THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 22, Issue of May 30, pp. 19980–19985, 2003

Printed in U.S.A.

NMR Elucidation of Early Folding Hierarchy in HIV-1 Protease*

Received for publication, February 14, 2003, and in revised form, March 17, 2003
Published, JBC Papers in Press, March 18, 2003, DOI 10.1074/jbc.M301615200

Neel S. Bhavesh‡, Ragini Sinha, P. M. Krishna Mohan, and Ramakrishna V. Hosur§

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

Folding studies on proteases by the conventional hydrogen exchange experiments are severely hampered because of interference from the autolytic reaction in the interpretation of the exchange data. We report here NMR identification of the hierarchy of early conformational transitions (folding propensities) in HIV-1 protease by systematic monitoring of the changes in the state of the protein as it is subjected to different degrees of denaturation by guanidine hydrochloride. Secondary chemical shifts, H^N-H^α coupling constants, ^1H-^{15}N nuclear Overhauser effects, and ^{15}N transverse relaxation parameters have been used to report on the residual structural propensities, motional restrictions, conformational transitions, etc., and the data suggest that even under the strongest denaturing conditions (6 M guanidine) hydrophobic clusters as well as different native and non-native secondary structural elements are transiently formed. These constitute the folding nuclei, which include residues spanning the active site, the hinge region, and the dimerization domain. Interestingly, the proline residues influence the structural propensities, and the small amino acids, Gly and Ala, enhance the flexibility of the protein. On reducing the denaturing conditions, partially folded forms appear. The residues showing high folding propensities are contiguous along the sequence at many locations or are in close proximity on the native protein structure, suggesting a certain degree of local cooperativity in the conformational transitions. The dimerization domain, the flaps, and their hinges seem to exhibit the highest folding propensities. The data suggest that even the early folding events may involve many states near the surface of the folding funnel.

The folding of a protein is conceptually described in terms of a folding funnel (1–6). The narrow end of the funnel represents the folded native state, and the broad end represents the unfolded state consisting of millions of rapidly inter-converting conformers. As a protein folds from an unfolded state, it goes through one or more partially folded intermediates, which need to be characterized for elucidation of its folding pathways. Experimentally, this is a very challenging task. The most common and direct approach relies on kinetic pulse labeling experiments of amide protons coupled with hydrogen exchange at different time points along the folding reaction of the protein (reviewed in Ref. 7). However, as has been pointed out (8), this also has limitations. First, it is limited by the necessity of detectable protection against exchange. Second, any lack of protection does not necessarily imply absence of structure. Third, it biases the interpretation of the structure in intermediates toward the native state. Even so, useful information has been obtained in many protein systems (7). However, this approach is complicated in the case of proteases with autolytic property, because the autolytic reaction interferes with interpretation of the hydrogen exchange data. Therefore, in these systems it becomes necessary to look for alternative experimental avenues.

Several equilibrium, real-time NMR, and kinetic pulse labeling studies (8–22) have indicated that in many proteins the local structural features of the kinetic intermediates have many things in common with the partially unfolded states created by the use of chemical denaturants such as urea, guanidine, or extreme pH conditions (15–22). That means characterization of the partially unfolded states created by denaturants can provide useful insights into the structural features of the kinetic intermediates. Further, the progressive folding of a protein will be associated with significant changes in its internal dynamics at all time scales (pico- to milleseconds). The fully unfolded state is highly dynamic with motions occurring mostly on picosecond time scales. Any restriction in the motions implies transient ordering of the polypeptide chain. As the protein starts to fold, more and more structure-forming breaking events (milli- to microsecond time scale) occur, and these lead to an increase in the slow motions. Thus a systematic monitoring of these graded changes in the motional characteristics as well as in the residual structures along the polypeptide chain under different conditions of denaturation provides very valuable information on the hierarchy of folding propensities in a protein. Following these ideas, we describe below NMR identification of the hierarchy of folding propensities in HIV-1[1] protease, using a tethered dimer construct of the protein in which the two monomers are joined head-to-tail by a flexible linker GGSSG; this protein, hereafter referred to as HIVTD, folds similarly to the native homodimer *in vitro* (23–24) and also has similar activity toward the substrates.[2]

MATERIALS AND METHODS

NMR Experiments—Isotopically ^{15}N-labeled HIV-1 protease-tethered dimer for the NMR experiments was prepared as described earlier (25). ^1H-^{15}N HSQC spectra were recorded with 1.0–1.2 mM protein samples in NMR buffer (50 mM acetate, 150 mM dithiothreitol, 5 mM EDTA, pH 5.2) containing different concentrations of guanidine hydrochloride at 32 °C. For the high resolution data required for coupling constant measurements, 16,384 and 512 complex points were acquired along t_2 and t_1 dimensions, respectively. For the relaxation measurements 2,048 and 512 complex points were collected along the two

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of the TIFR Alumni Association Scholarship (2002–2003) for career development, supported by the TIFR endowment fund.

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

[1] The abbreviations used are: HIV, human immunodeficiency virus; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser enhancement.

[2] V. Prashar, M. Kumar, B. Pillai, A. Chatterjee, N. S. Bhavesh, R. Mittal, R. V. Hosur, and M. V. Hosur, unpublished results.

