ELSEVIER

Biophysical Chemistry 123 (2006) 1 – 10

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Following autolysis in proteases by NMR: Insights into multiple unfolding pathways and mutational plasticities

Amarnath Chatterjee, Ramakrishna V. Hosur *

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

Received 22 January 2006; received in revised form 20 March 2006; accepted 20 March 2006
Available online 30 March 2006

Abstract

Biophysical studies in proteases are severely hampered due to the auto-cleavage property of these enzymes. In this context, we develop here a kinetic model and an NMR-based strategy to use this very autolytic property to derive useful insights into multiple unfolding pathways and mutational plasticities in these proteins. The basic idea lies in the interpretation of the auto-cleavage-driven decay of the folded protein peaks in the HSQC spectra as a function of time. The different peaks are seen to decay at different rates. As unfolding is the rate-determining step in the auto-cleavage reaction, the NMR spectral changes reflect on local unfolding processes at the residue level. A formalism is presented to gain insights into unfolding free energies and evaluate local perturbations due to single point mutations. The model is applied to HIV-1 protease-tethered dimer as an example, considering mutations at a particular site. Significant perturbations are seen even at very remote areas from the site of the mutation.

Keywords: Real-time NMR; Autolysis; Mutational plasticity; Multiple unfolding pathways

1. Introduction

Insights into protein unfolding mechanisms and stabilities are commonly derived from a combination of several biophysical techniques such as real-time NMR involving hydrogen exchange, fluorescence, circular dichroism, mass spectrometry, and so on [1–9]. However, in proteases, the autolytic activity is a major impediment for such detailed studies. Nevertheless, it is also recognized that autolysis itself proceeds via unfolding in some sense and thus has the potential to throw light on the unfolding mechanisms in the protein. Following protein engineering methods, Vriend et al. [10] investigated the effects of different mutations on the overall stabilities and inactivation of thermolysin-like proteases and concluded that (i) partial unfolding is the rate-limiting step for protease inactivation and, (ii) unfolding can start from many parallel sites on the protein and they can produce different species which are *sufficiently unfolded* to undergo autolysis. Similar conclusions were arrived at by Rose et al. [11] and Mildner et al. [12] in the case of HIV-1 protease, based on activity measurements in different mutant proteins. Besides, MALDI, a technique which provides molecular mass information, used in conjunction with protein engineering experiments, has provided insights into sequence–autolysis relationships in proteases [13–16]. However, for most proteases which have multiple cleavage sites, the hierarchy of cleavages has been difficult to establish; some insights regarding intrinsic cleavage differences have been derived, however, using peptides corresponding to the potential autocleavage sites as substrates [17–20]. Furthermore, thermodynamic information such as stabilities and mutational plasticities has been difficult to derive for autolytic proteases.

In this context we presented earlier a simplistic model to derive useful information on proteases from the autolytic reaction itself [21]. In the present paper, we have discussed the time dependence of the intensities of the peaks in the real-time NMR spectra, on the relative magnitudes of the primary and secondary cleavage rates, and show that, under certain experimental conditions, reasonable estimates of primary cleavage rates can be extracted. These reflect on the possibility of a variety of unfolding pathways. Furthermore, we show that comparison of the rates in two mutant proteins differing at a particular site by a single residue provides insights into

* Corresponding author. Tel.: +91 22 2280 4545; fax: +91 22 2280 4610.
E-mail address: hosur@tifr.res.in (R.V. Hosur).

0301-4622/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2006.03.013

mutational plasticities in the protease. This could possibly be the only way to extract such information in proteases, which are not readily analyzable by NMR hydrogen exchange experiments, the most commonly used method for deriving local stability and plasticity information in proteins.

The model is applied to HIV-1 protease-tethered dimer, which functions similarly to the usual homodimeric enzyme, as an example. Three different mutant proteins, C95M, (C95M, C1095A) and (C95M, C1095F) have been used for comparing the effects of single point mutations; the numbering goes from 1 to 99 in the first monomer and from 1001 to 1099 in the second monomer. These three mutant proteins are referred hereafter as, HIVTD-C, HIVTD-A and HIVTD-F, respectively. These were prepared in different contexts (C95F is, for example, a drug-resistant mutation in the normal protease, C95M is observed in HIV-2 protease, while C95A prevents oxidation and enhances stability). The choice of mutations has no particular significance for the present study except that their crystal structures are very similar [22–24]. Needless to say that, such investigations can be extended to other mutant proteins prepared with specific purposes.

2. Materials and methods

2.1. Protein preparation

The original clone of the HIV-1 protease-tethered dimer obtained as a kind gift from Dr. J. W. Erickson, had C95M mutation only on the first monomer. The clones for HIVTD-A and HIVTD-F were obtained using standard PCR-based strategies. Protein expression and purification protocols were also the same as described earlier [21].

All the isotopically labelled (^{15}N) proteins required for this work were bacterially expressed and purified as described previously [22]. The yield of the protein in each case was around 40 mg/l. For NMR sample preparation, the protein solution was concentrated and exchanged with pH 5.2 buffer containing 50 mM Na-acetate, 5 mM EDTA, 150 mM dithiothreitol (DTT), by ultrafiltration at 4 °C; such a high DTT concentration was used to prevent cysteine oxidation during the course of the experiments.

2.2. Mass spectroscopy

MALDI–TOF mass spectrometry analyses were carried out with Micromass (UK) MALDI–TOF Spec 2E spectrometer equipped with a UV nitrogen laser (337 nm) and a dual microchannel microplate detector. The recombinant purified protein in the NMR buffer was concentrated to ~1 mM as required for NMR experiments, kept at 42 °C, and from this, aliquots were taken at regular time intervals and immediately diluted with acetic acid to ~20 μM. This was done for two reasons, first this stopped the autolysis reaction by unfolding the protein; second MALDI experiments require low concentrations of a protein sample. Further 1 μl of this protein aliquots was mixed with 1 μl of freshly prepared matrix solution (10 mg/ml of 2,5-dihydroxybenzoic acid in 3 : 2 0.1% TFA/acetonitrile) and a total of 1 μl of this mixture was placed on the stainless steel probe plate and allowed to dry at room temperature. The spectra were recorded in the positive reflector linear mode at an accelerated voltage of 20 kV in the range from 2500 to 25 000 Da. For each measurement, the spectra were externally calibrated using myoglobin and trypsinogen.

2.3. NMR experiments

The NMR spectra were recorded on a 600 MHz three-channel *Varian Unity Plus* spectrometer equipped with pulsed field gradients, with 1 mM samples of ^{15}N-labeled HIVTD's in NMR buffer. A series of 2D ^{15}N HSQC spectra were collected as a function of time at 42 °C. Each spectrum was a result of 70 complex t_1 increments, with 4 scans, for each *fid* (free induction decay), and took approximately 11 min. The kinetic data was collected over about 100 min for each protein. In every case the protein preparations were carried out in identical fashions so as to maintain the same time profile for the autolysis reaction at the start of the NMR data collection. One spectrum was also collected after about 48 h in each case to see how much of the folded protein was still present after continuous autocleavage of the protein.

