Proc Indian Natn Sci Acad, **70, A,** No. 4, July 2004, pp. 579–596

EXPLORING *'UNSTRUCTURED'* PROTEINS

NEEL S BHAVESH AND RAMAKRISHNA V HOSUR*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005 (India)

(Received 08 January 2003 ; Accepted 12 February 2003)

In the post genomic era, as more and more genome sequences are becoming known and hectic efforts are underway to decode the information content in them, it is becoming increasingly evident that flexibility in proteins plays a crucial role in many of the biological functions. Many proteins have intrinsic disorder either wholly or in specific regions. It appears that this disorder may be important for regulatory functions of the proteins, on the one hand, or, may help in directing the folding process to reach the compact native state, on the other. Nuclear Magnetic Resonance (NMR) has over the last two decades emerged as the sole, most powerful technique to help characterize these disordered protein systems. In this review, we first discuss the significance of disorder in proteins and then survey the NMR methods available for their characterization. A brief description of the results obtained on several disordered proteins is presented at the end.

Key Words: Flexible Proteins; Unfolded Proteins; HNN and HN(C)N; Multidimensional NMR

1 Introduction

The genome sequencing projects have provided new opportunities and challenges to the fields of structural biology and molecular biophysics. Genomic sequence data provide one step toward a broader and more complete understanding of the molecular basis of life. The next stage of this project is to characterize the structures and functions of the corresponding gene products, and to integrate this information in expanding our understanding of structure-function relationships, protein folding, evolution, and the general principles of biology. All these fall under the scope of structural genomics[1-6], which is the most vigorously pursued area of research at the present times. It is hoped that such an extensive investigation will help discover hitherto unknown biological functions and provide clues to the occurrence and control of many diseases.

1.1 The Structure – Function Paradigm

It has been the central dogma since long that the function of a protein is related to its significant and unique state, which is a well-defined three-dimensional structure called native structure. Pursuing this goal, three-dimensional structures of several hundreds of proteins have been solved using X-Ray crystallography, NMR spectroscopy and more recently electron microscopy. Protein Data Bank (PDB) is the single depositary for all these data. In February 2003 the number of entries for proteins in PDB was more than 18000[7].

Though the above paradigm about folded structures and functions remains valid, recent structural and genomic data have clearly shown that not all proteins have unique folded structures under normal physiological conditions. The disorder can be local as well as global. Local disorders have been observed in many protein structures and they have been functionally rationalized also[8]. In view of these, Dunker *et al.* proposed a Protein Trinity paradigm[9], according to which the biological function was thought to be an interplay between three thermodynamic states namely, ordered, molten globule and random coil. In this view, not just the ordered state, but any of the three states can be the native state of a protein[9]. Recently a Protein Quartet Model (Fig. 1) has been proposed for generalization of the structure-function paradigm[10]. According to this, a biological function arises as a result of interplay between four specific conformational forms, namely, ordered forms, molten globules, premolten globules, and random coils. It will not be an exaggeration to assume an ensemble existence of all the four states at any particular time, their relative abundance being governed by basic thermodynamics. Upon ligand binding or some

*Author for Correspondence
Phone : +91-22-2280-4545 Extn. 2488;
Fax : +91-22-2280-4610
E-mail : hosur@tifr.res.in

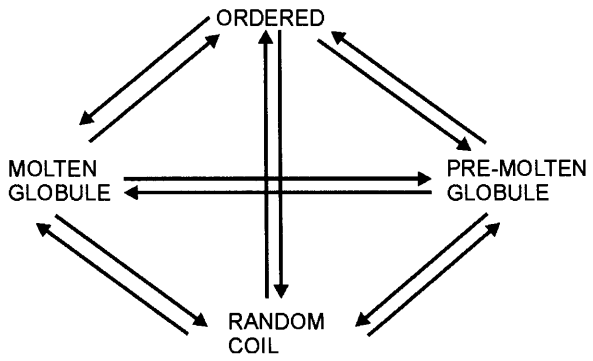

Fig. 1 Protein Quartet Model proposed for generalization of the structure-function paradigm[10].

signalling modification, concentration of one state may increase at the expense of the others. This can explain the fast regulatory steps involved in various biological functions.

1.2 The Natively Unfolded State

There are about 15,000 proteins in the Swiss Protein Database that have been predicted to contain disordered regions of at least 40 consecutive amino acids[11, 12]. PONDR VL-XT, an algorithm in a series of *p*redictors of *n*atural *d*isordered *r*egions (PONDRs) has been applied to about 30 genomes to assess for ≥40 consecutive residue disorders in proteins. The results were: bacteria = 6-33%, archea = 9-37%, eukaryotes = 35 –51%. With the growth in the information on unstructured proteins, their role in biological specificity, transport, regulation and disease is being realized[8, 10, 13-17]. A more complex signalling and regulation network is perceived to be a possible reason for the high occurrence of disorder in eukaryotes[9, 18]. A special term 'natively unfolded' was coined to describe the properties of *tau* protein[19] and since then a large number of proteins belonging to this special class have been reported[10]. A detailed study on 91 known 'natively unfolded' proteins has revealed that majority of natively unstructured proteins have 50 to 100 residues. A typical natively protein is characterized by: (a) a specific amino acid sequence with low overall hydrophobicity and high net charge; (b) hydrodynamic properties typical of a random coil in poor solvent or PMG conformation; (c) low level of ordered structure; (d) the absence of a tightly packed core; (e) high conformational flexibility; (f) ability to adopt relatively rigid conformation in the presence of natural ligand; and (g) a 'turn out' response to environmental changes with the structural complexity increase at high temperature or at extreme pH[17].

1.3 Biological Importance of being Natively Unfolded

From the functional point of view intrinsically disordered proteins have advantages like increased binding specificity at the expense of thermodynamic stability, regulation by proteolysis and the ability to recognize a range of proteins. The large occurrence of unstructured proteins in multicellular metazoans[20] has made researchers to envisage that unstructured proteins may be less sensitive to environmental perturbations and, therefore, may impart stability to complex regulatory networks that might otherwise be sensitive to changes in cellular conditions[15]. Two kinds of natively unfolded proteins can be distinguished: first, those in which the unfolded protein performs biological function in its natively unfolded state[8,14,21,22] and, second, those in which a natively unfolded protein acquires a compact structure upon binding to its ligand or any co-factors (synergistic folding)[23]. Excellent reviews have been written on this subject recently covering all the systems, uptil now discovered[13,15,16,18]. With increasing number of intrinsically unfolded proteins and detailed information on their biological function, need for an appropriate database was felt. Thus a web site *http://DisProt.wsu.edu* has been created by a group in Washington state university for deposition of data and advanced search on the pattern of PDB[18].

1.4 Unfolded State, Denatured State and Protein Folding

Detailed characterization of the unfolded state and consequent identification of the folding initiation sites in a given protein provide valuable insight into its folding mechanism[24]. Well-formed or transient residual structures in the unfolded state can be possible candidates for folding initiation sites[25]. Unfolded or partly unfolded states of globular proteins can be created by use of denaturants, which disrupt the non-covalent interactions and propel them to lose their biological activity. This state is referred to as the 'denatured state'. However, often, denaturation is not accompanied by complete unfolding of the protein, and the denatured state is an ensemble of conformations between native and completely unfolded states[10]. In the denatured state floppy molecular chains explore a far smaller number of

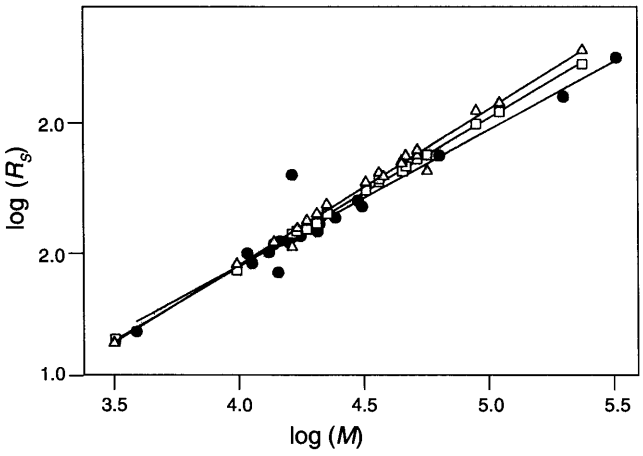

Fig. 2 Relation between hydrodynamic dimensions, R_s and molecular mass, M of denatured proteins produced by urea (□), and guanidine (Δ) and also of natively unfolded proteins (●). Interestingly, all the three types of denatured/unfolded proteins show a linear dependence with nearly identical slopes and intercepts. The data used to plot are taken from ref. [17] and fitting was performed using Sigmaplot 5.

different shapes than is theoretically available to them: between 100 and 1,000, whereas many millions are possible. They are like dancers executing a series of practicized moves rather than thrashing about at random.

Fig. 2 shows the relation between hydrodynamic characteristics and molecular mass of denatured proteins produced by urea, and guanidine and also of natively unfolded proteins. Interestingly, all the three types of denatured/unfolded proteins show a linear dependence with nearly identical slopes and intercepts[17]. This indicates that the study of urea and guanidine denatured states of globular proteins and of their pathway of renaturation, can provide valuable insights into *in vivo* folding of natively unfolded proteins.

2 Experimental Methods other than NMR for Studying *unstructured* Proteins

A number of physical techniques have been used for characterization of unfolded proteins to different degrees of detail[26]. In the following we give a brief description of these various methods, and NMR spectroscopy, which provides the maximal information is dealt with separately in the next section.

2.1 Protease Digestion

Protease digestion is one of the oldest techniques to get insight into protein structure and flexibility[27]. The digestion of possible cut site on protein is dependent on its flexibility and surface exposure. Proteolysis requires at least 10 residue long unfolded region[28]. Hypersensitivity to proteases is indicative of protein disorder[9]. However, this technique lacks residue specific resolution and requires use of other techniques for correct interpretation of data.

2.2 X-Ray Crystallography

X-Ray crystallography has been the method of choice for elucidation of 3D structures at atomic resolution of structured proteins. About 80 % of the structures deposited in Protein Data Bank are solved by X-Ray diffraction[7]. Missing electron density in the structure determined by X-Ray crystallography implies a disorder in the protein. B-factors also provide qualitative insights in to the dynamism in proteins. For an intrinsically unfolded protein or the denatured state of a globular protein, however, it is almost impossible to obtain a single crystal for detailed investigation.

