Biochemistry

 Volume 40, Number 49 December 11, 2001

Accelerated Publications

An Efficient High-Throughput Resonance Assignment Procedure for Structural Genomics and Protein Folding Research by NMR

Neel S. Bhavesh, Sanjay C. Panchal, and R. V. Hosur*

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

Received August 13, 2001; Revised Manuscript Received October 14, 2001

ABSTRACT: Sequence specific resonance assignment is the primary requirement for all investigations of proteins by NMR methods. In the present postgenomic era where structural genomics and protein folding have occupied the center stage of NMR research, there is a high demand on the speed of resonance assignment, whereas the presently available methods based either on NOESY or on some triple-resonance experiments are rather slow. They also have limited success with unfolded proteins because of the lack of NOEs, and poor dispersion of amide and carbon chemical shifts. This paper describes an efficient approach to rapid resonance assignment that is suitable for both folded and unfolded proteins, making use of the triple-resonance experiments described recently [HNN and HN(C)N]. It has three underlying principles. First, the experiments exploit the ^{15}N chemical shift dispersions which are generally very good for both folded and unfolded proteins, along two of the three dimensions; second, they directly display sequential amide and ^{15}N correlations along the polypeptide chain, and third, the sign patterns of the diagonal and the sequential peaks originating from any residue are dependent on the nature of the adjacent residues, especially the glycines and the prolines. 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. 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. Following the amide and ^{15}N assignments, the other proton and carbon assignments can be obtained in a straightforward manner, from the well-established three-dimensional triple-resonance experiments. We have successfully tested the new approach with different proteins in the molecular mass range of 10−22 kDa, and for illustration, we present here the backbone results on the HIV-1 protease-tethered dimer (molecular mass ∼ 22 kDa), both in the folded and in the unfolded forms, the two ends of the folding funnel. We believe that the new assignment approach will be of great value for both structural genomics and protein folding research by NMR.

Sequence specific resonance assignment in proteins has followed in the past either the NOESY-based approach (*1*) or the heteronuclear approach based on triple-resonance experiments (*2*). Presently, for most proteins which can be produced by protein engineering methods, the latter has been the method of choice because of the ease and lack of ambiguity with which the assignments can be obtained, and it has also enabled study of much larger proteins than could be done with the NOESY-based method (*3*−*6*); NOESY is of course still the central experiment for structure calculations following the assignments. The heteronuclear approach for

* To whom correspondence should be addressed. E-mail: hosur@tifr.res.in.

10.1021/bi015683p CCC: $20.00
Published on Web 11/10/2001

assignment relies on a set of standard three-dimensional experiments, such as HNCA (*7*), HN(CO)CA (*8*), CBCANH (*9*), and CBCA(CO)NH (*10*) which employ one-bond magnetization transfers along the polypeptide chain in doubly labeled proteins (^{13}C and ^{15}N), and in most cases, the magnetization on the amide protons is detected. The assignment procedure, in general, involves a 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 three-dimensional (3D) spectra (*2*, *11*). Using these, assignments of several globular proteins are being documented every year.

While the above successes are indeed commendable, the present trends of structural genomics and protein folding, which have occupied the center stage in NMR research, have made greater demands. Structural genomics which deals with the determination of structures of all native proteins in a given species, the proteins being produced by recombinant means, is a high-throughput effort (*12*). On the other hand, protein folding requires characterization of all kinds of partially folded and unfolded species (*13*, *14*), which is challenging because of the poor chemical shift dispersion of amide and carbon resonances. In this context, the method of assignment briefly described above, a standard at the present times, has certain rate-limiting steps. (a) Repeated scanning through the ^{15}N planes of the 3D spectra to locate peaks at the desired chemical shifts is a very time-consuming process. (b) Equivalence of carbon chemical shifts produces ambiguities. (c) Residue type identification along the chain required to remove ambiguities is a slow process and requires many NMR experiments. (d) The sequential walk stops whenever a proline is encountered, and then a new start has to be made from another point along the sequence. However, a priori, no known starting points are available in this exercise, and they have to be chosen almost randomly; their sequence specific identification comes at a much later stage. As a consequence, assignments are often made in short stretches, and they have to be put together in a self-consistent manner to obtain complete assignments (*15*). (e) Finally, in such a procedure, the assignment of a residue, say X, in the stretch PXP, where X is flanked on both sides by prolines, becomes impossible. NOE-based procedures can be used (*1*), but these require full side chain proton assignments, which is a long process. To take care of the above limitations and enhance the speed of assignment automation, algorithms have been developed (*15*−*20*). These are semiautomatic algorithms, employ different combination of experiments, and of course have different limitations with regard to generality, dependence on statistical analysis of chemical shifts, etc. With regard to partially folded or unfolded proteins, it is evident from the above discussion that neither the NOESY-based nor the heteronuclear-based approaches or a combination thereof would be able to provide rapid assignments.

In this total context, we develop here a new systematic approach based on the recently described 3D experimental procedures, HNN and HN(C)N (*21*). While the basic requirement for the experiments is a doubly labeled protein (^{13}C and ^{15}N), uniform deuteration, in addition, will greatly enhance the sensitivity and resolution because of the enhanced relaxation times of the various nuclei. The 3D spectra display direct correlations between resonances of the amides and the ^{15}N nuclei of i, $i-1$, and $i+1$ residues, in the ^{15}N plane of residue i. This alleviates the problem, inherent to the conventional method, of repeated scanning through the planes of the 3D spectra. More importantly, the peak patterns consist of positive and negative signs, and these depend characteristically on the nature of the i, $i-1$, and $i+1$ residues, especially the glycines and the prolines. On the basis of these patterns, proline and glycine neighbors can be readily identified. These provide many fixed points along the sequence, which serve as starting points and/or check points for the sequential walks along the polypeptide chain. Because of the many check points one can generally identify, the backbone H^N and ^{15}N assignments can be largely completed without having to obtain the residue specific side chain assignments explicitly, which is a significant advantage indeed. These features lead to a new protocol for resonance assignment, which is fast, and minimizes the amount of experimental data required. The procedure can be easily automated. Following the backbone amide and ^{15}N assignment, the other assignments of protons and carbons can be readily obtained from the standard suite of 3D spectra. We have successfully tested this approach with different proteins in the molecular mass range of 10−22 kDa, but here we show, as an illustration, the backbone results on the HIV-1 protease-tethered heterodimer (molecular mass ∼ 22 kDa), both in the folded and in the unfolded forms.