2.4. Intensity measurements

The peaks belonging to several backbone amide protons and tryptophan side chains (assigned earlier for the folded protein [25]) in the HSQC spectra of the proteins, recorded as a function of time, were integrated using *FELIX 97.0* software (Molecular Simulation Inc., San Diego, CA). The daughter peaks of W6 and W42 arising as result of autolysis were distinguished by the fact that in the real-time experiment, the daughter peak of W6 which lies closer to the primary cleavage site appears first and is consequently more intense than that of W42 all through the experiment. Next, the intensities of the backbone peaks belonging to the folded protein in each spectrum were divided by the total (folded + unfolded) peak intensity of W42 side chain within the same spectrum, to normalize intensity variations from one spectrum to another; the total intensity (folded + unfolded) of W42 side chain peaks was nearly constant (indicating conserved magnetization) throughout the kinetic experiment.

2.5. Data fitting

Our model requires that the intensities decay as sum of exponentials. However, when the decay is slow, the decay may appear like a linear decay. To quantitate the goodness of fits, we fitted the normalized HSQC peak intensities to straight lines, single exponentials as well as sums of exponentials (up to five) using *SIGMAPLOT* (Version 8.0, SPSS Inc.). Between five and seven points were used for the different peaks. It turned out that the single exponential fitting yielded the best χ^2 and in some cases where this was similar to those with the other fits, the single exponential fits had the lowest percentage fitting errors; the χ^2 values varied between 0.70 and 0.99 for the different peaks for the single exponential fits.

3. Results and discussion

3.1. Theory

3.1.1. Model for kinetics of autolysis

Autolysis occurs at specific sites in most proteases. The most vulnerable site of cleavage may be called the primary site, and the products of primary cleavage can undergo further cleavages at other sites which may be termed secondary sites, and so on. Our model of autolysis shown schematically in Fig. 1 can be enunciated as follows: (i) The unfolding process leading to the species susceptible to primary cleavage can occur along many parallel pathways starting from many different sites, in fact every residue. Thus in the ensemble there will be a population of molecules unfolding by each path. The protein solution consists of a mixture of fully folded native species and several partially folded species some of which are vulnerable to auto-cleavage. (ii) Unfolding is a slow process compared to the enzymatic cleavage reaction. Unfolding at the initiation site is considered to be the slowest along the whole pathway in every case, and therefore is the rate-limiting step. This will be referred to as 'initiator unfolding rate' and all the residues along the path will effectively unfold at this rate. (iii) The product following primary cleavage, is not unfolded at some other sites and hence can unfold further (secondary unfolding) and some of these may undergo secondary cleavages; this is a fast process. This can be represented as:

$$N_i \xrightarrow{k_{i1}^{\mathrm{in}}} P_{i1}^{j} \xrightarrow{k_{i2}^{j}} P_{i2} \tag{1}$$

where N_i represents the native protein population in the ensemble wherein unfolding starts at residue i leading to a species susceptible to primary cleavage, P_{i1}^{j} is the first generation product from primary cleavage of N_i in which the environment around residue j is unaltered, P_{i2} is the second generation product where j is also irreversibly unfolded, and k_{i1}^{in} and k_{i2}^{j} are the respective rate constants of the primary and secondary reactions; the superscript 'in' is used to indicate 'initiator unfolding rate'. (iv) P_{i1}^{j} will not fold back to the native protein as far as residue i and the residues that unfolded along with it, are concerned, and so also, P_{i2} will not fold back to P_{i1}^{j}. (v) The primary cleavage can be further described by a fundamental kinetic scheme as shown below,

$$N_i \underset{R_{if}}{\overset{R_{iu}}{\rightleftarrows}} U_{i1} \xrightarrow{R_{ic}} P_{i1}^{j} \tag{2}$$

where U_{i1} is the species *sufficiently unfolded* to undergo cleavage, R_{iu}, R_{if} are the unfolding and folding rate constants, respectively, and R_{ic} is the rate constant for enzymatic cleavage reaction. Similar equation can be written for unfolding of P_{i1}^{j} to U_{j2} and subsequent cleavage to fragmented products. The sequential reactions in Eq. (1) can be described by the rate equations,

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = -k_{i1}^{\mathrm{in}} N_i$$

$$\frac{\mathrm{d}P_{i1}^{j}}{\mathrm{d}t} = k_{i1}^{in} N_i - k_{i2}^{j} P_{i1}^{j}$$

$$\frac{\mathrm{d}P_{i2}}{\mathrm{d}t} = k_{i2}^{j} P_{i1}^{j} \tag{3}$$

The solutions of these equations can be readily obtained.

$$N_i(t) = N_i^0 e^{-k_{i1}^{\mathrm{in}} t}$$

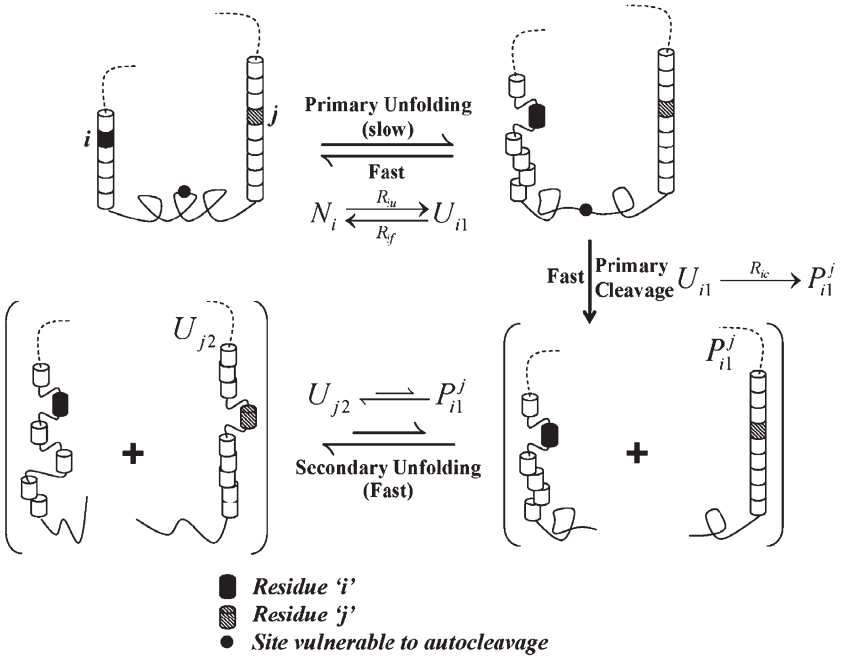

Fig. 1. Schematic showing the kinetic scheme used to derive the autolysis-based model.

$$P^j_{i1}(t) = \frac{k^{\text{in}}_{i1}}{k^j_{i2} - k^{\text{in}}_{i1}} \left(e^{-k^{\text{in}}_{i1} t} - e^{-k^j_{i2} t} \right) N^0_i$$

$$P_{i2}(t) = \left[\frac{k^{\text{in}}_{i1} k^j_{i2}}{k^j_{i2} - k^{\text{in}}_{i1}} \left(\frac{e^{-k^j_{i2} t}}{k^j_{i2}} - \frac{e^{-k^{\text{in}}_{i1} t}}{k^{\text{in}}_{i1}} \right) + 1 \right] N^0_i \quad (4)$$

where N^0_i refers to the initial population.

