Real time NMR monitoring of local unfolding of HIV-1 protease tethered dimer driven by autolysis

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

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

Received 8 March 2001; revised 9 April 2001; accepted 13 April 2001

First published online 4 May 2001

Edited by Thomas L. James

Abstract Structural studies in proteases have been hampered because of their inherent autolytic function. However, since autolysis is known to be mediated via protein unfolding, careful monitoring of the autolytic reaction has the potential to throw light on the folding–unfolding equilibria. In this paper we describe real time nuclear magnetic resonance investigations on the tethered dimer construct of the human immunodeficiency virus-1 protease, which have yielded insights into the relative stabilities of several residues in the protein. The residues lying along the active site (bottom, side and top of the active site) and those in helix have lower unfolding free energy values than the other parts of the protein. The residue level stability differences suggest that the protein is well suited to adjust itself in almost all the regions of its structure, as and when perturbations occur, either due to ligand binding or due to mutations. © **2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.**

Key words: Human immunodeficiency virus-1 protease; Autolysis; Local unfolding; Real time nuclear magnetic resonance

1. Introduction

Human immunodeficiency virus (HIV)-1 protease, an important enzyme for the AIDS virus [1], is a 22 kDa homodimer that self-assembles from two identical polypeptide chains, 99 residues each, and resembles other monomeric aspartyl proteases [2]. Extensive X-ray crystallographic structural studies on the protein and its complexes with a variety of inhibitors have revealed that the structure of the protein contains a conserved catalytic site at the interface of the two monomers, a flap region formed by two loops, one each from the two monomers, and the two monomers are held together by a short β-sheet formed by the amino-terminal of one monomer and the carboxyl-terminal of the other [3]. This non-covalent association of two identical subunits is very critical for the protein to remain in folded state and hence for its function.

Understanding protein folding/unfolding is of paramount importance in designing proteins that are resistant to denaturation, or in engineering proteins with new functions. It is also important for designing effective inhibitors for enzyme actions. However, HIV-1 protease, because of its inherent autolytic activity (discussed in [4,5]), has evaded a detailed investigation on the solution structure, stability and folding/unfolding equilibria. In the present study, we have used this very autolytic activity to investigate folding–unfolding equilibria at residue level, and this has been achieved by real time monitoring of the changes in the nuclear magnetic resonance (NMR) spectra. For this we have used a tethered dimer (TD) construct wherein the C-terminal of one monomer is covalently joined to the N-terminal of another monomer by a GGSSG linker [6,7]. Previous studies on TDs have shown that the overall three-dimensional structure of the TD remains largely unaffected when compared to the homo-dimeric protein [8,9], and thus the TD serves as a good surrogate enzyme to study the effects of various hetero-dimer formations on the stability and functions of the enzyme.

2. Materials and methods

2.1. NMR sample and real time experiments

The isotopically labelled (^{15}N) protein sample for the NMR experiments was prepared as described before [10]. The NMR experiments were recorded on a 1 mM sample of ^{15}N-labelled HIV-1 TD in NMR buffer: 50 mM Na-acetate, pH 5.2, 5 mM ethylenediaminetetraacetic acid and 150 mM dithiothreitol. A series of two-dimensional ^{15}N heteronuclear single quantum coherence (HSQC) [11] spectra were collected as a function of time at 32°C. Each spectrum was a result of 70 complex t_1 increments, with eight scans for each free induction decay, and took approximately 30 min. The kinetic data was collected over about 19 h.

2.2. Intensity measurements and normalisation

The peaks in the HSQC spectra of the protein, recorded as a function of time, were integrated by FELIX 97.0 software (Molecular Simulations Inc., San Diego, CA, USA). The intensities of peaks belonging to the folded protein in each spectrum were divided by the total intensity (folded+unfolded) of the W42 side-chain peak within the same spectrum, to normalise intensity variations from one spectrum to another; the total intensity (folded+unfolded) of W42 side-chain peaks was nearly constant (indicating conserved magnetisation) throughout the kinetic experiment. Subsequently, the intensities were further normalised with respect to the first spectrum of the kinetic data.

