Autoprocessing of HIV-1 protease is tightly coupled to protein folding

John M. Louis, G. Marius Clore and Angela M. Gronenborn

In the Gag-Pol polyprotein of HIV-1, the 99-amino acid protease is flanked at its N-terminus by a transframe region (TFR) composed of the transframe octapeptide (TFP) and 48 amino acids of the p6^{pol}, separated by a protease cleavage site. The intact precursor (TFP-p6^{pol}-PR) has very low dimer stability relative to that of the mature enzyme and exhibits negligible levels of stable tertiary structure. Thus, the TFR functions by destabilizing the native structure, unlike proregions found in zymogen forms of monomeric aspartic proteases. Cleavage at the p6^{pol}-PR site to release a free N-terminus of protease is concomitant with the appearance of enzymatic activity and formation of a stable tertiary structure that is characteristic of the mature protease as demonstrated by nuclear magnetic resonance. The release of the mature protease from the precursor can either occur in two steps at pH values of 4 to 6 or in a single step above pH 6. The mature protease forms a dimer through a four-stranded β-sheet at the interface. Residues 1–4 of the mature protease from each subunit constitute the outer strands of the β-sheet, and are essential for maintaining the stability of the free protease but are not a prerequisite for the formation of tertiary structure and catalytic activity. Our experimental results provide the basis for the model proposed here for the regulation of the HIV-1 protease in the viral replication cycle.

Retroviral proteases including HIV-1 protease, which range in size from 99 to 125 amino acids, are symmetric homodimers. Their active site is formed along the interface between the two subunits containing the two catalytically important Asp residues[1]. A single copy of the HIV-1 protease is synthesized as part of a polyprotein, Gag-Pol (Fig. 1), flanked by the highly variable p6^{pol} sequence at its N-terminus and by the reverse transcriptase (RT) at its C-terminus[2,3]. The initial critical step in the maturation of the Gag-Pol precursor is the folding/dimerization of the protease domain and formation of an active site that is capable of catalyzing the hydrolysis of the peptide bonds at crucial sites to release the mature protease. The mature protease has optimal enzymatic activity and specificity, and thus, is responsible for the processing of the Gag and Gag-Pol polyproteins at precise sites to release a large number of mature proteins required for viral assembly and maturation, and for subsequent replication[2,4] (Fig. 1).

A major effort directed toward structure-assisted drug design has led to the development of potent inhibitors that bind to the active site of the mature protease, and several of these drugs are currently in use for the treatment of AIDS[5]. Unfortunately, use of these protease inhibitors leads to selection for drug-resistant variants of the enzyme within a short time[6]. In contrast to the wealth of studies focusing on the active site, studies aiming to elucidate the pathway that leads to formation of the mature protease from the Gag-Pol polyprotein and the *in vivo* mecha-

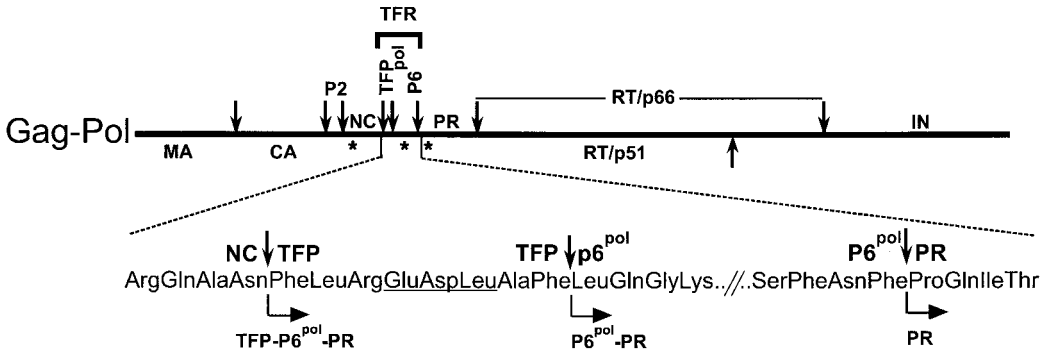

Fig. 1 Structural organization of Gag-Pol polyprotein in HIV-1. Straight arrows indicate the specific sites of cleavage by the viral protease. The amino acids spanning the NC-TFP, TFP-p6^{pol} and p6^{pol}-PR junctions are shown. The transframe region (TFR) consists of the transframe octapeptide TFP and 48 amino acids of p6^{pol}. Additional novel protease cleavage sites are represented by an asterisk in NC, in p6^{pol} and in PR[16,18,36]. The underlined Glu-Asp-Leu tripeptide of TFP is a specific inhibitor of HIV-1 protease[7]. Bracketed arrows indicate the N-termini of the protease precursor (TFP-p6^{pol}-PR), the intermediate p6^{pol}-PR and the mature PR. Nomenclature of HIV-1 proteins is according to Leis *et al*. 1988 (ref. 37). Abbreviations: MA, matrix; CA, capsid; PR, protease; NC, nucleocapsid; RT, reverse transcriptase; RN, RNase H; IN, integrase.