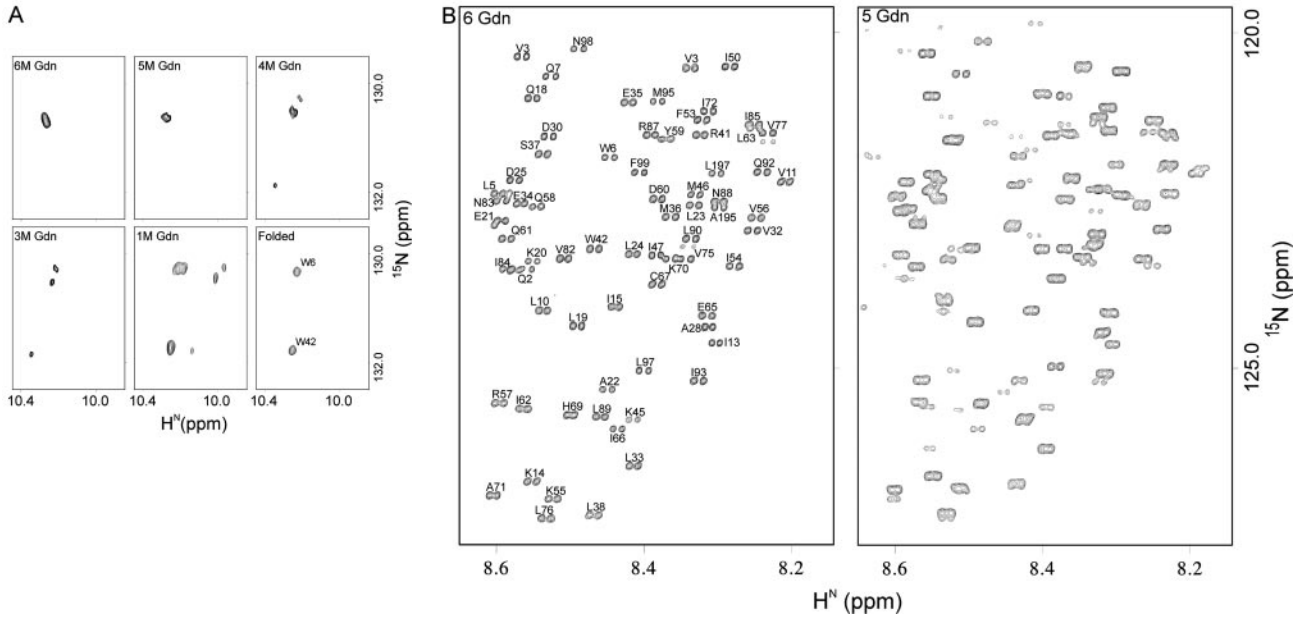

FIG. 1. *A*, ^{1}H-^{15}N correlation peaks of tryptophan side chain as a function of guanidine concentration. The same region of the fully folded protein in a complex with an inhibitor (acetyl pepstatin) is also shown for comparison. As the guanidine concentration is decreased from 6 to 1 M, the protein folds and acquires activity. This causes autocleavage of the protein, resulting in additional correlation peaks for the tryptophan side chain. *B*, comparison of a section of the HSQC spectra of the protein in 6 M (*left panel*) and 5 M (*right panel*) guanidine conditions. The peaks present in the 6 M spectrum are also present in the 5 M spectrum at almost identical positions, allowing easy transfer of assignments. These have been marked only in the 6 M spectrum. The 5 M spectrum contains additional peaks, which must be attributed to other partially folded forms of the protein. Both the spectra are processed in identical fashions.

dimensions. For R_2 measurements, the following Carr-Purcell-Meiboom-Gill (CPMG) delays were used: 8.352, 25.056, 41.760, 66.816, 83.357, 116.928, 141.984, and 192.096 ms and spectra duplicated at 25.056 and 83.357 ms. R_2 values were extracted by fitting the peak intensities to the equation $I(t) = B \exp(-R_2 t)$. ^{1}H-^{15}N heteronuclear NOE data were collected with a 3-s relaxation delay and 2-s presaturation of the protons. The equilibrium experiment was performed with a 5-s relaxation delay. The NOEs were calculated as peak intensity ratios, I_{sat}/I_{eq}, where I_{sat} is the peak intensity in the spectrum with proton saturation and I_{eq} is the peak intensity in the equilibrium experiment. The experiments were carried out using the pulse sequences described by Farrow *et al.* (26). All the experiments were performed on a 600 MHz Varian Unity plus spectrometer, and the data were processed using FELIX on a Silicon Graphic, Inc. work station. Analysis of the primary structure of HIVTD to locate hydrophobic clusters was carried using the program HCA_Draw (27).

RESULTS AND DISCUSSION

NMR Spectral Features of the Denatured States of HIVTD—We first monitored the changes in the state of the protease when guanidine denaturant concentration was systematically decreased from 6 to 1 M (Fig. 1*A*). At 6 M, the ^{1}H-^{15}N HSQC spectrum was characteristic of an unfolded state. As the guanidine concentration was reduced to 5 M, the 6 M peaks were still present, but additional peaks appeared (see below). The protein was, however, still intact (Fig. 1*A*). At 3 M guanidine, a coexistence of folded (or partially folded) and unfolded species was seen; as the concentration was further reduced, the protein started showing protease activity. Characteristic tryptophan side chain peaks of auto-cleavage products were seen in the HSQC spectrum and, also, auto-cleavage products were seen by gel electrophoresis. The autolytic activity increased progressively as the guanidine concentration was further reduced. Therefore, to gain mechanistic insight into the early hierarchical folding events in the protein we investigated the structural and dynamic characteristics of the protein at 6 and 5 M guanidine concentrations, under which conditions the protein still remains intact.