2.3. Gel electrophoresis

The recombinant purified protein in the NMR buffer was concentrated to ~1 mM as required for NMR experiments, kept at 32°C, and from this, aliquots were taken at regular time intervals and analysed on 12% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE).

*Corresponding author. Fax: (91)-22-215 2110.
E-mail: hosur@tifr.res.in

Abbreviations: HIV, human immunodeficiency virus; HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulphate

0014-5793/01/$20.00 © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)02426-7

3. Results and discussion

3.1. NMR monitors of the autolytic reaction and protein unfolding

In the present work, we have exploited the autolytic reaction of the HIV-1 protease TD to monitor local unfolding characteristics of the protein by real time NMR. The autolytic reaction was found to be concentration and temperature dependent, being slow at low protein concentration and low temperature. Thus we were able to record a HSQC spectrum of the fully folded protein using a concentration of ~200 μM and a temperature of 25°C. This is shown in Fig. 1. However, the stability of the protein was not large enough to be able to record all the triple resonance spectra using a doubly (^{13}C, ^{15}N) labelled sample for sequence specific assignments. Hence, the NH and ^{15}N assignments of the free protein required for the present analysis, were obtained by comparison with the published spectra of protein–inhibitor complexes. Chemical shifts of many peaks were seen to be conserved in the spectra of many protein–inhibitor complexes, such as with DMP-323 (symmetric) [12], KNI-272 [13] and pepstatin-A (asymmetric) [10]. These invariant peaks, which have been identified in Fig. 1 by specific assignments, are our monitors in this study.

