrated and unhydrated α-helix.[18]

In addition to folded proteins, ROA is also valuable in the study of unfolded proteins,[29] especially for the identification of PPII helix in such states. Although originally defined for the conformation adopted by polymers of L-proline, the PPII helix can be supported by amino acid sequences other than those based on L-proline and has been recognized as a common structural motif within the longer loops in the X-ray crystal structures of many proteins.[24] It consists of a left-handed helix with threefold rotational symmetry in which the ϕ,ψ angles of the constituent residues are restricted to values around −75°, +145°, corresponding to a region of the Ramachandran surface adjacent to the β-region. The extended nature of the PPII helix precludes intrachain hydrogen bonds, with protein X-ray crystal structures[24] and modeling studies[25] indicating that the structure is stabilized instead by main chain hydrogen bonding with water molecules, with side chains modulating PPII helix-forming propensities in various ways. This produces a rather flexible, adaptable, structure that is becoming increasingly recognized as a major conformational element of disordered polypeptides and unfolded proteins in aqueous solution,[30,31] and which may be important in the functional role of intrinsically unfolded sequences.[5,6,32,33] It can be distinguished from random coil in polypeptides using UVCD,[30,33] VCD,[34,35] infrared and Raman,[36] and ultraviolet resonance Raman.[37] Although these techniques in isolation have difficulty in identifying PPII helix when other conformational elements are present, combinations have been used to provide quantitative estimates of the PPII content relative to other conformational elements in peptides.[38,39] ROA has proved especially useful in identifying PPII helix,[40] even in proteins, and is proving valuable for the study of PPII structure in unfolded and partially folded proteins and its possible role in amyloid fibril formation in certain protein misfolding diseases.[16,29,41]

NLM analysis

Following our demonstration that useful structural relationships among proteins may be obtained by analyzing their ROA spectra using the well-known method of PCA,[13,17] we went on to show that the more advanced multivariate analysis method called NLM,[42] which may be readily implemented using STATISTICA,[21] yields even better results.[19] Multivariate analysis techniques start by considering each digitized spectrum to be a vector from the origin to a point in a multidimensional space, with the axes representing the digitized wavenumbers and the projected length of a given vector onto one of the coordinate axes being the spectral intensity at the wavenumber to which this axis is assigned. Points close to or distant from each other in this multidimensional space are then taken to represent similar or dissimilar spectra, respectively. Like PCA, NLM and related MS methods seek to create a lower dimensional space (two and three dimensions were considered[19]) in which the relative positions of the points approximately preserve the relationships between the spectra, thereby providing a more easily comprehended representation. The advantage of these methods over PCA is that they aim to best represent the relationships between all spectra rather than just describing the gross overall variance (which can lead to poor representation of some sample members),[42] so that characterization of spectral and hence structural similarities is optimized. As the methods focus on interpoint distances, within such plots the choice of the two or three axes is arbitrary and the distribution of points can be arbitrarily rotated or translated in unison within the lower dimensional space. In short, the methods minimize the following function (although the means of doing so may differ):

$$S = \sum_i \sum_{<j} w_{ij}\left(d_{ij} - \hat{d}_{ij}\right)^2$$

where d_{ij} and $\hat{d}_{ij}$ are the interpoint (i.e., interspectra) distances in the full multidimensional space and the reduced dimensionality space respectively, and w_{ij} is a weighting factor chosen according to the specific form of MS or NLM undertaken. Various weighting schemes can be employed, leading to emphasis on whether small or large distances are best reproduced or, as here, whether the monotonic order of distances is preserved, in which case

$$w_{ij} = w = \frac{1}{\sum_i \sum_{<j} d_{ij}^2}$$

and $S^{1/2}$ is the STRESS of the representation (STandardized REsidual Sum of Squares).[42]

Figure 1 shows a two-dimensional (2D) NLM plot, updated from the one shown previously,[19] for a set of 85 polypeptide, protein and virus ROA spectra (mostly ICP but some SCP) in aqueous solution over the range 702–1773 cm^{-1}. A list of most of the actual polypeptides, proteins and viruses is provided as supplementary mate-