Fig. 1*B* shows comparison of a section of the ^{1}H-^{15}N HSQC spectrum of the protein at 6 and 5 M guanidine concentrations. The number of peaks in the 6 M spectrum is as must be expected from the equivalence of residues in the two halves of the tethered dimer. All the peaks present in the 6 M spectrum are also present in the 5 M spectrum at almost identical positions, barring a few which show small shifts. This allowed an easy transfer of assignments (6 M assignments have already been reported (25). Besides, there are several weaker peaks, which suggest the presence of some partially folded forms. Keeping in mind that the denatured and partially folded states are highly dynamic and heterogeneous, one can envisage that the partially folded states could be in slow exchange with the "6 M denatured state." Though intuitively one may assign these additional peaks to their nearest neighbors, there could be more than one partially folded species existing simultaneously in solution and the peaks could correspond to different ones. The presence of these peaks indicates that the state identified by the conserved peaks in the two spectra would have differences in the dynamic characteristics under 6 and 5 M guanidine conditions. This is also evident from the fact that the peak resolution in the 5 M spectrum is much less, which must be attributed to reduced coupling constants or increased line width because of conformational exchange or both. We have monitored these differences using a variety of NMR parameters, which provide valuable insight into the folding conformational transitions in the protein.

Local Structural Preferences in 6 M Guanidine: Folding Nuclei—A large body of evidence in the literature indicates that the denatured state is not a random coil but contains some residual structure or at least some local structural preferences in an otherwise heterogeneous dynamic model (28). Certain regions of (ϕ,ψ) dihedral angle space may exhibit higher probabilities than others in the Ramachandran plot (29). These are believed to be the folding nuclei or the regions where the initial folding events occur in the polypeptide chain.

Detailed characterization of the residual structures in the denatured states of proteins is generally obtained from residue-wise (C^{α}, H^{α}, C^{β}, CO) chemical shift deviations from random

coil values (secondary shifts), H^N-H^α coupling constants, amide proton temperature coefficients, and in favorable cases the amide proton protections against deuterium exchange (30). Although both the secondary chemical shifts and H^N-H^α coupling constants reflect the secondary structural propensities, the former are much more sensitive and reflect even very small population differences in the (ϕ,ψ) space. On the other hand, the amide temperature coefficients (less than ~7 ppb/K) indicate hydrogen bonding and thus report on the presence of persistent structures (30). In the present case we monitored the carbon secondary chemical shifts, the H^N-H^α coupling constants, and the amide proton temperature coefficients in HIVTD, and a part of their analysis has been described earlier (25). As mentioned in that study, the sequence-corrected (31) C^α chemical shifts were the most diagnostic of secondary structures because of their insensitivity to the choice of the reference set of values for the random coil chemical shifts. These results (taken from Ref. 25) along with the complete results for coupling constants and amide proton temperature coefficients are shown in Fig. 2.

In Fig. 2*A*, the C^α secondary shifts are seen to be large for many residues and show substantial sequence dependence. Positive deviations indicate preferences in (ϕ,ψ) values belonging to the α-helical structures; the negative values indicate preferences for the (ϕ,ψ) values belonging to the β structures. Thus, we see β propensities for the residues, Pro-1, Val-3, Leu-5, Gln-7, Glu-21, Asp-25, Glu-34, Ser-37, Lys-43-Pro-44, Phe-53, Lys-55, Val-56, Ile-66, His-69, and Pro-79-Ile-84, and α propensities for the residues, Arg-8, Leu-10, Leu-24, Gly-40, Gln-61, Leu-63, Glu-65, Cys-67, Gly-68, Asn-88, Leu-90, and Gln-92. The secondary structure elements in the native protein are shown on the top of the figure for ready comparison. Clearly, all the β propensities and the α propensity in the region Asn-88-Gln-92 are native-like, and the remaining preferences, which are α type, are non-native.

At this stage we may ask whether tethering of the two monomers by the flexible linker GGSSG in HIVTD has any influence on the structural propensities. One may expect to observe these influences at the C-terminal of the first monomer and the N-terminal of the second monomer. In the NMR spectra we do observe separate peaks for the residues Pro-1, Gln-2, Val-3, Thr-96, Leu-97, and Asn-98 from the two monomers; the secondary shifts for these residues are also slightly different, as indicated by filled bars in Fig. 2*A*. However, the shifts for the C-terminal residues are very small for both the monomers, but the shifts for the N-terminal residues Pro-1 and Val-3 are large for both. This indicates that the observed propensities are primarily dictated by the intrinsic sequence and the tethering has only a small influence, if at all. However, as we shall discuss below, the tethering may have some influence on the motional characteristics, because of local interactions of the residues around the linker.