The real time NMR monitoring of the autolytic reaction

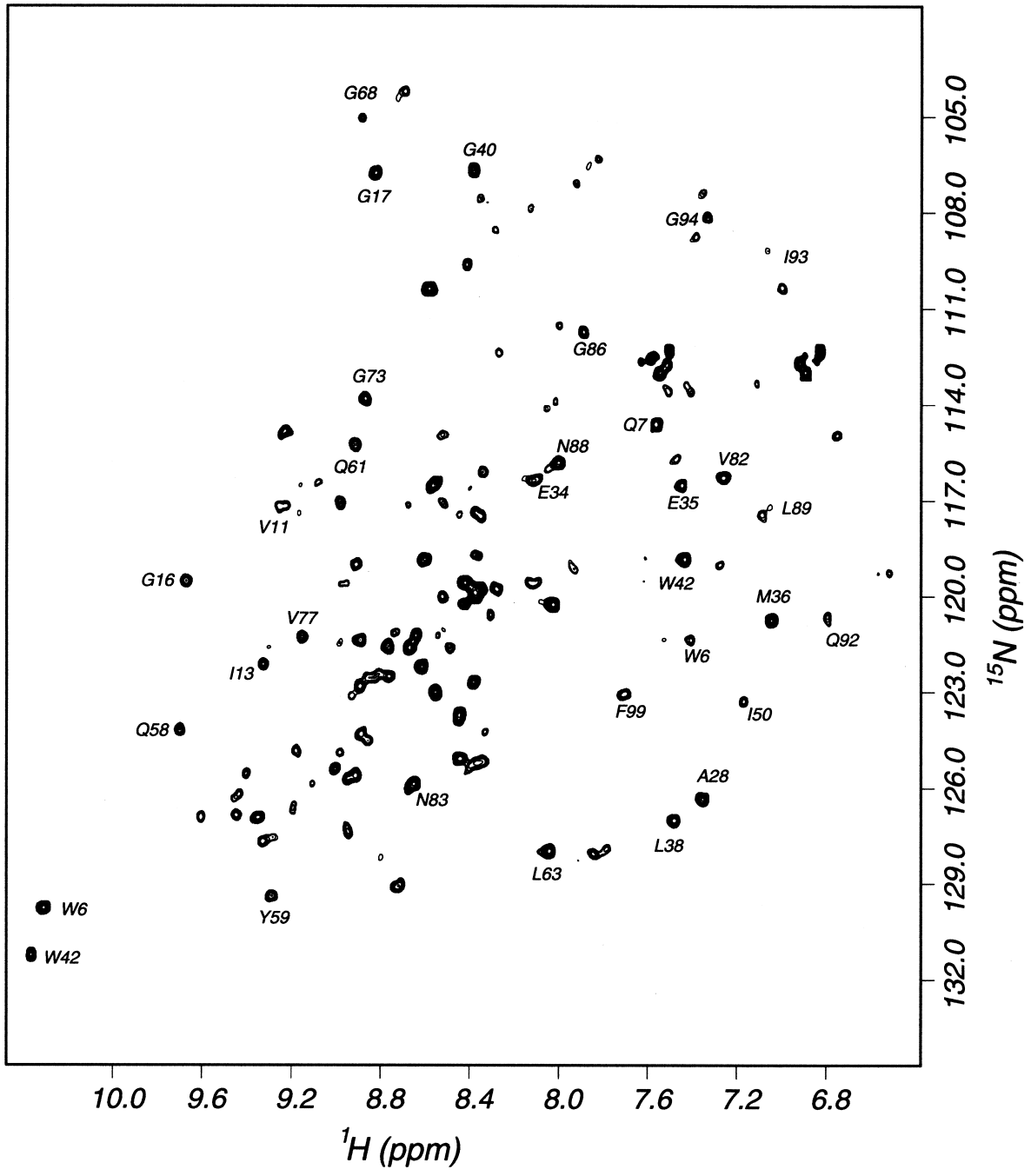

Fig. 1. ^{1}H–^{15}N HSQC spectrum of HIV-1 protease TD with sequence specific assignments of those peaks which were monitored during the kinetic experiment of autolysis.

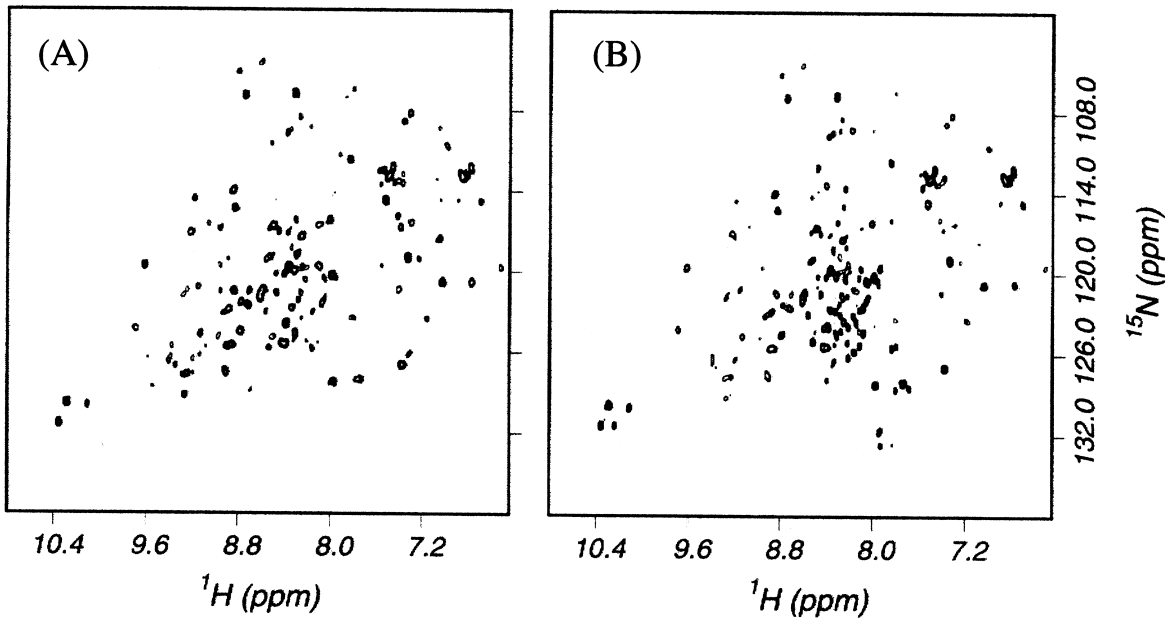

Fig. 2. The first (A) and last (B) HSQC spectra of the real time NMR data on HIV-1 protease TD (1 mM, 32°C). Each spectrum in the kinetic experiment took approximately 30 min.

was carried out at 32°C and a protein concentration of 1 mM. Fig. 2 shows the first and the last HSQC spectra of the kinetic data to display the spectral changes through the experiment. Compared to Fig. 1, a number of additional peaks are seen which arise as a result of autolysis. These, by virtue of their position in the central narrow chemical shift range for protons, indicate that they belong to unfolded protein fragments. Most of the folded protein peaks have lost some intensity and these report, as we shall show below, on the unfolding process driven by the autolytic reaction.