DOI 10.1002/prot

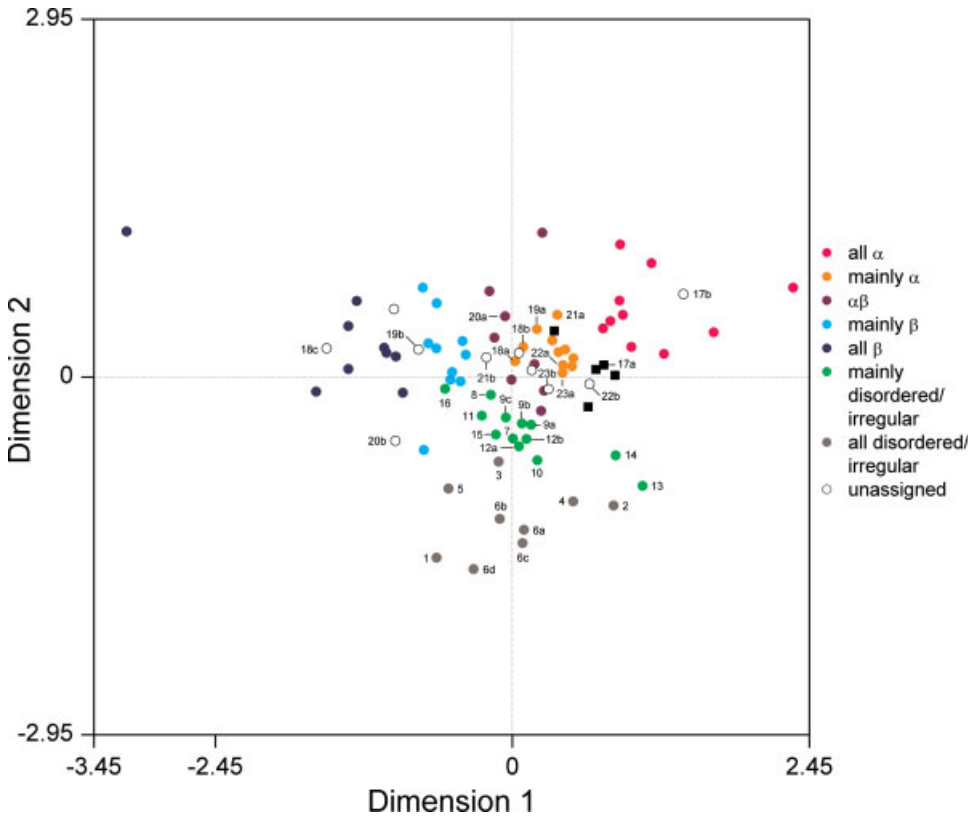

Figure 1

Two-dimensional NLM plot of a set of 85 peptide, protein and virus ROA spectra (a list of most of the samples is provided as supplementary material to Ref. 19). More complete definitions of the structure classes are: all α > ~60% α-helix with little or no β-sheet; mainly α > ~35% α-helix and a small amount (< ~15%) or no β-sheet; αβ, significant amounts (> ~15%) of both α-helix and β-sheet; mainly β > ~35% β-sheet and a small amount (< ~15%) or no α-helix; all β > ~45% β-sheet with little or no α-helix; mainly disordered/irregular, little α-helix or β-sheet; all disordered/irregular, no α-helix or β-sheet. The black squares identify a subset of mainly α proteins containing mostly α-helix and disordered structure with little or no β-sheet.

rial to Ref. 19. The spectra separate into clusters corresponding to different structure classes (as determined mostly by X-ray or NMR for proteins with well-defined native folds, and UVCD and/or NMR for polypeptides, and natively unfolded proteins). The orientation of the axes has been chosen so as to provide a display with increasing α-helix content to the right, increasing β-sheet content to the left, and increasing disordered or irregular structure (in effect structure containing no extended α-helix or β-sheet) from top to bottom. The polypeptide and protein positions are color-coded with respect to the seven different structure classes listed on the figure, which provide a useful initial classification that follows naturally from the 2D NLM clustering. Since the α-helix and β-sheet contents are inversely correlated (the larger the amount of one, the smaller the amount of the other), the dimension 1 values associated with α-helix and β-sheet have opposite signs. The dimension 2 values associated with disordered/irregular structure and the total amount of α-helix and β-sheet are similarly inversely correlated (the larger the amount of α-helix or β-sheet or both, the smaller the amount of disordered/irregular structure). Disordered/irregular sequences can have any structure other than α-helix or β-sheet, including PPII helix, turns, and unassociated β-strand.

Marked in Figure 1 are disordered polypeptides, a series of alanine-based oligomers, natively unfolded proteins, proteins containing both folded and unfolded domains, and proteins that are partially unfolded due to denaturation together with their native progenitors. Also marked are three proteins that have provided X-ray crystal and solution NMR structures showing a folded state comprising mainly disordered/irregular sequences with little or no secondary structure. The full list is provided in Table I. As discussed below, from the position of a particular system with respect to the clusters corresponding to the various structure classes, it is possible to gain insight into the different types of residual structure contained within it.

The positions of the members of the series of alanine peptides Ala_3–Ala_6 (6a–d) show a trend towards the position of acetyl-OOAla_7OO-amide (1), where O represents