3.2. NMR spectral changes due to autolysis

The heteronuclear ^{1}H–^{15}N HSQC spectrum of a protein represents a finger print of the protein. Every correlation peak belongs to one non-proline residue and its position is sensitive to the chemical environment around the residue in the three dimensional structure of the protein. The intensity of the peak is a monitor of the protein concentration, although variations in the intensities do occur due to other factors such as different exchange rates of the amide protons with the solvent, different relaxation rates of the amide and nitrogen nuclei, etc. However, for a given protein under given conditions, changes in the intensity of any particular peak would reflect only the changes in the concentration of the species contributing to that peak. This fact can be utilized to monitor the autolytic reaction of the protein.

In the above scheme of autolytic reactions, as the environment around residue i unfolds, leading to a species susceptible to primary cleavage reaction, the protein undergoes cleavage. Since in the cleaved product, residue i and so also those which unfolded along the particular path cannot fold back to the native form, there will be a depletion of the 'folded peak' intensities of all those residues at the unfolding rate of i, which is the slowest being the initiator. A particular residue may lie along many different pathways initiated at different residues. Thus the net depletion of the population contributing to the intensity of a folded peak will be the sum of the contributions from the individual pathways.

$$N_i(t) = N^0_i e^{-k^{\text{in}}_{i1} t} + \sum_{l=1}^{n_i} N^0_l e^{-k^{\text{in}}_{l1} t} \quad (5)$$

where, k^{in}_{i1} is the initiator unfolding rate constant of residue i, n_i is the number of unfolding pathways initiated at other residues (indexed by l) that include i along their paths. At the same time, this primary cleavage may not cause unfolding at many other residues, and therefore let us consider residue j, for example, to be in folded condition. Under these conditions, the intensity of the peak due to residue j in the HSQC spectrum remains unaltered. The intensity of the peak belonging to j can change only due to unfolding in its own environment in the native protein (as described above), and/or due to unfolding of the product P^j_{i1} both of which will produce new peaks in the spectrum, at the cost of the folded peak intensity of j. Thus, for every j peak, the time dependence of the protein population contributing to the peak intensity as the autolytic reaction progresses can be easily derived as,

$$N_j(t) = N^0_j e^{-k^{\text{in}}_{j1} t} + \sum_{l=1}^{m_j} N^0_l e^{-k^{\text{in}}_{l1} t} + \sum_{i<j} \frac{k^{\text{in}}_{i1}}{k^j_{i2} - k^{\text{in}}_{i1}} \left(e^{-k^{\text{in}}_{i1} t} - e^{-k^j_{i2} t} \right) N^0_i \quad (6)$$

Here, m_j is the number of unfolding pathways initiated at other residues (indexed by l) that include j along their paths. The first two terms represent primary unfolding of environment around residue j and the third term arises from secondary unfolding of all those species, in which 'j residue unfolding' did not occur in the primary reaction. We assume at this stage, for the sake of simplicity, that all the individual pathways are equally probable and hence their initial populations N^0_i, N^0_j are all equal (say, N^0); note, however, that this assumption does not really affect our conclusions, except that it may modify some of the forthcoming equations.

While the above discussion focused on population changes in the ensemble, the individual peak intensities in the HSQC spectra will vary, even in a properly folded protein, because of differences in the exchange rates of the amides with the solvent and relaxation rates of the ^{15}N nuclei. Therefore, the time dependence of intensity of j peak, $I_j(t)$, can be written as,

$$I_j(t) = \lambda_j N^0 \left[e^{-k^{\text{in}}_{j1} t} + \sum_{l=1}^{m_j} e^{-k^{\text{in}}_{l1} t} + \sum_{i<j} \frac{k^{\text{in}}_{i1}}{k^j_{i2} - k^{\text{in}}_{i1}} \left(e^{-k^{\text{in}}_{i1} t} - e^{-k^j_{i2} t} \right) \right] \quad (7)$$

where λ_j is an adjustable parameter to take care of the residue wise intensity variations in the folded protein spectra. Explicit estimation of the contributions of the secondary cleavages to $I_j(t)$ would require monitoring the time profiles of the additional peaks arising due to unfolding of P^j_{i1}. This, however, requires assignment of those peaks, which is hard to obtain. Two special cases can be distinguished:

Case A. $k^j_{i2} \gg k^{\text{in}}_{i1}$ and $k^j_{i2} t \gg 1$.

This implies that the secondary unfolding after the primary cleavage is much faster than the primary unfolding process. This situation applies mostly to residues that are close to the primary cleavage site and their local unfolding can be nearly instantaneous after the primary cleavage resulting in chemical shift changes for their peaks in the HSQC spectra. For these residues, Eq. (7) reduces to,

$$I_j(t) = \lambda_j N^0 \left[e^{-k^{\text{in}}_{j1} t} + \sum_{l=1}^{m_j} e^{-k^{\text{in}}_{l1} t} \right] \quad (8)$$

Thus, in addition to the primary unfolding of the residue j, the primary unfolding of several other residues l leading to primary cleavage will contribute to the depletion rate of the peak intensity of j. If the primary unfolding rate of residue j is extremely small, then it is actually the primary events of the other residues which will dictate the depletion of the folded peak of j as a function of time.

Case B. $k_{i2}^{j} \ll k_{i1}^{\mathrm{in}}$.