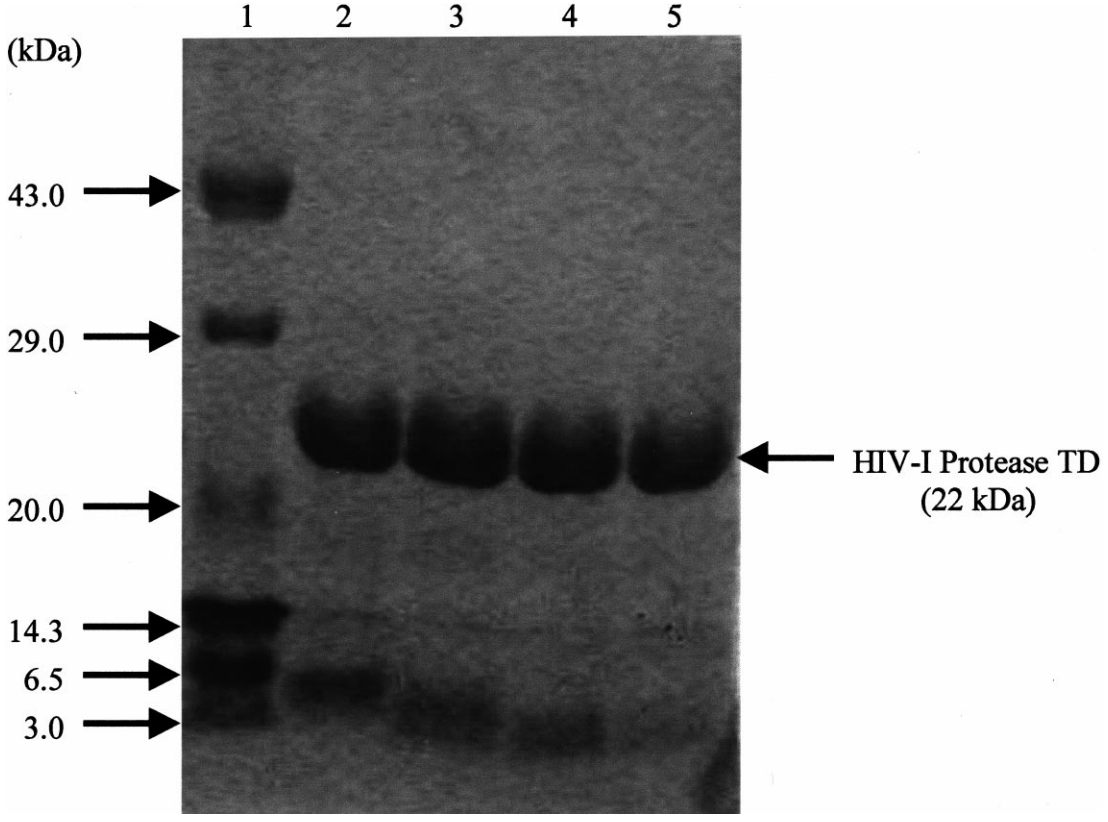

Fig. 3. 12% SDS–PAGE gel photograph showing the autolysis reaction of HIV-1 protease TD. Lane 1 shows the molecular weight markers. Lanes 2, 3, 4 and 5 show the bands for aliquots of reaction mixture at 3, 6, 12 and 24 h time intervals. Aliquots of the sample were taken from a protein solution at the NMR concentration at the indicated time intervals.

3.2. PAGE analysis

Fig. 3 shows the gel analysis of the autolytic reaction. The protein concentrations here are the same as in the real time NMR experiments. The 22 kDa band is the folded protein and the lower molecular weight band appears as a consequence of autolysis. This band decreases with time due to further chewing up of the protein into smaller fragments, which are not observable in the gel.

