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_2O-based minimal medium containing ^{15}N-NH_4Cl 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 $9.3(\pm0.4) \times 10^{-11}$ and $12.9(\pm0.5) \times 10^{-11}$ 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 $D_s^{momomer}/D_s^{dimer}$ 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/$^{13}C\alpha$ 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 $^{13}C\alpha$, $^{13}C\beta$ 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]

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

where $\Gamma_{2,i}^{\mathrm{obs}}$ and $\langle \Gamma_{2,i}^{\mathrm{calc}} \rangle$ 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, $\langle \Gamma_{2,i}^{\mathrm{calc}} \rangle$ is averaged over the members of each N_e ensemble. For the ensemble of ensembles average PRE Q-factor, Q_{ee}, $\langle \Gamma_{2,i}^{\mathrm{calc}} \rangle$ 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]:

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

where N is the number of distinct residue pairs (i, j), and $d_{i,j}^{\mathrm{precursor}}$ and $d_{i,j}^{\mathrm{mature}}$ 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_2O 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

doi:10.1038/nature07342

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

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

Nature 455, 693–696 (2008)

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($\pm$0.4) $\times$ 10^{-11} and 12.9($\pm$0.5) $\times$ 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 $\pm$ 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.'