2.3 Circular Dichroism

Circular dichroism (CD) provides secondary and tertiary structural information on a globular basis[29,30]. Far UV CD spectra from 190 nm to 230 nm provide estimates of secondary structure while near-UV CD show sharp peak from aromatic groups when protein is ordered. The combined use of far UV and near UV CD is used to distinguish whether a protein is ordered, molten globule or disordered. With the development of synchrotron based CD, it is possible to get insight into the folding motifs[31]. But CD lacks residue specific information and is only semi-quantitative. It fails to provide structural information for proteins containing ordered and disordered regions. Motional averaging limits the use of CD to investigate the structural propensities in the denatured state.

2.4 Size Determination

Determination of stokes's radius also helps in detection of disorder present in a protein. This hydrodynamic parameter is determined by gel-filtration, viscometery, SAXS, SANS, NMR spectroscopy, sedimentation and dynamic and static light scattering. The molecular volume, V is given as

$$V = \frac{M\bar{V}^{\Delta}}{N_O}$$

where, M is molecular weight, $\bar{V}^{\Delta}$ is the specific volume of a pure sample of the protein and N_O is the Avogadro's number. Abnormally large radii for a given

number of residues indicate presence of disorder. It is well documented that hydrodynamic radii of globular proteins increase by 15-20 % upon transformation to molten globule state[32].

2.5 Fluorescence Spectroscopy

Fluorescence spectroscopy provides valuable information on the local structural features around a fluorophore and more importantly on the dynamics in the disordered proteins. Line shape, quantum yield and position of λ_{max} in the emission and excitation spectra, fluorescence and anisotropy life time of fluorophore present on a protein probe the environment of the fluorophore, which in turn reflects on the structure and dynamics of proteins[33]. Tryptophan is an intrinsic fluorophore present in proteins and in case where it is absent or the number of tryptophans is large, protein is labelled with synthetic fluorophore for fluorescence measurements. Fluorescence resonance energy transfer (FRET) between an acceptor and a donor which are synthetically attached at specific sites on the protein can be used to derive the distance between the two. Time resolved measurement of FRET provides additional insight into the dynamics of unfolded proteins. Analysis of FRET data, however, becomes difficult when there are several donor-acceptor pairs present on the protein[34, 35].

3 NMR Methodology of Investigating Unfolded or Partly Folded Proteins

In addition to its normal application for determination of 3D structure at atomic resolution of folded proteins in solution, Nuclear Magnetic Resonance (NMR) spectroscopy is unparalleled in its ability to provide detailed structural and dynamical information of unfolded and flexible proteins[36]. The proton chemical shift dispersion and line width give first-hand information on the state of proteins; poor proton chemical shift dispersion and sharp lines are indicative of disorder. Fig. 3 shows the amide proton and ^{15}N dispersions in the HSQC spectra for a properly folded HIV-I protease (A) and for the same protein unfolded by 6 M guanidine hydrochloride (B). Clearly the amide proton dispersion in the unfolded protein is very narrow compared to that in the folded protein. The ^{15}N chemical shift dispersion is, however, fairly good in both the states.

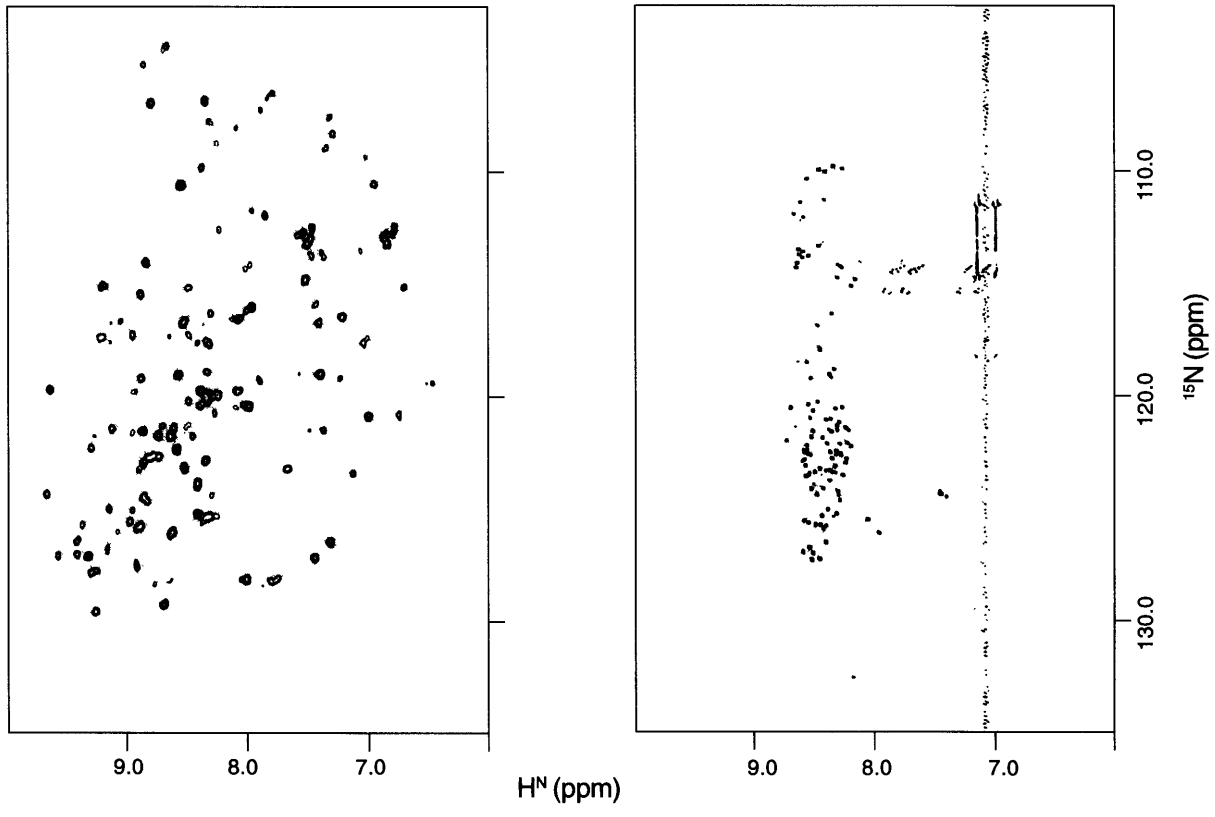

Fig. 3 Amide proton and ^{15}N dispersions in the HSQC spectra for a properly folded HIV-I protease (A) and for the same protein unfolded by 6 M guanidine hydrochloride (B)[70]

3.1 Resonance Assignment

Sequence specific resonance assignment is the first step in detailed NMR characterization of proteins. With the advent of ultra high field spectrometers, developments in NMR methodology, and advancements in isotope labelling techniques many proteins have been assigned. The assignment procedure has followed either the NOESY based approach[37] or the triple resonance experiments based heteronuclear approach[38] or combinations thereof. These have been reviewed extensively in the past[39-43] and we do not wish to repeat that here. Briefly, the assignment procedure, in general, involves sequential walk along the backbone of the chain, making use of H^N, C^α, C^β, and C' chemical shifts in the different planes of the 3D spectra[38, 44]. The side chains are assigned using a different set of heteronuclear 3D experiments. Sequential and long range NOE connectivities add further support to the assignments and also provide structural informations. However, this protocol has difficulties with unfolded or partly folded proteins because of poor chemical shift dispersions of proton and aliphatic carbon atoms. Even so, a few proteins have been successfully investigated following these same strategies as have been used for folded proteins. ^{15}N and ^{13}C filtered three and four dimensional NOESY and TOCSY based techniques were used in all these investigations. However, this is extremely slow and error prone. Thus, new methodologies are appearing in the literature which exploit the relatively better ^{15}N and ^{13}C' chemical shift dispersions in unfolded proteins. We describe here only those methods, briefly, for ready reference and utility. These techniques provide the backbone H^N and ^{15}N assignments and following them the other carbon and proton assignments can be readily obtained from the standard 3D experiments commonly used for the folded proteins[39-43].

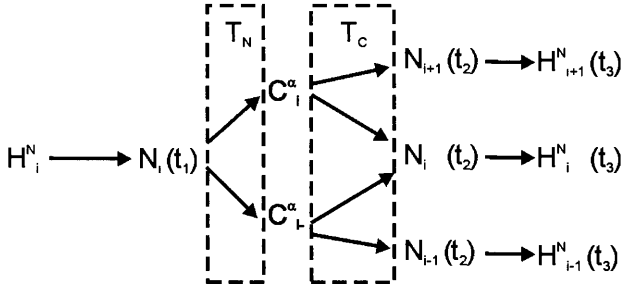

Fig. 4 Schematic diagram showing the magnetization transfer pathway in the HNN experiment. T_N and T_C are the delays during which the transfers indicated by the arrows take place in the pulse sequences[46]

3.1.1 HNN

HNN is a three dimensional triple resonance experiment which employs the magnetization transfer pathway as shown in Fig. 4. This basic transfer pathway was also used earlier by other authors[45]. However, several improvements in the implementation of the pathway resulted in substantial gains in sensitivity and spectral resolution[46] and also resulted in special features extremely useful from assignment point of view.

The HNN spectrum has the following characteristic features: First, the experiment exploits the ^{15}N chemical shift dispersions which are generally very good, along two of the three dimensions, second, the spectrum directly displays sequential amide and ^{15}N correlations between three consecutive residues along the polypeptide chain, and, third, the sign patterns of the diagonal and sequential peaks originating from any residue are dependent on the nature of the adjacent residues, especially the glycines and the prolines. Fig.5 shows the characteristic patterns for the various distinct triplets of residues. These lead to so called 'triplet fixed points' which serve as starting points and/or check points during the course of sequential walks, and explicit side chains assignment becomes less crucial for unambiguous backbone assignment. Fig. 6 shows a schematic walk protocol through the HNN spectrum and an experimental demonstration is shown in Fig. 7. These features significantly enhance the speed of data analysis, reduce the amount of experimentation required and thus result in a substantially faster and unambiguous assignment[47]. The HNN technique has been further extended to four dimensions by including C^α labelling along one of the four dimensions (unpublished results).

3.1.2 HN(C)N

HN(C)N is again a three dimensional triple resonance experiment which employs the magnetization transfer pathway as shown in Fig. 8. This basic transfer pathway was also used earlier by other authors[48-51]. However, again, several improvements in the implementation of the pathway resulted in substantial gains in sensitivity, resolution and special spectral features[46].

