LETTERS

Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease

Chun Tang[1]†, John M. Louis[1], Annie Aniana[1], Jeong-Yong Suh[1] & G. Marius Clore[1]

HIV-1 protease processes the Gag and Gag-Pol polyproteins into mature structural and functional proteins, including itself, and is therefore indispensable for viral maturation[1,2]. The mature protease is active only as a dimer[3–5] with each subunit contributing catalytic residues[6]. The full-length transframe region protease precursor appears to be monomeric yet undergoes maturation via intramolecular cleavage of a putative precursor dimer[5,7–11], concomitant with the appearance of mature-like catalytic activity[7,9]. How such intramolecular cleavage can occur when the amino and carboxy termini of the mature protease are part of an intersubunit β-sheet located distal from the active site is unclear. Here we visualize the early events in N-terminal autoprocessing using an inactive mini-precursor with a four-residue N-terminal extension that mimics the transframe region protease precursor[5,12]. Using paramagnetic relaxation enhancement, a technique that is exquisitely sensitive to the presence of minor species[13–16], we show that the mini-precursor forms highly transient, lowly populated (3–5%) dimeric encounter complexes that involve the mature dimer interface but occupy a wide range of subunit orientations relative to the mature dimer. Furthermore, the occupancy of the mature dimer configuration constitutes a very small fraction of the self-associated species (accounting for the very low enzymatic activity of the protease precursor), and the N-terminal extension makes transient intra- and intersubunit contacts with the substrate binding site and is therefore available for autocleavage when the correct dimer orientation is sampled within the encounter complex ensemble.

The regulation of HIV-1 protease autoprocessing is modulated by the N-terminal flanking transframe region (TFR) sequence (Fig. 1a)[2]. The catalytic activity of the monomeric protease precursor is approximately three orders of magnitude less than that of the mature protease dimer (which has a monomer–dimer equilibrium dissociation constant $K_d < 10$ nM)[2,5]. The appearance of mature-like catalytic activity and stable dimer formation is directly correlated with a single rate-limiting step comprising intramolecular (first order) cleavage of a putative transient dimeric precursor species at the p6[Pol]–protease (PR) junction[7,9,10]. Mutations within the latter that prevent cleavage lead to the production of an N-terminally extended 17-kDa protease precursor species, and cause a severe defect in Gag polyprotein processing and the complete loss of viral infectivity *in vivo*[17,18]. Subsequent cleavage at the C terminus of protease at the PR–reverse transcriptase (RT) junction (Fig. 1a) occurs via an intermolecular (second order) reaction catalysed by a fully active protease dimer[19]. Mutations within the PR–RT junction that block C-terminal cleavage do not significantly affect either enzymatic activity and dimerization of the protease *in vitro*[19,20] or processing of HIV-1 precursor proteins, virus maturation, viability and morphology *in vivo*[20], indicating that the presence of the C-terminal reverse transcriptase sequence has negligible influence on the protease precursor[19,20]. Thus, only autoprocessing at the N

terminus of protease at the p6[Pol]–PR junction is an absolute prerequisite for stable protease dimer formation, the appearance of mature catalytic activity and complete processing of viral precursors. Before cleavage at the p6[Pol]–PR junction, intermediate precursor forms may be liberated by intramolecular cleavage at competing sites (for example, p2–NC and TFP–p6[Pol]; see Fig. 1a) that become available for productive binding and hydrolysis[11], but these precursors will show the same low catalytic activity as that of the p6[Pol]–PR precursor[9,10].

As little as a four-residue extension at the N terminus of protease, corresponding to the C-terminal residues of p6[Pol], in conjunction with a D25N mutation result in an effectively monomeric species[5,12]. Disruption of the native protease dimer by N-terminal extension is due to removal of the protons on the secondary amine of the N-terminal proline residue, disrupting the interstrand hydrogen bond between the amine of the N-terminal proline of one subunit and the C-terminal carbonyl oxygen of the second subunit[6]. C-terminal extension, however, does not have an impact on this interstrand hydrogen bond because the secondary amine of Pro 1 is preserved. Therefore, we made use of the mini-precursor, bearing only the N-terminal cleavage site, to visualize the early transient events involved in autoprocessing of the protease at the p6[Pol]–PR junction that is required for the formation of a fully active, stable protease dimer.
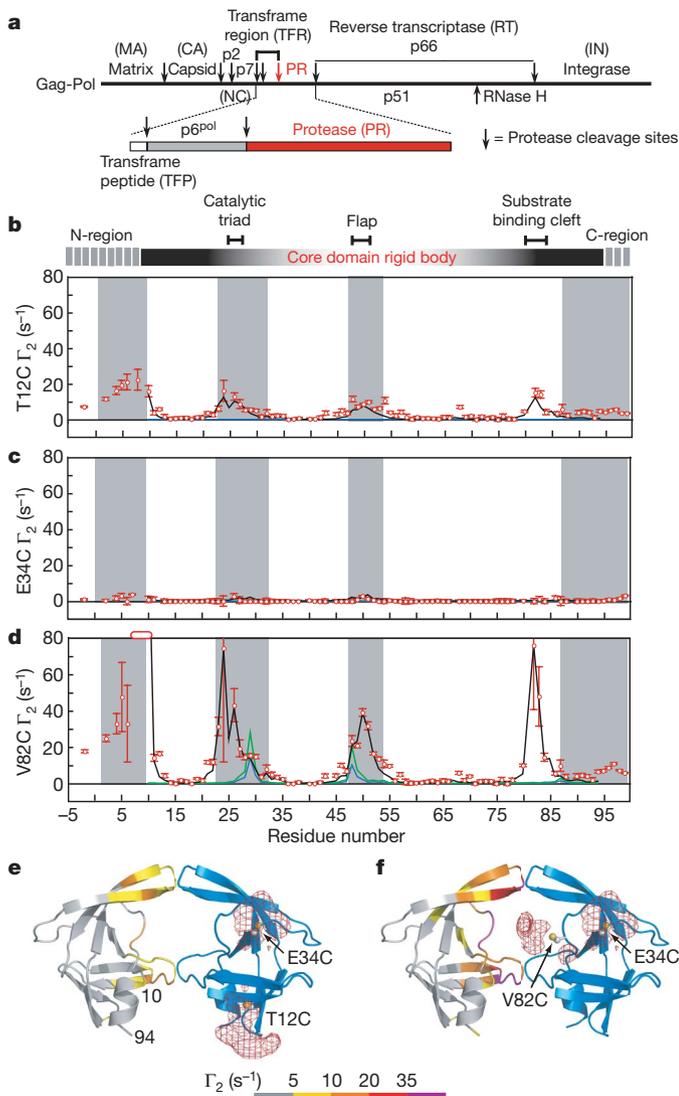
The optimized mini-precursor protease construct [SFNF]PR(D25N) comprises a four-residue N-terminal extension (Ser-Phe-Asn-Phe) derived from the TFR (Fig. 1a), a D25N mutation to abolish all residual catalytic activity, and C67A and C95A mutations to remove surface cysteines (Supplementary Fig. 1a)[9,10,12]. The corresponding active [SFNF]PR(D25) mini-precursor construct undergoes autoprocessing during expression to release the mature protease (see Methods). NMR analysis of [SFNF]PR(D25N) shows that it is monomeric (with an upper limit of ~10% dimer from translational diffusion measurements); the secondary and tertiary structures of the mature protease are preserved with the exception of the N- and C-terminal strands which form an intersubunit four-stranded anti-parallel β-sheet in the mature dimer; and residues −4 to 9 and 95–99 are disordered and highly mobile (see Methods and Supplementary Fig. 1b–e).