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

Correspondence should be addressed to A.M.G. *email: gronenborn@nih.gov*

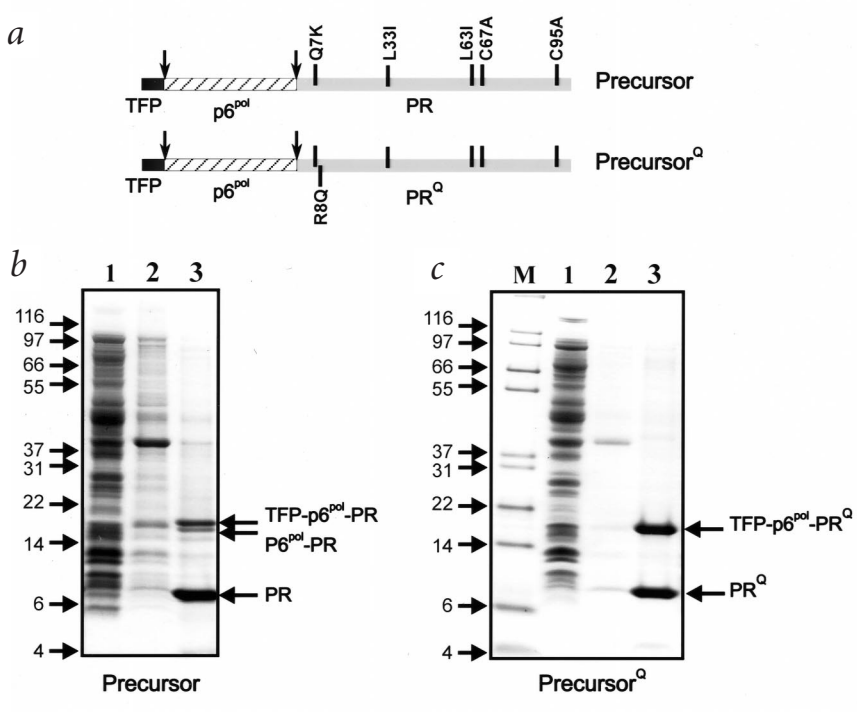

Fig. 2 Protease precursor constructs and expression in *E. coli*. ***a***, Precursor constructs used in this study. Vertical arrows indicate the specific sites of cleavage by the viral protease. The precursor (TFP-p6^{pol}-PR) encodes an intact sequence of TFP, flanked by p6^{pol} and PR. Precursor^{Q} is identical to the precursor construct, except for an additional mutation R8Q in the PR domain. The corresponding mature proteases are denoted PR^{Q} and PR, respectively. Expression of proteins and SDS-PAGE analyses of the subcellular fractions for ***b***, the precursor construct, and for ***c***, Precursor^{Q}. Cells were grown in LB medium and induced for expression using IPTG. Lane 1 shows the supernatant derived after lysis of induced cells in 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 10 mM DTT and 100 μg ml^{-1} lysozyme. The insoluble pellet attained from this step was treated with 2 M urea solution in buffer containing 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 10 mM DTT, 0.5 % NP-40. The supernatant and the pellet (inclusion bodies) derived from this step are shown in lanes 2 and 3, respectively. Proteins were subjected to 10–20% gradient SDS-PAGE followed by Coomassie staining. M denotes molecular mass standards in kilodaltons. The intermediary precursor, p6^{pol}-PR, is indicated.

nism of regulation of HIV-1 and other related viral proteases have been difficult, given the complexity of the the viral biology and intricacies of Gag-Pol polyprotein processing (Fig. 1; for a review, see ref. 4).

The transframe region (TFR) flanking the N-terminus of the protease may represent a proregion similar to that found in zymogen forms of cellular aspartic proteases. However, the role of the transframe region is not fully understood. The TFR consists of two domains: the N-terminal transframe octapeptide (TFP), which is conserved in all variants of HIV-1, and a 48- or 60-amino acid p6^{pol}. These two domains are separated by protease cleavage sites[3,7]. Although the isolated 68-amino acid TFR has no overall stable secondary or tertiary structure, a small potential for helix formation at its N-terminus was shown by NMR[8]. An early study has shown that deletion of the p6^{pol} domain in the Gag-Pol polyprotein leads to a significantly higher rate of processing activity of the Gag-ΔPol precursor[9]. A later study shows that the p6^{pol} and the nucleocapsid (NC) domains (Fig. 1) have opposite effects on protease maturation. Specifically, the presence of p6^{pol} decreases protease processing whereas the addition of the NC domain to p6^{pol}-protease strikingly increases the processing activity, possibly by influencing protease dimerization[10]. Apart from a report that TFP and its derivatives are hydrophilic competitive inhibitors of the protease[7], no specific structural feature has been ascribed to TFR in regulating the autocatalytic maturation of the protease from the Gag-Pol precursor.

Studies of the maturation of the protease from a model precursor containing only the two native cleavage sites flanking the protease domain and fused to the maltose binding protein (MBP) at its N-terminus showed that the reaction takes place in two independent sequential steps. The first step involves an intramolecular cleavage at the N-terminus of the protease domain concomitant with a large increase in mature-like enzymatic activity and the appearance of the transient protease intermediate containing the flanking C-terminal polypeptide[11,12]. The second step is the intermolecular cleavage at the C-terminus of the protease, releasing the mature enzyme[13]. Using a full-length native p6^{pol}-protease precursor (Fig. 1), we confirmed that the rate-limiting intramolecular cleavage at the p6^{pol}-protease site is indeed concomitant with the increase in mature-like catalytic activity. This finding rules out the possibility that MBP influences the observed kinetics of the maturation of the model precursor (data not shown). The above results are consistent with the observation that blocking the cleavage at the p6^{pol}-protease site leads to considerably reduced cleavage at the C-terminus of protease and impairs Gag polyprotein processing and viral infectivity[14–17].