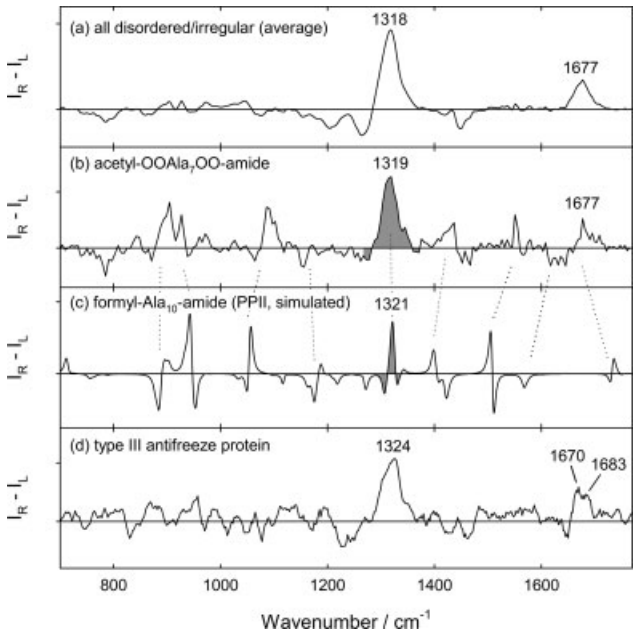

Figure 2

*(**a**) Average of the standardized backscattered ROA spectra (ICP and SCP) for the three polypeptides and two proteins (1–5) in the all disordered/irregular structure class identified in Figure 1 (adapted from Ref. 19). (**b**) ICP ROA spectrum of the alanine-based peptide acetyl-OOAla$_7$OO-amide in water at pH 4.6. (adapted from Ref. 40). (**c**) Ab initio simulated ROA spectrum of formyl-Ala$_{10}$-amide constrained to the PPII-helical geometry. (**d**) SCP ROA spectrum of the Type III antifreeze protein from ocean pout in water at pH 7.3. The intensity scales are arbitrary. All the experimental spectra were recorded in aqueous solution at concentrations of ~50 mg/mL.*

ornithine, with increasing chain length. This is in harmony with suggestions from other work that PPII propensity in alanine peptides increases with increasing chain length and becomes clearly established from Ala$_4$ and above.[36] The displacement of acetyl-OOAla$_7$OO-amide towards the left-hand side of the plot may indicate a small β-sheet propensity, which would be consistent with previous conclusions that PPII helix is accompanied by a small contribution from β structure in this peptide.[43]

Disordered peptides and the ab initio simulated ROA spectrum of PPII helix

It is instructive to display [Fig. 2(a)] the average ROA spectrum of the three polypeptides and two proteins (1–5) in the all disordered/irregular cluster.[19] This average spectrum is very similar to that previously suggested to be characteristic of PPII helix (especially the strong positive ROA band at ~1318 cm^{-1} and the weaker positive band at ~1677 cm^{-1}), which as mentioned above, is thought to be a major conformational element in such structure.

Our confidence in the assignment of the main features of this ROA spectrum to PPII helix is reinforced by an ab initio simulation of the ROA spectrum of the model oligomer formyl-Ala$_{10}$-amide constrained to the PPII-helical geometry. This simulated spectrum is displayed in Figure 2(c), together with the ROA spectrum of acetyl-OOAla$_7$OO-amide (1) in Figure 2(b). A very similar peptide, acetyl-XXAla$_7$OO-amide, where X represents diaminobutyric acid, has been found, using a combination of NMR and CD, to take up a predominantly PPII-helical conformation in aqueous solution.[43] The close similarity of these two ROA spectra [together with that of the average all disordered/irregular ROA spectrum in Fig. 2(a)] provides good evidence that PPII helix is indeed a major generic conformational element in such structure.

The simulated ROA spectrum of formyl-Ala$_{10}$-amide displayed in Figure 2(c) is the one showing the best overall agreement with the experimental ROA spectrum of acetyl-OOAla$_7$OO-amide. It is obtained from the structure closest to that of a regular PPII helix. The simulated conventional Raman spectrum (not shown) also agrees well with the experimental Raman spectrum of acetyl-OOAla$_7$OO-amide. The strong positive ~1319 cm^{-1} ROA band, highly characteristic of PPII structure, in the experimental ROA spectrum of acetyl-OOAla$_7$OO-amide is well-reproduced in the simulated spectrum. The corresponding normal mode calculation provides an assignment of this band to C$^{\alpha}$—H deformations coupled along the entire molecule (with little contribution from amide III-type N—H deformations). We have found from simulated spectra of other conformations of formyl-Ala$_{10}$-amide that this band, which forms part of a computed negative–positive–negative triplet for the PPII-helical conformation, is particularly sensitive to the geometry of the peptide backbone and only reproduces the experimental band for conformations close to that of PPII helix.

ROA in the amide I region, which originates mostly in the C—O stretch, is also very sensitive to the peptide backbone structure. The positive peak at ~1677 cm^{-1} characteristic of PPII helix, which sometimes appears as part of a couplet with a small negative band at lower wavenumber, is also well-reproduced by the simulation on formyl-Ala$_{10}$-amide in the PPII-helical conformation in which it is generated by amide I vibrations coupled along the entire molecule (the relatively large systematic shift to higher wavenumbers is typical for normal vibrations computed using the B3LYP DFT functional[44,45]).

Most of the other spectral features in the experimental ROA spectrum of acetyl-OOAla$_7$OO-amide are also well-reproduced in the simulated spectrum of formyl-Ala$_{10}$-amide in the PPII-helical conformation, which provides assignments to mostly C$_{\alpha}$—H deformations mixed with CH$_3$ group vibrations, also often coupled throughout the molecule. These bands, which are moderately sensitive to conformation, tend to be suppressed in the heteropolypeptide sequences in protein ROA spectra due to contributions from many other sidechains and their associated conformational freedom. In fact conformational freedom

DOI 10.1002/prot

of sidechains, together with backbone flexibility, needs to be incorporated into the simulation to reproduce the observed ROA bandshapes,[44,45] the computed ones (using a half-height bandwidth of 10 cm^{-1}) being too narrow from the single-geometry simulation used here.