Because enzymatically active protease is dimeric, and the rate-limiting step in autoprocessing is unimolecular[7,9], transient self-association of the precursor must occur to initiate autoprocessing. To visualize this phenomenon we measured intermolecular paramagnetic relaxation enhancements (PREs) by introducing a spin label via conjugation to three engineered surface-exposed cysteine residues: T12C, E34C and V82C (one at a time). These sites are frequently mutated in viable HIV-1 variants[2]. T12C and V82C are located at the periphery of the substrate-binding cleft in the mature dimer, whereas E34C is relatively far removed from the dimer interface (Fig. 1e, f). In a rapidly exchanging system, the PRE ^1H_N-Γ_2 rates[21] are population-weighted averages of the PRE rates of the species present[13,14]. Because the PRE rate

---

[1]Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.
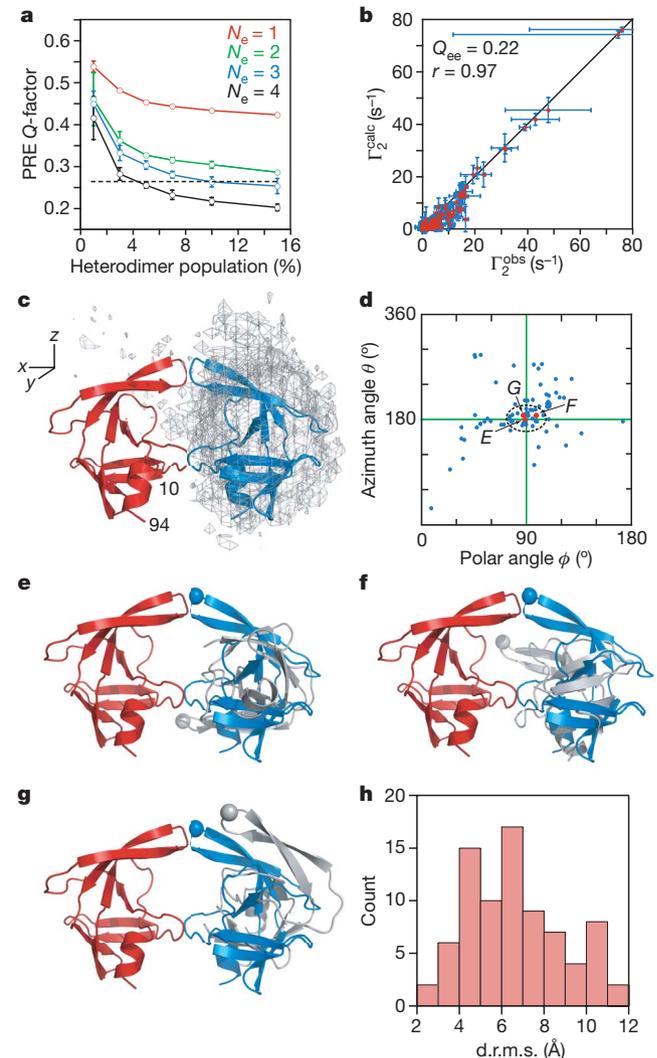†Present address: Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, USA.

for a paramagnetic centre-proton pair is proportional to the $<r^{-6}>$ average of the distance between them, and the PRE effect is large owing to the high magnetic moment of an unpaired electron, the PRE in the fast exchange regime is very sensitive to the presence of lowly populated (<5%), highly transient species in solution providing there are paramagnetic centre-proton distances in the minor species that are shorter than in the predominant species[13–15].

PREs were measured on a 1:1 mixture of 0.2 mM U-[^{2}H/^{13}C/^{15}N]-labelled ^{SFNF}PR(D25N) and spin-labelled ^{SFNF}PR(D25N) at natural isotopic abundance. Because ^1H_N-Γ_2 rates are measured using ^1H–^{15}N correlation-based experiments[21], the observed ^1H_N-Γ_2 rates arise solely from intermolecular interactions between the spin-labelled

protein and the isotopically labelled protein (Fig. 1b–d). For the E34C spin label, no ^1H_N-Γ_2 rates greater than 5 s^{-1} are observed (Fig. 1c); this sample therefore provides a negative control, excluding the existence of solvent PRE effects arising from diffusion and random elastic collisions, or from direct intermolecular interactions between the spin

Figure 1 | Intermolecular PRE profiles. a, Organization of the Gag-Pol polyprotein[1,2]. b–d, Intermolecular PREs observed on U-[^{2}H/^{13}C/^{15}N]-labelled ^{SFNF}PR(D25N) originating from a spin label conjugated to T12C (b), E34C (c) and V82C (d) of ^{SFNF}PR(D25N) at natural isotopic abundance. Residues broadened beyond detection are denoted by open bars. Error bars represent 1 s.d. Γ_2 rates back-calculated from the structure of the mature dimer (for the core residues 10–94) at populations of 1% and 2% are shown as blue and green lines, respectively. Average Γ_2 rates derived from the top 20 structures of the $N_e = 4$ simulated annealing calculations at a population of 5% heterodimer are shown as black lines. Grey shaded areas delineate residues that are buried at the dimer interface in the mature protease. e, f, Observed intermolecular PREs originating from the spin label attached to T12C (e) and V82C (f) colour-coded on a ribbon diagram of the mature dimer[24] (spin label attached to the blue subunit). Atomic probability density maps[25] (plotted at a threshold of 10% of maximum) showing the distribution of the spin-label oxygen radicals are shown as red meshes.