This implies that the secondary unfolding after the primary cleavage is very slow compared to the primary unfolding itself and this situation applies mostly to residues far removed from the cleavage site, and perhaps engaged in stable secondary and/or tertiary structure formations away from the cleavage site. For these residues, Eq. (7) reduces to,

$$I_j(t) = \lambda_j N^0 \left[e^{-k_{j1}^{\mathrm{in}} t} + \sum_{l=1}^{m_j} e^{-k_{l1}^{\mathrm{in}} t} - \sum_{i<j} \left(e^{-k_{i1}^{\mathrm{in}} t} - e^{-k_{i2}^{j} t} \right) \right] \tag{9}$$

However, if data analysis is restricted to very short time points (initial rate approximation), such that,

$$k_{i1}^{\mathrm{in}} t \ll 1; \quad k_{i2}^{j} t \ll 1 \tag{10}$$

Then, it is easy to see that the contribution of the third term will be very small and hence can be neglected, i.e. the change in folded peak intensity with time will be largely dictated by the first two terms, which represent primary unfolding processes only. Then,

$$[I_j(t)]_{\mathrm{initial}} = \lambda_j N^0(t) \left[e^{-k_{j1}^{\mathrm{in}} t} + \sum_{l=1}^{m_j} e^{-k_{l1}^{\mathrm{in}} t} \right] \tag{11}$$

The right-hand sides of Eqs. (8) and (11) are identical, except that in the latter case the data collection has to be restricted to short time periods. A priori, there is no easy way of figuring out how many independent pathways contribute to the unfolding of residue j. As a lowest order approximation, the data can be fitted to multiple exponentials and the best fit can be interpreted to indicate the number of pathways with different rate constants and the individual rate constants can be extracted. It could happen that some of the independent pathways have the same rate constant, and then these will not be distinguished by the fitting procedure.

3.3. Unfolding free energies

From Eq. (2), under conditions $R_{jf} \gg R_{jc}$, which is generally satisfied (folding which is intramolecular occurs on ms time scale, while cleavage which is intermolecular occurs on sec to min time scale), it follows that,

$$k_{j1}^{\mathrm{in}} = K_j^{\mathrm{in}} R_{jc} \tag{12}$$

where, K_j^{in} is the equilibrium constant for the unfolding reaction, $N_j \underset{R_{jf}}{\overset{R_{ju}}{\rightleftharpoons}} U_{j1}$. This leads to the equation

$$-RT \ln k_{j1}^{\mathrm{in}} = -RT \ln K_j^{\mathrm{in}} - RT \ln R_{jc} \tag{13}$$

or

$$-RT \ln k_{j1}^{\mathrm{in}} = \Delta G_j^{\mathrm{in}} - RT \ln R_{jc} \tag{14}$$

Here the superscript 'in' helps to remind that there are many free energy contributors to the overall unfolding free energy of the environment around residue j. If the rate constant is degenerate for a few pathways, then the corresponding free energy change (calculated per mol) accounts for the total ensemble population representing those pathways. If the NMR data for any peak can be fitted to multiple exponentials, say, p, as discussed above, then there will be as many different rate constants contributing to the decay of the peak due to j. Now, for all of these, the enzymatic cleavage rate constant R_{jc}, being the intrinsic rate constant for a particular chemical reaction, can be assumed to be the same. Then, it follows that,

$$-RT \sum_{q=1}^{p} [\ln k_{q1}^{\mathrm{in}}]_j = \sum_{q=1}^{p} [\Delta G_q^{\mathrm{in}}]_j - pRT \ln R_{jc} \tag{15}$$

The total unfolding free energy change per mol for residue j will be,

$$\Delta G_j = \frac{1}{p} \sum_{q=1}^{p} [\Delta G_q^{\mathrm{in}}]_j \tag{16}$$

3.4. Mutational perturbations

For any two proteins, differing by a single point mutation, unfolding processes around a particular residue i, which is reasonably far from the mutation site, would be the same, provided, of course, the structures of the two proteins are similar. Further, the intrinsic catalytic rate (R_{ic}) can also be considered to be the same in the two cases except when the cleavage site is close to the mutation site along the sequence; note that we are referring here to closeness along the sequence and not on the structure, since the structure has to get unfolded around the cleavage site for the cleavage to occur. Consequently, comparison of the rates as per Eq. (15) will eliminate the second term on the right hand side and thus yields exclusively a comparison of the free energy changes associated with unfolding around that residue in the two proteins. Considering two mutants A and B,

$$\left(RT \sum_{q=1}^{p} [\ln k_{q1}^{\mathrm{in}}]_i \right)_B - \left(RT \sum_{q=1}^{p} [\ln k_{q1}^{\mathrm{in}}]_i \right)_A = \Delta G_{iA} - \Delta G_{iB} = \Delta\Delta G_i(A,B) \tag{17}$$

$\Delta\Delta G_i$ (A,B) represents the stability of environment around residue i in A in comparison to that in B. A positive value indicates higher relative stability and vice versa.

4. Application to HIV-1 protease-tethered dimer as an example

4.1. Autolysis-driven changes in HSQC spectra

We have applied the above ideas to HIV-1 protease-tethered dimer (HIV-1 PRTD), which functions similarly to the natural homodimeric enzyme, using representative mutants, (C95M), (C95M, C1095A) and (C95M, C1095F), which have been referred to as HIVTD-C, HIVTD-A, and HIVTD-F, respectively, in the Introduction. The studies can, of course, be extended to other mutants as well. The HSQC spectra of the three proteins

with freshly prepared samples were qualitatively very similar; overlays of the spectra of HIVTD-A and HIVTD-F on that of HIVTD-C are given in Fig. 2A and B, respectively. The nearly exact overlaps of the spectra indicate that the three mutant proteins have very similar structural features. Indeed, as mentioned before, the crystal structure of HIVTD-C, HIVTD-A and HIVTD-F are nearly identical [22–24]. The spectra had also many common peaks compared to the HSQC spectrum of the protein–pepstatin A complex. The complex was stable for several months and it was thus possible to obtain complete protein assignments there [25]. Then, a comparison of the spectra of the free proteins with those of the complex allowed us to obtain several assignments for the free proteins by simple inspection and the unambiguously assigned peaks have been labeled in the figure.

As must be expected, because of the autolysis reaction, the peaks belonging to the native protein lose intensity with time. The most interesting feature of these changes is that the different native peaks lose intensity at different rates in every protein and furthermore, for a given peak, the decay rates are different for the different mutant proteins; this is in accordance with the parallel unfolding pathways discussed above. For the measurement of these rates, the spectra have to be recorded rapidly, and this of course will be at the cost of some resolution. Fig. 3A shows one of the HSQC spectra of the kinetic data on the three proteins. We observe that the central region is very crowded and unusable for detailed quantitative analysis. In this region, neither can the assignments be obtained unambiguously nor can the peak intensities be measured reliably. Because of this, our analysis is restricted to the distinct peripheral peaks in the HSQC spectrum whose assignments are marked in the figure. Fig. 3B shows typical decay profiles for a few selected peaks.