From our success with folded proteins, we envisage that at the minimal level of experimentation, the present assignment approach in conjunction with ^{15}N-resolved H−H TOCSY and NOESY experiments would constitute the core approach for rapid determination of protein structures. The successful application to unfolded proteins, on the other hand, is invigorating for the study of protein folding mechanisms (*14*). Various partially unfolded states can be readily assigned, and their relaxation dynamics investigations throw light on the dynamic residual structures at various stages of folding, which in turn enable identification of nucleation sites, early events in folding, etc. (*13*, *14*, *22*−*25*). Although there have been some reports in the literature about following protein folding systematically by NMR (reviewed in ref *14*), they have been relatively few compared to the number of structure determinations of folded proteins, and this is primarily because of the difficulties encountered in assigning unfolded and partially folded proteins. Even where there has been success, extensive experimentation has been necessary. We emphasize here that, in the present approach, only two experiments [HNN and HN(C)N] would prove to be sufficient for both folded and unfolded states of a given protein.

MATERIALS AND METHODS

Protein Preparation. HIV-1 protease-tethered dimer (HIV-TD) cloned in the pET 11a expression vector was obtained as a gift from M. V. Hosur (Bhabha Atomic Research Centre, Mumbai, India). This clone carried the C95M mutation in one of the monomers. We also prepared a double mutant wherein the equivalent cysteine in the second monomer was also mutated to alanine (C195A mutation) in a different context. Both the proteins have almost identical crystal structures (M. V. Hosur and co-workers, unpublished results). For NMR experiments, we used either of the two proteins. ^{15}N-labeled and doubly ^{13}C- and ^{15}N-labeled samples of HIV-TD single (C95M) and double (C95M/C195A) mutants were produced by overexpression in minimal medium and purified using the procedure described previously (*26*).

NMR Samples. For experiments with the unfolded protein, the double mutant was used. The protein was concentrated to ~1 mM and exchanged with pH 5.2 NMR buffer consisting of 50 mM sodium acetate, 5 mM EDTA, 150 mM DTT, and 6 M guanidine hydrochloride, by ultrafiltration. For experiments with the folded protein, the single mutant was used. We may mention here that there is no special reason for these different choices. Since the free protein is susceptible to autolysis in the absence of an inhibitor, a complex of the doubly labeled protein with an inhibitor, acetyl pepstatin, was prepared at a low protein concentration by slow addition of a 2-fold excess of acetyl pepstatin dissolved in methanol, and the solution was repeatedly concentrated and diluted 3−4-fold with NMR buffer to remove excess acetyl pepstatin and autolyzed fragments of the protein. The final concentration of the complex in the NMR sample was ~2 mM.

NMR Experiments. The NMR experiments were carried out on a Varian Unity Plus 600 MHz spectrometer. Two-dimensional (2D) HSQC spectra were recorded with 512 complex t_1 increments, 2048 t_2 points, and four scans for each fid. A 3D HNN spectrum of the unfolded protein in 6 M guanidine hydrochloride was recorded with the following parameters: 68 complex points along t_1 (^{15}N) and t_2 (^{15}N) and 1024 complex points along t_3 (H^N), four scans for each fid, and $T_N = T_C = 32$ ms. The data were processed using Felix 97 software (Molecular Simulations Inc.). A 3D HN-(C)N spectrum of the same sample was recorded using parameters identical to those for the HNN spectrum. The T_{CC} delay was set to 9 ms. The acquisition time for each of the two experiments was approximately 29 h. The experimental parameters for the folded protein−acetyl pepstatin complex for both experiments were as follows: $T_N = T_C =$ 28 ms, 64 increments along both t_1 and t_2, 1024 complex points along t_3, and 12 scans for each fid. T_{CC} was set to 9 ms. The acquisition time for each experiment was approximately 80 h. We also recorded the triple-resonance experiments [HNCA, HN(CO)CA, CBCANH, CBCA(CO)-NH, and HNCO] on both the samples using "protein pack" on the Varian spectrometer. All the spectra were recorded at 32 °C.

The signal-to-noise ratio on our Varian Unity Plus 600 MHz spectrometer is ~540 on a standard sample. This is lower by a factor of 2−10 compared to the currently available sensitivities, and hence, we had to use a larger number of scans for each fid than would be required on the modern high-field spectrometers.

RESULTS AND DISCUSSION

Magnetization Transfer in HNN and HN(C)N Experiments. The pulse sequences and the theoretical details of HNN and HN(C)N experiments have been described previously (*21*), and we present here only the salient features. Figure 1 (A and B) traces the magnetization transfer pathway through the two experimental sequences.

The amide magnetization originating from the *i*th residue is partly transferred in the end to the amides of residues $i - 1$ and $i + 1$. What remains on *i* itself results in the diagonal peak, and what is transferred to the two neighboring residues results in the cross-peaks in the ^{15}N plane of the *i*th residue. Thus, the peaks appear at the following coordinates in the two spectra. HNN: $F_1 = N_i$, $(F_3, F_2) = (H_i, N_i)$, (H_{i-1}, N_{i-1}), (H_{i+1}, N_{i+1}); $F_2 = N_i$, $(F_3, F_1) = (H_i, N_i)$, (H_i, N_{i-1}), (H_i, N_{i+1}). HN(C)N: $F_1 = N_i$, $(F_3, F_2) = (H_i, N_i)$, (H_{i-1}, N_{i-1}); $F_2 = N_i$, $(F_3, F_1) = (H_i, N_i)$, (H_i, N_{i+1}).