We speculated that the low catalytic activity of the native p6^{pol}-protease precursor that is observed preceding its maturation may be an intrinsic feature of the dimeric precursor. Alternatively, the low activity may be an apparent effect of the equilibrium between the unfolded or partially folded form of the protein and the folded, enzymatically active dimer predominantly shifted toward the unfolded state. To address these critical events of the viral replication cycle, we developed expression systems for purifying idealized precursor proteins of the protease fused to the intact TFR containing the native cleavage sites to permit studies of the autocatalytic maturation of the protease by kinetics and NMR. Based on our experimental results, we suggest a plausible mechanism for the regulation of protease activity in the virus replication cycle.

Idealized precursor proteins of protease fused to the native TFR

The wild type mature protease undergoes rapid autoproteolysis at ambient temperature, resulting in discrete cleavage products and loss of enzymatic activity even at very low (nanomolar) protein concentrations. The major site of autoproteolysis in the mature enzyme is between Leu 5 and Trp 6; the other two minor sites are at Leu 33–Glu 34 and Leu 63–Ile 64 (refs 18,19). A mutant enzyme bearing the substitutions Q7K, L33I and L63I retains the specificity and kinetic properties of the wild type enzyme[19]. This mutant enzyme, free of any inhibitor, is highly stabilized against autoproteolysis at concentrations up to 1 mM at pH 5.8, thus permitting structural studies by NMR.

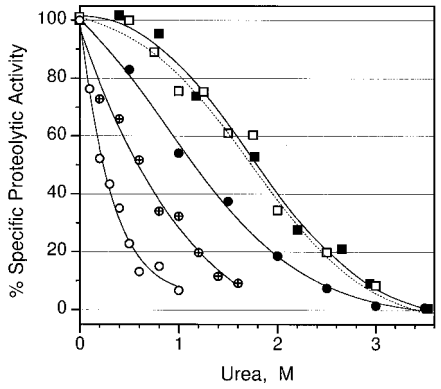

Fig. 3 Urea denaturation curves for wild type mature protease (□), PR (■), PR^{Q} (●) and ΔPR (⊕) as determined by measuring changes in enzymatic activity in buffer A at 25 °C. For p6^{pol}-PR precursor (○; Fig. 1) the enzymatic activity reflects the ability to undergo autocatalytic maturation. ΔPR was measured at final enzyme and substrate concentrations of 1.1 μM and 600 μM, respectively.

The newly devised precursor constructs consist of a protease domain (termed PR), containing the mutations Q7K, L33I, L63I, C67A and C95A to restrict autoproteolysis as well as cysteine thiol oxidation, fused to the intact TFR (Figs 1, 2*a*). The expression in *Escherichia coli* and the extent of autoprocessing of the protease precursor and its variant precursor^{Q}, which bears an additional mutation R8Q in the PR domains are shown (Fig. 2*b,c*). Both precursors contain the three intact domains TFP, p6^{pol} and PR (or PR^{Q}), separated by two native PR cleavage sites. Lanes 1–3 in Fig. *2b,c* show the supernatant upon lysis of cells, and the supernatant and the pellet derived after 2 M urea treatment of the initial insoluble pellet, respectively. A comparison of lanes 2 and 3 reveals that 2 M urea does not solubilize the expressed proteins in the inclusion bodies. In addition, the expressed precursor proteins undergo processing in *E. coli* to release the mature protease domains (lane 3, Fig. 2*b,c*). The processed mature PR is the predominant species in the case of the precursor, whereas for the precursor^{Q}, approximately equal amounts of precursor and mature enzyme (PR^{Q}) are observed. The precursor forms may represent the fraction of total expressed protein that is either misfolded or does not undergo maturation, and accumulates in the insoluble fraction.

Catalytic activity and structure stability of mature wild type and mutant proteases

The kinetic parameters k_{cat} and K_m for wild type protease and mutant PR-catalyzed hydrolysis of the substrate, as well as the inhibition constant (K_i) for the hydrolytic reaction with the inhibitor, are shown in Table 1. Mutant PR exhibits only a two- to three-fold increase in k_{cat}/K_m compared to the wild type enzyme. The measured K_i for the hydrolytic reaction was also similar for the enzymes tested using an inhibitor with a reduced peptide bond (RPB inhibitor), an analog of the Gag CA-p2 cleavage site substrate (Table 1). Protein stability experiments, as measured by enzymatic activity *versus* urea concentration, reveal that 50% of enzymatic activity is lost at 1.75 M urea for the wild type protease and at 1.8 M urea for PR (Fig. 3). Denaturation curves obtained in this fashion superimpose extremely well onto those obtained by monitoring protein fluorescence or circular dichroism[13,20], demonstrating that loss of