Natively unfolded proteins

All of the natively unfolded proteins (4, 7–14) fall below the zero line in the 2D NLM plot of Figure 1, with increasing negative values of dimension 2 corresponding to increasing proportions of disordered/irregular structure. Phosvitin (4) lies well below most of the others, consistent with its classification (Table I) as all disordered/irregular, the others being classified as mainly disordered/irregular. Most cluster around the vertical center line corresponding to the zero of the dimension 1 axis, suggesting a similar propensity for small amounts (including zero) of α-helical or β-sheet structure. The structures of β- and κ-casein, which fall within the mainly disordered/irregular cluster (seven and eight, respectively), have been modeled using sequence-based structure prediction algorithms constrained by experimental vibrational spectroscopic data.[46] The resulting open, flexible, three-dimensional structures may be useful as an aid to visualizing the general structural characteristics of some natively unfolded proteins.

The wheat protein ω-gliadin (13) and the T-A-1 peptide (14), a high molecular weight glutenin subunit of wheat gluten, are displaced considerably to the right of the vertical center line, which suggests a bias towards α-structure. This emphasizes their structural kinship with the wheat protein A-gliadin (17a), which although natively unfolded, contains a large amount of hydrated α-helix in the C-terminal region with smaller amounts of disordered structure (mainly PPII) located in the N-terminal region,[47] and consequently appears within the mainly α region. In a 60% methanol/40% aqueous solution, which encourages formation of additional α-helix, the position of A-gliadin shifts to well inside the all α region (17b).

Notice that the positions of the brain protein human α-synuclein (9a), and the A30P (9b) and A53T (9c) mutants associated with familial cases of Parkinson's disease, are very similar, which reinforces the suggestion that the fibrillogenic properties of the mutants are due to perturbations of the physicochemical properties of the amino acid sequences rather than to any significant structural changes.[48] Likewise the positions of the brain protein human tau46 (12a) and the P301L mutant (12b) associated with inherited frontotemporal dementia.[48]

Especially interesting is the cluster of five proteins, indicated by black squares, previously identified from the way in which it shifted in a three-dimensional (3D) NLM plot in the direction of increasing disorder out of a cluster of mainly α proteins in which it falls in the 2D NLM plot of Figure 1.[19] This cluster comprises A-gliadin (17a) and the intact helical rod-shaped or filamentous plant viruses potato virus X (PVX), narcissus mosaic virus (NMV), papaya mosaic virus (PMV), and tobacco rattle virus. Neither X-ray crystal nor solution NMR structures are available for any of these; but their ICP ROA spectra have been measured previously. The average of the standardized ROA spectra of A-gliadin, PVX, NMV, PMV was found to be dominated by bands from the average all α and all disordered/irregular spectra.[19] All the virus ROA spectra are dominated by bands from the coat proteins; however, no X-ray crystal structures have been reported for the corresponding isolated coat protein subunits, presumably due to lack of suitable crystals. This, together with the fact that the viruses cluster with A-gliadin in 2D and 3D NLM plots, suggests that they all have similar structures which are similar to that of A-gliadin, containing significant amounts of hydrated α-helix and disordered structure. It has been suggested that this combination of structural elements may have functional significance, such as facilitating disorder-to-order transitions (and vice versa) and suppressing aggregation.[19]

Proteins containing both folded and unfolded domains

The prion protein, with its folded C-terminal half (which is predominantly α-helical) and its completely unfolded N-terminal half is the paradigm for a protein containing both folded and unfolded domains. Our set of protein ROA spectra contains both the full-length ovine prion protein PrP^{25-233} (18a) and the N-terminal truncated version PrP^{94-233} (18b). Both proteins appear on the left-hand side of the mainly α region, but the full-length protein appears lower down, which is consistent with the presence of the unfolded N-terminal tail and which is thought to contain significant amounts of PPII-like structure.[49]

Partially unfolded denatured proteins

Proteins that are partially unfolded due to denaturation are expected to contain more residual structure than proteins that are unfolded in their native states. Our set of protein ROA spectra contains several denatured proteins, and this expectation is indeed born out by their positions in the NLM plot of Figure 1. Hen lysozyme, for example, which appears within the mainly α region (19a) in its native state, shifts to the border of the mainly β and all β regions (19b) in the denatured state generated by reducing all its disulphide bonds.[50] Similarly ribonuclease A, when reduced,[50] shifts from its native state position (20a) in the αβ region to a position (20b) well below that of reduced hen lysozyme suggesting that, although it contains some β-structure, it also contains

significantly more disordered structure than reduced hen lysozyme. A reduced state of the N-terminal truncated ovine prion protein $PrP^{94–233}$ (18c) falls well within the all β region, consistent with the appearance of its ROA spectrum,[17] which is characteristic of a structure rich in rather flat, regular, up-and-down β-sheet.