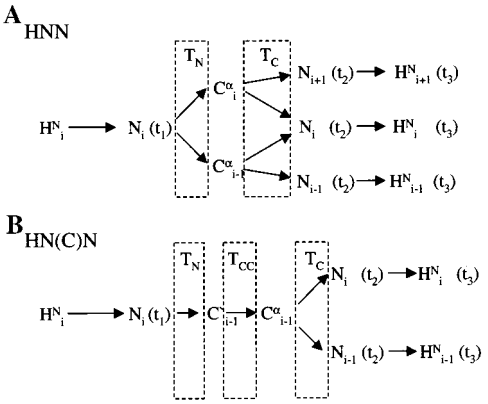

FIGURE 1: Schematic diagram showing the magnetization transfer pathway in the HNN and HN(C)N experiments. T_N, T_C, and T_{CC} are the delays during which the transfers indicated by the arrows take place in the pulse sequences (*21*).

It is clear that if any of the neighboring residues is a proline, then that peak will not appear in the spectrum.

Peak Patterns for Proline and Glycine Neighbors. We consider here triplets of residues since the HNN and HN-(C)N experiments display correlations among three consecutive residues at a time. In the HNN spectrum, every F_1-F_3 and F_2-F_3 plane contains the diagonal peak ($F_1 = F_2 = i$) and two sequential peaks to residues $i - 1$ and $i + 1$. On the other hand, in the HN(C)N spectrum, the F_2-F_3 plane contains the diagonal ($F_1 = F_2 = i$) peak and one sequential peak to residue $i - 1$, whereas the F_1-F_3 plane contains the diagonal peak and one sequential peak to residue $i + 1$. Thus, although the HN(C)N sequence generates only *i* and $i - 1$ correlations, the F_1-F_3 and F_2-F_3 planes taken together help in identifying a triplet of consecutive residues.

Analytical expressions for the intensities and signs of the various diagonals and cross-peaks were derived previously following product operator methods (*21*). It turns out that the evolutions of the magnetization components are slightly different for glycine and non-glycine residues, because of the absence of the β-carbon in the former. This results in different combinations of positive and negative signs for the various self-peaks and cross-peaks in the different planes of the 3D spectra. Further, the absence of an amide proton for a proline results in the absence of the corresponding peak. In addition, ^{15}N chemical shifts also display certain residue type dependence (*27*). Glycines are distinctly upfield compared to others. Interestingly, the average values of the shifts for the different residue types are similar in both folded and unfolded proteins, though the spreads are larger in the folded proteins. Thus, there will be different patterns of peaks for different triplets of residues containing glycines and prolines; we hasten to add that chemical shift-based distinction is only a guideline.

Four categories of triplets of residues may be distinguished: (I) PXZ, ZXP, and PXP, (II) PXG, PGX, PGG, GXP, GGP, XGP, and PGP, (III) XGZ, GXZ, ZXG, GGZ, XGG, GXG, and GGG, and (IV) ZXZ′, where X, Z, and Z′ can be any residue other than proline and glycine. Category I has prolines but no glycines. Category II has combinations

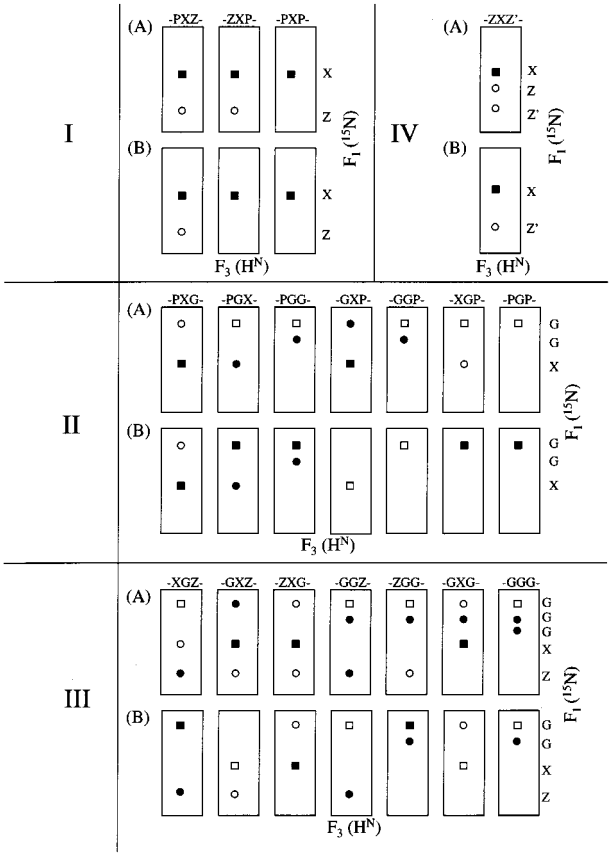

FIGURE 2: Schematic patterns in the F_1-F_3 planes at the F_2 chemical shift of the central residue in the triplets mentioned on the top of each panel, in the HNN (A) and HN(C)N (B) spectra for various special triplet sequences of categories I−IV (see the text). X, Z, and Z′ are any residues other than 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.

of glycines and prolines. Category III has glycines but no prolines. Category IV is a general one, not containing glycines and prolines, and has been included so that the special patterns could be distinguished from the general pattern. The expected peak patterns for each of the above cases in the F_1-F_3 planes of the HNN and HN(C)N spectra, at the F_2 chemical shift of the central residue, under the optimally chosen experimental conditions of magnetization transfer delays ($T_N = 28-32$ ms, $T_C = 28-32$ ms; see Figure 1) are schematically shown in Figure 2. In each of the planes, the peaks are aligned at the amide (F_3) chemical shift of the central residue. The choice of the relative ^{15}N chemical shifts of the Z, Z′, and X residues is quite arbitrary. In reality, the positions of the positive and negative peaks can get altered as per the relative chemical shifts. The important things to consider are (i) the sign of the self-peak or the diagonal ($F_1 = F_2$) peak and (ii) the signs of the sequential peaks relative to that of the diagonal peak. The patterns in Figure 2 can be readily understood from the following salient features of the HNN and HN(C)N spectra.