HIV-1 protease has three autolysis sites, at Leu_5–Trp_6, Leu_{33}–Glu_{34}, and Leu_{63}–Ile_{64}. From the crystal structure [14,15], two of the three autolysis sites (Leu_{33}–Glu_{34} and Leu_{63}–Ile_{64}) are seen to be in β-strands that are engaged in extensive interactions with neighbouring β-strands on both the monomers. The third site (Leu_5–Trp_6) is in the N-terminal region, part of which is in a four-stranded intramolecular β-sheet. Thus for autolysis to occur the protein has to unfold and render the peptide bonds of these three sites accessible to the folded enzyme.

Of the three sites of autolysis, that at Leu_5–Trp_6 would appear to be the most vulnerable, since it alone relies totally on intermolecular interactions and these interactions are most vulnerable on the energy scale. This is further confirmed by our NMR data. The first kinetic ^{15}N-HSQC spectrum displays a partner peak to the folded W6 side-chain (indole ring ^{15}N) peak, but such a peak is absent for the folded W42 side-chain peak (Fig. 2A). This situation is true for the first 100 min during which the W6 side-chain partner (unfolded) peak increases in intensity but the W42 side-chain partner (unfolded) peak is absent. In the third kinetic spectrum the W42 side-chain peak makes its first appearance. This indicates that the Leu_5–Trp_6 region adopts an extended conformation required for hydrolysis first, while the other regions, particularly near W42, remain intact. The other two regions comprising autolysis sites Leu_{33}–Glu_{34} and Leu_{63}–Ile_{64} are engaged in intramolecular interactions and thus are more stable within the TD. After the first cleavage, the protein unfolds further with time, and the other two autolysis sites at residues 33 and 63 start to become accessible.

3.3. Kinetic scheme of autolysis

The kinetic scheme of autolysis reaction in most proteases [16,17] is described as shown below:

$$\mathrm{N} + \mathrm{N} \underset{k_f}{\overset{k_u}{\rightleftarrows}} \mathrm{N} + \mathrm{U} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{NU} \xrightarrow{k_c} \mathrm{N} + \mathrm{P} \quad (1)$$

where N is the folded protein, U is the unfolded protein susceptible to autolysis, NU is the protein–substrate (folded–unfolded) complex and P is the autolysed product. k_u is the unfolding rate, k_f is the folding rate, k_1 is the association rate of NU complex, k_2 is the dissociation rate of the complex and k_c is the catalytic rate. The above scheme can be written simplistically as a combination of two schemes as below:

$$\mathrm{N} \underset{k_f}{\overset{k_u}{\rightleftarrows}} \mathrm{U} \quad (2)$$

$$\mathrm{N} + \mathrm{U} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{NU} \xrightarrow{k_c} \mathrm{N} + \mathrm{P} \quad (3)$$

The schemes in both Eqs. 2 and 3 are well established schemes. Eq. 2 represents a two state folding/unfolding equilibrium and Eq. 3 represents a Michaelis–Menten enzyme kinetic scheme. Now, for Eq. 3, if $[U] \ll K_M$ (which is true in the present case), then, the rate of depletion of U [18] is given by:

$$\frac{d[U]}{dt} = -\frac{k_c}{K_M}[N][U] \quad (4)$$

Considering steady state for U in Eq. 1 and using Eq. 4 we obtain:

$$k_u[N] - k_f[U] - \frac{k_c}{K_M}[N][U] = 0 \quad (5)$$

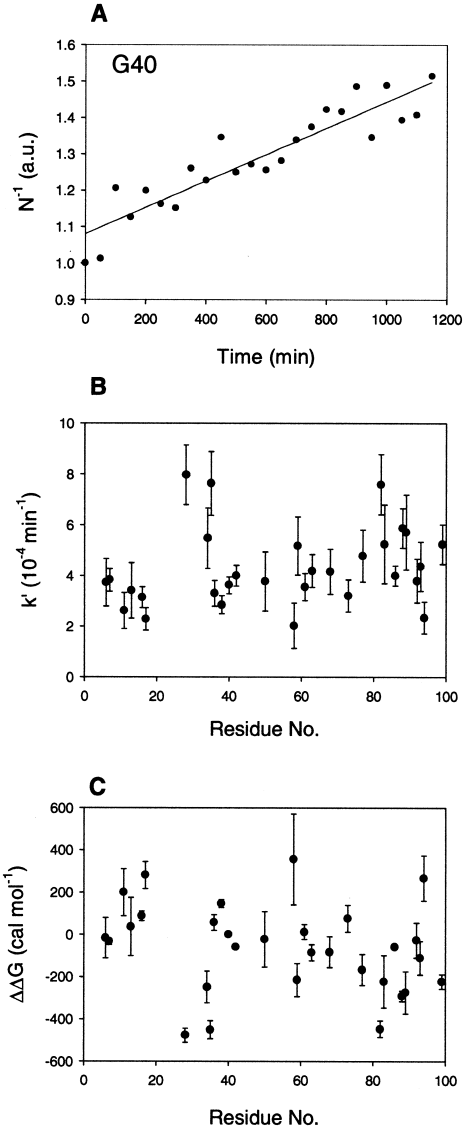

Fig. 4. A: Inverse of normalised peak intensities of the backbone ^{1}H–^{15}N correlation peak of G_{40} as a function of time through the kinetic experiment, and their fitting to Eq. 12 (solid line). The fitting was done using SigmaPlot 5.0. The R_{sqr} value for the fit was 0.84. B, C: Plots of rate constant (k') (B) and relative free energy value ($\Delta\Delta G$) (C) derived from the curve fitting parameters, for 30 residues monitored during the unfolding process. $\Delta\Delta G$'s are calculated with respect to G_{40}. The errors in the individual values have been indicated. The average error is 0.065 kcal mol^{-1} and the variation in $\Delta\Delta G$ values is clearly much larger.

The N ↔ U equilibrium constant K is obtained from Eq. 5 by simple algebra as:

$$K = \frac{[U]}{[N]} = \frac{k_u}{k_f + \frac{k_c}{K_M}[N]} \quad (6)$$

Now, from Eq. 2 the rate of depletion of N is given by:

$$\frac{d[N]}{dt} = -k_u[N] + k_f[U] \quad (7)$$

Substituting for [U] from Eq. 6 into Eq. 7 we obtain:

$$\frac{d[N]}{dt} = -\frac{k_u\frac{k_c}{K_M}[N]^2}{k_f + \frac{k_c}{K_M}[N]} \quad (8)$$

For most proteins the folding rates are of the order of 10^3–10^5 s^{-1} (reviewed in [19]), and the catalytic rates occur in minute time scales [5]. The autolysis rates are even slower. Thus, in the denominator of Eq. 8, $k_f \gg (k_c/K_M)[N]$ and hence:

$$\frac{d[N]}{dt} = -\frac{k_u}{k_f}\frac{k_c}{K_M}[N]^2 \quad (9)$$

or

$$\frac{d[N]}{dt} = -k'[N]^2 \quad (10)$$

where

$$k' = K\frac{k_c}{K_M} \text{ and } K = \frac{k_u}{k_f} \quad (11)$$

Integrating Eq. 10 with respect to time t and applying the condition $N = N_0$ at time $t = 0$ one gets:

$$\frac{1}{[N]} = \frac{1}{[N]_0} + k't \quad (12)$$

Thus, a plot of 1/[N] versus t gives a straight line with slope k', which allows estimation of the equilibrium constant.

3.4. Analysis of real time NMR data

In the HSQC spectra, the intensity of a peak is directly related to the protein concentration. However, different peaks may have different intensities due to unequal exchange rates with the solvent of the individual residues. Even so, the changes in the normalised peak intensities (see Section 2) as a function of time still provide useful kinetic information discussed in Section 3.3. Further, the various peaks in the HSQC spectra may have different decay profiles depending on the rates of local unfolding events contributing to the generation of autolysis susceptible states and these report, therefore, on the local equilibrium parameters.

We monitored 30 peaks in the HSQC spectrum (Fig. 1) of the HIV TD protein, for which we could obtain unambiguous assignments. It is fortuitous that, the residues corresponding to these peaks are well spread out over the whole sequence of the molecule and hence provide valuable monitors for the entire protein. The intensity decay profiles of the different backbone as well as the tryptophan side-chain cross peaks in the HSQC spectrum could be fitted to Eq. 12; Fig. 4A shows the fitting for residue G_{40} as an illustration. Fig. 4B shows the derived backbone rates (k') for the 30 different residues distributed over the length of the protein. Clearly, there is a significant gradation in the rates, indicating residue specific local changes.