The HN(C)N spectrum shows H^N-^{15}N correlations between three consecutive residues $i - 1$, i and $i + 1$ in the different planes of the 3D spectrum. Under the choice of standard experimental data acquisition parameters, the patterns of peaks

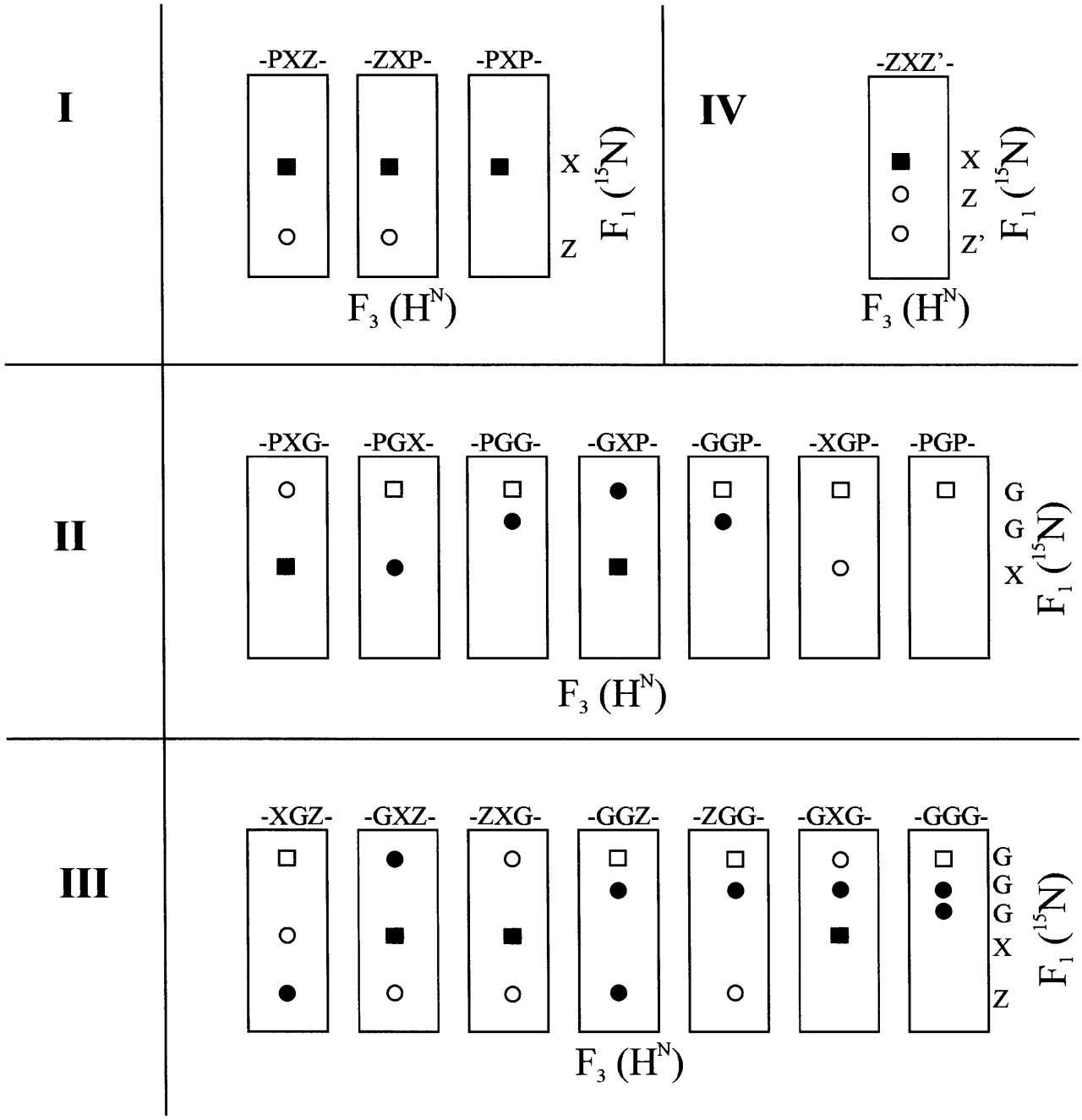

Fig. 5 Schematic patterns in the F_1-F_3 planes at the F_2 chemical shift of the central residue in the triplets mentioned on the top of each panel, in the HNN spectra for various special triplet sequences of categories I-IV. X, Z, Z' is any residue other than glycine and proline. Squares are the diagonal peaks and circles are the sequential peaks. Filled and open symbols represent positive and negative signals respectively. In all cases the peaks are aligned at the F_3 (H^N) chemical shift of the central residue[47]

depending on whether the *i* th residue is a glycine or otherwise are shown in Fig. 9. From these, it is clear that whenever there is a glycine or a proline, there is a distinct change in the sign patterns. These serve as triplet fixed points as in the case of HNN spectrum. The sequential walk through the HN(C)N is schematically shown in Fig. 10. It relies exclusively on the (F_1, F_3) planes of the HN(C)N spectrum and the special peak patterns described above serve as start points and/or check points during the sequential walk[52]. The essential ingredients of the sequential walk protocol are indicated as, 'start', 'continue', 'check' and 'break' in the Figure. An illustrative application is shown in Fig. 11.

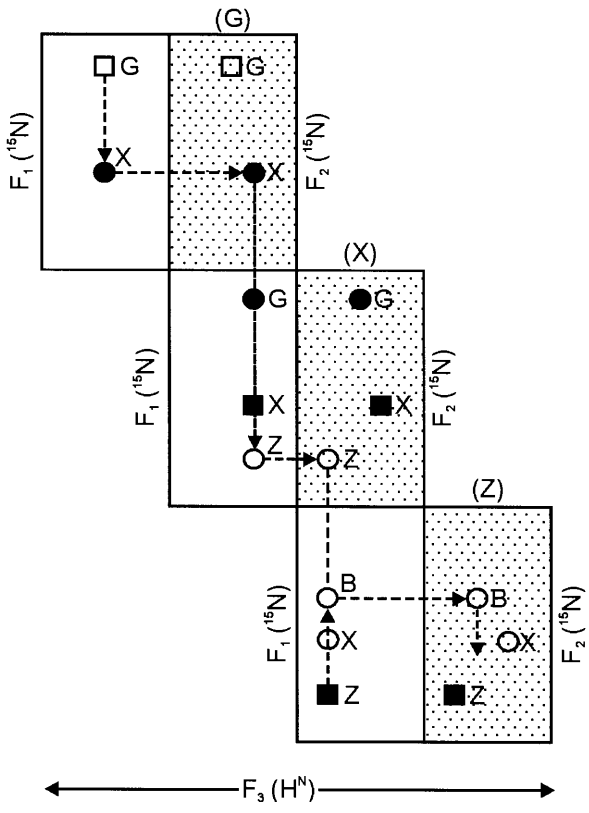

Fig. 6 The protocol for sequential walk through the HNN spectrum using an illustrative sequence PGXZB where X, Z, B can be any residue other than glycine and proline. Pairs of F_1-F_3 and F_2-F_3 (light shaded) planes belonging to the three residues, G, X and Z are stacked in an appropriate alignment so that the F_1-F_3 plane of one residue stacks over the F_2-F_3 plane of the neighbouring residue. Squares are diagonal peaks and circles are sequential peaks. Filled and open symbols are positive and negative peaks respectively. Note that the sequence chosen includes the PGX and GXZ special triplets and the signs of the peaks have been drawn accordingly. The dashed line indicates the sequential walk. The vertical line at the amide position of a particular residue goes from the diagonal peak to the sequential peak in the F_1-F_3 plane and identifies the ^{15}N chemical shift of the sequentially connected residue (G to X, X to Z and Z to B in the G, X and Z planes respectively), whereas, the horizontal line going from the F_1-F_3 plane to the F_2-F_3 plane of a given residue enables identification of the amide chemical shift of the sequentially connected residue. Note that the diagonal peaks in one plane become sequential peaks in the vertically neighbouring plane[47]

3.1.3 3D (H)N(CO-TOCSY)NH, 3D(H)CA(CO-TOCSY)NH and 3D (H)CBCA(CO-TOCSY)NH

Three heteronuclear 3D NMR experiments namely, (H)N(CO-TOCSY)NH, (H)CA(CO-TOCSY)NH and (H)CBCA(CO-TOCSY)NH which make use of favourable ^{15}N chemical shift dispersion in ^{13}C, ^{15}N labelled unfolded proteins have been developed[53]. The magentization transfer pathways employed by these are shown in Fig 12. They exploit the slow transverse ^{15}N relaxation of unfolded proteins and use carbonyl carbon homonuclear isotropic mixing to transfer magnetization sequentially along amino acid sequence. The backbone carbonyl carbons have long transverse relaxation, since they have no directly bound protons and relax almost exclusively due to chemical shift anisotropy (CSA) making them good choice for magnetization transfer through the backbone using $^3J_{C'C'}$ mediated isotropic mixing. These approaches have been successfully used to obtain assignment of recombinant human prion protein, hPrP(23-230).

3.1.4 3D (HCA)-CO(CA)NH

3D (HCA)-CO(CA)NH experiment correlates carbonyl carbons with amide ^{15}N and ^1H nuclei of the *i* and *i+1* residues[54]. The magnetization transfer pathway of this experiment is shown in Fig 13. It uses the well dispersed ^{13}CO resonances and provides a means for sequential assignment of backbone resonances in analogy to the HNCA experiment, i.e., pairs of ^{13}CO chemical shifts are linked by common correlations to amide proton and nitrogen chemical shifts. This has been used for obtaining resonance assignment in apomyoglobin under different unfolding conditions, relying on favourable dispersions of ^{15}N and ^{13}CO chemical shifts[55-57]. Further improvement in resolution of peaks has been obtained by extending to the fourth dimension in 4D (H)CACOCANH spectrum, with C^α chemical shifts.