In the HNN experiment which generates correlations from residue *i* to both residues $i - 1$ and $i + 1$, the sign of the self-peak for glycine is always the opposite of that of any other residue; the actual signs will depend on how the spectra are phased. We have given the diagonal of glycine a negative sign. To emphasize this sign distinction, the diagonal peaks are shown with a different symbol (square) in the figure. This enables unambiguous discrimination between triplets having G's and triplets having X's as the central residues. For example, the patterns for PGX and PXG in HNN seem similar, but in the former, the diagonal peak is negative and the sequential positive; the reverse is true in the latter case. The sign of the sequential peak at the $i - 1$ position will be positive or negative depending upon whether that residue is glycine or otherwise. Similarly, the sign of the sequential peak to residue $i + 1$ will be positive or negative depending on whether the *i*th residue is glycine or otherwise.

In the HN(C)N experiment which generates *i* to $i - 1$ correlations, the signs of the self-peaks and sequential peaks are always opposite. The actual signs are dictated by whether the $i - 1$ residue is a glycine or otherwise, and of course by the phasing of the spectra. Again, we have made the diagonal negative for a glycine at the $i - 1$ position. Consequently, in the F_1-F_3 plane at the F_2 chemical shift of residue *i*, the signs of peaks *i* and $i + 1$ are dictated by the nature of the residues at positions $i - 1$ and *i*, respectively.

Comparison of the patterns in HNN and HN(C)N spectra enables ready discrimination of the $i - 1$ and $i + 1$ neighbors and this provides directionality to the assignment. The patterns in HN(C)N resolve some of the possible ambiguities in the HNN patterns and vice versa. For example, ZXP and PXZ can be readily distinguished from the HN(C)N spectrum, whereas they look similar in the HNN spectrum. Similarly, PGG and GGP patterns are similar in HNN, but are distinctly different in HN(C)N. PGP and XGP would look similar in HN(C)N, but they can be readily distinguished from the HNN spectrum, and so on.

Triplet Fixed Points. For a given protein with a known amino acid sequence, it would be possible to identify many special triplet sequences simply by inspecting the various F_1-F_3 planes in the 3D HNN and HN(C)N spectra of the protein. In case a particular pattern occurs more than once in the spectra, ^{15}N chemical shifts (*27*) may be helpful in identifying favorable cases specifically. For example, an AGP stretch can be readily distinguished from an SGP or a TGP stretch on the basis of the chemical shifts, but it may be more difficult to distinguish a VXP from a LXP. The triplets, which are unique, would be unambiguous fixed points, and several others, which may have some ambiguity, would serve as ambiguous fixed points. Both of these can be used for a variety of purposes, the most important being as starting points or check points during sequential walks along the polypeptide chain. Whenever complete assignments are hampered for some reasons, by any of the methods, the fixed points will serve at least as partial monitors and enable local characterization of the protein.

Sequential Walk through an HNN Spectrum. While an F_1-F_3 plane at the F_2 chemical shift of residue *i* in the HNN spectrum displays self-correlations and sequential correlations to ^{15}N chemical shifts of residues $i - 1$ and $i + 1$ at the amide position of *i*, the F_2-F_3 plane at the F_1 chemical shift of *i* displays the three correlations at their respective amide positions; note that both the F_1 and F_2 dimensions have ^{15}N chemical shifts. In any given F_1-F_3 plane, distinction between the peaks of $i - 1$ and $i + 1$ residues can be obtained by comparison with the identical plane in the HN(C)N spectrum. This leads to a new strategy for rapid assignment