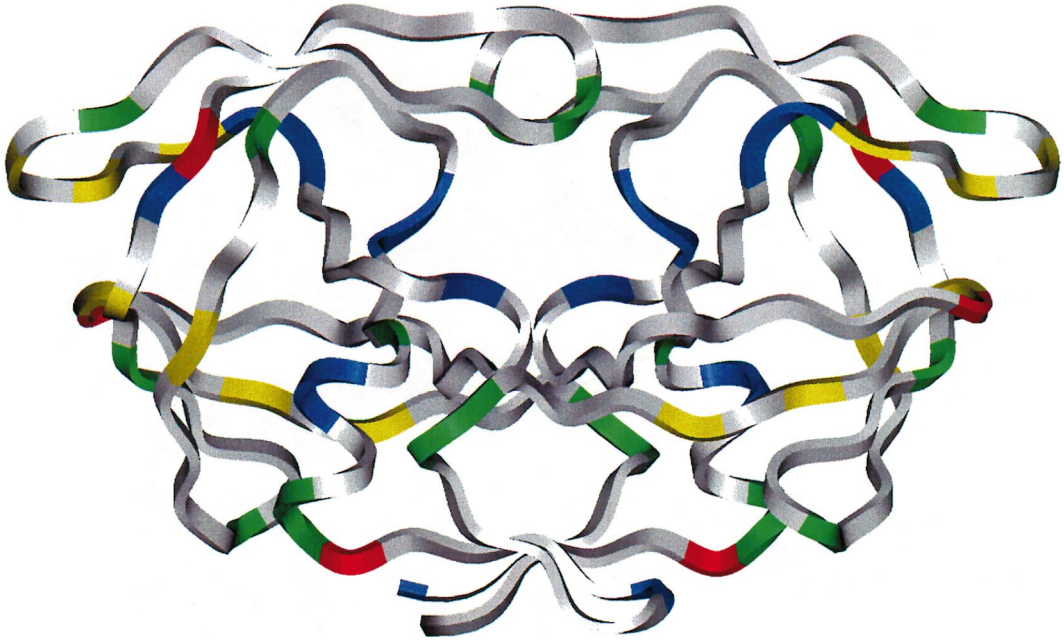

Fig. 5. Display of relative stabilities of the 30 residues on the structure of the HIV-1 protease TD in a colour coded manner. The residues in both the halves of the dimer have been marked. The order of the stabilities is: blue (−0.5 to −0.2 kcal mol^{-1}) < green (−0.2 to 0.0 kcal mol^{-1}) < yellow (0.0 to 0.2 kcal mol^{-1}) < red (0.2 to 0.4 kcal mol^{-1}). Thus the residues lining the active site and in the helix are relatively more susceptible to unfolding compared to those in other parts of the protein.

From Eq. 11 it follows that the folding/unfolding equilibrium constant (K), and hence the free energy changes, can be calculated from the knowledge of k', provided the value of k_c/K_M is known. For autolysis, this is a difficult quantity to measure. Nonetheless, assuming that this quantity is a constant for the protein, we have calculated relative free energy changes associated with local unfolding as:

$$\Delta\Delta G_i = -RT[\ln K_i - \ln K_r] = -RT[\ln k'_i - \ln k'_r] \quad (13)$$

where the subscript r refers to a chosen reference residue. We have chosen G_{40} as the reference since this has the least error in k'. The residue-wise variations in relative local unfolding energies are plotted in Fig. 4C. Clearly, the variations are much larger than the average error (65.5 cal mol^{-1}) and hence can be considered very significant.

Fig. 5 shows the residue-wise relative free energy data on the crystal structure of the HIV-1 TD molecule (PDB ID: 1G6L) [9] in a colour coded manner. The lower the relative free energy change for a given residue, the lower is its stability and vice versa. The residues (N88, L89, I93) in the α-helix have lowest relative free energies compared to those in other parts of the protein. Next on the scale are the residues present in the β-sheets and loops. It is interesting to note that all the residues within a particular secondary structure element do not behave similarly. The residues lining the active site (top, bottom and side) are seen to be less stable compared to those in the other parts of the protein. This points to a flexible nature of the protein, which may have important implications for the accommodation of different substrates in its active site.