The residue-wise H^N-H^α coupling constants in HIVTD in 6 M guanidine measured from the fine structures of the peaks in the HSQC spectra are shown in Fig. 2*B*. The values range from 5.5 to 9.0 Hz with an average close to 7 Hz. This is typical of random coils, and the small variations observed are within the ranges expected from sequence dependence of these coupling constants (32). Thus the observed coupling constants here do not show any specific secondary structural preferences, which is in contrast to the results from secondary chemical shifts. However, such differences have been seen in many proteins, leading to the belief that the coupling constants are less sensitive to small variations in the (ϕ,ψ) populations and hence are less diagnostic (30).

Fig. 2*C* shows the residue-wise amide proton temperature

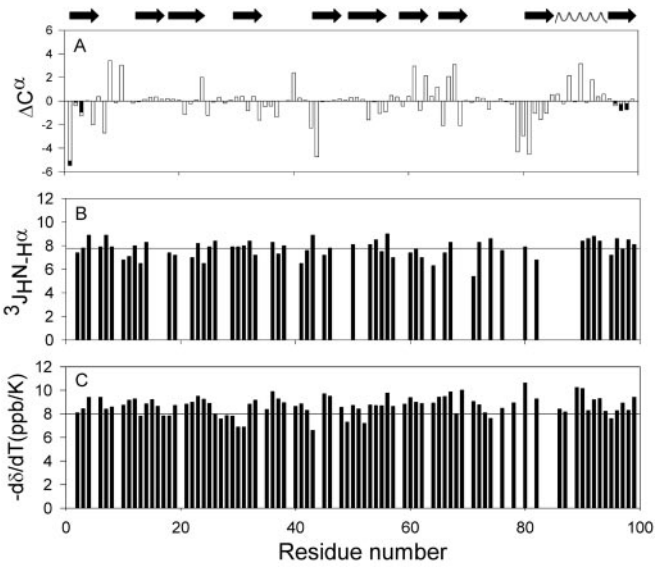

FIG. 2. **NMR parameters in 6 M guanidine, pH 5.2, temperature 32 °C.** *A*, C^α chemical shift deviations from random coil values (data taken from Ref. 25). *B*, H^N-H^α coupling constants for the different residues measured from the peak fine structures in the high resolution HSQC spectrum (Fig. 1*B*). *C*, residue-wise amide proton temperature coefficients. In *panel A*, for a few residues at the N and C-terminals, both *filled* and *open bars* are shown that correspond to the two monomeric halves of the tethered dimer (see text). Such a distinction is not seen in *panels B* and *C*. The secondary structural elements along the polypeptide chain in the native protein are shown on the *top*. *Thick arrows* indicate β-structures. The *wavy symbol* indicates the α-helix.

coefficients in the HIVTD protein in 6 M guanidine. For most residues the values are in the range 7–9 ppb/K, with an average at 8 ppb/K. This is typical of amide protons exposed to the solvent and not involved in any kind of H-bonds. Thus it appears that there are no persistent structures in the protein in 6 M guanidine.

Examination of the local structural preferences derived from the secondary chemical shifts in the protein in the light of its primary structure reveals an interesting correlation: many of the preferences are seen to be associated with prolines. In the primary sequence, prolines are located at positions 1, 9, 39, 44, 79, and 81. Of these, Pro-9 and -39 seem to induce α-helical propensity in the neighboring residues. On the other hand, all the other prolines are associated with β propensities. Thus, it appears that in HIVTD, the prolines may play a major role in dictating the initial folding events when the folding starts from an unfolded state. The different native and non-native folding nuclei described above, along with the locations of the prolines, are depicted on the native protein structure in Fig. 3. Interestingly, these cover the residues at the dimer interface and in the hinge region of the native protein structure. This suggests that native-like folding of active site residues and perhaps active site formation may be an early event in the folding of HIVTD.

Motional Characteristics in 6 M Guanidine—^{1}H-^{15}N heteronuclear NOEs and ^{15}N transverse relaxation rates (R_2) provide most valuable information on the backbone motions of the individual NH vectors at different time scales. Although the heteronuclear NOEs are most sensitive to motions on the pico- to nanosecond time scales, the transverse relaxation rates are sensitive also to slow motions and conformational transitions occurring on milli- to- microsecond time scale. In many instances these have provided valuable insights into sequence-dependent motional restrictions and flexibilities in denatured proteins, which in turn provide clues to the folding mechanisms (33–38). A close correlation was observed between the regions of nanosecond-restricted motions and the locations of hydro-

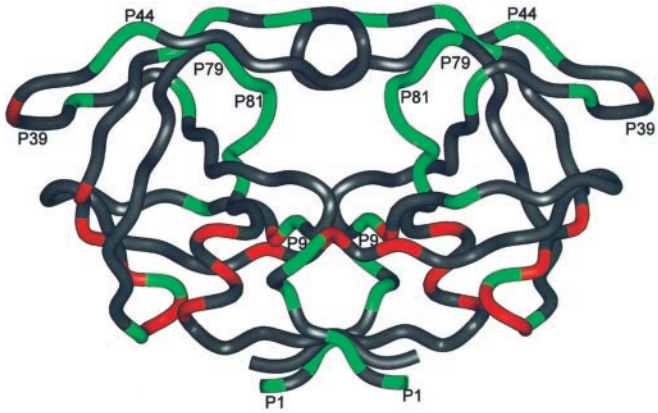

FIG. 3. **Structural preferences in 6 M guanidine or the folding nuclei in HIVTD displayed on the structure (23) of the folded protein in a color-coded manner: *green*, β-type secondary structural propensities, *red*, α-helical-type propensities.** The positions of prolines, which seem to influence the structural propensities are also shown.