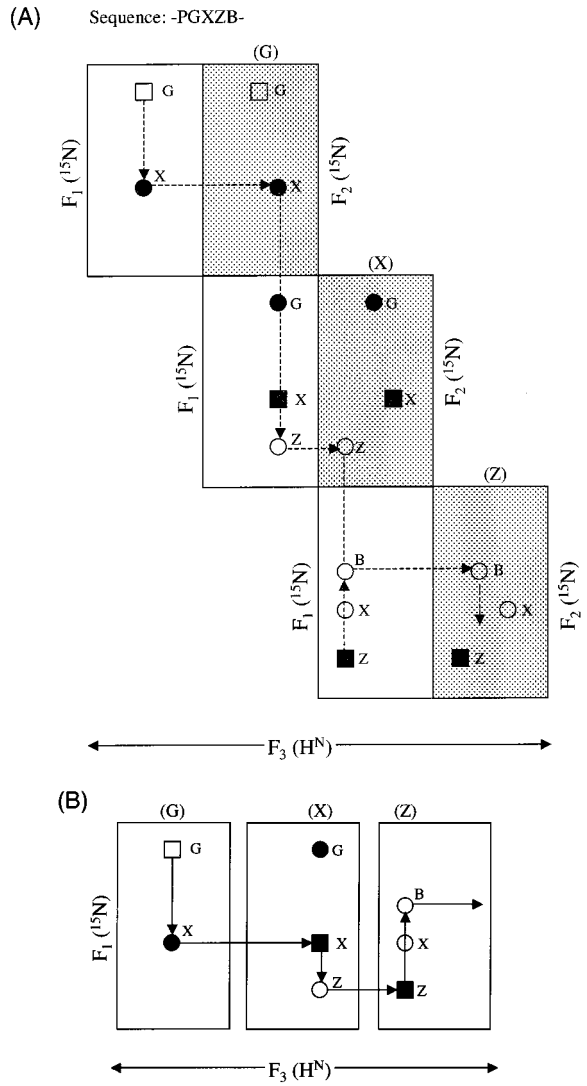

FIGURE 3: (A) Protocol for a sequential walk through the HNN spectrum using a PGXZB illustrative sequence where X, Z, and B can be any residue other than glycine and proline. Pairs of F_1-F_3 and F_2-F_3 (dotted) planes belonging to the three residues (G, X, and Z) are stacked in an appropriate alignment so that the F_1-F_3 plane of one residue stacks over the F_2-F_3 plane of the neighboring residue. Squares are diagonal peaks, and circles are sequential peaks. Filled and open symbols are positive and negative peaks, respectively. Note that the chosen sequence includes the PGX and GXZ special triplets, and the signs of the peaks have been drawn accordingly. The dashed line represents 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 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 neighboring plane. (B) Simplified schematic diagram in which the H^N identification from the F_2-F_3 planes is implicitly assumed, and the sequential walk is shown in the F_1-F_3 planes.

of the amide and ^{15}N chemical shifts of the individual residues, sequence specifically, as shown in Figure 3A. In Figure 3B, a simplified version is shown where H^N identification from the F_2-F_3 planes has been implicitly assumed. The sequential walk can start from any of the fixed points, which can be identified as described above.

A note of caution is in order at this stage. Because of the positive and negative combinations of the diagonal and sequential peaks in the HNN and HN(C)N spectra, cancellation of intensities can occur whenever two neighboring residues have very nearly equal ^{15}N chemical shifts. In the HN(C)N spectrum, diagonal peak cancellations can also occur if the central residues of two different patterns (for instance, X in GXZ and ZXG) have nearly identical ^{15}N shifts. Although these situations can be readily identified by referring to the HNN spectrum, one has to be watchful of such eventualities during the sequential walks. In any case, it is desirable to record these HNN and HN(C)N spectra with high resolution along the two ^{15}N dimensions. This can easily be achieved since ^{15}N spectral widths are small, and high acquisition times can be readily obtained even with a fairly small number of t_1 and t_2 increments.

Application to the HIV-1 Protease-Tethered Dimer. We have successfully tested our approach with different proteins in the molecular mass range 10−22 kDa, and we choose here, for illustration, the HIV-1 protease-tethered heterodimer (HIV-TD) (molecular mass of 22 kDa), both in the folded form (protein−acetyl pepstatin complex) and in the unfolded form (in 6 M guanidine hydrochloride). The protein has two equivalent halves (except for position 95) joined head to tail covalently by a GGSSG linker. The crystal structure of the free protein has recently been reported (*28*). Amazingly, in this construct, the free protein has the flaps closed, in contrast to the wild-type homodimeric protein wherein the flaps are in the open state in the free protein but are closed when it is bound to an inhibitor. Yet the tethered dimer is enzymatically active, suggesting significant dynamism in the structure of the protein. The free protein is also highly susceptible to autolysis. This would have implications for folding−unfolding equilibria (*29*). To date, there are no reports on the folding and unfolding characteristics of this protein. The successful application of the new method for rapid assignment presented here, for the folded as well as the unfolded state of the protein, would be of great value for all the studies described above.

Table 1 lists the amino acid sequence and the identifiable special triplet sequences of the three categories (I−III) in the protein. It is clear from the table that triplets belonging to categories I and II and a few belonging to category III would be less ambiguous. Many others would be more ambiguous, but would nevertheless be very valuable check points. Altogether, there are 41 check points distributed over the whole length of the sequence. We identified a large number of these triplet fixed points in our spectra at the outset itself. Figure 4A shows the F_1-F_3 planes representing some of those fixed points in the HNN spectrum of the unfolded protein, and the corresponding planes from the HN(C)N spectrum are shown in Figure 4B. Similar planes for the folded protein are shown in panels C and D of Figure 4. Note that the specific triplet identifications given here have come after the sequential walks (see below), but they have been included here for better appreciation of the patterns.

Starting from the several fixed points thus identified and the H^N identification for the respective ^{15}N chemical shifts, from the F_2-F_3 planes, we were able to obtain complete assignments, for the unfolded protein, as per the protocol used to generate in Figure 3B. An illustrative sequential walk through the HNN spectrum is shown in Figure 5A, and a