4.2. MALDI monitoring

We have tried to characterize the autolytic behavior of HIV-1 protease with the aid of MALDI (Fig. 4). The molecular weight of HIV1-PRTD is ~21.9 kDa. Earlier it has been shown by mutational studies, activity assays and peptide substrate studies that the primary cleavage sites are most probably L5-W6 and the secondary cleavage sites are at L33-E34 and L63-I64 [19,26–30]. In the present case, since we have a tethered dimer, we have another set of primary and secondary cleavage sites, namely, L1005-W1006 for primary and L1033-E1034 and L1063-I1064 for secondary cleavage. Accordingly, we should have, in the MALDI spectra, species corresponding to MW of ~21.9 kDa for the native protein, ~10.2, ~11.7 and 21.4 kDa for primary cleavage products, and MW of less than 7 kDa for secondary cleavage products. Indeed, in our MALDI data, all such species are seen. In the initial 2 h, we have essentially the native protein and the primary cleavage products (peaks: N^+, P^+, N^{2+}, P^{2+}) while the secondary products (peaks: S^+, Sm^+) show markedly lesser presence (Fig. 4A); here, all the secondary cleavage products which can be many in number have been grouped in the symbols S^+, S^{m+} ($m=1, 2, 3$). However, as the reaction proceeds for more than 24 h, the secondary products show up more prominently (Fig. 4B).

4.3. Primary cleavage rates and mutational plasticity

As discussed in the previous sections, the primary autocleavage rates can be obtained from the 'initial rate approximation' applied to the decay profiles of the intensities of the folded peaks in the HSQC spectra of the proteins. We observed that the autolysis reaction proceeds extremely slowly and even after 48 h, a large number of peaks, especially those which we are considering here, had fair amount of intensity (Fig. 3C) indicating that there was still a substantial fraction of folded protein in the solution. The MALDI data also indicated that the extent of secondary cleavage was very small even after several hours after protein preparation. Thus it seems reasonable that, the HSQC spectra recorded for about 1–2 h at 11-min intervals, following a 'dead time' of approximately 4 h required

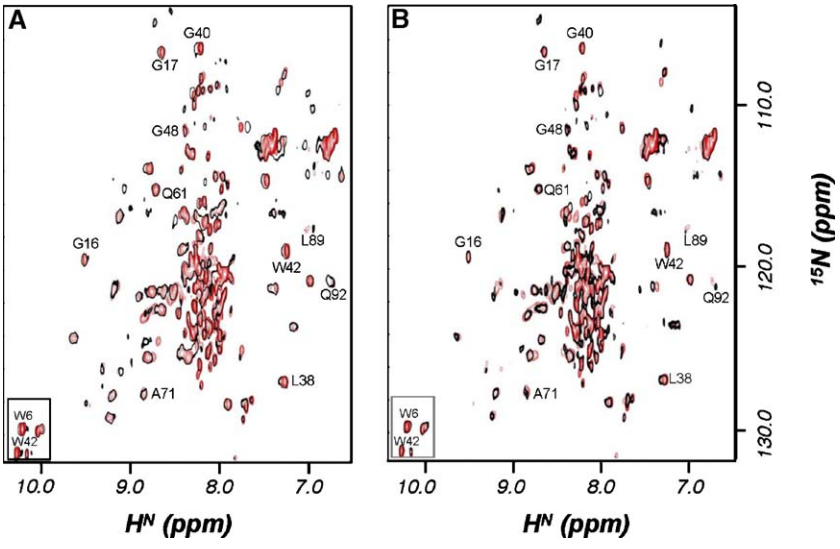

Fig. 2. ^1H–^{15}N HSQC spectrum overlay of HIVTD-C (red contours) over (A) HIVTD-A (black contours) and (B) HIVTD-F (black contours). The residues monitored have been labeled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

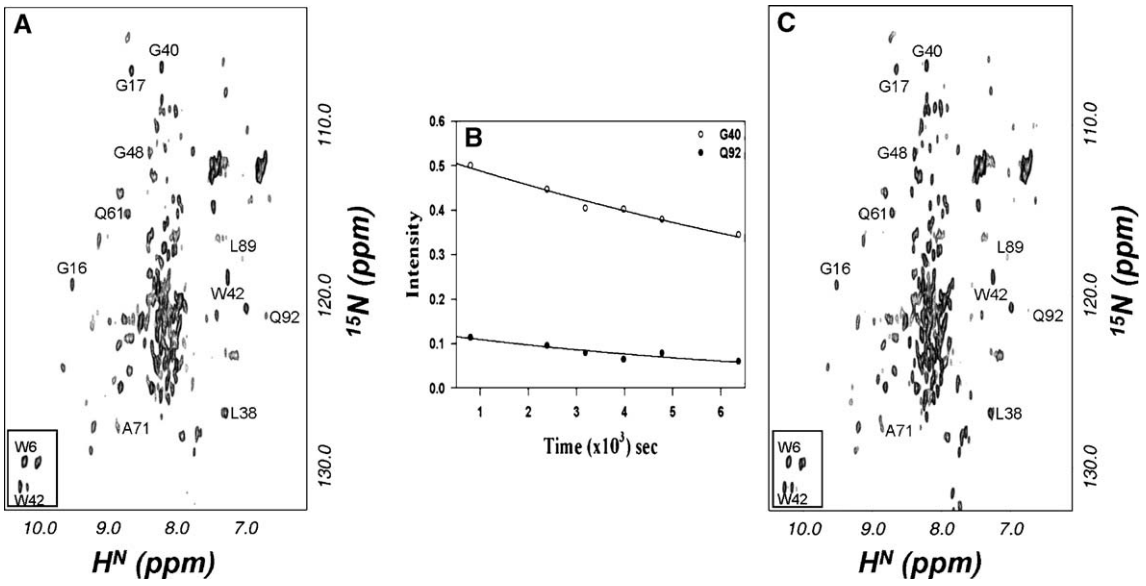

Fig. 3. Kinetic NMR data. (A) First time point (after a dead time of 4 h required for sample concentration at 4 °C) $^{1}H–^{15}N$ HSQC spectrum of HIVTD-F of the real time NMR experiment. (B) Illustrative decay profiles of two selected native peaks from HIVTD-F spectra; solid lines represent the single exponential fits. (C) $^{1}H–^{15}N$ HSQC spectrum of HIVTD-F after 48 h from the start of the kinetic experiment. The assignments of the peaks used in the present study are marked in A and C. The peaks enclosed in the box correspond to the tryptophan side chains; both intact (labeled) and cleaved products are seen. The daughter peak of W6 is stronger than that of W42 (see Materials and methods).

to concentrate the protein from low concentration to NMR concentration (~0.8 mM), which was carried out at 4 °C to minimize autolysis during this step, can be taken to satisfy the condition of initial rate approximation. We analyzed the NMR data for about 10 residues (these exclude residues near the mutation site) in the three mutant proteins to extract the primary cleavage rates from multi-exponential fits. We fitted the data to single exponentials and sums of exponentials (up to five) as well as to straight lines, to check for the validity of the model, the best fits and to extract the cleavage rates from the derived parameters. It turned out that in all the cases single exponential gave the best fit. The fitting statistics in the form of χ^2 is presented in Table 1. The percentage fitting errors for the derived rate constants were also lowest for the single exponentials. These errors along with the rate constants derived are displayed in Fig. 5A. This implies that within the experimental errors, for each residue considered, the rate constants for the overlapping pathways are similar. Interestingly, the rate constants for the different residues within the same protein are different (well beyond the fitting errors), and like wise, the rate constants for any given residue in the three mutant proteins are also significantly different. The first observation consolidates the previous observations that: (i) 'unfolding is the rate-limiting step as apposed to the enzymatic cleavage reaction', and (ii) 'there are parallel unfolding pathways with different rate constants operating in the ensemble'. The second observation indicates remote stability perturbations in the protein structure by the mutations.