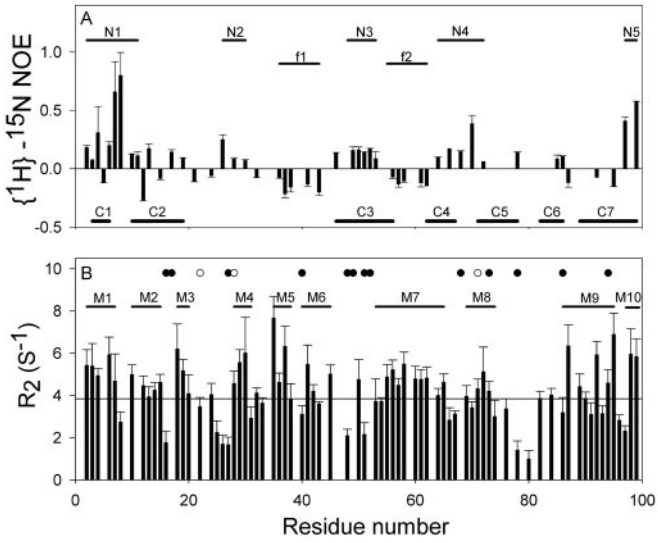

FIG. 4. *A*, residue-wise ^{1}H-^{15}N heteronuclear NOEs in 6 M guanidine at 32 °C. C1–7 are the hydrophobic clusters calculated from the sequence using the program HCA_Draw (27). N1–5 are the domains exhibiting positive NOEs and thus restricted motion on nano-to-picosecond time scale; f1,-2 are domains exhibiting negative NOEs, implying high flexibilities. *B*, residue-wise ^{15}N R_2 values in 6 M guanidine, 32 °C. The *horizontal line* indicates the average. *Error bars* are indicated on each vertical bar. M1–10 identify regions of restricted motion and/or conformational exchange. *Filled* and *open circles* identify the positions of glycines and alanines, respectively, along the sequence of the protein.

phobic clusters (35–38), which led to the suggestion that transient hydrophobic clustering caused by local side chain interactions may drive the early folding events. In denatured apomyoglobin, a correlation was also observed between the regions of greater flexibilities and locations of the small amino acid residues glycine and alanine, which led to the suggestion that these residues serve as molecular hinges in the folding process of the protein (38).

The ^{1}H-^{15}N heteronuclear NOEs for the different residues in HIVTD in 6 M guanidine are shown in Fig. 4*A*. As against a typical flat bell-shaped profile expected of fully denatured proteins, we observe a definite pattern of positive and negative NOEs, which seems to suggest the presence of motionally restricted (positive NOE) as well as highly flexible (negative NOE) regions along the polypeptide chain. Assuming that a group of at least three residues of similar motional properties in close vicinity may be taken to define a local domain, we identified domains of restricted motions (N1-N5), and domains of flexibilities (f1 and f2)(Fig. 4*A*); note that N5 has only two residues, but this has been included because of the large magnitude of the NOE. These regions of restricted motions are indicative of side chain interactions, which cause local transient ordering and thus would play important roles in directing the folding process of the protein. Interestingly, both the N- and C-terminal residues exhibit motional restrictions. One may think that because these groups of residues are engaged in a β-sheet in the dimerization domain of the native protein structure there may be a propensity for native-like dimer formation even in 6 M guanidine. Alternatively, it is possible that because of the tethering of the two monomers in HIVTD, which brings the C-terminal of the first monomer close to the N-terminal of the second monomer, there may be local interactions, resulting in restriction of motions. In fact, both mechanisms may be operative, the latter driving the former, because the polypeptide chain has an intrinsic tendency to form a specific dimer.

The ^{15}N transverse relaxation rates (R_2) for the different residues in HIVTD in 6 M guanidine are shown in Fig. 4*B*. Here again we observe substantial sequence-dependent variation, indicating variations in motional flexibilities. The values range from 1.0 to 7.8 s^{-1}, with an average around 3.8 s^{-1}. Low R_2 values indicate high flexibilities, whereas large R_2 values indicate restricted motions on nanosecond time scales and also possible contributions from conformational transitions on milli- to microsecond time scale. As in the case of ^{1}H-^{15}N NOEs, we identified domains of restricted motions, M1–10, on the basis of the R_2 values. Of these, M1, -2, -4, -8, and 10 overlap with N1, -2, -4, and -5 domains observed from NOE, but M5, -6, and -7 overlap with f1, -2 domains of high flexibility observed from NOE. This suggests that conformational transitions occurring on the milli- to microsecond time scale, which are not sensed by the ^{1}H-^{15}N NOEs, make significant contributions to the R_2 values. The average R_2 value in all the domains M1–10 is roughly the same, implying that all of them have contributions from slow motions. Incidentally, in denatured apomyoglobin (8 M urea, pH 2.3) where there was no evidence for milli- to microsecond time scale motions, the domains of restricted motion derived from both ^{1}H-^{15}N NOE and ^{15}N R_2 were identical (38).