3.2 Structural Propensities and Residual Structures in Denatured Proteins

3.2.1 Secondary Chemical Shifts and CSI

$^1H^\alpha$ and ^{13}C chemical shifts of C', C^α and C^β are useful indicators of secondary structure in folded proteins[58-61] as they are primarily determined by the backbone ϕ, ψ dihedral angles[62]. In order to detect the presence of residual structures in unfolded proteins, chemical shift deviations from random coil values for $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ resonances are calculated. A positive deviation of $^{13}C^\alpha$ and $^{13}C'$ chemical shifts from their random coil values is indicative of presence of helical segment while converse is true for $^{13}C^\beta$ chemical shifts. However, there are more than one set of random coil values published in the literature and these differ because of the experimental conditions used in arriving at those values. Wishart *et al.* 1995 used 1M urea, pH 5 and 25 °C for their experiments on the peptides chosen, whereas Schwarzinger *et al.*[63,64] derived another set

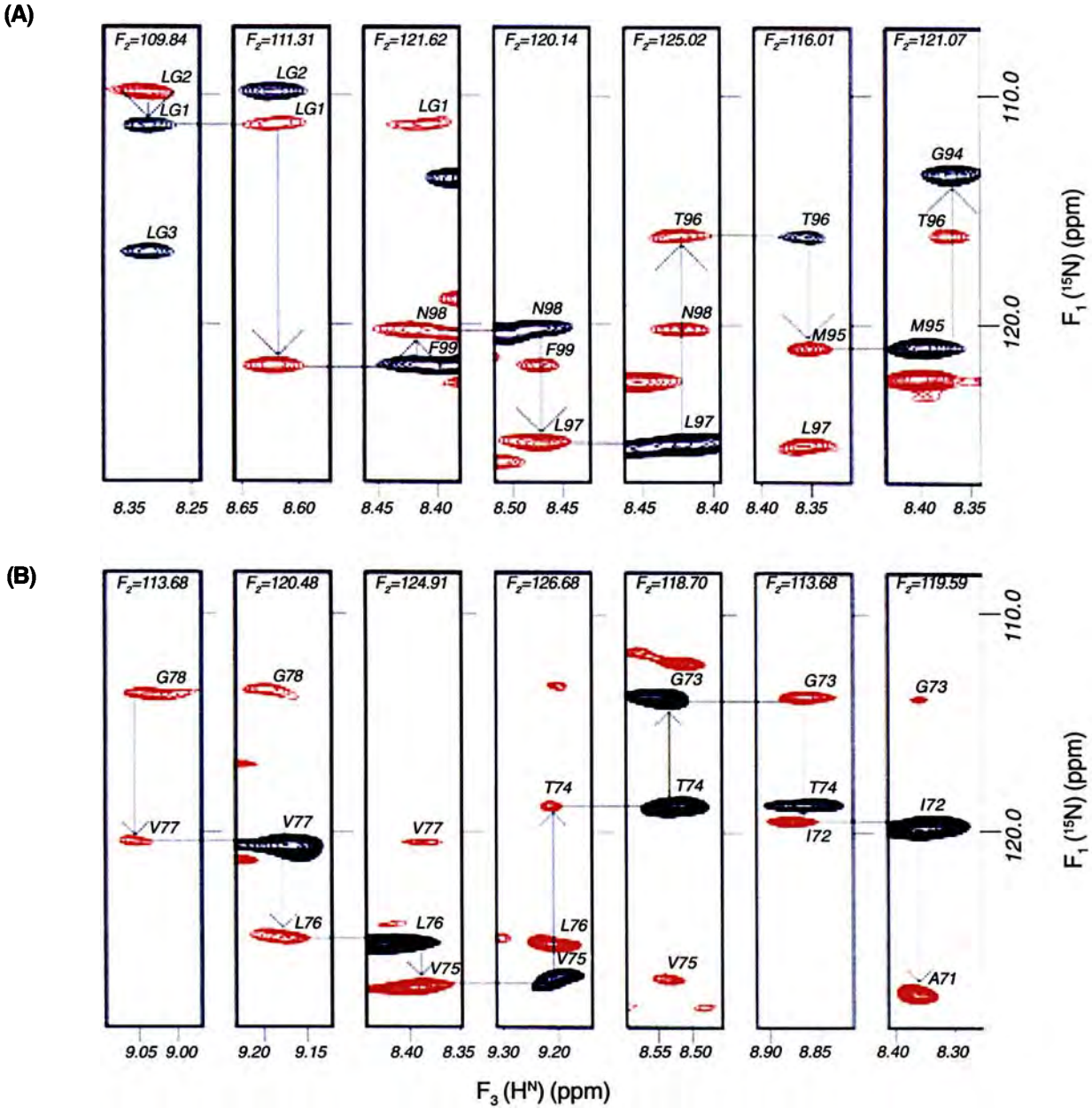

Fig. 7 Illustrative sequential walks through the HNN spectra of HIV-1 protease for unfolded protein (A) and folded protein (B). Black and red contours are positive and negative peaks respectively. Several of the planes are fixed points: GGS, FGG, GMT in (A) and VGP, GTV, IGT in panel (B)[47]

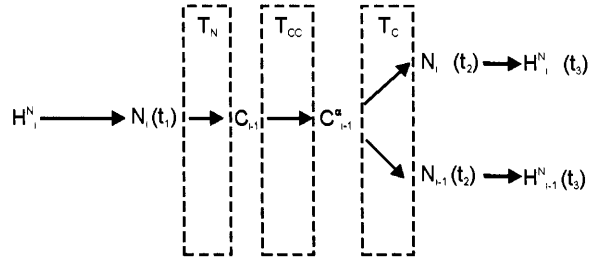

Fig. 8 Magnetization transfer pathway employed in the HN(C)N experiment. The magnetization originates on the amide proton of the i^{th} residue and is finally detected on the amide protons of the i and i - 1 residues. t_1, t_2, t_3 are the time variables of the three dimensional experiment. T_N, T_{CC} and T_C are the delay periods for effecting the magnetization transfers[46]

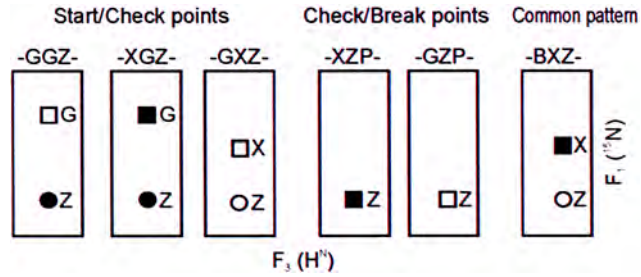

Fig. 9 Schematic peak patterns in the F_1-F_3 planes of HN(C)N spectrum and the corresponding different triplets of residues. In every strip the peaks occur at the H^N chemical shift (F_3) of the central residue in the triplet indicated on the top. B and X can be any residue other than P and G; Z can be any residue other than P. Squares are diagonal (F_1=F_2) peaks and circles are sequential peaks. Filled and open symbols represent positive and negative signs respectively. The patterns involving G and P serve as start and /or check points during a sequential assignment walk through the spectrum. The pattern, which does not involve a G or a P, is the most common one occurring through the sequential walk[52]

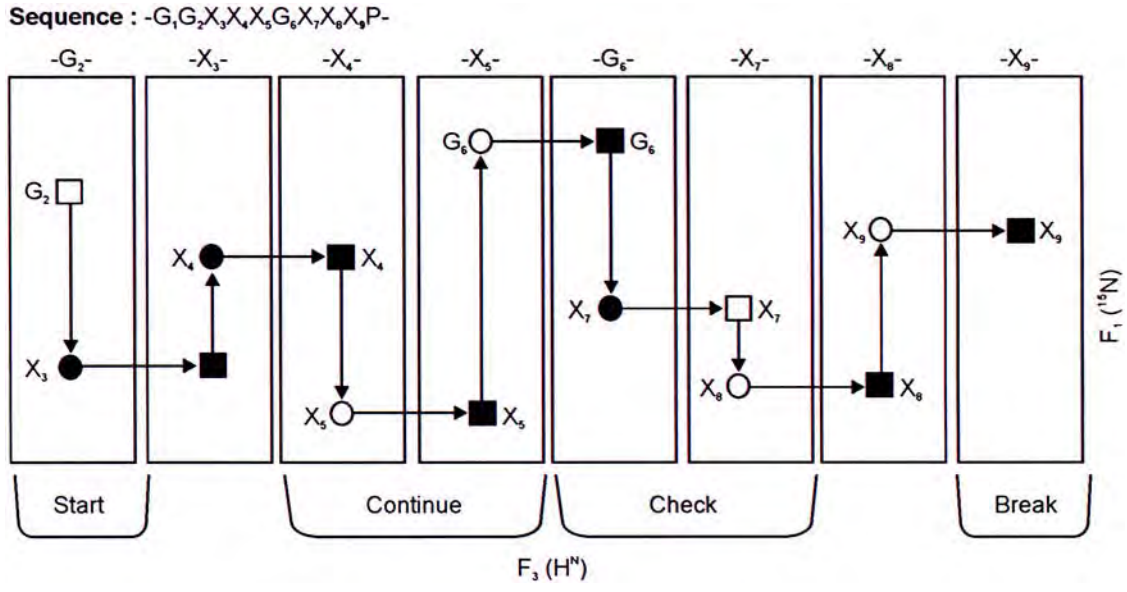

Fig. 10 The schematic sequential walk protocol through the F_1-F_3 planes of the HN(C)N spectrum of a protein. An arbitrary amino acid sequence is chosen to illustrate the start, continue, check and break points during the sequential walk. The peak patterns are drawn as per the schematic in Fig. 9 and the residue identified on the top of each strip identifies the central residue of the triplet[52]

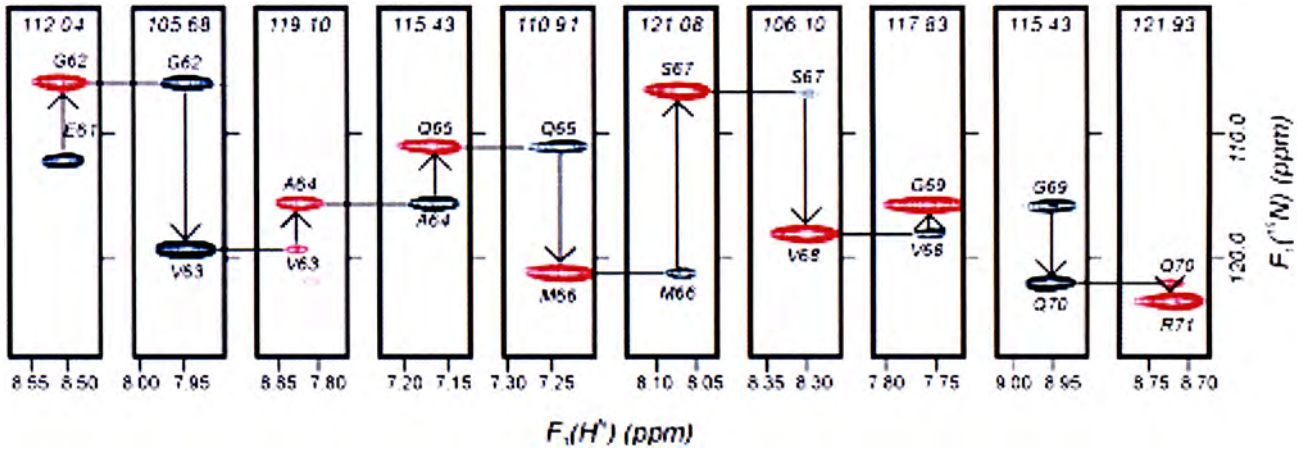

Fig. 11 An illustrative stretch of sequential walk through the HN(C)N spectrum of the FKBP protein. A sequential peak in one plane joins to the diagonal peak in the adjacent plane on the right. Note that the panels of G62, V63, G69 and V70 constitute the check points in this sequential walk. The numbers at the top identify the F_2 chemical shifts, which help identification of the diagonal peaks[52]