In conclusion, we have described here a novel approach, which uses the very problematic autolytic activity of a HIV TD protease protein to derive the relative stabilities of 30 different residues. The relative free energies indicate easy access of the unfolded states by the native protein at residue level. This implies that the protein can adjust itself in every portion of its structure to respond to the perturbations caused either by ligand binding or by mutations. From the point of view of drug design, the above results may suggest the possible vulnerable sites to target for the destabilisation of the protein. It is clear that the catalytic site is not the only site that can be targeted for this purpose. The independent local folding/unfolding events implicitly seen here, have important implications for the understanding of the 'folding problem' in general.

Acknowledgements: The facilities provided by the National Facility for High Field NMR, supported by the Government of India, are gratefully acknowledged. We thank Dr. M.V. Hosur of Bhabha Atomic Research Centre, Mumbai for the clone of the TD protein.

References

[1] Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M. and Sigal, I.S. (1988) Proc. Natl. Acad. Sci. USA 85, 4686–4690.
[2] Davies, D.R. (1990) Annu. Rev. Biophys. Biophys. Chem. 19, 189–215.
[3] Wlodawer, A. and Erickson, J.W. (1993) Annu. Rev. Biochem. 62, 543–585.
[4] Grant, S.K., Deckman, I.C., Culp, J.S., Minnich, M.D., Brooks, I.S., Hensley, P., Debouck, C. and Meek, T.D. (1992) Biochemistry 31, 9491–9501.
[5] Mildner, A.M., Rothrock, D.J., Leone, J.W., Bannow, C.A., Lull, J.M., Reardon, I.M., Sarcich, J.L., Howe, W.J., Tomich, C.C., Smith, C.W., Heinrikson, R.L. and Tomasselli, A.G. (1994) Biochemistry 33, 9405–9413.
[6] Cheng, Y.E., Yin, F.H., Foundling, S., Blomstrom, D. and Kettner, C.A. (1990) Proc. Natl. Acad. Sci. USA 87, 9660–9664.
[7] Dilanni, C.L., Davis, L.J., Holloway, M.K., Herber, W.K., Darke, P.L., Kohl, N.E. and Dixon, R.A.F. (1990) J. Biol. Chem. 265, 17348–17354.
[8] Bhat, T.N., Baldwin, E.T., Liu, B., Chen, Y.E. and Erickson, J.W. (1994) Nat. Struct. Biol. 1, 552–556.
[9] Pillai, B., Kannan, K.K. and Hosur, M.V. (2001) Proteins 43, 57–64.
[10] Panchal, S.C., Pillai, B., Hosur, M.V. and Hosur, R.V. (2000) Curr. Sci. 79, 1684–1695.
[11] Bodenhausen, G. and Ruben, D.J. (1980) Chem. Phys. Lett. 69, 185–189.
[12] Yamazaki, T., Nicholson, L.K., Torchia, D.A., Stahl, S.J., Kaufman, J.D., Wingfiled, P.T., Domaille, P.J. and Campbell-Burk, S. (1994) Eur. J. Biochem. 219, 707–712.
[13] Freedberg, D.I., Wang, Y., Stahl, S.J., Kaufman, J.D., Wingfiled, P.T., Kiso, Y. and Torchia, D.A. (1998) J. Am. Chem. Soc. 120, 7916–7923.
[14] Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.H. (1989) Science 245, 616–621.
[15] Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A. (1989) Science 246, 1149–1152.
[16] Vriend, G., Berendsen, H.J.C., Van den Burg, B., Venema, G. and Eijsink, V.G.H. (1998) J. Biol. Chem. 273, 35074–35077.
[17] Mitchinson, C. and Wells, J.A. (1989) Biochemistry 28, 4807–4815.
[18] Atkins, P.W. (1990) in: Physical Chemistry, Chap. 26, pp. 804–805, Oxford University Press, Oxford.
[19] Plaxco, K.W., Simons, K.T., Ruczinski and Baker, D. (2000) Biochemistry 39, 11177–11183.