In the light of earlier reports (35–38) that hydrophobic interactions restrict the motions, we calculated the hydrophobic clusters along the sequence of the protein using the program HCA_Draw (27). These are shown by black bars (C1–7) at the bottom in Fig. 4*A*. The domains N1, -3, -4, and -5 in Fig. 4*A* and domains M1–3, M7–10 in Fig. 4*B* overlap with many of the hydrophobic clusters, though partially in a few cases. This positive correlation between restricted motion and hydrophobic clustering is in accordance with earlier observations on other proteins and suggests that side chain interactions do play a major role in local transient ordering of the protein chain. At the same time, N2 in Fig. 4*A* and M4–6 in Fig. 4*B* do not overlap with any cluster. These observations seem to lend partial support to the model of protein folding in which an initial hydrophobic collapse occurs following removal of denaturant; in the present case C1, -2, -3, -7, and parts of C4, -5 along the polypeptide chain are involved. The motional restriction in N2 and in M4–6 seems to occur through some other kind of local transient structure formation.

Many residues in HIVTD exhibit extremely low R_2 values, less than 2 s^{-1}, (Fig. 4*B*). These are indicative of very high flexibility; interestingly, this coincides with the positions of glycines and alanines (shown by filled and open circles respectively) along the chain. This lends support to the model proposed by Schwarzinger *et al.* (38) that Gly and Ala resi-

dues act as molecular hinges in the folding mechanism of a protein.

Conformational Transitions along the Folding Funnel—When the denaturing conditions are made slightly milder, the protein treks along the folding funnel. As mentioned before, at 5 M guanidine, the 6 M peaks in the HSQC spectrum are nearly conserved, and additional peaks corresponding to other partially folded conformers appear. This indicates that the secondary structural propensities of the conserved species, between 6 and 5 M guanidine, could only be marginally different. However, there could be dynamics differences, which would reflect on the conformational transitions on the folding funnel. We have monitored these effects from the changes in the residue-wise H^N-H^α coupling constants and ^{15}N transverse relaxation rates as described below. The differences in these parameters are also qualitatively evident in Fig. 1*B* itself.

Fig. 5*A* shows the changes in the coupling constants along the sequence of the protein as we go from 6 to 5 M guanidine. The changes are clearly not random, and there seems to be a decrease in the coupling constant for all the residues (barring a few). This would indicate a slight increase in the helical propensity for most residues in 5 M guanidine. Comparing with Fig. 2*A*, we observe that the stretch Thr-12–Asp-60, which has very little structural propensity in 6 M guanidine, barring a few residues (Glu-21, Leu-24, Asp-25) near the active site and a few near the hinge region (Glu-34, Ser-37, Gln-40, Lys-43, Pro-44), acquires some preferences in 5 M guanidine. This is a reflection on the hierarchy of the folding events.

The ^{15}N transverse relaxation rates showed a substantial increase in 5 M guanidine conditions for 75% of the residues, and the average R_2 value was higher by about 30%. A large portion of this increase may possibly come from assembly of the two monomeric halves into the dimer, which may be facilitated because of the head-to-tail covalent linkage through the linker in HIVTD. This transient assembly results in increased line widths. Such a conclusion is based on the observation that as the guanidine concentration is further reduced the protein starts acquiring autolytic activity (Fig. 1), implying that a proper specific dimer is being formed even in the presence of certain concentrations of the denaturant. However, as can be seen in Fig. 5*B* the changes in the R_2 values in going from 6 to 5 M guanidine are sequence-specific, and this reflects on the hierarchy of conformational transitions or folding propensities in the protein. In the completely unfolded state, which represents an ensemble of rapidly inter-converting conformers, there are very few exchange contributions to the transverse relaxation rate. As the protein starts folding searching for the stable configurations, conformational transitions start getting slower and then they contribute to the line widths reflected in the R_2 values. This can happen in a sequence-specific manner. As more and more contacts get established, the contributions of slow exchanges would tend to increase and spread along the sequence. Thus, under conditions which are still *near* the unfolded state where the protein is still dynamic and heterogeneous (note, the fully folded state also has minimal conformational exchange contributions to the R_2 values), the relative magnitudes of the changes in R_2 values (ΔR_2) in going from 6 to 5 M guanidine, along the sequence of the polypeptide chain, would reflect on the susceptibilities to local transient order formation or susceptibilities to folding conformational transitions. This will continue until persistent stable structures are formed, and then again there would be minimal exchange contributions. Referring to Fig. 5*B*, we can group the residues on the basis of increase in R_2 values (s^{-1}) to three ranges, 3.0–4.5, 1.5–3.0, and 0.0–1.5. This can be taken to identify the residue-level hierarchy of folding propensities. However, the third