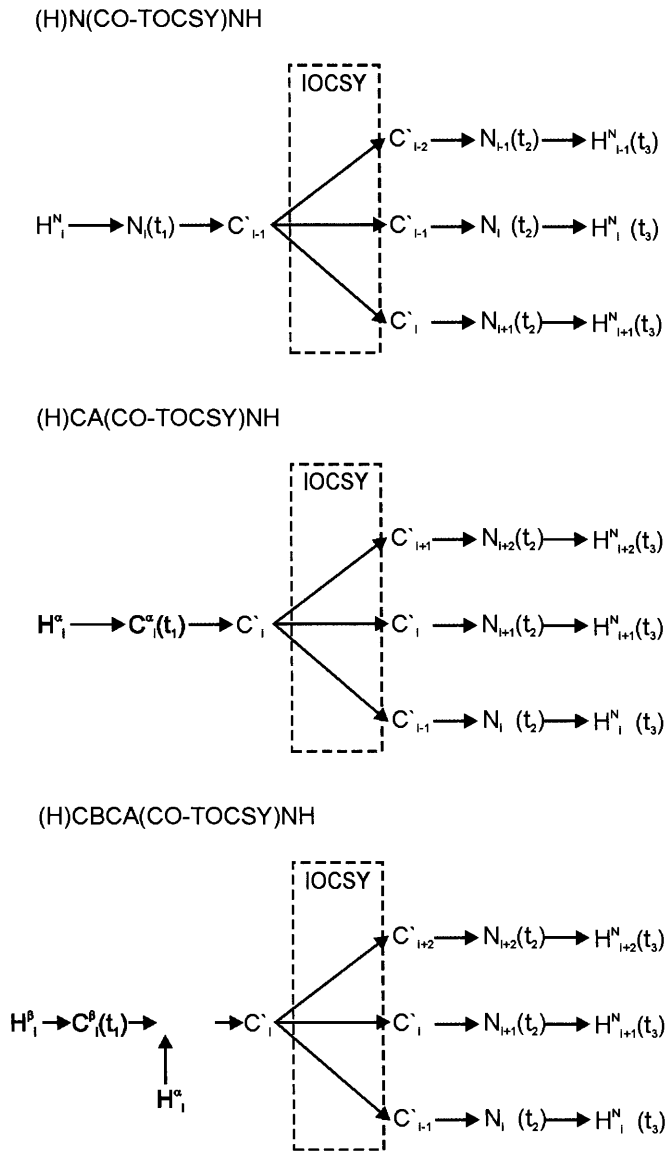

Fig. 12 Magnetization transfer pathway in the three-dimensional triple resonance experiment (A) (H)N(CO-TOCSY)NH, (B) (H)CA(CO-TOCSY)NH and (C) (H)CBCA(CO-TOCSY)NH. The arrows indicate the transfer of magnetization through the pulse sequences[53]

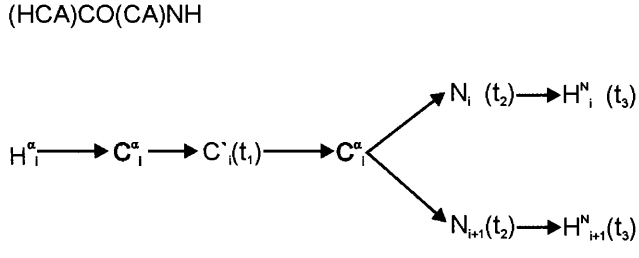

Fig. 13 Magnetization transfer pathway in the three-dimensional triple resonance experiment (HCA)-CO(CA)NH. The arrows indicate the transfer of magnetization through the pulse sequence[54]

of values from peptide spectra recorded in 8M urea, pH 2.3 and 20 °C so as to obtain a better data set for their apomyoglobin protein. Schwarzinger *et al.*[64] also observed that corrections will have to be applied for sequence effects and provided a set of rules for these corrections. $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shifts are better indicators of backbone conformational propensities than $^{13}C^{\beta}$ secondary shifts.

3.2.2 H^N-H^{α} Coupling Constants

Coupling constants provide valuable secondary structural information in proteins. The $^3J_{HN,H\alpha}$ coupling constant is sensitive to the dihedral angle ϕ, and thus provides a probe for backbone conformational

preferences[37]. β-structures are characterized by large H^N-H^α coupling constant values in the range 8-10 Hz, while α-helical structures are characterized by values in the range 3-5 Hz. In unfolded proteins, however, the heterogeneity and conformational averaging leads to average values of 6-7.5 Hz[65]. Nonetheless, values significantly different from these average random coil values would indicate definite propensities for the structures.

The most common method for measuring the H^N-H^α coupling constants relies on the HNHA experiment[66] or the HNCA-J experiment[67]. In the former the coupling constants are derived from the ratios of the diagonal to cross peak intensities in the different ^{15}N planes of the 3D spectrum, and in the latter they are measured from peak multiplet structures. While these work well for folded proteins with well-resolved peaks, they have serious problems for unfolded proteins where chemical shift dispersion is poor and reliable estimation of the peak intensities or separations is difficult. Relaxation losses also contribute to the uncertainties in intensity measurements in HNHA. Recently these coupling constants have been measured from a high resolution ^1H-^{15}N HSQC spectrum, where this information is contained in the fine structure of the correlation peaks[68-70]. This requires less time for data acquisition and analyses and is well suited for unfolded proteins where lines are very sharp.

3.2.3 Amide Proton Temperature Coefficient

The temperature dependence of the H^N chemical shift, that is temperature coefficient, provides an estimate of the involvement of the amide proton in hydrogen bonding[71,72]. Random coil temperature coefficients determined for residue X in a series of GGXGG unstructured peptide models at pH 5 over the temperature range 278 to 318 K are around -8 ppb/K[73]. Lowered temperature coefficients of the residues showing a deviation of ≥ 1 ppb/K, indicates the involvement of these residues, at least transiently, in hydrogen bonding. The H^N temperature coefficients cannot distinguish between intra- and inter-molecular hydrogen bonds, which can complicate the analyses in aggregated form of unfolded proteins.

3.2.4 Hydrogen Exchange

An H→D HX experiment checks if any of the assigned amide protons shows protection from hydrogen exchange. This is an extremely useful parameter to understand the compactness and hydrogen bonding in disordered or partially folded proteins. The exchange reaction is generally described by the following model:

$$(N - H)_{cl} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} (N - H)_{op} \xrightarrow{k_{rc}} (N - D)$$

The protein is believed to be in two sets of conformations, closed and open, in equilibrium with each other. However, solvent exchange occurs only from the open state of the protein. The exchange characteristics are a reflection on the solvent accessibilities of the individual amide protons in the folded protein, on the one hand, and on the folding/unfolding equilibria on the other[74,75]. The hydrogen bonded amide protons exchange much slower compared to the non-hydrogen bonded ones. Thus hydrogen-deuterium (H/D) exchange studies give considerable insight into structure, stability, folding, dynamics and intermolecular interactions in protein systems in solution. Under conditions where $k_{cl} >> k_{op}$, the measured exchange rate k_{ex} is $K_{op}k_{rc}$ where K_{op} is the equilibrium constant for structural opening and k_{rc} is the intrinsic exchange rate. Then the free energy change for structural opening is given by

$$\Delta G_{HX} = -RT \ln\left(\frac{k_{ex}}{k_{re}}\right)$$

Several studies suggest that the most stable residues exchange following a global unfolding, while the residues with intermediate stability exchange due to local fluctuations in the native state or following a partial unfolding, and the least stable residues exchange directly with the solvent in the native state. Accordingly, the unfolding free energy values measured at the various amide sites would actually reflect a series of different unfolding events through the structure of the protein.

3.2.5 Nuclear Overhausser Effect

Nuclear Overhausser effect (NOE) provides inter atomic distance information in unfolded proteins as it does in folded proteins. It provides information on the secondary structures and other long-range interaction present in proteins. The $d_{\alpha N}(i, i+1)$, $d_{NN}(i, i+1)$ and $d_{\beta N}(i, i+1)$ NOEs provide the information on ϕ and ψ dihedral angles preferences while medium range $d_{\alpha N}(i, i+2)$, $d_{\alpha N}(i, i+3)$ and $d_{\beta N}(i, i+3)$ NOEs indicate the presence of secondary structures. Though an uphill task to assign due to poor resonance dispersion, distributions of structures, anisotropic tumbling and spin diffusion, the long-range NOEs point to definite presence of structural elements in an ensemble of disordered protein conformations. 3D NOESY-HSQC

and 3D HSQC-NOESY-HSQC are routinely used for NOE connectivities, but overlap of proton resonances limits their use for unfolded proteins[65]. For overcoming these problems, NOESY based triple resonances experiments have been developed, which exploit the dispersion of ^{15}N and $^{13}C'$ resonances to help resolve ambiguities in aliphatic ^{1}H and ^{13}C chemical shifts[76]. However, quantitative interpretation of NOE in unfolded proteins is difficult due to conformational averaging.

3.2.6 Paramagnetic Relaxation

Paramagnetic relaxation probes like nitroxide, which enhance nuclear spin relaxation by their paramagnetic centre, enable measurement of long-range distances. This approach has been used for characterization of long-range structural elements in a denatured state of staphylococcal nuclease[77-79] and apomyoglobin[80]. Conformational averaging may bias this approach towards underestimation of contact distances.

3.2.7 Residual Dipolar Couplings

Residual dipolar couplings (RDC) between nuclei, observable in partially aligned molecules in dilute orienting media, cause a significant improvement in the precision of structure determination of biological macromolecules in aqueous solutions. In addition to providing new constraints, which enhance the experimental structural input, the RDCs are unique in defining the orientations of specific dipole-dipole vectors in the molecular frame and thus provide long-range constraints. This becomes especially useful in defining domain orientations in multi-domain proteins. The historical developments, theoretical basis and the applications to different protein and nucleic acid segments have been recently reviewed [81]. Although there are complexities involved in converting RDCs measured in unfolded proteins into sets of orientational constraints[82], the structural information given by RDCs is independent of distances. The approach has been successfully used for characterization of topology of urea denatured staphylococcal nuclease. The measurement of RDCs has indicated the presence of long range native like structures unaffected by local dynamism. Many of the residues were found to maintain a relatively fixed orientation with reference to a single reference axis.