Analysis of the k_{i1}^{in} rates as per Eq. (17) yielded the relative residue wise local unfolding free energy changes, $\Delta\Delta G_i$ (A, C), $\Delta\Delta G_i$ (F, C), and $\Delta\Delta G_i$ (F, A) where C, A and F inside the parentheses refer to the proteins HIVTD-C, HIVTD-A and HIVTD-F respectively. These are displayed in Fig. 5B. Some estimates of the errors in $\Delta\Delta G_i$ were obtained as below: First, the error in the cleavage rate was related to a maximum possible error (δG_j^{in}) in ΔG_j^{in} (although this is not explicitly measured) by differentiating Eq. (14) and taking the absolute value,

$$\delta G_j^{\mathrm{in}} = RT\frac{\mathrm{d}k_{j1}^{\mathrm{in}}}{k_{j1}^{\mathrm{in}}} \tag{18}$$

Since there was only one rate constant found for each residue, there is only one free energy term on the right hand side of Eq. (16)"' and hence we drop the superscript 'in' hereafter for the free energies. Then, the error ($\delta\delta G_i$) in $\Delta\Delta G_i$ was calculated as,

$$\delta\delta G_i(A,B) = \left[\{(\delta G_i^A)^2 + (\delta G_i^B)^2\}/2\right]^{1/2} \tag{19}$$

Though this simple procedure may not be the most rigorous way of estimating the errors, these do give a reasonable feel for the interpretability of the differences in the $\Delta\Delta G_i$ values. The derived errors are also shown in the Fig. 5B.

The data in Fig. 5B help to compare the relative stability patterns in HIVTD-C, HIVTD-A and HIVTD-F. For example, a negative number for a given residue in the column $\Delta\Delta G_i$ (A, C) indicates lesser stability in HIVTD-A compared to that in HIVTD-C and vice versa. Thus, environments around residues 38, 40, 42, 71 and 92 have lower stability in HIVTD-A compared to HIVTD-C and the magnitude of destabilization ranges from 1.0 kJ/mol to 3.4 kJ/mol. On the other hand, environments around residues 16, 17, 48, 61 and 89 are more stable in HIVTD-A compared to HIVTD-C, the stability

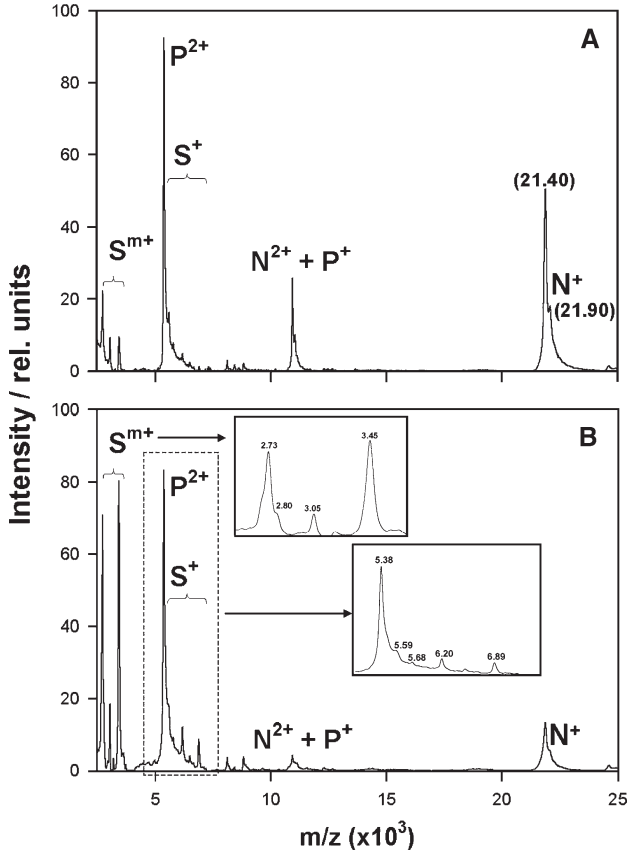

Fig. 4. MALDI–TOF analysis of HIVTD-C autolytic activity. (A) MALDI spectrum of HIVTD-C after 2 h of autolysis at 42 °C. Peaks due to singly $[N]^+$, $[P]^+$, $[S]^+$; doubly $[N]^{2+}$, $[P]^{2+}$; and other $[S]^{m+}$ charged species of the protease and its autolytic products are seen. The peaks $[N]^+$, $[P]^+$ and $[S]^{m+}$ (m=1,2,3) correspond to the intact protein, primary cleavage product and secondary cleavage products, respectively. (B) MALDI spectrum of HIVTD-C after 24 h of autolysis at 42 °C. Insets show blow ups with peaks labeled by the respective m/z values.

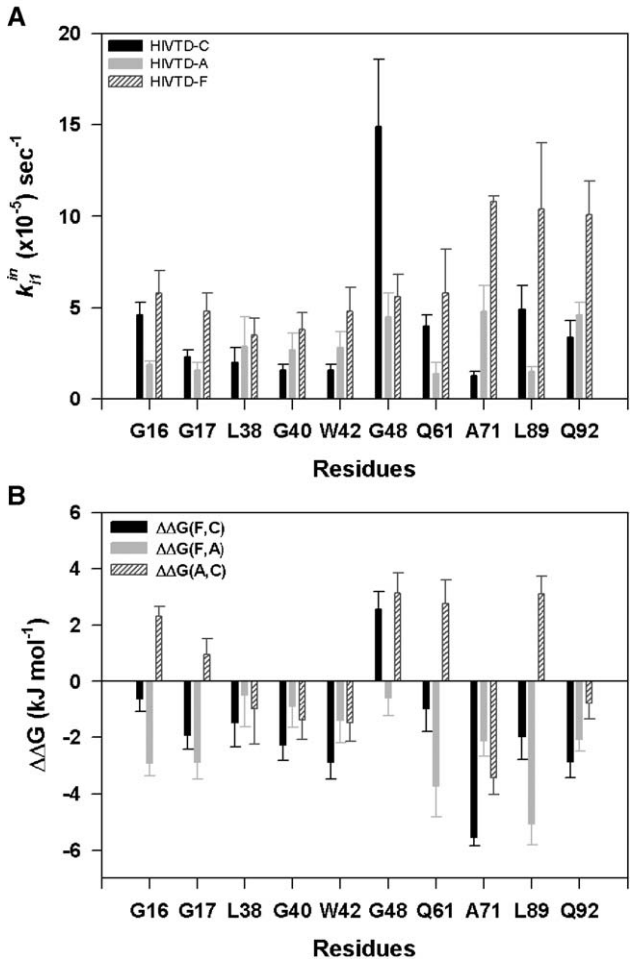

Fig. 5. (A) Exponential decay rates for different residues derived for the three mutant proteins of HIVTD at 42 °C. (B) $\Delta\Delta G$ values calculated for the different residues in the three mutant pairs (HIVTD-A, HIVTD-C), (HIVTD-A, HIVTD-F), (HIVTD-C, HIVTD-F).