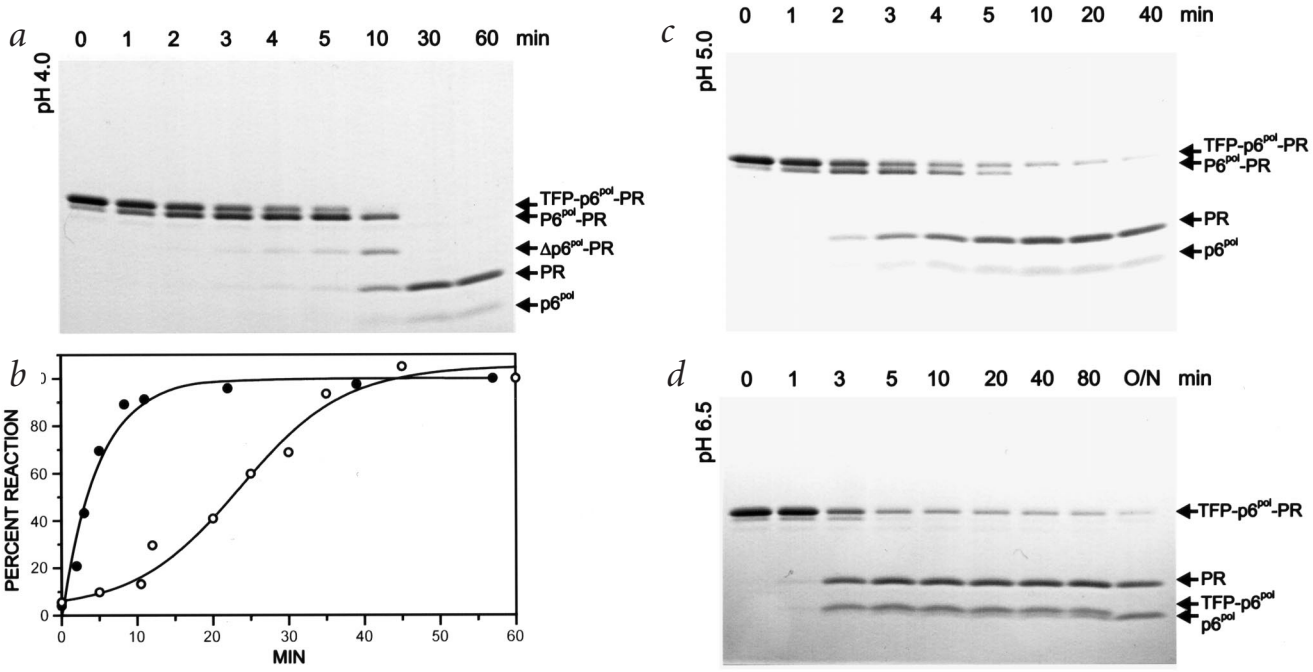

Fig. 4 Time course of the autocatalytic maturation of the precursor ***a***, at pH 4.0 in 50 mM sodium formate and 0.5 M urea, ***c***, at pH 5.0 in 50 mM sodium acetate and 0.5 M urea and ***d***, at pH 6.5 in 25 mM sodium phosphate buffers at 25 °C. Aliquots of the reaction mixture were drawn at the desired time and subjected to 10–20% SDS-PAGE in Tris-Tricine buffer. Protein bands were visualized by Coomassie brilliant blue G250 staining. The relative mobilities of the full length precursor and product peaks are indicated by arrows. ***b***, Time course for autoprocessing of the precursor monitored by following the appearance in enzymatic activity at pH 6.5 (●); and at pH 4.0 (○). The lag in the reaction time course coincides with accumulation of the transient intermediate p6^{pol}-PR that is subsequently converted to the mature PR.

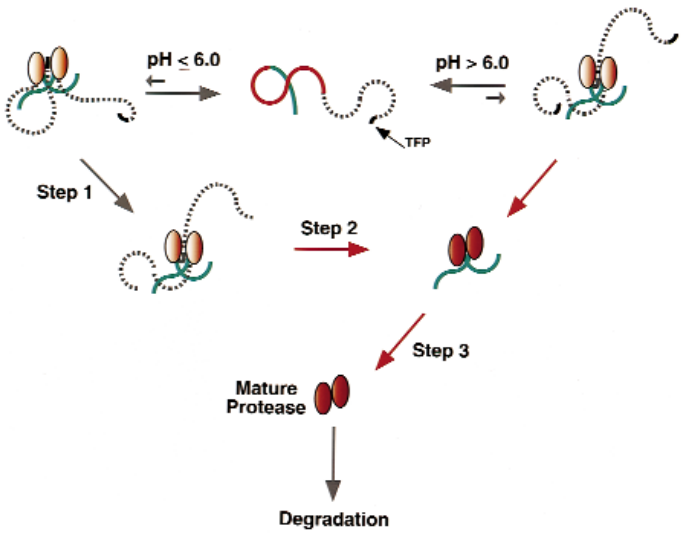

Fig. 5 A proposed model for the steps involved in the maturation of HIV-1 protease. Solid red lines and ovals represent the unfolded and folded forms of the PR domain, respectively. Arrows in red leading to dark red ovals indicate the increase in mature-like catalytic activity. Solid black and green, and dashed gray lines denote TFP, ΔRT and p6^{pol} regions, respectively. We propose that cleavage at the TFP-p6^{pol} site may also occur by an intramolecular mechanism, as an increase in mature-like catalytic activity is concomitant only with cleavage at the p6^{pol}-PR site (step 2).

enzymatic activity is tightly coupled to loss of stable tertiary structure. Overall, a comparison of the results obtained for the wild type enzyme and for PR indicates a considerable similarity in the environments of their active sites and in their structural stability.