Figure 2 | Ensemble simulated annealing and the protease mini-precursor encounter complex ensemble. a, PRE Q-factor as a function of ensemble size and population of heterodimer. Dashed line denotes the expected Q-factor when agreement between observed and calculated Γ_2 rates is comparable to the experimental error in the measurements. b, Correlation between observed and calculated Γ_2 rates for $N_e = 4$ and a heterodimer population of 5%. Q_{ee} is the ensemble of ensembles average PRE Q-factor for the 20 calculated $N_e = 4$ ensembles and r the correlation coefficient. Error bars in a and b represent 1 s.d. c, Atomic probability density map[25] (grey mesh, plotted at a threshold of 20% of maximum) showing the distribution of the spin-labelled subunit relative to the isotopically labelled subunit (red ribbon) in the ^{SFNF}PR(D25N) encounter complexes. The location of the second subunit in the mature dimer is shown as a blue ribbon. d, Orientations in spherical coordinates of the vector joining the centre of masses of the two interacting molecules in the encounter complexes relative to the coordinate system shown in c with the z axis corresponding to the C_2 symmetry axis of the mature dimer. The ϕ,θ angles for the mature dimer are located at the crosshair. e–g, Representative encounter complexes (labelled and denoted by red dots in d) corresponding to the structures with the closest spherical angles (e), the smallest d.r.m.s. (f) and the smallest atomic r.m.s. displacement (g) relative to the mature dimer. The Cα atom of Gly 51 at the tip of the flap is shown as a sphere to guide the eye. The isotopically labelled and spin-labelled subunits are shown in red and grey, respectively; the blue subunit corresponds to the orientation relative to the red subunit seen in the mature dimer. h, Histogram of the d.r.m.s. metric for the $N_e = 4$ structures (total of 20 × 4 = 80 conformers) at a population of 5% heterodimer.

label and the U-[^{2}H/^{13}C/^{15}N]-labelled protein. The PRE profiles for the T12C (Fig. 1b) and V82C (Fig. 1d) spin labels are similar but the magnitude for the latter is 4- to 8- fold greater than for the former. Within the ordered core of the precursor (residues 10–94), large intermolecular PREs are observed for residues 21–30, 46–55 and 80–85 located at or close to the dimer interface. Residues 21–30 encompass the catalytic triad, residues 46–66 correspond to the flap region which gates the active site, and residues 80–81 and 83–84 are located in the substrate binding cleft (Fig. 1e, f). In addition, the N-terminal region experiences sizeable PREs from the T12C (Fig. 1b) and V82C (Fig. 1d) spin labels. These data demonstrate that transient self-association of the precursor involves residues located at the dimer interface in the mature dimer. A similar intermolecular PRE profile is observed from V82C spin-labelled, full-length TFR–PR(D25N) precursor to U-[^{2}H/^{13}C/^{15}N]–^{SFNF}PR(D25N), indicating that the transient dimerization interface is preserved on further N-terminal extension of the protease precursor (Supplementary Fig. 2a).

Back-calculation of the PREs from the structure of the mature dimer shows that almost zero PRE values are expected for the T12C and E34C spin labels at a population of 1–2% mature heterodimer (Fig. 1b, c). For the V82C label, small PRE values at a population of 1–2% mature heterodimer are predicted for residues 27–30 and 48–50 (Fig. 1d, blue line). The mature dimer does not predict the large observed PRE values observed for residues 20–26, 30–35 and 80–83. Furthermore, in the mature dimer residues 80–83 of one subunit are located on the opposite side of the dimer interface from residues 80–83 of the other subunit, and thus the large intermolecular PREs observed from the V82C spin label to residues 80–83 would require a ~180° rotation of one subunit relative to its position in the mature dimer. Thus, the upper limit of the total population of mature dimer (heterodimer and homodimer) cannot exceed 2–4%.

Transient interactions between ^{SFNF}PR(D25N) precursor monomers were visualized semi-quantitatively using rigid-body simulated annealing calculations[14,16,22] to optimize the agreement between observed and calculated Γ_2 rates arising from the T12C, E34C and V82C spin labels simultaneously (see Methods). The flexible N- and C-terminal regions (residues −4 to 9 and 95–99, respectively) were excluded from the calculations. A single conformer representation ($N_e = 1$) for the transient dimer does not account for the PRE data and even at a heterodimer population of 15% the PRE Q-factor[23] (see Methods for definition) has a value of greater than 0.4 (Fig. 2a). Thus, the dimeric ^{SFNF}PR(D25N) precursor is an ensemble of multiple encounter complexes. For $N_e \geq 2$, the average PRE Q-factor decreases rapidly as the heterodimer population is increased above 1%, levelling off at a population of ~5% (Fig. 2a). The best results are obtained with $N_e = 4$, and larger ensemble sizes are unjustified and would result in over-fitting the data. For $N_e = 4$, the PRE Q-factors at a heterodimer population of 3–5% are close to the expected PRE Q-factor based on experimental error (Fig. 2a), consistent with translational diffusion data (Methods and Supplementary Fig. 1d). Given a total protein concentration of 0.4 mM, the apparent K_d for self-association is therefore 3–6 mM. A comparison of the calculated and observed PRE profiles and a correlation plot of observed versus calculated Γ_2 rates for $N_e = 4$ at a heterodimer population of 5% are shown in Fig. 1b–d and Fig. 2b, respectively.