variation being in the range, 0.9–3.1 kJ/mol. In contrast, in HIVTD-F, all of the 10 residues except residue 48 have reduced local stabilities compared to HIVTD-C, and again there is a wide range of free energy variation. The overall reduction in stability in HIVTD-F could be the result of introduction of large bulky group in the protein structure. Between HIVTD-F and HIVTD-A, all of the 10 common residues have reduced local stability in the former. All the pair-wise relative stabilities around individual residues in HIVTD-A, HIVTD-F and

Table 1
χ^2 fitting statistics of the normalized intensity decay profiles for the residues monitored in the three proteins, HIVTD-C, HIVTD-A, and HIVTD-F

	χ^2								
	HIVTD-C			HIVTD-A			HIVTD-F		
	Linear fit	Single exponential fit	Double exponential fit	Linear fit	Single exponential fit	Double exponential fit	Linear fit	Single exponential fit	Double exponential fit
G16	0.93	0.95	0.93	0.96	0.97	0.96	0.79	0.81	0.81
G17	0.97	0.99	0.98	0.70	0.71	0.70	0.85	0.87	0.86
L38	0.68	0.70	0.68	0.70	0.71	0.71	0.68	0.70	0.70
G40	0.82	0.83	0.83	0.74	0.75	0.73	0.96	0.98	0.98
W42	0.87	0.88	0.86	0.82	0.82	0.82	0.71	0.75	0.74
G48	0.89	0.92	0.89	0.67	0.70	0.69	0.84	0.84	0.84
Q61	0.95	0.96	0.95	0.94	0.97	0.96	0.96	0.97	0.97
A71	0.91	0.92	0.90	0.84	0.85	0.85	0.98	0.99	0.98
L89	0.74	0.75	0.72	0.84	0.85	0.84	0.64	0.70	0.70
Q92	0.75	0.76	0.75	0.88	0.89	0.89	0.82	0.86	0.85

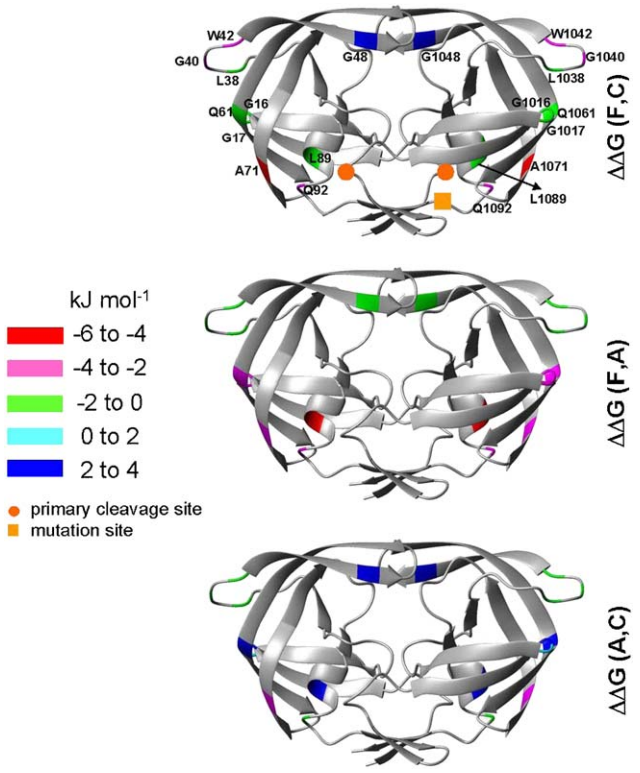

Fig. 6. Relative local stabilities (ΔΔG) of different residues in HIVTD-A, HIVTD-F and HIVTD-C, displayed in a colour coded manner on the native structure of HIVTD-C. The colour codes of increasing stability (kJ/mol) are the following: red (−6 to −4), magenta (−4 to −2), green (−2 to 0), cyan (0 to 2), blue (2 to 4). The orange ball indicates the cleavage site and the orange square identifies the mutation site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

HIVTD-C are displayed in a color-coded manner on the structure of the native protein in Fig. 6. The residues in both the halves of the dimer have been marked keeping in mind the chemical shift equivalence of the residues. Color increments are used for every 2.0 kJ/mol, which is much larger than the average error (0.8 kJ/mol) and thus represent significant differences in the stabilities. From the different colour patterns it is evident that there are substantial differences in the local stabilities of the mutant proteins. In other words, the effects of single point mutation at position 95 are significantly felt at very remote places and the magnitudes of the effects are also very varied, depending upon the nature of the mutation. These would have significant implications for mutation-induced alterations in substrate-binding efficacies, as well as in drug-binding efficacies of the protease.

5. Conclusions

The present study attempts to develop an NMR-based theoretical model to derive useful insights into unfolding characteristics in proteases from an analysis of the kinetics of the autolytic cleavage reaction. It relies on the interpretation of the decay rates of the folded protein peaks in the HSQC spectra of proteases as a function of time. A formalism has been presented to extract the differences in the free energy changes for unfolding of the environment around any given residue far removed from the mutation site, between two proteins differing by a single point mutation. As an example, the model has been applied to HIV-1 protease-tethered dimer, using three different mutant proteins carrying mutations at a particular site in the dimerization domain. The results show interesting variations in the relative stabilities. While these results provide the 'proof of principle' for the new ideas in the model, they are also of much significance from the point of view of protein unfolding concepts, in general, as they demonstrate existence of multiple unfolding pathways under native conditions. Further, with regard to the protein under investigation, these results have significant implications in the context of drug resistant mutations, so commonly observed. It is evident that non-active site mutations can cause stability changes around the hinge region of the dimer, the binding sites of the drugs, etc., and thus can influence their binding efficacies.

Acknowledgements

We thank the National Facility for High Field NMR at the Tata Institute of Fundamental Research for all the NMR and computational 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. AC is a recipient of TIFR Alumni Association Scholarship for Career Development supported by TIFR Endowment Fund. We thank Dr. Neel S. Bhavesh for useful discussion in the early phases of this work.