Autocatalytic maturation of protease precursor

The precursor (TFP-p6^{pol}-PR) contains two native protease cleavage sites, TFP-p6^{pol} and p6^{pol}-PR (Fig. 2*a*). Upon shifting the pH from <2.8 to between pH 4 and pH 6.5 (see Methods), the precursor undergoes a sequential two step reaction to release mature PR (Figs 4, 5). At pH 5.0 (Figs 4*c*, 5), which is optimal for catalytic activity of mature PR and the autocatalytic maturation reaction, the first cleavage occurs at the TFP-p6^{pol} site to generate the intermediate precursor p6^{pol}-PR. In the second step, p6^{pol}-PR is converted to mature PR. The reaction at pH 4.0 is identical to that at pH 5.0, except that a small yet noticeable amount of an intermediary cleavage product is observed that migrates between the p6^{pol}-PR and PR (Fig. 4*a*). The cleavage occurs between residues Leu 23 and Gln 24 of p6^{pol} to generate this intermediary product (ref. 16; our unpublished results). At pH 6.5 (Figs 4*d*, 5), the two step sequential order of precursor processing is reversed; the first cleavage occurs at the p6^{pol}-PR site to release the mature PR and the TFP-p6^{pol} domains. TFP-p6^{pol} is then cleaved by mature PR to release p6^{pol} and TFP domains. Similar to the model precursor[11] and the native p6^{pol}-PR precursor (Fig. 1), TFP-p6^{pol}-PR exhibits very low catalytic activity. The rate of increase in mature-like catalytic activity is independent of the cleavage at the TFP-p6^{pol} site and is identical to the rate of cleavage at the p6^{pol}-PR site. A plot of the rate of appearance of enzymatic activity *versus* time is characterized by a lag period followed by a first-order process, indicating that cleavage at the N-terminus of the PR domain (p6^{pol}-PR site) is required for catalytic activity (Fig. 4*b*; ○). At pH 6.5, no lag period is observed due to the preferential cleavage at the p6^{pol}-PR site first, resulting in the release of the mature PR domain (Figs 4 *b*,*d*, 5).

The order in which the stepwise maturation of the precursor occurs at different pH values complements our findings that competitive inhibition of the mature protease by TFP and its analogs increases with decreasing pH (pK_a 3.8)[7]. TFP inhibits PR with a K_i of 147 ± 4 μM in 50 mM sodium formate buffer, pH 4.2, which is close to the value observed with the wild type protease[7]. Decreasing pH leads to a higher affinity of TFP for the active site of the protease, promoting the processing at the TFP-p6^{pol} site before the p6^{pol}-PR site. The most likely explanation for the difference in preferential cleavage at different pH values lies in the difference in the protonation states of the two ionizable amino acids in the TFP sequence, namely Glu and Asp residues. It appears that productive binding of the TFP sequence to the nascent active site is preferred for the neutral rather than positively charged state of the Glu and Asp side chains.

Preliminary studies of the inhibition of the maturation reaction of the precursor at pH 5.0 indicate that the two cleavage steps (Fig. 5, steps 1 and 2) leading to the formation of the mature PR are inhibited differently by the RPB inhibitor that is specific to the mature PR (Fig. 6 and Table 1). The concentration of the RBP inhibitor at 50% cleavage inhibition (IC_{50}) at the TFP-p6^{pol} site (step 1) is more than two orders of magnitude larger than that for the inhibition of the hydrolytic reaction catalyzed by the mature PR. This result is consistent with *in vivo* studies showing that even in the presence of a potent protease inhibitor, Gag-Pol is only cleaved in the vicinity of the N-terminus of TFR (TFP-p6^{pol}), leading to the accumulation of the Pol intermediate, p6^{pol}-PR-RT-IN (Fig. 1)[7,21]. Thus the maturation of PR from the Gag-Pol polyprotein *in vivo* may resemble the *in vitro* two step process described for the TFP-p6^{pol}-PR precursor at pH ≤ 6.0 (Fig. 4*c*). Inhibition of processing (K_i) at the p6^{pol}-PR site (Fig. 5, step 2) is within the range of values reported for the inhibition of the wild-type protease[11].

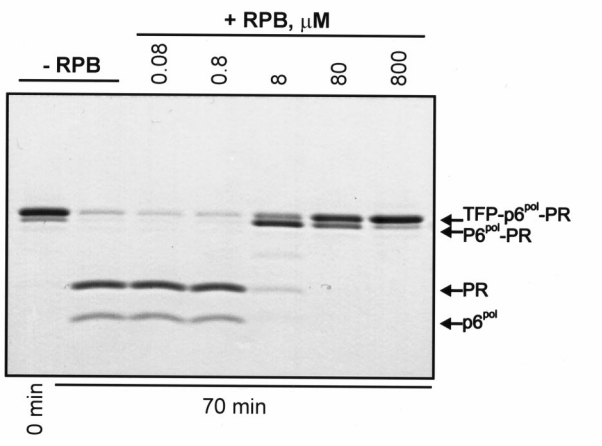

Fig. 6 Inhibition of precursor maturation by RPB inhibitor. The maturation reaction was initiated in either the absence (control) or the presence of increasing amounts of inhibitor in 100 mM sodium acetate buffer, pH 5.0 (similar to Fig. 4*c* but in the absence of urea) and 1.29 μM precursor. Reactions were terminated after 70 min of incubation at 25 °C and subjected to SDS-PAGE. Proteins were visualized by Coomassie staining.