The distribution of the spin-labelled monomer relative to the isotopically labelled monomer in the computed ensemble of ^{SFNF}PR(D25N) encounter complexes is shown in Fig. 2c. The predominant interactions between the two monomers involve the same residues that comprise the dimer interface in the mature dimer, and one subunit of the mature dimer is embedded within the ensemble distribution of the spin-labelled subunit. The orientation of the subunits in the encounter complex ensembles can be described by spherical angles describing the orientation of the vector joining the centre of masses of the two subunits to the coordinate axis frame. Many members within the calculated ensemble are clustered around the values corresponding to the mature dimer (Fig. 2d). This is reflected in the distribution of the distance root mean square (d.r.m.s.; see Methods) deviation metric where over one-half of the

ensemble members have d.r.m.s. values less than 6 Å (Fig. 2h). However, the structures with spherical angles close to the mature dimer (indicated by arrows in Fig. 2d) and low d.r.m.s. values have a widespread range of relative self-rotations, as illustrated by three examples comprising the ensemble members with the closest spherical angles to the mature dimer (Fig. 2e), the smallest d.r.m.s. (Fig. 2f) and the smallest Cα atomic r.m.s. displacement (Fig. 2g). The difference from the mature dimer in rotation angle about the axis joining the centre of masses of the two subunits ranges from 13° (Fig. 2g) to 135° (Fig. 2e), with an intermediate rotation angle of 70° for the structure in Fig. 2f (see Supplementary Fig. 3 for definitions). One can therefore conclude that the actual occupancy of a structure within the encounter complex ensemble corresponding to the mature dimer is very small.

To probe the conformational space sample by the disordered N-terminal flanking sequence of the ^{SFNF}PR(D25N) precursor we introduced a spin label on a Cys residue inserted immediately after the N-terminal serine ($^{S(C)FNF}$PR(D25N)). PRE measurements were carried out on a 1:1 mixture of 0.2 mM U-[^{2}H/^{13}C/^{15}N]–^{SFNF}PR$_{D25N}$ precursor and 0.2 mM spin-labelled, natural isotopic abundance $^{S(C)FNF}$PR(D25N) to detect intermolecular PREs, and on a sample of 0.2 mM spin-labelled, U-[^{2}H/^{13}C/^{15}N]-labelled $^{S(C)FNF}$PR(D25N) to observe both inter- and intramolecular PRE effects. Although the overall PRE profiles for the two samples are similar (although differences in detail are apparent), the magnitude of the PREs for the second sample is much larger than for the first, reflecting the contribution from intramolecular PREs (Fig. 3a). The N-terminal residues −4 to 9, and residues comprising the active site, flap and substrate-binding cleft, display large inter- and intramolecular PREs (Fig. 3a, c). The intermolecular PREs involving residues 82–84 are fully consistent with the large intermolecular PREs observed on the N-terminal residues from spin-labelled V82C (Fig. 1d). These data indicate that the N-terminal tail can insert itself into the active site and make transient contact with both subunits in the encounter complex ensemble. The spin label is located four residues proximal to the scissile peptide bond, and the observation that large PREs are observed for both sides of the active site (see Fig. 3c) suggests that the tail shuttles back and forth within the substrate binding cleft formed by the two subunits in the context of a dimer. Such translational movement is a functional requirement, as the protease precursor cuts the N-terminal transframe region in two major locations before cleaving its C terminus (Fig. 1a)[9,10]. This is confirmed by the observation of a very similar intermolecular PRE profile from full-length TFR–PR(D25N) spin-labelled at position −44, four residues downstream from the TFP–p6^{pol} cleavage site at residues −48/−49, to U-[^{2}H/^{13}C/^{15}N]–^{SFNF}PR(D25N) (Supplementary Fig. 2b).

The C-terminal region of the ^{SFNF}PR(D25N) precursor was spin-labelled at N98C. The resulting intermolecular PREs are much smaller than those with the spin label at the N terminus, but the PRE profiles are similar (Fig. 3b, d). Thus, the C-terminal flexible region can also make intermolecular contacts with the active site and substrate-binding cleft in the context of the precursor encounter complex ensemble. Because the N- and C termini are highly mobile, intermolecular PREs between the N- and C termini will be significantly attenuated. Nevertheless, intermolecular PREs are observed on residues 95–97 from the spin label at the N terminus (Fig. 3a), and on residues 5–8 (Fig. 3b) from the spin label at the C terminus (Fig. 3b). Small intermolecular PREs are also observed from the N98C spin label to the C-terminal region (residues 95–99). These observations might suggest the existence of transient, loose interactions between the N- and C termini that may partially approximate a portion of the intersubunit β-sheet in the mature dimer.