The NLM plot of Figure 1 reveals that proteins in denatured (but nonreduced) molten globule states (A-states) generated at low pH contain more residual α-helical structure than proteins in denatured reduced states, presumably because the retained disulphide bonds help to stabilize α-helical sequences and inhibit chain association into β-sheet. The native states of bovine α-lactalbumin (21a) and equine lysozyme (22a) have the same "lysozyme" type of fold and accordingly fall in the mainly α region. However, A-state bovine α-lactalbumin (21b) falls within the αβ region whereas A-state equine lysozyme (22b) remains well within the mainly α region. This reflects the striking difference between the corresponding A-state ROA spectra,[51] with that of equine lysozyme retaining the strong positive ~1340 cm^{-1} band characteristic of hydrated α-helix that appears in the native state ROA spectra of both proteins but which is completely absent from the ROA spectrum of A-state bovine α-lactalbumin in which clear β-sheet signatures have appeared in its place. This highlights the remarkable sensitivity of ROA to the "complexity of order" in partially unfolded denatured states, which can be quite different in proteins having very similar native state structures.

As a further example, a molten globule A-state of human lysozyme shows very different characteristics to those of both A-state bovine α-lactalbumin and A-state equine lysozyme. Both native (23a) and A-state (23b) human lysozyme fall within the mainly α region. However, the A-state protein is a little lower down, which is consistent with the appearance of the ROA spectra in which the positive ~1340 cm^{-1} hydrated α-helix band in the native protein (which is weaker than the corresponding bands in native bovine α-lactalbumin and equine lysozyme) is absent from the A-state spectrum, being replaced by significant positive ROA intensity at ~1318 cm^{-1} characteristic of PPII structure. This PPII structure has been suggested to be responsible for the fibrillogenic character of A-state human lysozyme.[41]

Folded proteins with little or no α-helix or β-sheet

Although not natively unfolded, the soybean Bowman–Birk inhibitor (16)[52] is placed in the mainly disordered/irregular structure class since X-ray crystal and NMR solution structures show it to be dominated by "loop-type" structure together with, according to PDB X-ray structure 1pi2, five short stretches of β-strand contained in two small sections of antiparallel β-sheet. The unusually large number of seven disulphide links hold this small single-chain protein in its well-defined fold. Its position on the left of the main disordered/irregular cluster in the NLM plot (Fig. 1) is consistent with the presence of more well-defined β-sheet than the proteins in the cluster that are properly "natively unfolded". Like the other proteins in the cluster, its ROA spectrum[53] is dominated by a positive ROA band at ~1320 cm^{-1} characteristic of PPII structure, significant amounts of which may be present in the long loops. However, the spectrum is qualitatively different from the others in that it contains sharp detailed structure throughout as found in most folded proteins. Most of the 61 residues in structure 1pi2 congregate in the β/PPII region of the Ramachandran plot, with similar numbers in the PPII (right) and β (left) halves of the region.

The X-ray crystal and NMR solution structures of the Type III antifreeze protein from ocean pout (15)[54] also show it to be dominated by "loop-type" structure plus a small section of double-stranded β-sheet which serves to stabilize the fold, there being no disulphide links in this small single-chain protein. The X-ray structure 1hg7 reveals a small amount of α-helix (6.1%) and 3_{10}-helix (7.6%); whereas the NMR structure 1kde reveals only 3_{10}-helix (4.6%), which may suggest that the α-helix is only formed in the stabilizing crystal environment. The protein shows a simple ROA spectrum [Fig. 2(d)] dominated by PPII bands, but lacking the extra structure present in the ROA spectrum of the Bowman–Birk inhibitor, which is consistent with the absence of disulphide links and hence the more "rheomorphic"[48] character of its structure. This is reflected in the Ramachandran plots in which most of the 66 residues in 1hg7 and the 65 residues in 1kde congregate in the β/PPII region with even fewer in the α region in 1kde than in 1hg7. Both 1hg7 1kde show significantly more residues on the right-hand side of the β/PPII region than on the left, corresponding to more PPII than β propensity. This greater PPII versus β propensity of the antifreeze protein compared with the Bowman–Birk inhibitor is probably responsible for its position more to the right (close to the dimension 1 zero line) in the NLM plot of Figure 1 where it clusters with most of the natively unfolded proteins, which suggests that the latter have similar structural propensities to those of the antifreeze protein.