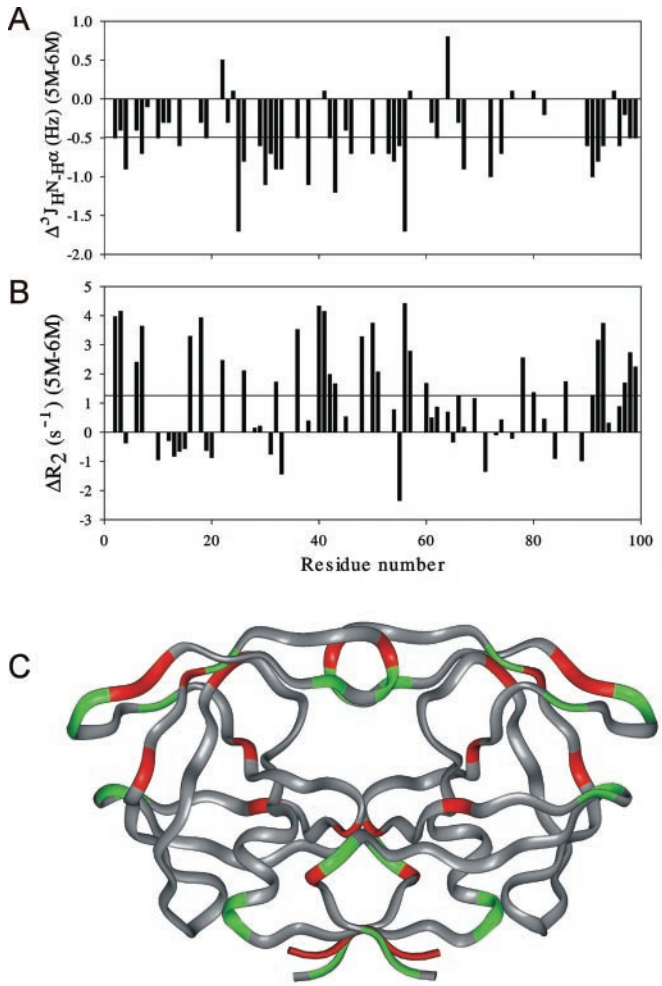

FIG. 5. *A*, changes in H^N-H^α coupling constants on decreasing the guanidine concentration from 6 to 5 M. The *horizontal line* indicates the average change. *B*, changes in R_2 values on decreasing the guanidine concentration from 6 to 5 M. The *horizontal line* indicates the average change. *C*, propensities of conformational transitions or folding events identified by different colors on the basis of the magnitudes of the R_2 changes, displayed on the structure of the native protein. *Green* residues have higher folding propensities than the *red* ones.

range (0.0–1.5) is roughly similar to the errors in R_2 measurements and hence may not be considered significant. Thus, it follows that residues Gln-2, Val-3, Gln-7, Gly-16, Gln-18, Met-36, Gly-40, Arg-41, Gly-48, Ile-50, Val-56, Gln-92, Ile-93 (ΔR_2 = 3.0–4.5 s^{-1}) have the highest propensities for conformational transitions, followed by residues Trp-6, Ala-22, Thr-26, Val-32, Trp-42, Lys-43, Gly-51, Arg-57, Asp-60, Gly-78, Leu-97, Asn-98, Phe-99 (ΔR_2 = 1.5–3.0 s^{-1}) at the second level. Many of these residues overlap with the hydrophobic clusters in Fig. 4*A*, indicating that conformational transitions may be driven by changes in the side chain interactions among hydrophobic residues when the denaturant concentrations are changed. Fig. 5*C* shows these hierarchical folding regions in a color-coded manner on the structure of the native protein. It is noteworthy that the residues involved in these hierarchical conformational transitions are either contiguous or are in close spatial proximity on the structure, suggesting a kind of cooperativity in the folding process. These conformational transitions could involve both formation of native contacts as well as destruction of the non-native contacts. Interestingly, the dimerization domain, the flaps, and their hinges show the highest propensity to conformational transitions in the HIVTD protein. It is, of course, possible that the effects observed for the dimerization domain may be a consequence of the tethering of the two monomers, which facilitates

bringing the C-terminal of the first monomer near the N-terminal of the second monomer to engage in the formation of a transient order, as in the native protein structure. All the above conformational transitions may lead to the formation of partially folded species as seen in Fig. 1*B*.

One might wonder whether the present data would throw any light on whether the folding of the protease is a simple two-state or a complex multistate process; this is of significance for understanding the mechanism of folding of the protein. Conventionally, this is discerned by monitoring the folding using a variety of probes such as circular dichroism, fluorescence, calorimetry, IR spectroscopy, etc. In the event that it is a two-state process, the folding profiles from all the techniques would be completely superimposable (39). In the present case, generation of a complete folding profile by any of the conventional techniques, either in the kinetic or in the equilibrium experiments, is severely hampered because of the autocleavage property of the protein. Nevertheless, we observe that in 5 M guanidine, the HSQC spectra show many extra peaks compared with the 6 M spectrum and most of these peaks do not correspond to the native state of the protein. This means that there is at least one other non-native state that exchanges slowly with the 6 M denatured state. It is more likely that there are several non-native states that are slightly different from the 6 M denatured state, and these arise from different protein molecules at the surface of the folding funnel, following different paths along the folding funnel. This is consistent with the different extents of residue-level folding propensities along the sequence of the polypeptide chain. Thus we believe that the protein has a complex folding mechanism and even the early events involve many states near the surface of the folding funnel. Different molecules may follow different paths, and there may be local cooperativity of folding transitions along each path.