The PRE data presented here demonstrate that although the HIV-1 protease precursor is predominantly monomeric, transient encounter complex dimers are formed using the same interface as that of the mature dimer but with a wide range of relative subunit orientations. Only a very small fraction of the encounter complexes adopt the same subunit orientation as in the mature protease, accounting for the very low enzymatic activity of the precursor. This small subset, which may be partially stabilized by transient, loose interactions involving the N- and
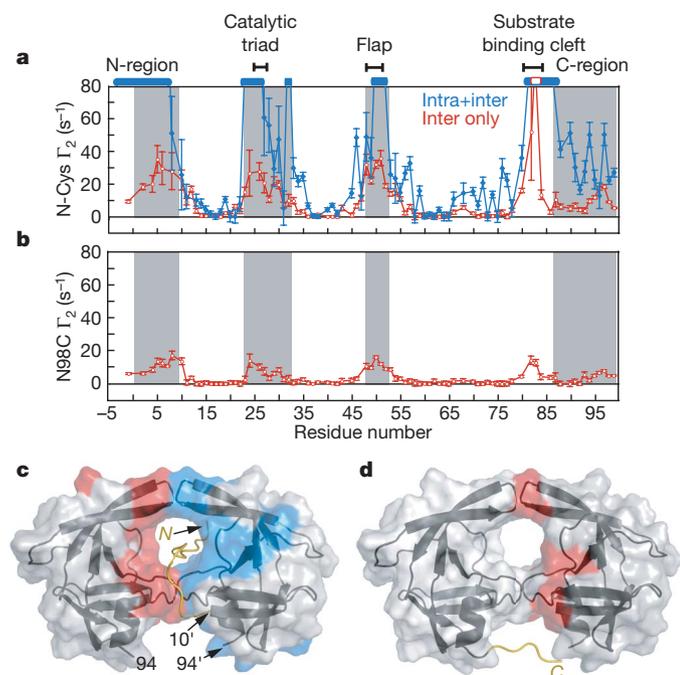
Figure 3 | PRE profiles with spin labels attached at the N- and C termini of the ^{SFNF}PR(D25N) mini-precursor. a, Intermolecular PREs (red) observed for a 1:1 mixture (0.2 mM each) of N-terminal spin-labelled $^{S(C)FNF}$PR(D25N) at natural isotopic abundance and U-[^{2}H/^{13}C/^{15}N]–^{SFNF}PR(D25N), and the sum of the inter- and intramolecular PREs (blue) observed for 0.2 mM N-terminal spin-labelled U-[^{2}H/^{13}C/^{15}N]–$^{S(C)FNF}$PR(D25N). Residues broadened beyond detection are denoted by open bars. **b**, Intermolecular PREs observed for a 1:1 mixture (0.2 mM each) of U-[^{2}H/^{13}C/^{15}N]–^{SFNF}PR(D25N) and C-terminal spin-labelled (at N98C) ^{SFNF}PR(D25N) at natural isotopic abundance. Grey shaded areas in **a** and **b** delineate residues that are buried at the dimer interface in the mature protease. Error bars in **a** and **b** represent 1 s.d. **c, d**, Inter- and intramolecular PREs with Γ_2 rates >10 s^{-1} colour-coded in red and blue, respectively, onto the molecular surface of the mature protease dimer originating from the N-terminal (**c**) and the C-terminal (**d**) spin labels. The intramolecular PRE rates are given by the difference in PRE rates between the blue and red profiles in **a**. Cartoons of modelled N-terminal (residues −4 to 9) and C-terminal (residues 95–99) regions bearing the spin labels are included in **c** and **d**, respectively.

C-terminal regions, can accommodate transient insertion of the N-terminal region including the N-terminal cleavage site in the substrate binding cleft, thereby providing a structural model for autoprocessing at the N terminus of the protease leading to the formation of a stable dimer with mature catalytic activity.

METHODS SUMMARY

Sample preparation and NMR spectroscopy. Protein expression, mutagenesis, purification and conjugation of engineered surface cysteine residues to 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) are described in the Methods. Samples for NMR were in 20 mM sodium phosphate buffer, pH 5.8. NMR experiments were collected at 20 °C at a ^{1}H spectrometer frequency of 600 MHz. ^{1}H$_N$ PRE data were acquired using a two-dimensional ^{1}H–^{15}N correlation-based pulse scheme with an interleaved two time-point measurement[21].
Simulated annealing calculations. Conjoined rigid-body/torsion angle dynamics simulated annealing calculations on the basis of the PRE data were carried out using Xplor-NIH[22] as described[14].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 July; accepted 12 August 2008.

1. Louis, J. M., Weber, I. T., Tozser, J., Clore, G. M. & Gronenborn, A. M. HIV-1 protease: maturation, enzyme specificity, and drug resistance. *Adv. Pharmacol.* **49**, 111–146 (2000).

2. Louis, J. M., Ishima, R., Torchia, D. A. & Weber, I. T. HIV-1 protease: structure, dynamics and inhibition. *Adv. Pharmacol.* **55**, 261–298 (2007).

3. Wlodawer, A. & Erikson, J. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585 (1993).

4. Wlodawer, A. & Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).

5. Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M. & Louis, J. M. Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J. Biol. Chem.* **278**, 43311–43319 (2003).

6. Miller, M. *et al.* Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **246**, 1149–1152 (1989).

7. Louis, J. M., Nashed, N. T., Parris, K. D., Kimmel, A. R. & Jerina, D. M. Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. *Proc. Natl Acad. Sci. USA* **91**, 7970–7974 (1994).

8. Co, E. *et al.* Proteolytic processing mechanisms of a miniprecursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry* **33**, 1248–1254 (1994).

9. Louis, J. M., Wondrak, E. M., Kimmel, A. R., Wingfield, P. T. & Nashed, N. T. Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. *J. Biol. Chem.* **274**, 23437–23442 (1999).

10. Louis, J. M., Clore, G. M. & Gronenborn, A. M. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature Struct. Biol.* **6**, 868–874 (1999).

11. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M. & Kaplan, A. H. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.* **78**, 8477–8485 (2004).

12. Ishima, R., Torchia, D. A. & Louis, J. M. Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor. *J. Biol. Chem.* **282**, 17190–17199 (2007).

13. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227–1230 (2006).

14. Tang, C., Iwahara, J. & Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383–386 (2006).

15. Volkov, A. N., Worall, J. A., Holtzmann, E. & Ubbink, M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proc. Natl Acad. Sci. USA* **103**, 18945–18950 (2006).

16. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).

17. Tessmer, U. & Kräusslich, H.-G. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J. Virol.* **72**, 3459–3463 (1998).

18. Ludwig, C., Leiherer, A. & Wagner, G. Importance of protease cleavage sites within and flanking human immunodeficiency virus type 1 transframe protein p6* for spatiotemporal regulation of protease activation. *J. Virol.* **82**, 4573–4584 (2008).

19. Wondrak, E. M., Nashed, N. T., Haber, M. T., Jerina, D. M. & Louis, J. M. A transient precursor of the HIV-1 protease: isolation, characterization and kinetics of maturation. *J. Biol. Chem.* **271**, 4477–4481 (1996).

20. Cherry, E. *et al.* Characterization of human immunodeficiency virus type-1 (HIV-1) particles that express protease-reverse transcriptase fusion proteins. *J. Mol. Biol.* **284**, 43–56 (1998).

21. Iwahara, J., Tang, C. & Clore, G. M. Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184**, 185–195 (2007).

22. Schwieters, C. D., Kuszewski, J. & Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).

23. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against ^{1}H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126**, 5879–5896 (2004).

24. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **73**, 1391–1396 (1991).