Rabbit metallothionein (5)[55] is another example of a small protein with a well-defined native fold dominated by a disordered/irregular polypeptide sequence, but this time with no α-helix or β-sheet at all. Rather than forming disulphide bridges, all 20 cysteine residues in this cysteine-rich protein bind seven divalent metal ions that, inter alia, serves to stabilize the native fold. The PDB X-ray structure 4mt2 shows the 62 residues of this protein to be distributed roughly equally around the α, turn and β/PPII regions of the Ramachandran surface, with significantly more density on the PPII than on the β side of the β/PPII region. This distribution of ϕ,ψ angles, espe-

DOI 10.1002/prot

cially the larger number of residues in the α and turn regions than in the Bowman–Birk and antifreeze proteins, may be responsible for the strong positive band peaking at ~1313 cm^{-1}, with a broad shoulder on the low-wavenumber side, which dominates the ROA spectrum.[53]

CONCLUSIONS

Although the discussion of structural trends and propensities from multivariate analysis of protein ROA data in this paper is qualitative, the results are nonetheless useful and informative. In particular, the NLM plot of the ROA data provides an immediately comprehensible "snapshot" of the main structural trends and propensities of a large set of peptides and proteins in folded, partially unfolded, and completely unfolded states. Further refinement of the NLM and other multivariate analysis techniques applied to ROA spectral data should ultimately place the analysis of structural propensities on a quantitative footing with respect to the specific types of structural elements present. The use of ROA to identify and characterize proteins containing significant amounts of unfolded structure, for which X-ray crystallography is often inapplicable, will inter alia be valuable in structural genomics/proteomics.

There are striking differences between the structural characteristics of proteins that are unfolded in their native states and those that are unfolded due to denaturation. The former tend to cluster in the mainly disordered/irregular region of the NLM plot of Figure 1 and appear to contain a significant amount of the extended PPII-helical conformation; whereas the latter appear in other regions and can contain significant amounts of β structure in the case of reduced proteins and α-helix in the case of acid molten globules.

The presence of PPII-type structure in unfolded peptides and proteins is the subject of much debate among theoreticians. Simulations with a modified AMBER force field have suggested that the PPII conformation is favored in $(Ala)_7$-type peptides in water at ambient temperature[56] with water structure as a possible stabilizing influence.[57,58] Other simulations, however, suggest that PPII is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins, which behave more like statistical coils.[59] These latter simulations successfully reproduced the radius of gyration for acetyl-XX$(Ala)_7$OO-amide observed using small-angle X-ray scattering, and which is significantly less than the value expected for an ideal PPII helix in solution.[60] Our ability to closely simulate the observed ROA spectrum of acetyl-OO$(Ala)_7$OO-amide using a model oligopeptide in the PPII conformation, together with the correlation of dominant PPII bands in the ROA spectrum of a Type-III antifreeze protein with the majority of residues occupying the PPII region of the Ramachandran surface in both X-ray crystal and NMR solution structures, provides strong support for the importance of PPII-type structure in unfolded peptides and proteins, and in disordered/irregular sequences within folded proteins. Our results might be reconciled with the observation of an unexpectedly small radius of gyration by the suggestion of a flexible, fluctuating structure that on the level of individual residues explores a wide basin around the backbone angles corresponding to the ideal PPII geometry, but that experimental averaging pushes the observed values of the angles to what is interpreted as the PPII geometry.[60] This flexibility could explain the broadness of the characteristic positive ~1319 cm^{-1} PPII ROA peak in acetyl-OO$(Ala)_7$OO-amide. Also relevant could be simulations demonstrating that, on account of short persistence lengths of less than five residues, the observation of generic sequence-independent radii of gyration (for a given chain length) for unfolded proteins can be reconciled with spectroscopic evidence for the presence of sequence-specific residual structure.[61] Further discussion of this controversy may be found in a recent review,[62] and in a new paper reporting a combined VCD, infrared, and Raman study of acetyl-XX$(Ala)_7$OO-amide in D_2O at pD 2.2, which concluded that it contained ~50% of the PPII conformation, together with β-turns and β-strand.[63]

The flexible adaptable character of PPII structure appears to be essential for the function of some proteins (a "careful disorderliness"[15]), with the facile interconversion between PPII and hydrated α-helix possibly exploited in certain order–disorder transitions.[19] However, since PPII structure can be highly fibrillogenic in some situations,[41,48] the physicochemical properties like charge and hydrophobicity of the amino acids in protein sequences in natively unfolded proteins have been chosen by natural selection to inhibit such aggregation.[64,65] On the other hand the reduced proteins, in which the stabilizing influence of the disulphide bonds on the native state is lost, possess a high propensity for β-sheet formation, none more so than reduced ovine $PrP^{94–233}$ (which may be an important factor in prion disease). This may be a manifestation of the suggestion that β-sheet formation, and corresponding aggregation characteristics including amyloid fibril formation, is a generic property of polypeptide chains in the absence of a stabilizing environment.[66]

REFERENCES

1. Rose GD, editor. Unfolded proteins. Adv Protein Chem 2002;62.
2. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J Mol Biol 1999;293: 321–331.
3. Uversky VN, Gillespie JR, Fink AL. Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins 2000;41: 415–427.
4. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.

5. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41:6573–6582.
6. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev 2005;6:197–208.
7. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev 2004;104:3607–3622.
8. Sue SC, Chang CF, Huang YT, Chou CY, Huang TH. Challenges in NMR-based structural genomics. Physica A 2005;350:12–27.
9. Barron LD, Bogaard MP, Buckingham AD. Raman scattering of circularly polarized light by optically active molecules. J Am Chem Soc 1973;95:603–605.
10. Hug W. Raman optical activity. In: Chalmers JM, Griffiths PR, editors. Handbook of vibrational spectroscopy, Vol 1. Chichester: Wiley; 2002. pp 745–758.
11. Barron LD, Hecht L, Blanch EW, McColl IH. Raman optical activity comes of age. Mol Phys 2004;102:731–744.
12. Barron LD. Molecular light scattering and optical activity, 2nd ed. Cambridge: Cambridge University Press; 2004.
13. Zhu F, Isaacs NW, Hecht L, Barron LD. Raman optical activity: a tool for protein structure analysis. Structure 2005;13:1409–1419.
14. Zhu F, Isaacs NW, Hecht L, Tranter GE, Barron LD. Raman optical activity of proteins, carbohydrates and glycoproteins. Chirality 2005;18:103–115.
15. Barron LD. Structure and behaviour of biomolecules from Raman optical activity. Curr Opin Struct Biol 2006;16:638–643.
16. Barron LD, Zhu F, Hecht L, Barron, LD. Structure and behavior of proteins from Raman optical activity. In: Uversky VN, Permyakov EA, editors. Protein structures: methods in protein structure and stability. New York: Nova Science Publishers, in press.
17. McColl IH, Blanch EW, Gill AC, Rhie AGO, Ritchie MA, Hecht L, Nielsen K, Barron LD. A new perspective on β-sheet structures using vibrational Raman optical activity: from poly(L-lysine) to the prion protein. J Am Chem Soc 2003;125:10019–10026.
18. McColl IH, Blanch EW, Hecht L, Barron LD. A study of α-helix hydration in polypeptides, proteins and viruses using vibrational Raman optical activity. J Am Chem Soc 2004;126:8181–8188.
19. Zhu F, Tranter GE, Isaacs NW, Hecht L, Barron LD. Delineation of protein structure classes from multivariate analysis of protein Raman optical activity data. J Mol Biol 2006;363:19–26.
20. Hecht L, Barron LD, Blanch EW, Bell AF, Day LA. Raman optical activity instrument for studies of biopolymer structure and dynamics. J Raman Spectrosc 1999;30:815–825.
21. Statistica, Version 6. Tulsa OK: StatSoft Inc.; 2004.
22. Gaussian 03, Revision C.02. Wallingford CT: Gaussian Inc.; 2004.
23. Bouř P, Sopková J, Bednárová L, Maloň P, Keiderling TA. Transfer of molecular property tensors in Cartesian coordinates: a new algorithm for simulation of vibrational spectra. J Comput Chem 1997;18:646–659.
24. Adzhubei AA, Sternberg MJE. Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol 1993;229:472–493.
25. Creamer TP, Campbell MN. Determinants of the polyproline II helix from modelling studies. Adv Protein Chem 2002;62:263–282.
26. Bouř P, Keiderling TA. Partial optimization of molecular geometry in normal coordinates and use as a tool for simulation of vibrational spectra. J Chem Phys 2002;117:4126–4132.
27. Klamt A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvent phenomena. J Phys Chem 1995;99:2224–2235.
28. Nafie LA, Che D. Theory and measurement of Raman optical activity. In: Evans M, Kielich S, editors. Modern nonlinear optics, Part 3, Vol. 85. New York: Wiley; 1994. pp 105–149.
29. Barron LD, Blanch EW, Hecht L. Unfolded proteins studied by Raman optical activity. Adv Protein Chem 2002;62:51–90.
30. Shi Z, Woody RW, Kallenbach N. Is polyproline II a major backbone conformation in unfolded proteins? Adv Protein Chem 2002; 62:163–240.
31. Bochicchio B, Tamburro AM. Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability and functions. Chirality 2002;14:782–792.
32. Siligardi G, Drake AF. The importance of extended conformations and, in particular, the P_{II} conformation for the molecular recognition of peptides. Biopolymers 1995;37:281–292.
33. Csizmók V, Bokor M, Bánki P, Klement E, Medzihradszky KF, Friedrich P, Tompa K, Tompa P. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. Biochemistry 2005;44:3955–3964.
34. Keiderling TA, Xu Q. Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. Adv Protein Chem 2002;62:111–161.
35. Dukor RK, Keiderling TA. Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. Biopolymers 1991;31:1747–1761.
36. Schweitzer-Stenner R, Eker F, Griebenow K, Cao X, Nafie LA. The conformation of tetraalanine in water determined by polarized Raman, FT-IR and VCD spectroscopy. J Am Chem Soc 2004;126: 2768–2776.
37. Asher SA, Mikhonin AV, Bykov S. UV Raman demonstrates that α-helical polyalanine peptides melt to polyproline II conformations. J Am Chem Soc 2004;126:8433–8440.
38. Schweitzer-Stenner R, Measey T, Hagarman A, Eker F, Griebenow K. Salmon Calcitonin and amyloid β: two peptides with amyloidogenic capacity adopt different conformational manifolds in their unfolded states. Biochemistry 2006;45:2810–2819.
39. Schweitzer-Stenner R, Measey T, Kakalis L, Jordan F, Pizzanelli S, Forte C, Griebenow K. Conformations of alanine-based peptides in water probed by FTIR, Raman, vibrational circular dichroism, electronic circular dichroism, and NMR spectroscopy. Biochemistry 2007;46:1587–1596.
40. McColl IH, Blanch EW, Hecht L, Kallenbach NR, Barron LD. Vibrational Raman optical activity characterization of poly(L-proline) II helix in alanine oligopeptides. J Am Chem Soc 2004;26:5076–5077.
41. Blanch EW, Morozova-Roche LA, Cochran DAE, Doig AJ, Hecht L, Barron LD. Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. J Mol Biol 2000;301:553–563.
42. Krzanowski WJ. Principles of multivariate analysis, a user's perspective. Oxford: Oxford University Press; 1998.
43. Shi Z, Olsen CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99:9190–9195.
44. Kapitán J, Baumruk V, Kopecký V, Bouř P. Conformational flexibility of L-alanine zwitterion determines shapes of Raman and Raman optical activity spectral bands. J Phys Chem A 2006;110:4689–4696.
45. Kapitán J, Baumruk V, Bouř P. Demonstration of the ring conformation in polyproline by the Raman optical activity. J Am Chem Soc 2006;128:2438–2443.
46. Kumosinski TF, King G, Farrell HM. An energy-minimized casein submicelle working model. J Protein Chem 1994;13:701–714.
47. Blanch EW, Kasarda DD, Hecht L, Nielsen K, Barron LD. New insight into the solution structures of wheat gluten proteins from Raman optical activity. Biochemistry 2003;42:5665–5673.
48. Syme CD, Blanch EW, Holt L, Jakes R, Goedert M, Hecht L, Barron LD. A Raman optical activity study of rheomorphism in caseins, synucleins and tau. Eur J Biochem 2002;269:148–156.
49. Blanch EW, Gill AC, Rhie AGO, Hope J, Hecht L, Nielsen K, Barron LD. Raman optical activity demonstrates Poly(L-proline) II helix in the N-terminal region of the ovine prion protein: implications for function and misfunction. J Mol Biol 2004;343:467–476.
50. Wilson G, Hecht L, Barron LD. Residual structure in unfolded proteins revealed by Raman optical activity. Biochemistry 1996;35: 12518–12525.
51. Blanch EW, Morozova-Roche LA, Hecht L, Noppe W, Barron LD. Raman optical activity characterization of native and molten globule states of equine lysozyme: comparison with hen lysozyme and