CONCLUSIONS

We have attempted here to obtain useful insights into the early events in the folding pathway of HIV-1 protease, using a tethered dimer construct as its representative. From a number of NMR parameters determined under conditions of guanidine denaturation, the early folding nuclei have been identified. These include native as well as non-native secondary structures, and the native structures span the residues at the active site, the hinge, and the dimerization domain. Interestingly, proline residues seem to have significant influence on the local structural preferences in the denatured state of the protein. Hydrophobic interactions between the side chains seem to cause substantial restrictions on the motional properties along the chain, and these may be important in driving the folding process. It is also observed that glycines and alanines cause enhanced conformational flexibilities and thus may act as molecular hinges in the folding as suggested earlier (38). As the denaturant concentration is progressively reduced, partially folded species start appearing and the protein begins to acquire autolytic activity. The investigations at 6 and 5 M guanidine denaturing conditions show that the dimerization domain, the flaps, and the hinges or the elbow of the protein have the highest propensity to conformational transitions. The residue-wise graded propensities also suggest a certain degree of local cooperativity in these conformational transitions. Finally, from the observation of more than one non-native species even in 5 M guanidine and the sequence-wise hierarchy of the conformational transitions, it appears that the protein folds following a complex mechanism with different molecules following different pathways. To our knowledge this is the first experimental description of the early events in the folding mechanism of HIV-1 protease.

Acknowledgments—We thank the National Facility for High Field NMR at the Tata Institute of Fundamental Research 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, Mumbai.

REFERENCES

1. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol.* **4,** 10–19.
2. Plotkin, S. S., and Onuchic, J. N. (2002) *Q. Rev. Biophys.* **35,** 111–167
3. Baldwin, R. L., and Rose, G. D. (1999) *Trends. Biochem. Sci.* **24,** 77–83
4. Baldwin, R. L. (2002) *Science* **295,** 1657–1658
5. Dobson, C. M., and Karplus, M. (1999) *Curr. Opin. Struct. Biol.* **9,** 92–101
6. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000) *Trends Biochem. Sci.* **25,** 331–339
7. Englander, S. W. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 213–238
8. Raschke, T. M., Kho, J., and Marqusee, S. (1999) *Nat. Struct. Biol.* **6,** 825–831
9. Dyson, H. J., and Wright, P. E. (1996) *Annu. Rev. Phys. Chem.* **47,** 369–395
10. Killick, T. R., Freund, S. M. V., and Fersht, A. R. (1999) *Protein Sci.* **8,** 1286–1291
11. Balbach, J., Steegborn, C., Schindler, T., and Schmid, F. X. (1999) *J. Mol. Biol.* **285,** 829–842
12. Dobson, C. M., and Hore, P. J. (1998) *Nat. Struct. Biol.* **5** (suppl.), 504–507
13. Barbar, E. (1999) *Biopolymers* **51,** 191–207
14. Shortle, D., and Ackerman, M. S. (2001) *Science* **293,** 487–489
15. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) *Science* **249,** 1544–1548
16. Jennings, P. A., and Wright, P. E. (1993) *Science* **262,** 892–896
17. Raschke, T. M., and Marqusse, S. (1997) *Nat. Struct. Biol.* **4,** 298–304
18. Dabora, J. M., Pelton, J. G., and Marqusee, S. (1996) *Biochemistry* **35,** 11951–11958
19. Arai, M., and Kuwajima, K. (2000) *Adv. Protein Chem.* **53,** 209–282
20. Chamberlain, A. K., and Marqusee, S. (2000) *Adv. Protein Chem.* **53,** 283–328
21. Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A., and Baum, J. (1993) *Biochemistry* **32,** 5681–5691
22. Schulman, B. A., Redfield, C., Peng, Z. Y., Dobson, C. M., and Kim, P. S. (1995) *J. Mol. Biol.* **253,** 651–657
23. Pillai, B., Kannan, K. K., and Hosur, M. V. (2001) *Proteins* **43,** 57–64
24. Kumar, M., Kannan, K. K., Hosur, M. V., Bhavesh, N. S., Chatterjee, A., Mittal, R., and Hosur, R. V. (2002) *Biochem. Biophys. Res. Commun.* **294,** 395–401
25. Bhavesh, N. S., Panchal, S. C., Mittal, R., and Hosur, R. V. (2001) *FEBS Lett.* **509,** 218–224
26. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003
27. Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. (1987) *FEBS Lett.* **224,** 149–155
28. Richards, F. M., Eisenberg, D. S., and Kuriyan, J. (eds) (2002) *Adv. Prot. Chemistry: Unfolded Proteins,* Vol. 42, Academic Press, San Diego, CA
29. Ramachandran, G. N., and Sasisekharan, V. (1968) *Adv. Protein Chem.* **23,** 283–438
30. Dyson, H. J., and Wright, P. E. (2001) *Methods Enzymol.* **339,** 258–270
31. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) *J. Am. Chem. Soc.* **123,** 2970–2978
32. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) *J. Mol. Biol.* **274,** 152–159
33. Kazmirski, S. L., Wong, K. B., Freund, S. M. V., Tan, Y. J., Fersht, A. R., and Daggett, V. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4349–4354
34. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) *Science* **295,** 1719–1722
35. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry* **39,** 2894–2901
36. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) *Biochemistry* **40,** 3561–3571
37. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol.* **5,** 148–155
38. Schwarzinger, S., Wright, P. E., and Dyson, H. J. (2002) *Biochemistry* **41,** 12681–12686
39. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Folding,* pp. 517–518, W. H. Freeman & Co., New York