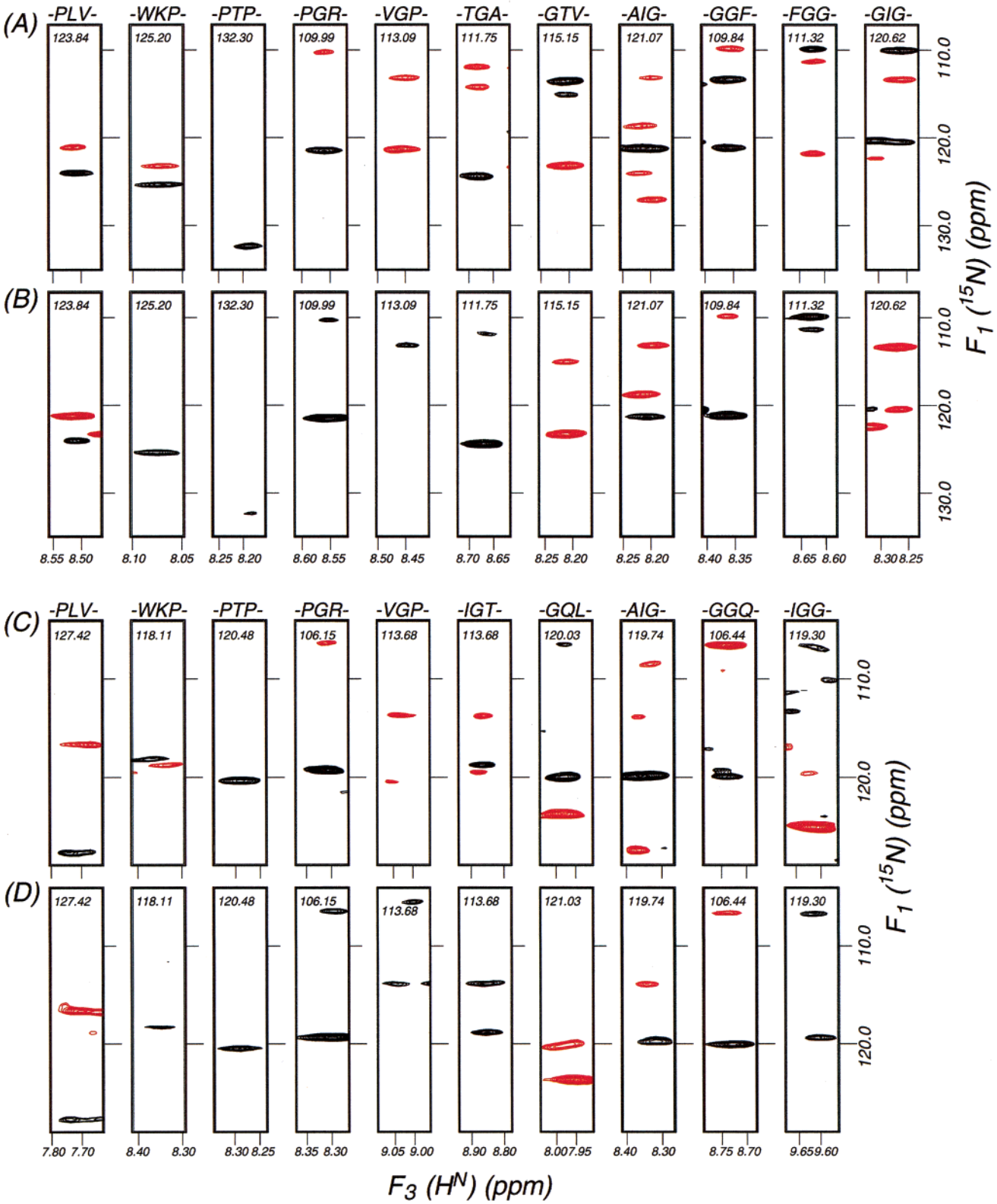

FIGURE 4: F_1-F_3 planes through HNN and HN(C)N spectra of the unfolded and folded HIV-1 protease-tethered heterodimer; in the latter case, the protein is in a complex with acetyl pepstatin. Panels A and B show data for the unfolded protein, while panels C and D show data for the folded protein. One triplet each from the types listed in Table 1 has been chosen for display, except for GIG, for which only the pattern in the unfolded protein has been shown since the same in the folded protein was not distinct. The specific triplet identification indicated on the top of each strip has actually been realized after the sequential walk starting from the various fixed points (see the text). The numbers inside each panel are the F_2 chemical shifts which help to identify the diagonal ($F_1 = F_2$) peaks. Black contours are positive peaks, and red contours are negative peaks.

summary of the connectivities is shown in Figure 5C. We may mention here that the standard procedure based on HNCA, HN(CO)CA, etc., was unsuccessful in providing the assignments for the unfolded protein because of the extensive degeneracy of the C^{α} and C^{β} chemical shifts. Similarly, for the folded protein in the complex with acetyl pepstatin, we

Table 1: Amino Acid Sequence of the HIV-1 Protease-Tethered Dimer and the Special Triplet Sequences, Belonging to Categories I−III, Therein

triplet type[a]	cateogry	sequence present in the HIV-1 protease-tethered dimer
PXZ	I	PQV, PLV, PKM, PVN
ZXP	I	QRP, SLP, WKP
PXP	I	PTP
PGX	II	PGR
XGP	II	SGP, VGP
XGZ	III	TGA, IGT, IGR, IGM, IGC(A)
GXZ	III	GQL, GAD, GRW, GFI, GTV, GRN, GMT
ZXG	III	KIG, DTG, MIG, AIG, LVG, IIG, QIG, NFG, SSG
GGZ	III	GGQ, GGI, GGF, GGS
ZGG	III	$(IGG)_3$, FGG
GXG	III	GIG

[a] X and Z can be any residue other than G and P.

could obtain all but three sequential connectivities in the HNN spectrum as per the protocol used to generate Figure 3B. An illustrative sequential walk through the HNN spectrum of the folded protein is shown in Figure 5B, and a summary of the connectivities is shown in Figure 5D. These assignments then enabled a rapid analysis of the HNCA, HN(CO)CA, HNCO, CBCANH, and CBCA(CO)NH spectra in a straightforward manner which elucidated the C^{α}, C^{β}, and C′ assignments (BioMagResBank accession number 5062 for the unfolded protein). Figure 6 shows the amide and ^{15}N assignments in the HSQC spectra for both the folded protein in the complex (A) and the unfolded protein in 6 M guanidine hydrochloride (B). Examination of the two spectra reveals that the amide proton dispersion in the unfolded protein is very narrow compared to that of the folded protein, and successful assignment of such a spectrum is indeed an achievement. The successful application to both folded and unfolded states of the protein also suggests the possibility of detailed characterization of various stable intermediates during the folding reaction of the protein.