25. Schwieters, C. D. & Clore, G. M. Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR* **23**, 221–225 (2002).

METHODS

Vector construction and protein sample preparation for NMR studies. Mutations T12C, E34C, V82C and N98C within the 99-amino-acid-long HIV-1 protease sequence and the mutation to insert a Cys in the flanking SFNF sequence (C-terminal residues of p6^{pol} within the transframe region; see Fig. 1a and Supplementary Fig. 1) to generate $^{S(C)FNF}$PR(D25N) were introduced in the ^{SFNF}PR(D25N) template[5] using the appropriate forward and reverse primers and the QuikChange kit and protocol (Stratagene). The $^{S(C)FNF}$PR(D25N) construct was used because we were unable to obtain efficient spin-labelling of a precursor protein bearing an N-terminal cysteine. Mutations A(−44)C (fifth residue of p6^{pol}) and V82C (in the protease sequence) were also introduced in the full-length TFR–PR(D25N) construct (that is, TFP–p6^{pol}–PR(D25N); see Fig. 1a and Supplementary Fig. 1a) using the same protocol. (The TFR is 56 residues in length and adopts a random coil conformation.) The newly introduced mutations were verified both by DNA sequencing and mass spectrometry. (Note that the ^{SFNF}PR(D25N) template, in addition to the D25N mutation which eliminates all traces of catalytic activity, and the C67A and C95A mutations which remove all additional surface cysteine residues other than that to which the spin label is going to be attached, also contains three other mutations, Q7K, L33I and L63I; the latter three mutations restrict autoproteolysis of the mature protease dimer, and have been shown to have indiscernible effects on structure, stability and catalytic activity of the mature dimer[10].)

Escherichia coli BL21 (DE3) host cells bearing the appropriate vector were grown in Luria-Bertani medium or in D$_2$O-based minimal medium containing ^{15}N-NH$_4$Cl and ^{13}C$_6$,^{2}H$_7$-glucose as the sole nitrogen and carbon sources, respectively, at 37 °C, and induced for expression. Proteins were purified from inclusion bodies using an established protocol as described previously involving size-exclusion chromatography under denaturing conditions followed by reverse-phase HPLC[9,26]. Peak fractions (~0.5 mg ml^{-1}) were stored in aliquots at –70 °C. Alternatively, two aliquots (2.5 mg) of the proteins were lyophilized and stored at –20 °C.

A total of 2.5 mg of the lyophilized protein was dissolved in 1.2 ml of 4 M guanidinium-HCl, 1.7 mM HCl, pH 1.6. Spin-label conjugation was carried out by dissolving 0.5 mg of 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) (catalogue number I709500; Toronto Research Chemicals) in 10 μl of ethanol, followed by the addition of 140 μl of 1 M Tris-HCl, pH 8, and adding the resulting mixture to the protein solution. After incubation for 1 h at room temperature, 30 μl of 1 M dithiothreitol was added and the incubation continued for another 1.5 h. The sample was loaded onto a Superdex-75 column (1.6 × 60 cm, GE HealthCare) equilibrated in 4 M guanidinium-HCl, 20 mM sodium formate, pH 2.6, at a flow rate of 1.4 ml min^{-1} at room temperature. Peak fractions were pooled and the concentration was estimated by measuring absorbance at 280 nm. The extent of labelling was 100% as determined by MALDI–TOF analysis on a Voyager-DE instrument (Perceptive Biosystems). Spin-labelling does not perturb the structure of the ^{SFNF}PR(D25N) mini-precursor as judged by NMR spectroscopy. The three mutations within the protein core, T12C, E34C and V82C, are frequently mutated in viable HIV-1 variants and are therefore not expected to alter significantly the catalytic properties of the protease[2]. It should be noted that V82C is located close to the substrate binding cleft comprising residues 80–81 and 83–85, but its side chain points outwards towards solvent. In the one instance where kinetic data are available for a mutation at position 82 (V82A), only a modest 10–15% decrease in k_{cat}/K_m relative to wild type is observed, and structural differences between wild-type protease and the V82A mutant are insignificant, with an r.m.s. deviation between the two crystal structures of only 0.12 Å for all main chain atoms[27]. Thus, the presence of a bulky spin label at position 82 would not be expected to result in a major perturbation in catalytic activity.

After extensive dialysis against 7 mM HCl, 1.4 mg each of the conjugated protein and the U-[^{2}H/^{13}C/^{15}N]-labelled ^{SFNF}PR(D25N) protein were mixed and adjusted to a final concentration of 0.25 mg ml^{-1} protein, 35% acetonitrile and 0.05% trifluoroacetic acid. The solution was dialysed against 2 l of 7 mM HCl and 4 l of 20 mM sodium phosphate, pH 5.8, each for a period of 1.5–2 h and concentrated to ~400 μM using Amicon Ultra-4 (10,000 MWCO) devices. Protein concentration (mg ml^{-1}) was determined spectrophotometrically using ε (0.1%) = 1.097 at 280 nm.

Control active ^{SFNF}PR(D25) mini-precursor protease construct. The ^{SFNF}PR(D25N) precursor construct does not undergo autoprocessing owing to the substitution of the active site Asp 25 by Asn. To verify that ^{SFNF}PR(D25N) represents a suitable model system we examined the autoprocessing activity of the corresponding ^{SFNF}PR(D25) precursor; that is, the precursor without the active site mutation. Most of the expressed protein undergoes maturation at the N terminus (between Phe-Pro) of the protease in the control ^{SFNF}PR(D25) precursor to produce the mature protease as expected. This was confirmed by

subjecting an aliquot of the purified (dissolved) inclusion bodies to electrospray-mass spectrometry. The measured mass of 10,728 Da clearly corresponds to the PR(D25) mature protease (calculated mass of 10,728.3 Da). Under identical conditions of analysis for ^{SFNF}PR(D25N), which is devoid of catalytic activity, only the full-length protein corresponding to a mass of 11,222 Da (calculated mass of 11,222.8 Da) is observed consistent with previous observations from studies using the inactive full-length TFR–PR(D25N) precursor, which does not undergo maturation[5], as compared to the active TFR–PR(D25) precursor, which exhibits time-dependent processing at the p6^{pol}–protease junction to release the mature protease[9,10].