References

[1] S.D. Hoeltzli, C. Frieden, Real-time refolding studies of 6-19F-tryptophan labeled *Escherichia coli* dihydrofolate reductase using stopped-flow NMR spectroscopy, Biochemistry 35 (1996) 16843–16851.
[2] S.D. Hoeltzli, C. Frieden, Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-19F-tryptophan-labeled *Escherichia coli* dihydrofolate reductase, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 9318–9322.
[3] M.P. Lillo, B.K. Szpikowska, M.T. Mas, J.D. Sutin, J.M. Beechem, Real-time measurement of multiple intramolecular distances during protein folding reactions: a multisite stopped-flow fluorescence energy-transfer study of yeast phosphoglycerate kinase, Biochemistry 36 (1997) 11273–11281.
[4] S. Cavagnero, Z.H. Zhou, M.W. Adams, S.I. Chan, Unfolding mechanism of rubredoxin from *Pyrococcus furiosus*, Biochemistry 37 (1998) 3377–3385.
[5] J. Woenckhaus, R. Kohling, P. Thiyagarajan, K.C. Littrell, S. Seifert, C.A. Royer, R. Winter, Pressure-jump small-angle X-ray scattering detected kinetics of staphylococcal nuclease folding, Biophys. J. 80 (2001) 1518–1523.
[6] B.M. Kolakowski, L. Konermann, From small-molecule reactions to protein folding: studying biochemical kinetics by stopped-flow electrospray mass spectrometry, Anal. Biochem. 292 (2001) 107–114.
[7] S. Benjwal, S. Verma, K.H. Rohm, O. Gursky, Monitoring protein aggregation during thermal unfolding in circular dichroism experiments, Protein Sci. 15 (2006) 635–639.
[8] M. Roy, P.A. Jennings, Real-time NMR kinetic studies provide global and residue-specific information on the non-cooperative unfolding of the beta-trefoil protein, interleukin-1beta, J. Mol. Biol. 328 (2003) 693–703.

[9] S.W. Englander, Protein folding intermediates and pathways studied by hydrogen exchange, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 213–238.
[10] G. Vriend, H.J. Berendsen, B.B. van den, G. Venema, V.G. Eijsink, Early steps in the unfolding of thermolysin-like proteases, J. Biol. Chem. 273 (1998) 35074–35077.
[11] J.R. Rose, R. Salto, C.S. Craik, Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions, J. Biol. Chem. 268 (1993) 11939–11945.
[12] A.M. Mildner, D.J. Rothrock, J.W. Leone, C.A. Bannow, J.M. Lull, I.M. Reardon, J.L. Sarcich, W.J. Howe, C.S. Tomich, C.W. Smith, The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties, Biochemistry 33 (1994) 9405–9413.
[13] H. Li, V.F. Thompson, D.E. Goll, Effects of autolysis on properties of mu- and *m*-calpain, Biochim. Biophys. Acta 1691 (2004) 91–103.
[14] R. Theberge, L.H. Connors, M. Skinner, C.E. Costello, Detection of transthyretin variants using immunoprecipitation and matrix-assisted laser desorption/ionization bioreactive probes: a clinical application of mass spectrometry, J. Am. Soc. Mass Spectrom. 11 (2000) 172–175.
[15] C.E. Parker, D.I. Papac, K.B. Tomer, Monitoring cleavage of fusion proteins by matrix-assisted laser desorption ionization/mass spectrometry: recombinant HIV-1IIIB p26, Anal. Biochem. 239 (1996) 25–34.
[16] B. Thiede, W. Hohenwarter, A. Krah, J. Mattow, M. Schmid, F. Schmidt, P. R. Jungblut, Peptide mass fingerprinting, Methods 35 (2005) 237–247.
[17] H. Kumura, S. Murata, T. Hoshino, K. Mikawa, K. Shimazaki, Autolysis of the proteinase from *Pseudomonas fluorescens*, J. Dairy Sci. 82 (1999) 2078–2083.
[18] P. Tompa, P. Buzder-Lantos, A. Tantos, A. Farkas, A. Szilagyi, Z. Banoczi, F. Hudecz, P. Friedrich, On the sequential determinants of calpain cleavage, J. Biol. Chem. 279 (2004) 20775–20785.
[19] K. Partin, H.G. Krausslich, L. Ehrlich, E. Wimmer, C. Carter, Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase, J. Virol. 64 (1990) 3938–3947.
[20] P. McCelland, J.A. Lash, D.R. Hathaway, Identification of major autolytic cleavage sites in the regulatory subunit of vascular calpain: II. A comparison of partial amino-terminal sequences to deduced sequence from complementary DNA, J. Biol. Chem. 264 (1989) 17428–17431.
[21] S.C. Panchal, N.S. Bhavesh, R.V. Hosur, Real time NMR monitoring of local unfolding of HIV-1 protease tethered dimer driven by autolysis, FEBS Lett. 497 (2001) 59–64.
[22] M. Kumar, K.K. Kannan, M.V. Hosur, N.S. Bhavesh, A. Chatterjee, R. Mittal, R.V. Hosur, Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations, Biochem. Biophys. Res. Commun. 294 (2002) 395–401.
[23] B. Pillai, K.K. Kannan, M.V. Hosur, 1.9A X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR, Proteins 43 (2001) 57–64.
[24] V. Prashar, M.V. Hosur, 1.8A X-ray structure of C95M/C1095F double mutant of tethered HIV-1 protease dimer complexed with acetyl pepstatin, Biochem. Biophys. Res. Commun. 323 (2004) 1229–1235.
[25] S.C. Panchal, B. Pillai, M.V. Hosur, R.V. Hosur, HIV-1 protease tethered heterodimer–pepstatin–A complex: NMR characterization, Curr. Sci. (2000) 1684–1695.
[26] P.L. Darke, R.F. Nutt, S.F. Brady, V.M. Garsky, T.M. Ciccarone, C.T. Leu, P.K. Lumma, R.M. Freidinger, D.F. Veber, I.S. Sigal, HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins, Biochem. Biophys. Res. Commun. 156 (1988) 297–303.
[27] M. Kotler, R.A. Katz, W. Danho, J. Leis, A.M. Skalka, Synthetic peptides as substrates and inhibitors of a retroviral protease, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 4185–4189.
[28] H.G. Krausslich, R.H. Ingraham, M.T. Skoog, E. Wimmer, P.V. Pallai, C. A. Carter, Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides, Proc. Natl. Acad. Sci. U. S. A. 86 (1989) 807–811.
[29] D.D. Loeb, C.A. Hutchison III, M.H. Edgell, W.G. Farmerie, R. Swanstrom, Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases, J. Virol. 63 (1989) 111–121.
[30] A.G. Tomasselli, M.K. Olsen, J.O. Hui, D.J. Staples, T.K. Sawyer, R.L. Heinrikson, C.S. Tomich, Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease, Biochemistry 29 (1990) 264–269.