CONCLUDING REMARKS

We have described in this paper a novel approach for obtaining rapid assignment of NMR resonances in doubly ^{13}C- and ^{15}N-labeled proteins. The method is based on the recently described HNN and HN(C)N triple-resonance experiments. The success of the method has been demonstrated with a 22 kDa protein, HIV-1 protease, both in the folded and in the unfolded forms, which have implications for structural genomics and protein folding studies. However, for the general utility of the techniques on a routine basis, certain comments about the sensitivity and range of applications of the techniques would be in order. As discussed in the original description of the techniques, the experimental pulse sequences employ two successive transfers through relatively small one- and two-bond coupling constants, because of which the pulse sequences are long, and transverse relaxation during these long periods causes loss of sensitivity. While this is not a serious problem in unfolded proteins and smaller proteins, it may become a matter of concern for larger folded proteins which have fast relaxation rates. Recognizing this factor, we had stated in the original paper that for folded proteins, deuteration may become necessary for reducing the relaxation losses through the pulse sequences. While deuteration benefits are indeed significant, the results with

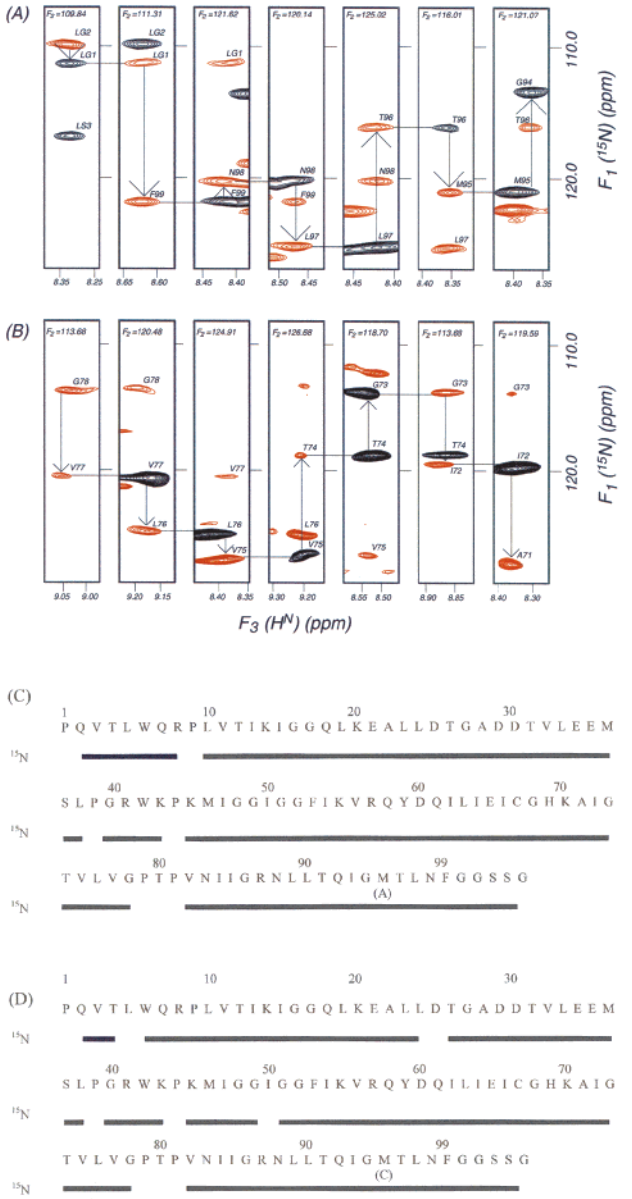

FIGURE 5: Illustrative sequential walks through the HNN spectra for the unfolded protein (A) and folded protein (B). Black and red contours are positive and negative peaks, respectively. Several of the planes are fixed points: GGS, FGG, and GMT in panel A and VGP, GTV, and IGT in panel B. Summaries of ^{15}N sequential connectivities for unfolded and folded proteins are shown in panels C and D, respectively.

nondeuterated proteins presented here, and those which have not been presented here, indicate, however, that our earlier statement was a rather cautious and conservative one. In the case presented here, the experiments have been performed on an old Varian Unity Plus 600 MHz NMR spectrometer with a signal-to-noise ratio of ∼540, on a standard sample, whereas in modern NMR spectrometers, ratios of 1300 are easily available; with the availability of "cryoprobes", signal-to-noise ratios as high as 6000 have been realized on 600 MHz spectrometers. At higher magnetic field strengths of up to 900 MHz (in proton frequency units), as are currently available, the sensitivities are even higher. Thus, the sensitivity of the HNN and HN(C)N techniques is not an issue, and experiments can be successfully carried out on large proteins even without deuteration.