NMR experiments. All NMR data were acquired at 20 °C on a Bruker DRX600 spectrometer equipped with a z-gradient triple resonance cryoprobe.

Measurement of translational diffusion coefficients (D_s) by pulse field gradient NMR[28] was carried out using the Watergate BPP-LED pulse scheme described previously[29]. The translational diffusion coefficient D_s is derived from a linear least-squares fit to a plot of $\ln[I(f)/I(f_0)]$ versus $(f^2 − f_0^2)$:

$$\ln[(I(f)/I(f_0)] = -(\gamma\delta G_{max})^2(f^2 - f_0^2)(\Delta - \delta/3 - \tau/2)D_s$$

where $I(f)$ and $I(f_0)$ are the intensities of the NMR signal at fractional gradient strengths of f and f_0; f_0 is the fractional gradient strength of the reference spectrum (0.1); f is the fractional gradient strength with values of 0.2, 0.3, 0.4, 0.5 and 0.6 times G_{max}, the maximum gradient strength (70 × 10^{-4} T cm^{-1}); γ is the gyromagnetic ratio of ^{1}H (2.6752 × 10^8 s^{-1} T^{-1}); Δ = 15.4 ms; δ = 5 ms (gradient duration); and τ = 0.2 ms. The overall diffusion delay is 10 ms. The value of the scaling factor $(\gamma\delta G_{max})^2(\Delta - \delta/3 - \tau/2)$ is 1.19 × 10^{10} s m^{-2}. The values of D_s were 9.3(±0.4) × 10^{-11} and 12.9(±0.5) × 10^{-11} m^2 s^{-1} for the ^{SFNF}PR(D25N) precursor and the mature PR(D25N) dimer, respectively, at the same (0.4 mM) subunit concentration (Supplementary Fig. 1d). The ratio of the two D_s values (0.72 ± 0.04) is fully consistent with the expected value of 0.75 for a $D_s^{monomer}/D_s^{dimer}$ ratio[28], placing an upper limit of about 10% for the population of dimeric species.

^{15}N-{^{1}H} heteronuclear NOE measurements were carried out using a flip-back scheme as described[30]. Residues −4 to 9 and 95–99 of ^{SFNF}PR(D25N) have heteronuclear ^{15}N-{^{1}H} NOE values ranging from −1 to 0.5 indicating that they are disordered and highly mobile. Backbone assignments were derived using the following three-dimensional triple resonance experiments: HNCO, HN(CO)CA and CBCA(CO)NH[31,32]. The weighted mean backbone chemical shift difference between different constructs is given by $[\Delta\delta^2_{HN} + \Delta\delta^2_N/25 + \Delta\delta^2_{C\alpha}/4]^{1/2}$ as described previously[33]. A comparison of ^{1}H/^{15}N/^{13}Cα chemical shifts reveals significant perturbations relative to the corresponding mature dimeric PR(D25N) for residues located at the dimer interface (Supplementary Fig. 1b), but only minor perturbations relative to the equivalent monomeric PR(1–95) construct obtained by deletion of the C-terminal four residues (Supplementary Fig. 1c)[5]. Analysis of the chemical shift index (based on ^{13}Cα, ^{13}Cβ and ^{13}C′ shifts)[34] for ^{SFNF}PR(D25N) and PR(D25N) indicates that the secondary structure elements are preserved in the precursor with the exception of the N- and C-terminal strands which form an intersubunit four-stranded antiparallel β-sheet in the mature dimer (Supplementary Fig. 1e).

PRE ^{1}H$_N$-Γ$_2$ rates are given by the difference in R_2 relaxation rates between the paramagnetic (spin-labelled) and diamagnetic states of the protein. R_2 rates were determined from a two-time-point interleaved two-dimensional ^{1}H–^{15}N correlation-based experiment, as described previously[21]. The time interval between the two time points was 32 ms for the intermolecular PRE measurements and 4 ms for the intramolecular PRE measurements. The short time interval for the latter is used to minimize any errors in Γ$_2$ rates introduced by any potential diamagnetic contamination (that is, spin-labelling less than 100%)[21].

Tertiary structure of ^{SFNF}PR(D25N). To verify that the tertiary structure of the ordered region of ^{SFNF}PR(D25N) (that is, residues 10–94) is the same as that of an individual subunit of the mature protease, we made use of the CS-Rosetta chemical shift structure determination algorithm which uses a hybrid approach of chemical-shift-based fragment selection and ROSETTA Monte Carlo driven fragment assembly[35]. The resulting ten lowest energy models are essentially identical to the corresponding region of the mature dimer with a backbone r.m.s. deviation of only 1.3 ± 0.2 Å (Supplementary Fig. 1e).

PRE calculations and ensemble refinement. Because the electron relaxation rate τ_s of the free radical is much longer than that of the protein rotational correlation time τ_r[21], the PRE correlation time τ_c [= $(\tau_r^{-1} + \tau_s^{-1})^{-1}$] for the calculation of intermolecular PRE rates was assumed to be the same as τ_r (12 ns) for the mature protease dimer[36]. To account for the flexibility of the linker between the spin label and the protein backbone, a ten-conformer randomized ensemble was used to represent the conformational space sampled by the spin label. The randomized ensemble was generated by high-temperature simulated annealing and slow cooling in Xplor-NIH[22] subject to a target function compris-

ing stereochemical terms, a quartic van der Waals repulsion term to prevent atomic overlap between the spin label and the protein, and a multidimensional conformational database potential of mean force[37] describing the $\phi/\psi/\chi_1$ conformational space available to the surface cysteine residue to which the spin label was conjugated. Note that overlap between the members of the Cys spin-label ensemble is permitted as the ten-member ensemble represents a distribution of states. To ensure full sampling of the conformational space available to the spin label a different ten-conformer randomized ensemble was used for each structure calculation. Agreement between observed and calculated Γ_2 rates is given by the PRE Q-factor, Q_{PRE}:[23]

$$Q_{PRE} = \left[\sum_i \left\{ \Gamma_{2,i}^{obs} - p\langle \Gamma_{2,i}^{calc} \rangle \right\}^2 \bigg/ \sum_i \left(\Gamma_{2,i}^{obs} \right)^2 \right]^{1/2}$$

where $\Gamma_{2,i}^{obs}$ and $\langle \Gamma_{2,i}^{calc} \rangle$ are the observed and ensemble average calculated transverse Γ_2 rates for residue i, respectively, and p is the overall population of the encounter complex species. All members of an ensemble of size N_e are weighted equally. For the average Q-factor $<Q>$ for all calculated n ensembles, $\langle \Gamma_{2,i}^{calc} \rangle$ is averaged over the members of each N_e ensemble. For the ensemble of ensembles average PRE Q-factor, Q_{ee}, $\langle \Gamma_{2,i}^{calc} \rangle$ is averaged over all ensemble members and all ensembles[14].