 DOI 10.1002/prot

bovine α-lactalbumin. Biopolymers (Biospectroscopy) 2000;57:235–248.
52. Ikenaka T, Norioka S. Bowman-Birk family serine protease inhibitors. In: Barrett AJ, Salvesen G, editors. Protease inhibitors. Amsterdam: Elsevier; 1986. pp 361–374.
53. Smyth E, Syme CD, Blanch EW, Hecht L, Vašák M, Barron LD. Solution structure of native proteins with irregular folds from Raman optical activity. Biopolymers 2001;58:138–151.
54. Antson AA, Smith DJ, Roper DI, Lewis S, Caves LSD, Verma CS, Buckley SL, Lillford PJ, Hubbard RE. Understanding the mechanism of ice binding by type III antifreeze proteins. J Mol Biol 2001;305:875–889.
55. Braun W, Vašák M, Robbins AH, Stout CD, Wagner G, Kägi JHR, Wüthrich K. Comparison of the NMR solution structure and the x-ray crystal structure of rat metallothionein-2. Proc Natl Acad Sci USA 1992;89:10124–10128.
56. Gnanakaran S, Garcia AE. Validation of an all-atom protein force field: from dipeptides to larger peptides. J Phys Chem B 2003;107:12555–12557.
57. Garcia AE. Characterization of non-alpha helical conformations in Ala peptides. Polymer 2004;45:669–676.
58. Kentsis A, Mezei M, Gindin T, Osman R. Unfolded state of polyalanine is a segmented polyproline II helix. Proteins 2004;55:493–501.
59. Makowska J, Rodziewicz-Motowidło S, Baginska K, Vila JA, Liwo A, Chmurzynski L, Scheraga HA. Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. Proc Natl Acad Sci USA 2006;103:1744–1749.
60. Zagrovic B, Lipfert J, Sorin EJ, Millett IS, van Gunsteren WF, Doniach S, Pande VS. Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci USA 2005;102:11698–11703.
61. Tran HT, Wang X, Pappu RV. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behaviour of denatured proteins. Biochemistry 2005;44:11369–11380.
62. Shi Z, Chen K, Liu Z, Kallenbach NR. Conformation of the backbone in unfolded proteins. Chem Rev 2006;106:1877–1897.
63. Schweitzer-Stenner R, Measey TJ. The alanine-rich XAO peptide adopts a heterogeneous population, including turn-like and polyproline II conformations. Proc Natl Acad Sci USA 2007;104:6649–6654.
64. Uversky VN, Gillespie JR, Fink AL. Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
65. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003;424:805–808.
66. Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci 1999;24:329–332.