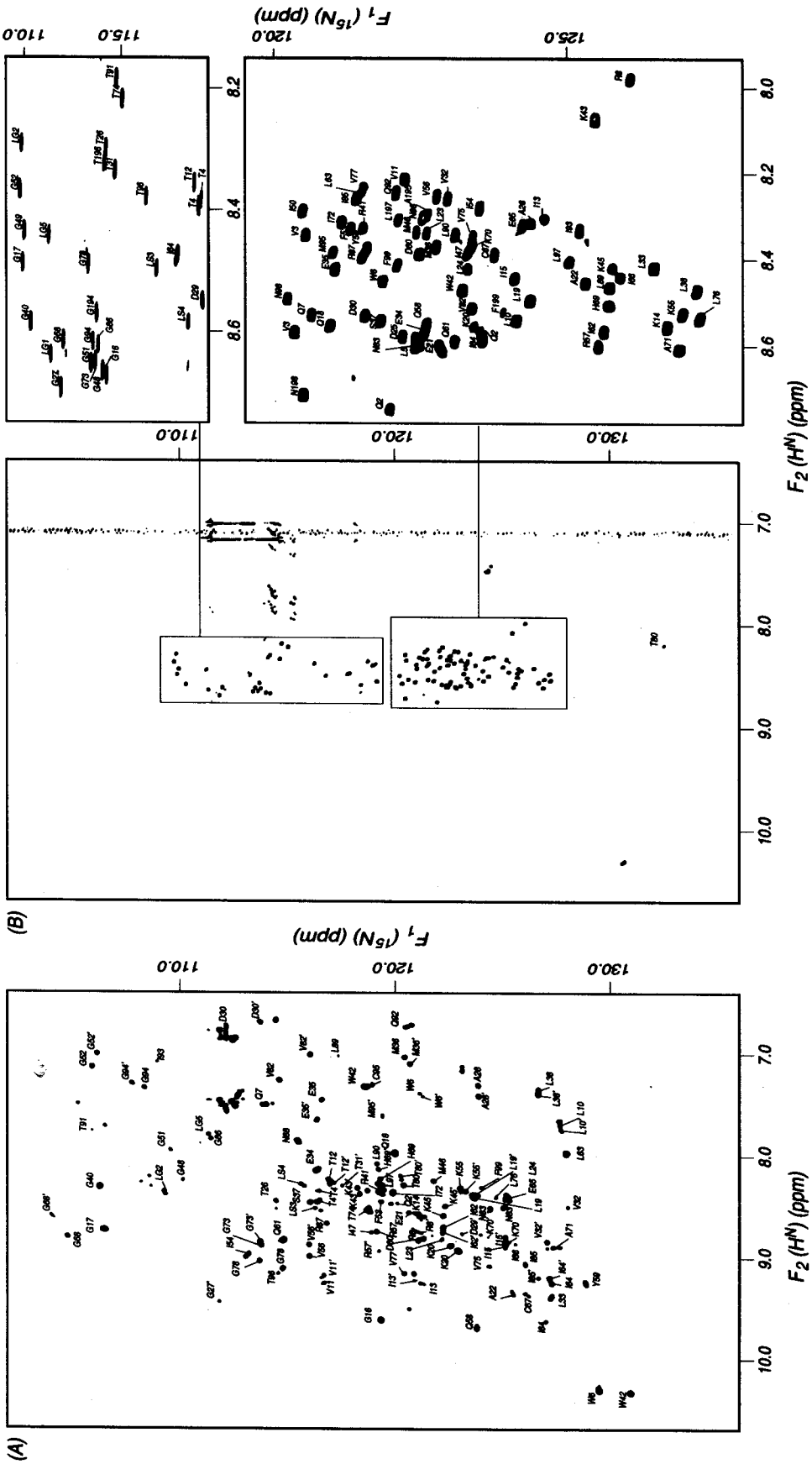

FIGURE 6: HSQC spectra with H^N and ^{15}N assignments for the folded (A) and unfolded (B) HIV-1 protease-tethered dimer. In the folded protein spectrum, separate peaks are seen for equivalent residues in the two halves for many cases and these are distinguished by primed and unprimed labels. These two sets are sequentially connected within themselves, although the continuity of this distinction is lost whenever they merge at some point and restart at another place.

Another important point deserves reiteration. Many large proteins may have certain contiguous regions which are very flexible. These regions behave like small proteins from the relaxation pint of view, and can be readily monitored by the approach described in this paper. A few fixed points in that area will help unambiguous assignments, and it is possible to carry out a variety of structural and dynamics studies for that portion of the protein.

ACKNOWLEDGMENT

We thank the National Facility for High Field NMR at TIFR, for all the facilities. The clone for the HIV-1 protease-tethered dimer was a kind gift from Dr. M. V. Hosur of Bhabha Atomic Research Center.

REFERENCES

1. Wuthrich, K. (1986) *NMR of Protein and Nucelic Acids*, John Wiley and Sons, New York.
2. Bax, A., and Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−138.
3. James, T. L., and Oppenheimer, N. J., Eds. (1994) *Methods in Enzymology*, Vol. 239, Academic Press, San Diego.
4. Wider, G., and Wuthrich, K. (1999) *Curr. Opin. Struct. Biol. 9*, 594−601.
5. Clore, G. M., and Gronenborn, A. M. (1998) *Curr. Opin. Chem. Biol. 2*, 564−570.
6. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR spectroscopy principles and practice*, Academic Press, San Diego.
7. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
8. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
9. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 99*, 201−207.
10. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
11. Kay, L. E., and Gardner, K. H. (1997) *Curr. Opin. Struct. Biol. 7*, 722−731.
12. Burley, S. K. (2000) *Nat. Struct. Biol. 7* (Suppl.), 932−934.
13. Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6−12.
14. Dyson, H. I., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 499−503.
15. Guntert, P., Salzmann, M., Braun, D., and Wuthrich, K. (2000) *J. Biomol. NMR 18*, 129−137.
16. Friedrichs, M. S., Mueller, L., and Wittekind, M. (1994) *J. Biomol. NMR 4*, 703−726.
17. Meadows, R. P., Olejniczak, E. T., and Fesik, S. W. (1994) *J. Biomol. NMR 4*, 79−96.
18. Zimmerman, D. E., Kulikowski, C. A., Huang, Y. P., Feng, W. Q., Tashiro, M., Shimotakahara, S., Chien, C. Y., Powers, R., and Montelione, G. T. (1997) *J. Mol. Biol. 269*, 592−610.
19. Atreya, H. S., Sahu, S. C., Chary, K. V., and Govil, G. (2000) *J. Biomol. NMR 17*, 125−136.
20. Moseley, H. N., and Montelione, G. T. (1999) *Curr. Opin. Struct. Biol. 9*, 635−642.
21. Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) *J. Biomol. NMR 20*, 135−147.
22. Dill, K. A., and Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
23. Shortle, D. (1993) *Curr. Opin. Struct. Biol. 6*, 24−30.
24. Fushman, D., Cahill, S., and Cowburn, D. (1997) *J. Mol. Biol. 266*, 173−194.
25. Meekhof, A. E., and Freund, S. M. V. (1999) *J. Mol. Biol. 286*, 579−592.
26. Panchal, S. C., Pillai, B., Hosur, M. V., and Hosur, R. V. (2000) *Curr. Sci. 79*, 1684−1695.
27. Peti, W., Smith, L. J., Redfield, R., and Schwalbe, H. (2001) *J. Biomol. NMR 19*, 153−165.
28. Pillai, B., Kannan, K. K., and Hosur, M. V. (2001) *Proteins 43*, 57−64.
29. Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) *FEBS Lett. 497*, 59−64.

BI015683P