The coordinates used in the Xplor-NIH[22] calculations were taken from the X-ray structure of the unliganded mature HIV-1 protease dimer (Protein Data Bank accession code 1HHP)[24]. Residues 10–94 were treated as a rigid body, and the flexible N- and C-terminal residues were not included in the calculations. The coordinates of the isotopically labelled subunit were held fixed, the initial positions of the spin-labelled subunit (at natural isotopic abundance) were randomized, and rigid-body simulated annealing was carried out against the PRE data sets for the spin label conjugated to the T12C, E34C and V82C sites simultaneously. The target function comprises a PRE restraint term[23], a quartic van der Waals repulsion term to prevent atomic overlap between the spin-labelled and isotopically labelled subunits, and a very weak radius of gyration term[38] to ensure that each member of the ensemble makes at least some intermolecular contacts[14,39]. Note that atomic overlap between ensemble members of spin-labelled subunits is permitted as these represent separate but rapidly interconverting configurations of the encounter complex species[14,39]. A grid search was performed varying the population of heterodimer and the ensemble size N_e used to represent the self-associated species[14]. For each ensemble size and population of encounter complex species, 100 calculations were carried out. Ensembles were ranked by PRE Q-factor and van der Waals repulsion energies, and the top 20 ensembles with the smallest PRE Q-factors were used for subsequent analysis[39]. Structures were rendered using PyMol (http://www.pymol.org) and re-weighted atomic probability density maps were generated using Xplor-NIH[22] as described[25].

d.r.m.s. metric. One metric we used to compare the precursor encounter complexes with the mature dimer was the distance root mean square (d.r.m.s.) metric defined by[40]:

$$d.r.m.s. = \frac{1}{N} \sum_{i,j} \left| d_{i,j}^{precursor} - d_{i,j}^{mature} \right|$$

where N is the number of distinct residue pairs (i, j), and $d_{i,j}^{precursor}$ and $d_{i,j}^{mature}$ are the distance matrices in a calculated precursor encounter complex structure and the mature HIV-1 protease dimer structure, respectively.

Spherical coordinate systems used to describe relative subunit orientation in the encounter complexes. Two spherical coordinate systems are used to describe the relative orientation of the subunits in the precursor encounter complexes[40]. The first (polar angle ϕ and azimuth angle θ) describes the orientation of the vector joining the centre of masses of the two subunits (shown as grey spheres in Supplementary Fig. 3a) to an external axis system with the z axis corresponding to the C_2 symmetry axis of the mature dimer. The second (polar angle α and azimuth angle β) describes the orientation of a vector joining the centre of mass of the second subunit to an arbitrarily chosen atom of the same subunit (Cα atom of Gly 51) relative to an axis system with the z' axis given by the vector joining the centre of masses of the two subunits (with the red subunit in Fig. 2 corresponding to the fixed reference subunit) (Supplementary Fig. 3a).

26. Wondrak, E. M. & Louis, J. M. Influence of flanking sequences on the dimer stability of human immunodeficiency virus type 1 protease. *Biochemistry* **35**, 12957–12962 (1996).
27. Mahalingam, B. *et al.* Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **271**, 1516–1524 (2004).
28. Altieri, A. S., Hinton, D. P. & Byrd, R. A. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **117**, 7566–7567 (1995).
29. Chou, J. J., Baber, J. L. & Bax, A. Characterization of phospholipid mixed micelles by translational diffusion. *J. Biomol. NMR* **29**, 299–308 (2004).
30. Grzesiek, S. & Bax, A. The importance of not saturating H$_2$O in protein NMR: application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).
31. Clore, G. M. & Gronenborn, A. M. Two-, three- and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 29–63 (1991).
32. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).
33. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. The CD4 determinant of downregulation by HIV-1 Nef directly binds to Nef: mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256–10261 (1996).
34. Wishart, D. S. & Sykes, B. D. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).
35. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105**, 4685–4690 (2008).
36. Katoh, E. *et al.* A solution NMR study of the binding kinetics and internal dynamics of an HIV-1 protease-substrate complex. *Protein Sci.* **12**, 1376–1385 (2003).
37. Clore, G. M. & Kuszewski, J. χ_1 rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J. Am. Chem. Soc.* **124**, 2866–2867 (2002).
38. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **121**, 2337–2338 (1999).
39. Tang, C., Ghirlando, R. & Clore, G. M. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **130**, 4048–4056 (2008).
40. Kim, Y. C., Tang, C., Clore, G. M. & Hummer, G. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc. Natl Acad. Sci. USA* **105**, 12855–12860 (2008).

CORRIGENDUM

Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease

Chun Tang, John M. Louis, Annie Aniana, Jeong-Yong Suh
& G. Marius Clore

In the online-only Methods of this Letter, the values of the translational diffusion coefficients (D_s) ascribed to the precursor and the mature dimer were inadvertently transposed. The corrected sentence should read:

'The values of D_s were $9.3(\pm0.4) \times 10^{-11}$ and $12.9(\pm0.5) \times 10^{-11}$ m^2s^{-1} for the mature PR(D25N) dimer and the ^{SFNF}PR(D25N) precursor, respectively, at the same (0.4 mM) subunit concentration (Supplementary Fig. 1d). The ratio of the two D_s values (0.72 ± 0.04) is fully consistent with the expected value of 0.75 for a $D_s^{dimer}/D_s^{monomer}$ ratio[28], placing an upper limit of about 10% for the population of dimeric species.'