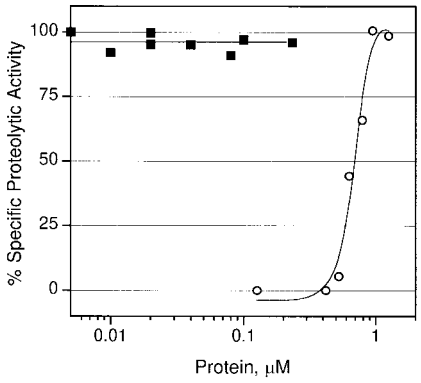

Fig. 7 Specific activity as a function of the dimeric enzyme concentration of precursor (○) and mature PR (■). The autoprocessing reaction was carried out in sodium phosphate buffer at pH 6.5 for 11 min followed by measurement of enzymatic activity using 410 μM substrate. For the precursor the enzymatic activity reflects the ability to undergo dimerization and autocatalytic maturation. No decrease in specific activity is observed for PR and PR^Q to as low as 5 and 3 nM, respectively.

Dissociation constant of protease precursor

Dissociation of the dimeric form of the mature PR was measured by following enzymatic activity as a function of enzyme dilution. The K_d of the mature PR dimer is less than 5 nM at pH 5.0, similar to that of the mature wild type protease (<10 nM; refs 13, 20) under identical conditions. Since the precursor undergoes cleavage at the $p6^{pol}$-PR site as the first step at pH 6.5, in contrast to its behavior at more acidic pH values, the simultaneous appearance of the mature-like enzymatic activity and the mature PR product as a function of precursor concentration is a direct measure of the precursor's ability to dimerize and undergo autocatalytic maturation. A plot of specific enzymatic activity *versus* precursor concentration shows that the apparent K_d for the precursor is ~680 nM (Fig. 7). The decreased stability in the presence of urea (Fig. 3) and >130-fold increase in the apparent K_d of the precursor compared to the mature PR indicate that the structural stability of the protease domain is significantly lower in the precursor.

NMR studies of precursor and mature proteases

Study of the precursor by NMR requires obtaining a sufficient quantity of the protein in an intact form. Previously we screened several of the drug resistance mutations of the protease introduced in the precursor construct with the aim of understanding the relationship between protease maturation and drug resistance. Coincidentally, the R8Q mutant, among several others that were analyzed for expression, showed an elevated level of precursor accumulation (Fig. 2, $precursor^Q$). The R8Q mutant was one of the first drug-resistant mutants to be reported[22], and Weber suggested that the mutation might alter the dimer interface interactions[23]. Although the mature R8Q mutant PR (PR^Q) is more sensitive toward urea denaturation than PR (Fig. 3), the dimer dissociation constant (K_d <3 nM), the kinetic parameters for PR^Q-catalyzed hydrolysis of the peptide substrate, and the inhibition constant for the hydrolytic reaction with inhibitor are comparable to that of PR and the wild type protease (Table 1). Likewise, $precursor^Q$ bearing the R8Q mutation undergoes time-dependent maturation comparable in kinetics and products to that of the precursor. We therefore employed uniformly ^{15}N-labeled $precursor^Q$ in our subsequent studies of protease maturation by NMR.

Rapid precursor maturation, leading to accumulation of products (TFP, $p6^{pol}$ and PR) precludes investigation of a sample of the intact uninhibited precursor by NMR over long periods of time. We nevertheless recorded a ^1H-^{15}N correlation spectrum on a sample containing $precursor^Q$ (identical to Fig. 8, lane -) that exhibited the expected features, namely a set of well-dispersed peaks from the mature PR superimposed on a set of resonances that show the typical narrow dispersion of a random-coil $p6^{pol}$ (data not shown). These features are consistent with ^1H-^{15}N correlation spectra of individually analyzed uninhibited mature PR and $p6^{pol}$ (ref. 8, and unpublished results). An SDS-PAGE analysis of $precursor^Q$ was prepared (Fig. 8) by refolding the protein in the absence (lane -) or presence of a six-fold molar excess of the RPB inhibitor (lane +). However, incomplete inhibition of the processing reaction precluded the use of this inhibitor in the preparation of the inhibited precursor on a large scale for NMR; for this reason we used the more potent inhibitor DMP323 (ref. 24; see Methods)

As can be easily appreciated, the ^1H-^{15}N correlation spectrum of $precursor^Q$ in the presence of excess DMP323 inhibitor (Fig. 9*a*) demonstrates all the hallmarks of a poorly structured, predominantly random-coil polypeptide[25]. It thus appears that the pro-sequence TFR prevents formation of the stable folded mature protease domain when residing on the same polypeptide chain, even in the presence of one of the tightest binding inhibitors. Although only very small amounts of mature protease–DMP323 complex are present in the NMR sample (Fig. 9*a*, see insert of gel), low-intensity narrow resonances can be observed, for instance in the tryptophan Nε region as well as a clearly visible glycine (Gly 40) at ~106 (^{15}N)/8.4 (^1H) p.p.m. The ^1H-^{15}N correlation spectrum of purified mature PR^Q complexed with DMP323 (Fig. 9*c*, blue contours) shows well-dispersed resonances and is very similar to spectra of DMP323 complexes with other mature protease variants studied by NMR[26,27]. This high degree of similarity permitted easy assignment by comparison, and several resonance identities are indicated.