3.3 Ensembles of Structures in the Unfolded State

Though various NMR parameters as listed above provide useful residue-level insight into the secondary structures and structural propensities in unfolded states of proteins, the calculation of 3D structures remains an almost utopian task, as, measured NOEs, *J*-couplings and RDCs are ensemble averaged quantities of rapidly inter-converting conformers that do not have as simple a relationship with structures(s) as in folded proteins. Choy and Forman-Kay[83] have developed an algorithm called ENSEMBLE for calculation of ensembles of structures. It optimises the population weights assigned to each structure on the basis of experimental properties derived from NOEs, *J*-couplings, ^{13}C chemical shifts, translational diffusion coefficients and tryptophan environment obtained from NMR and fluorescence spectroscopy. The approach was applied to drkN SH3 domain and five sets of state ensembles were calculated with all having similar average hydrodynamic properties. The percentage of residues in the β region decreased by 10% while there was 50 % loss in the β strand content. It also indicated presence some residual native and non-native structures.

3.4 Dynamics

NMR is unequaled in its ability to provide information on residue specific dynamics in proteins. In unfolded proteins, a restricted motion indicates higher propensity to form structure or presence of residual structure. The longitudinal and transverse relaxation rates (R_1 and R_2) of backbone ^{15}N nuclei as well as the ^{15}N-^{1}H steady-state heteronuclear NOE are useful probes of protein backbone dynamics and overall molecular tumbling motions[84,85]. While all the three relaxation parameters are sensitive to motions on a picosecond to nanosecond time scale, the ^{15}N-^{1}H NOE is the most sensitive, to these high frequency motions of the protein backbone. On the other hand, transverse relaxation is very sensitive to slow time scale (micro- to millisecond) motions and conformational exchange, in addition Negative values of heteronuclear NOEs indicate the occurrence of large-amplitude motions on a sub-nanosecond time scale, which are very frequent in unfolded proteins. Similarly, high R_2 values suggest the presence of significant conformational exchange contributions in proteins. A detailed analysis of relaxation data provides important insight into the possible nucleation sites of protein folding. It can also throw light on residues interacting with their ligands.

Model-free approach is the commonly used approach for analysis of relaxation data of folded proteins[86, 87]. This is based on the assumption of

separability of internal and globular motions. Thus, dynamics is described in terms of an overall rotational correlation time τ_m, an internal correlation time τ_c, and an order parameter S^2 describing the amplitude of the internal motions. It provides correlation between dynamics and a set of intuitive physical parameters. However, the model free approach has limitations with regard to unfolded proteins because of anisotropic tumbling and multitude of uncorrelated motions in these systems. On the other hand, another approach, namely, reduced spectral density analysis, which assumes a distribution of correlation times, is more appropriate for analysis of the ^{15}N relaxation data for unfolded proteins[65,88]. Three spectral density functions $J(0)$, $J(\omega_N)$ and $J(\omega_H)$ are calculated as described by Lefèvre *et al.*[89]. Of these, $J(\omega_H)$ is largely determined by heteronuclear NOEs and is most sensitive to higher frequency motions of the protein backbone. $J(\omega_N)$ is dominated by R_1 and $J(0)$ is dominated by both R_1 and R_2. Thus, $J(0)$ is sensitive to both nanosecond time scale motions and contributions from slower micro- to millisecond exchange processes. Assuming a linear correlation between $J(0)$ and $J(\omega_N)$ two empirical parameters α and β are calculated from the slope and y-intercepts. These are then used to calculate the time constants characterizing various motions of the protein by solving the following cubic equation in τ[89]:

$$2\alpha\omega^2\tau^3 + 5\beta\omega^2\tau^2 + 2(\alpha - 1)\tau + 5\beta = 0$$

where ω is the larmor frequency of ^{15}N nuclei. A similar analysis can be performed for $J(0)$ and $J(\omega_H)$ relation.

A general framework for studying dynamics of folded as well as unfolded states of protein has been presented by Prompers and Brüschweiler[90]. This does not require the separability assumption. The approach, named isotropic reorientational eigenmode dynamics (iRED) depicts correlated dynamics of different polypeptide parts together with mode-specific correlation times.

Cross-correlated dipole-dipole spin relaxation is used to explore molecular dynamics and structural properties[91-94]. It provides information directly on the dynamics of the interacting spins, if their relative orientations are known. More recently $^1H^{\beta}$-$^{13}C^{\beta}$ dipole-dipole cross-correlated spin relaxation has been used to probe dynamics of unfolded proteins[95]. The dipole-dipole cross-correlated spin relaxation rate for two ^1H-^{13}C dipoles, $\Gamma_{HC1,\,HC2}$, in the CH_2 group is obtained from the CBCA(CO)NNH spectrum. $\Gamma_{HC1,\,HC2}$ can be written in terms of spectral densities, which are evaluated in the context of a particular motional model[96]. $\Gamma_{HC1,\,HC2}$ is given by the relation

$$\Gamma_{HC1,\,HC2} = -\left(\frac{0.25}{T}\right) \ln\left\{4 * I_U(T) * \frac{I_D(T)}{I_C(T)^2}\right\}$$

where I_U, I_D and I_C are the intensities of the upfield, downfield and central components of the $^{13}C^{\beta}$ triplet respectively, and T is the duration of the ^{13}C constant time evolution during which cross-correlated spin relaxation occurs[95]. A large $|\Gamma_{HC1,\,HC2}|$ is indication of slow motion indicating some tertiary contacts. The buried residues also exhibit large values.

4 *'Residual structures'* in the Unfolded State: Specific Cases

The number of proteins which have been investigated in the unfolded state till date is rather small compared to that for folded proteins, because of the technical difficulties in the analysis of their spectral and complications in the interpretation of the NMR data. In the early days the same techniques as were used for the folded proteins were used for unfolded proteins as well and this had limited success for reasons already elaborated in the earlier sections. It is however envisaged that the newer methods, coupled with technological advances would facilitate larger number of investigations in the years to come. In the following, we briefly discuss the results obtained in the few proteins studied so far.

4.1 The Phage 434 Repressor Protein

This was the first protein for which sequence specific backbone assignment was obtained in the unfolded state[97-98]. The assignments in 7 M urea were obtained using 2D TOCSY relayed ^{15}N HSQC and 2D NOE relayed ^{15}N HSQC spectra. COSY was used for assignment of other aliphatic ^1H resonances. Comparing the assignment at 0 M urea and 7 M urea, coexistence of both the native and unfolded forms was observed at 4.2 M urea having exchange life time of about ~ 1s. Deviation from random roil chemical shifts indicated the presence of some residual structure. Using the high density of NOE constraints for segment 53 to 60 in the protein, structure calculation was done for the segment. The structure showed similar structural feature as in folded form and hydrophobic cluster was preserved. The authors speculated the segment as possible nucleation site for folding of this protein[99].

4.2 Fibronectin Binding Protein

A 130 residue fragment (D1-D4) taken from a fibronectin-binding protein of *Staphylococcus aureus*, which contains four fibronectin-binding repeats and is unfolded but biologically active at neutral pH, has been studied extensively by NMR spectroscopy. A combination of 3D NOESY-HSQC, 3D TOCSY-HSQC and 3D HSQC-NOESY-HSQC experiments was used to obtain sequence specific backbone assignments. Assignment for D2-D3 was first obtained to overcome the difficulties posed by high sequence homology between different D repeats. The secondary chemical shifts, $^3J_{HN\text{-}H\alpha}$ couplings and absence of long range NOEs were indicative of lack of persistent secondary and tertiary structures. ^{15}N relaxation analysis was done to understand the differences in dynamical properties of residues in different domains. It was finally concluded that the sequence involved in binding has a high propensity for populating the extended conformation. This is likely to allow a number of both charged and hydrophobic groups to be presented to fibronectin for highly specific binding[100].

4.3 Lysozyme

Structural and dynamical properties of oxidized and reduced forms of hen lysozyme in 8 M urea were studied[101]. ^{15}N resolved NOESY and TOCSY based experiments were used for backbone assignments in both oxidized and reduced forms. Resonances for the residues near the disulfide bridges were absent due to line broadening in oxidized form indicating slow motions. Deviation from random coil behaviour was observed for residues W62, W63, W108, W11 and W123 on the basis of amide and H^α chemical shifts and high $R_1\rho$ relaxation rate, indicating the presence of hydrophobic clusters. A number of medium range NOEs were observed in the region corresponding to A, B, D, and C-terminal 3_{10} helices in the native protein. This suggested that the oxidized form could resemble a molten globule[101]. Hennig *et al.* measured ^3J(C', Cγ), ^3J(N, Cγ), ^3J(C', C') and ^3J(C', C$^\beta$) coupling constants in denatured hen lysozyme in 8 M urea pH 2.0[102]. It provided insight into the side chain χ_1 torsion angle population in the denatured state. The fractional populations of the –60°, 60° and 180° χ_1 rotamers were derived from the measured coupling constants. This provides information on the side chain conformational preferences, which depends on the variations in the electrostatic and steric properties of the side chains. H^N and H^α chemical shifts and R_2 rates indicated the presence of extensive hydrophobic clusters in the denatured state of lysozyme[103].

4.4 FKBP

Detailed structural characterization has been done on the FK 506 binding protein (FKBP) unfolded in urea and guanidine hydrochloride[104]. Sequence specific ^1H, ^{15}N and ^{13}C resonance assignments for FKBP in 6.3 M urea and 2 M guanidine hydrochloride were obtained using 3D TOCSY-HSQC and 3D H(C)(CO)NH-TOCSY[105] experiments. The measured C^α and ^1H chemical shift, H^N-H^α coupling constants, chemical exchange and ^{15}N relaxation rates indicated extensive conformational averaging for FKBP in urea. The presence of medium range NOEs indicated presence of helical conformation for some residues. This disagreement between NOE and other parameters was explained considering the r^6 dependence of NOE; short ^1H-^1H distances corresponding to a particular secondary structure give strong NOEs even when present in a relatively small population of molecules while other parameters are averaged values. The medium range NOEs patterns were different in the FKBP in 2 M guanidine hydrochloride, possibly due to dependence of the helix formation on the salt concentration.