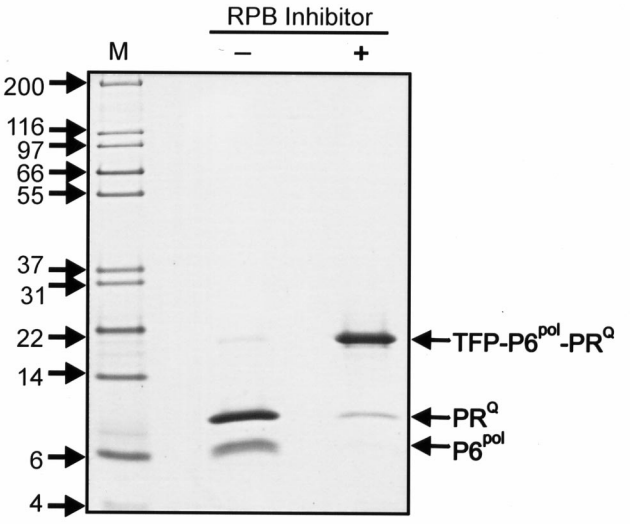

Fig. 8 Inhibition of $precursor^Q$ maturation. The precursor (5 μM) was subjected to the processing reaction in 50 mM sodium formate buffer, pH 4.6, 1 mM DTT, 1 mM EDTA, in the absence (lane -) and in a six-fold excess of the RPB inhibitor[34] (lane +) for 1 h at 25 °C. Samples were analyzed by SDS-PAGE and Coomassie staining. M denotes molecular mass standards in kilodaltons.

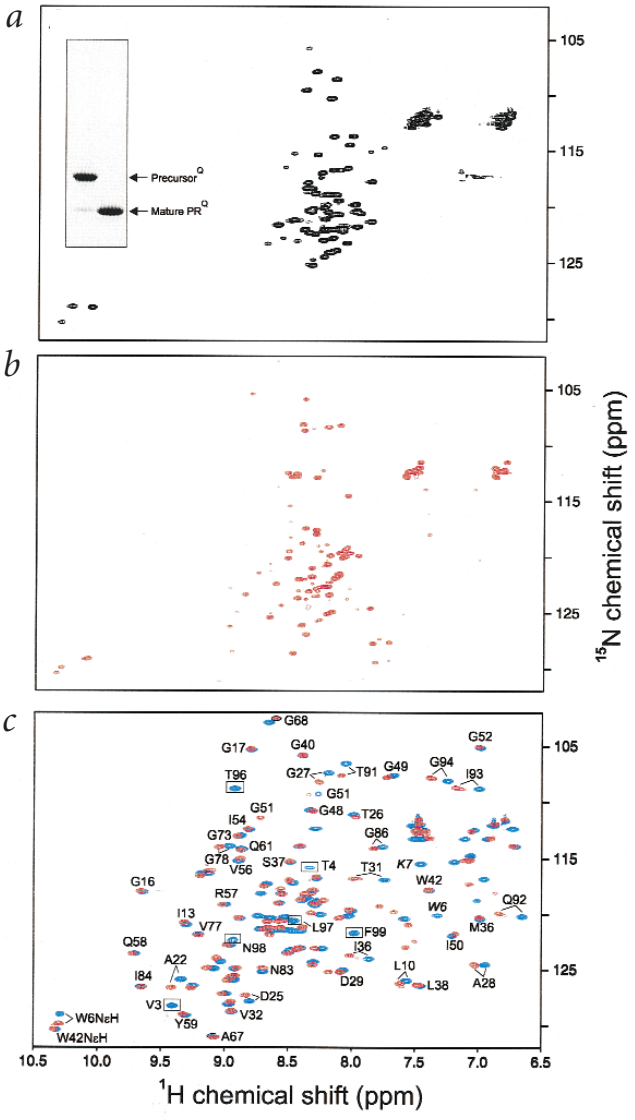

Fig. 9 ^{1}H-^{15}N correlation spectra of ***a***, precursor^{Q}, ***b***, ΔPR and ***c***, ΔPR and mature PR^{Q} in complex with DMP323 (red and blue contours, respectively), in 50 mM sodium acetate buffer, pH 5.2, at 25 °C. Inset in (*a*) illustrates a SDS-PAGE gel of an aliquot of the sample after measurement of the spectra of precursor^{Q} and mature PR^{Q}. There were no degradation products observed in either preparation, except for a minor mature protease product released from the precursor that occurs at the protein folding step even in the presence of excess inhibitor DMP323 (ref. 32). Uninhibited ΔPR (*b*) precipitates over time because of its structural instability. In (*c*), the residues of the exposed loop region (Leu 5–Lys 7), which is a major site for autoproteolysis, are shown in italic. Residues that participate in forming the terminal β-sheet are boxed.

Structural stability of mature protease lacking residues 1–4

Following our investigation of the precursor constructs containing additional sequences at the N-terminus, we decided to study N-terminal deletion constructs of the protease domain. This is particularly pertinent since it had been suggested that the four-stranded antiparallel β-sheet at the dimer interface of the protease was essential for protease structure and function[1]. In this sheet, the last two β-strands from each monomer form the inner strands and are sandwiched between the first two β-strands, which contain the N-terminal residues of the protein. Thus we prepared PR lacking its first four N-terminal residues (termed ΔPR) and tested whether ΔPR can form a tertiary structure competent in catalysis, even in the absence of the characteristic four-stranded β-sheet. In the absence of inhibitor, ΔPR is >90% unfolded compared to the folded dimer, as demonstrated by the ^{1}H-^{15}N correlation spectrum of this protein (Fig. 9*b*). This is consistent with an ~10-fold lower k_{cat} for the ΔPR-catalyzed hydrolysis and a k_{cat}/K_m ~50-fold lower compared to those for PR (Table 1). However, addition of substrate or inhibitor to ΔPR shifts the equilibrium from the unfolded state toward the stable folded dimer, as demonstrated by the ^{1}H-^{15}N correlation spectrum of the ΔPR–DMP323 complex (Fig. 9*c*, red contours). This clearly demonstrates that an active site capable of supporting a hydrolytic reaction can form in the absence of the interface interactions involving residues 1–4. This result also indirectly supports the intramolecular autocatalytic mechanism for the maturation reaction: it is indeed possible for the p6^{pol}-PR cleavage site sequence to fit into the active site with the scissile peptide bond (Phe-Pro) near the active site Asp residues without major reorganization of the PR structure[11]. The apparent K_d for the ΔPR dimer is >60-fold higher (~300 nM) than that of PR, and 50% denaturation of ΔPR occurs at 0.6 M urea (Fig. 3). Interestingly, these values of ΔPR are closer to those of the precursor than those of the mature PR, demonstrating that both deletion and addition of residues at the N-terminus results in a destabilization of the folded structure.