4.5 Apomyoglobin

Structural and dynamic properties of sperm whale Apomyoglobin have been extensively studied in different denaturing conditions[55-57]. Sequence specific resonance assignments for Apomyoglobin denatured under low salt condition at pH 2.3 was obtained using (HCA)-CO(CA)NH spectrum which uses the superior $^{13}C'$ chemical shift dispersion. The HNCO spectrum was used to distinguish interresidue and intraresidue peaks in (HCA)-CO(CA)NH spectrum. Standard TOCSY-HSQC, CBCA(CO)NH and C(CO)NH-TOCSY spectra were used for confirmation of the assignment. The presence of small population of native like helix in two out of three helical regions stabilized in the molten globule state was concluded on the basis of downfield shifted sequence corrected $^{13}C^\alpha$ and $^{13}C'$ secondary shifts, motional restriction from ^{15}N relaxation data and slow amide proton exchange rate[57].

Detailed studies have also been done on the pH 4 molten globule state of Apomyoglobin[55] and also its helix destabilizing mutants[57]. Standard triple resonance 3D experiments in conjunction with 3D HSQC-NOESY-HSQC were used to obtain assignments. Secondary chemical shifts, $H_i^N - H_{i+1}^N$NOEs,

temperature coefficients revealed the presence of large native like helical regions and this was supported by the dynamics data, especially R_2 and corresponding $J(0)$ values which are sensitive to slow motions[57]. Study of molten globule state at pH 4.1 of helix destabilizing mutant of apomyoglobin revealed significant differences in the distribution of helical segments in the protein. Though the overall helical content in other regions was same, a loss of helical structure was observed in the regions where destabilizing mutation was made. The acid denatured state of apomyoglobin has also been studied by paramagnetic spin labelling. Nitroxide coupled to mutant cysteine residues were used as probes of chain compaction and long-range tertiary contacts. Even in the highly denatured form, the protein had transient compact states in which there were native-like contacts between the N and C-terminal regions, though central region had random coil conformation[80].

In the urea (8M) and low pH denatured state of apomyoglobin, clusters of small amino acids such as glycine and alanine led to increased backbone mobility, suggesting their role as molecular hinges. Also local hydrophobic interactions persisted which caused some restriction of backbone motion on picosecond to nanosecond time scale[106].

The quench flow and CD experiments revealed altered folding pathway for the N132G/E136G mutant of apomyoglobin[56]. Pressure dependent folding studies on Apomyoglobin have revealed the existence of equilibrium mixture of native, intermediates, molten globule and unfolded conformers. The spectral changes were observed to be reversible with respect to pressure changes. Finally, loss of partial molar volume was found to parallel loss of conformational order[107].

4.6 Barnase

Barnase, a small extracellular ribonuclease from *Bacilllus amyloliquefaciens* has been characterized in its pH, urea and temperature denatured forms[68,69]. Sequence specific assignments for a small portion of the protein were obtained by following the exchange of ^{15}N-^{1}H heteronuclear magnetization at the transition point (pH 2.4) and further assignments were obtained from the standard triple resonance 3D experiments and NOESY-HMQC spectra. Transient structure formation in a small region was concluded in the pH denatured state by analysis of chemical shifts, NOEs and exchange rate[68]. Comparison of the different denatured states indicated that pH/temperature and urea denatured barnase are more "unfolded" than the pH denatured protein. The residual native and non-native structures in helical and β–sheet regions were speculated as initiation points in the folding of barnase.

4.7 Barstar

Barstar, an intracellular inhibitor of barnase, gets reversibly denatured in presence of 3 M urea at 278 K[108]. Standard 3D triple resonance and NOE based experiments were used for resonance assignment. The presence of some residual structure was inferred on the basis of chemical shifts, NOEs and coupling constants. On the basis of these the first and second helices were thought to be the potential initiation sites for the folding of barstar.

Barstar is also known to form a molten globule-like A form below pH 4. This form exists as a soluble 160 kDa aggregate of sixteen monomeric subunits, and appears to remain homogenous in solution. Its flexible region has been recently characterized[109]. New assignment methodology based on HNN and HN(C)N was used to obtain sequence specific assignment for 20 residues in its N terminal flexible region. Chemical shifts, temperature coefficients, exchange rate and dynamics data suggest that the A form of barstar is an aggregate with a rigid core, but with the N-terminal 20 residues of each of the monomeric subunits, in a highly dynamic random coil conformation which shows transient local ordering of structure. The N-terminal segment, anchored to the aggregated core, exhibits a free-flight motion.

4.8 Annexin

Detailed characterization of partially folded D2 domain of annexin I was done using ^{15}N relaxation data obtained at three magnetic fields, 500, 600 and 800 MHz[110]. Dynamical behaviour of different types of residual structures was explained on the basis of Lorentzian distribution of correlation times, which was used for representation of relaxation data. This approach yields a clearer picture of unfolded state dynamics than the Model-free approach, which uses discrete correlation times. High values of the width of the distribution highlight the heterogeneous dynamical behaviour of the interconverting structures.

4.9 Chymotrypsin Inhibitor 2

^{15}N relaxation and molecular dynamics (MD) have been used to get an insight into the folding pathway of Chymotrypsin inhibitor 2 (CI2) at atomic resolution[111]. The unfolded state of CI2 at 6.4 M guanidium hydrochloride contains some residual native helical

structure along with hydrophobic clustering in the centre of the chain. The lack of persistent non-native structure in the denatured state reduces barriers that must be overcome, leading to fast folding through a nucleation-condensation mechanism.

4.10 Fibronectin Domain

Meekhof *et al.* have attempted to map the energy landscape for the third fibronectin type III domain from human tenascin (TNfn3) through measurement of ^{15}N backbone dynamics and other structural parameters in 5 M urea[112]. Secondary chemical shifts indicate local preferences for the α-regions. Notable clusters of protected amides were observed for acidic residues indicating intramolecular hydrogen bonding. Few nascent turn like structures were also observed. On the basis of the dynamics data it was concluded that the deviation from random coil behaviour does not imply the formation of stable structural elements, but indicate conformational propensities only.

4.11 Staphylococcal nuclease

Extensive work has been done on the unfolded state of staphylococcal nuclease[78,79,82,113]. Measurement of residual dipolar couplings (D_{HN}) in urea denatured uniformly ^{2}H and ^{15}N labelled protein oriented in strained polyacrylamide gels presents a high correlation among the dipolar couplings for individual residues indicating a native like spatial positioning and orientation of chain segments. A systematic decrease in correlation coefficient was observed with increase in urea concentration in the plots of D_{NH} in urea against D_{NH} in water. Some long range interaction similar to native were also found to be present in the 8 M urea denatured form. Steric repulsion between the residues or weekened long range interactions are possible reasons for this partially fixed orientation. Earlier a *de novo* structure determination of the same protein using paramagnetic relaxation from 14 extrinsic spin labels revealed that many features of the folded arrangement of segment positions and orientations persist in this denatured state [79].

4.12 Plastocyanin

Potential folding initiation sites were speculated in apo-plastocyanin on the basis of detailed structural and dynamic characterization of its unfolded state under non-denaturing conditions[114]. Though there was comparative uniformity in backbone motions, the nanosecond timescale motional restriction for certain segments in the protein was associated with transient formation of non-native helical structures. Certain bias towards β- and α-regions of conformational space was also observed which is sufficient to reduce the search time for folding to a level reasonably consistent with those observed in nature.

4.13 HIV-1 protease

NMR identification of local structural preferences in HIV-1 protease in the 'unfolded state' at 6 M guanidine hydrochloride has been reported[70]. Analyses of the chemical shifts revealed presence of local structural preferences many of which were native like, and there were also some non-native structural elements. Three bond H^{N}-H^{α} coupling constants that could be measured for some of the N-terminal and C-terminal residues were consistent with the native like β structural propensities. Unusually shifted ^{15}N and amide proton chemical shifts of residues adjacent to some prolines and tryptophans also indicated presence of some structural elements. These conclusions were supported by amide proton temperature coefficients and NOE data. The preferred structural propensity was in residues at the dimer interface and this includes the cavity at the active site. A large number of these residues are hydrophobic in nature suggesting that hydrophobic clustering may be a strong driving force for the initial folding events of the protein. These observations suggested that active site formation was an early folding event in HIV-1 protease.

Acknowledgements

We thank the National Facility for High Field NMR at TIFR supported by Government of India. NSB is recipient of the TIFR Alumni Association Scholarship for career development supported by TIFR endowment fund.

References

1 C A Orengo, A E Todd and J M Thornton *Curr Opin Struct Biol* **9** (1999) 374

2 S A Teichmann, C Chothia and M Gerstein *Curr Opin Struct Biol* **9** (1999) 390

3 F S Domingues, W A Koppensteiner and M J Sippl *FEBS Lett* **476** (2000) 98

4 G T Montelione *Proc Natl Acad Sci USA* **98** (2001) 13488

5 S A Teichmann, A G Murzin and C Chothia *Curr Opin Struct Biol* **11** (2001) 354

6 A R Fersht and D Valerie *Cell* **108** (2002) 573

7 H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov and P E Bourne *Nucleic Acids Res* **28** (2000) 235

8 K W Plaxco and M Grob *Nature* **386** (1997) 657

9 A K Dunker, J D Lawson, C J Brown, R M Williams, P Romero, J S Oh, C J Oldfield, A M Campen, C M Ratliff, K W Hipps, J Ausio, M S Nissen, R Reeeves, C H Kang, C R Kissinger, R W Bailey, M D Grisworld, W Chiu, E C Garber and Z Obradovic *J Mol Graph Model* **19** (2001) 26

10 V N Uversky *Protein Sc* **11** (2002) 739

11 P Romero, Z Obradovic, C Kissinger, J E Villafranca, E Garner, S Guilliot and A K Dunker *Pac Symp Biocomput* **3** (1998) 437

12 A K Dunker, E Garner, S Guilliot, P Romero, K Albercht, J Hart, Z Obradovic, C Kissinger and J E Villafranca *Pac Symp Biocomput* **3** (1998) 473

13 P E Wright and H J Dyson *J Mol Biol* **293** (1999) 321

14 J N Adkins and K J Lumb *Protein: Str Funct Genet* **46** (2002) 1

15 H J Dyson and P E Wright *Curr Opin Struct Biol* **12** (2002) 54

16 P Tompa *Trends Biochem Sci* **27** (2002) 527

17 V N Uversky *Eur J Biochem* **269** (2002) 2

18 A K Dunker, C J Brown, J D Lawson, L M Iakoucheva and Z Obradovic *Biochemistry* **41** (2002) 6573

19 O Schweers, E Schönbrunn Hanebeck, A Marx and E Mandelkow *J Biol Chem* **269** (1994) 24290

20 L Lee, E Stollar, J Chang, J G Grossman, R O'Brien, J Ladbury, B Carpenter, S Roberts and B Luisi *Biochemistry* **40** (2001) 6580

21 G W Daughdrill, M S Chadsey, J E Karlinsey, K T Haughes and E W Dahlquist *Nature Struct Biol* **4** (1997) 285

22 S Huang, K S Ratliff and A Matouschek *Nature Struct Biol* **9** (2002) 301

23 S J Demarest, M Martinez-Yamout, J Chung, H Chen, W Xu, H J Dyson, R M Evans and P E Wright *Nature* **415** (2002) 549

24 V Daggett and A R Fersht *Trends Biochem Sci* **28** (2003) 18

25 P Hammarström and U Carlsson *Biochem Biophys Res Comm* **276** (2000) 393

26 F M Richards, D S Eisenberg, and J Kuriyan (Eds.) *Adv Prot Chemistry: Unfolded Proteins* Academic Press San Diego (2002) **42**

27 G Markus *Proc Natl Acad Sci USA* **54** (1965) 253

28 S J Hubbard, R J Beynon and J M Thornton *Protein Eng* **11** (1998) 349

29 G D Fasman *Circular Dichroism and the Conformational Analysis of Biomolecules* Plenum Press New York (1996)

30 N J Greenfield *Anal Biochem* **235** (1996) 1

31 B A Wallace and R W Janes *Curr Opi Chem Biol* **5** (2001) 567

32 V N Uversky, J R Gillespie and A L Fink *Protein: Str Funct Genet* **41** (2000) 415

33 J R Lakowich (Ed.) *Topics in Fluorescence Spectroscopy* Kluwer Academic/Pleneur Publishers New York NY **6** (2000)

34 L Lankiewicz, J Malicka and W Wiczk *Acta Biochem Pol* **44** (1997) 477

35 G S Lakshmikanth, K Sridevi, G Krishnamoorthy and J B Udgaonkar *Nature Struct Biol* **8** (2001) 799

36 H J Dyson and P E Wright *Nature Struct Biol* **5** (**Suppl**) (1998) 499

37 K Wuthrich *NMR of Protein and Nucelic Acids* John Wiley and Sons NY (1986)

38 A Bax and S Grzesiek *Acc Chem Res* **26** (1993) 131

39 T L James, N J Oppenheimer (Ed.) *Methods in Enzymology* Academic Press San Diego CL **239** (1994)

40 J Cavanagh, W J Fairbrother, A G Palmer and N J Skelton *Protein NMR Spectroscopy: Principles and Practice* Academic Press San Diego CL (1996)

41 G M Clore and A M Gronenborn *Curr Opin Chem Biol* **2** (1998) 564

42 G Wider and K Wuthrich *Curr Opin Struct Biol* **9** (1999) 594

43 A E Ferentz and G Wagner *Q Rev Biophy* **33** (2000) 29

44 L E Kay and K H Gardner *Curr Opin Struct Biol* **7** (1997) 722

45 R Weisemann, H Ruterjans, W Bermel *J Biomol NMR* **3** (1993) 113

46 S C Panchal, N S Bhavesh and R V Hosur *J Biomol NMR* **20** (2001) 135

47 N S Bhavesh, S C Panchal and R V Hosur *Biochemistry* **40** (2001) 14727

48 S Grzesiek, J Anglister and A Bax *J Magn Reson* **101B** (1993) 114

49 H Matsuo, E Kupce, H Li, G Wagner *J Magn Reson* B **113** (1996) 91

50 B Braken, A G Palmer, J Cavanagh *J Biomol NMR* **9** (1997) 94

51 T Ikegami, S Sato, M Walchli, Y Kyogoku, and M Shirakawa *J Magn Reson* **124** (1997) 214

52 A Chatterjee, N S Bhavesh, S C Panchal and R V Hosur *Biochem Biophys Res Comm* **293** (2002) 427

53 A Liu, R Reik, G Wider, C V Schroetter, R Zahn and K Wüthrich (2000) *J Biomol NMR* **16** 127

54 F Löhr and H Rüterjans *J Biomol NMR* **6** (1995) 189

55 D Eliezer, J Chung, H J Dyson and P E Wright *Biochemistry* **39** (2000) 2894

56 S Cavagnero, C Nishimura, S Schwarzinger, H J Dyson and P E Wright *Biochemistry* **40** (2001) 14459

57 J Yao, J Chung, D Eliezer, P E Wright and H J Dyson *Biochemistry* **40** (2001) 3561

58 S Spera and A Bax *J Am Chem Soc* **113** (1991) 5490

59 D S Wishart, B D Sykes and F M Richards *J Mol Biol* **222** (1991) 311

60 D S Wishart and B D Sykes *Methods Enzymol* **239** (1994) 363

61 D S Wishart and B D Sykes *J Biomol NMR* **4** (1994) 171

62 A C de Dios, J G Pearson and E Oldfield *Science* **260** (1993) 1491

63 S Schwarzinger, G J A Kroon, T R Foss, P E Wright and H J Dyson *J Biomol NMR* **18** (2000) 43

64 S Schwarzinger, G J A Kroon, T R Foss, J Chung, P E Wright and H J Dyson *J Am Chem Soc* **123** (2001) 2970

65 H J Dyson and P E Wright *Methods Enzymol* **339** (2001) 258

66 G W Vuister and A Bax *J Am Chem Soc* **115** (1993) 7772

67 G Montelione, M E Winkler, P Rauenbuehler and G Wagner *J Magn Reson* **82** (1989) 198

68 V L Arcus, S Vuilleumier, S M V Freund, M Bycroft and A R Fersht *Proc Natl Acad Sci USA* **91** (1994) 9412

69 V L Arcus, S Vuilleumier, S M V Freund, M Bycroft and A R Fersht *J Mol Biol* **254** (1995) 305

70 N S Bhavesh, S C Panchal and R V Hosur *FEBS Lett* **509** (2001) 218

71 G D Rose, L M Gierasch and J A Smith *Adv Protein Chem* **37** (1985) 1

72 H J Dyson and P E Wright *Annu Rev Biophys Biophys Chem* **20** (1991) 519

73 G Merutka, H J Dyson and P E Wright *J Biomol NMR* **5** (1995) 14

74 B M P Huyghues-Despointes, J M Scholtz, and CN Pace *Nature Struct Biol* **6** (1999) 910

75 S W Englander and M M Krishna *Nat Struct Biol* **8** (2002) 741

76 O Zhang, J D Forman-Kay, D Shortle and L E Kay *J Biomol NMR* **9** (1997) 181

77 O Zhang, L E Kay, D Shortle and J D Forman-Kay *J Mol Biol* **272** (1997) 9

78 J R Gillespie and D Shortle *J Mol Biol* **268** (1997) 158

79 J R Gillespie and D Shortle *J Mol Biol* **268** (1997) 170

80 M Lietzow, M Jamin, H J Dyson and P E Wright *J Mol Biol* **322** (2002) 655

81 A E de and N Tjandra *Progr NMR Spectrosc* **40** (2002) 175

82 D Shortle and M S Ackerman 2001 *Science* **293** (2001) 487

83 W Y Choy and J D Forman-Kay *J Mol Biol* **308** (2001) 1011

84 L E Kay, D A Torchia and A Bax *Biochemistry* **28** (1989) 8972

85 L E Kay *Nature Struct Biol* **5(Supp)** (1998) 513

86 G Lipari and A Szabo *J Am Chem Soc* **104** (1982) 4546

87 G Lipari and A Szabo *J Am Chem Soc* **104** (1982) 4559

88 A V Buevich and J Baun *J Am Chem Soc* **121** (1999) 8671

89 J F Lefèvre, K T Dayie, J W Peng and G Wagner *Biochemistry* **35** (1996) 2674

90 J J Prompers and R Brüschweiler *J Am Chem Soc* **124** (2002) 4522

91 L G Werbelow and D M Grant *Adv Magn Reson* **9** (1977) 189

92 R L Vold and R R Vold *Prog NMR Spectrosc* **12** (1978) 79

93 B Reif, M Hennig and C Griesinger *NMR Science* **276** (1997) 1230

94 D Yang, R Konrat and L E Kay *J Am Chem Soc* **119** (1997) 11938

95 Y Daiwen, Y K Mok, D R Muhandiram, J D Forman-Kay and L E Kay *J Am Chem Soc* **121** (1999) 3555

96 V A Daragan and K H Mayo *J Magn Reson Ser B* **107** (1995) 274

97 D Neri, G Wider and K Wuthrich *Proc Natl Acad Sci U S A* **89** (1992) 4397

98 D Neri, G Wider and K Wüthrich *FEBS Lett* **303** (1992) 129

99 D Neri, M Billeter, G Wider and K Wüthrich *Science* **257** (1992) 1559

100 C J Penkett, C Redfield, J A Jones, I Dodd, J Hubbard, R A G Smith, L J Smith and C M Dobson *Biochemistry* **37** (1998) 17054

101 H Schwalbe, K M Fiebig, M Buck, J A Jones, S B Grimshaw, A Spencer, S J Glaser, L J Smith and C M Dobson *Biochemistry* **36** (1997) 8977

102 M Hennig, W Bermel, A Spencer, C M Dobson, L J Smith and H Schwalbe *J Mol Biol* **288** (1999) 705

103 J Klein-Seetharaman, M Oikawa, S B Grimshaw, J Wirmer, E Duchardt, T Ueda, T Imoto, L J Smith, C M Dobson and H Schwalbe *Science* **295** (2002) 1719

104 T M Logan, Y Thériault and S W Fesik *J Mol Biol* **236** (1994) 637

105 T M Logan, E T Olejniczak, R X Xu and S W Fesik *J Biomol NMR* **3** (1993) 225

106 S Schwarzinger, P E Wright and H J Dyson *Biochemistry* **41** (2002) 12681

107 R Kitahara, H Yamada, K Akasaka and P E Wright *J Mol Biol* **320** (2002) 311

108 K B Wong, S M V Freund and A R Fersht *J Mol Biol* **259** (1996) 805

109 J Juneja, N S Bhavesh, J B Udgaonkar and R V Hosur *Biochemistry* **41** (2002) 9885

110 V Ochsenbein, J M Neumann, E Guittet and C V Heijenoort *Protein Sci* **11** (2002) 957

111 S Kazmirski, K B Wong, S M V Freund, Y J Tan, A R Fersht and V Daggett *Proc Natl Acad Sci USA* **98** (2001) 4349

112 A E Meekhof and S M V Freund *J Mol Biol* **286** (1999) 679

113 K Ye and W Jinfeng *J Mol Biol* **307** (2001) 309

114 Y Bai, J Chung, H J Dyson and P E Wright *Protein Sci* **10** (2001) 1056