Plausible mechanism of regulation of the protease in the viral life cycle

The results of our studies using different model and native precursor proteins of the HIV-1 protease suggest that in HIV-1 and related viruses, which exhibit similar organization of the Gag-Pol precursor, the transframe region flanking the N-terminus of the protease may function as a negative regulator for protein folding and dimerization. The low dimer stability of the protease precursor relative to that of the mature enzyme is an ideal way of preventing the emergence of enzymatic functions until assembly of the viral particle is complete. Depending on the pH of the environment in which Gag-Pol maturation takes place, removal of the transframe region can occur in either two sequential steps or a single step. Intramolecular cleavage at the p6^{pol}-PR site to release a free N-terminus of PR is critical for the formation of a stable tertiary structure of the PR and enzymatic activity. Subsequent processing of the other Gag-Pol cleavage sites will occur rapidly through intermolecular processes. In

Table 1 Kinetic parameters for protease-catalyzed hydrolysis of substrate and inhibition constants for the hydrolytic reaction with the inhibitor[1]

Protease	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($mM\ s^{-1}$)	K_i (nM)
Wild type	3.74 ± 0.12	278 ± 20	13.5 ± 1.4	40 ± 2
PR	5.5 ± 0.23	177 ± 18	31.0 ± 4.4	89 ± 13
PR^{Q}	5.2 ± 0.16	225 ± 15	23.1 ± 2.3	33 ± 4
ΔPR	0.44 ± 0.04	660 ± 107	0.66 ± 0.17	N.D.[2]

[1]Assays were performed in 100 mM sodium acetate buffer, pH 5.0, 1 mM EDTA, 1 mM DTT at 25 °C. Substrate: Lys-Ala-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2 (ref. 11). RPB inhibitor: Arg-Val-Leu-(r)Phe-Glu-Ala-Nle-NH_2 (ref. 13).
[2]N.D., not determined.

contrast to TFR, the reverse transcriptase (RT) domain flanking the C-terminus of the protease domain does not influence the catalytic activity of the protease[13,28]. The cleavage at Leu 5/Trp 6 (ref. 18) within the protease domain could be viewed as a final step in the protease-associated cascade of events, destabilizing the tertiary structure and promoting dissociation of the protease dimer, thereby downregulating the catalytic activity of PR in the viral life cycle (Fig. 5).

Comparison of the activation mechanism of retroviral proteases to other proteases

It appears that for HIV-1 protease, and possibly other viral aspartic proteases, activation is tightly coupled to folding. The proregion in retroviral proteases most likely destabilizes the folded dimeric structure in the polyprotein by interacting with the nascent dimer interface (Fig. 5). Once cleavage has occurred, the dimer becomes more stable and activity ensues. This is in contrast to most zymogens and their corresponding mature enzymes, in which the catalytic machinery is stably preformed and activation is achieved by a conformational change involving peptide bond cleavage to remove parts of the polypeptide chain protruding into or obstructing access to the active site. The mechanism of cleavage to remove the proregion is strikingly similar in both cases[29–31]. Viewed within the greater context of zymogen activation, the HIV-1 protease may represent the most extreme case of activation by conformational rearrangement, namely the transition from an unstructured, inactive precursor protein to a stably folded, active mature enzyme. This is in contrast to the function of the proregion in α-lytic protease, which has been shown to enhance the folding rate of the mature protease and is thought to interact with the transition state[32].

Methods

Protease constructs. Three new HIV-1 protease constructs for expression, using pET11a vector in *E. coli* BL21(DE3) (Novagen, Inc., Madison, Wisconsin), were utilized in this study (Fig. 2). The precursor and precursor^{Q} constructs encompass the region of 156 amino acids from the N-terminus of the TFP to the C-terminus of the protease (Fig. 1; Genbank accession number HIVHXB2CG). The PR domain in all constructs bears the mutations Q7K, L33I, L63I, C67A and C95A, designed to restrict autoproteolysis and cysteine thiol oxidation. Precursor^{Q} includes an additional mutation R8Q in the PR domain. The third construct is designed to express a truncated version of the mature PR domain, from residues 5–99 (ΔPR). A Gly residue was added to the N-terminus of residue 5 of ΔPR to allow nearly complete excision of the N-terminal methionine when expressed in *E. coli*[33]. The nucleotide sequence of the cloned DNA was confirmed by sequencing.

Isotopic labeling and purification. Cells were grown either in Luria–Bertani medium or in medium for ^{15}N labeling and induced for expression[20,27]. To purify the precursors, mature proteases and ΔPR, the insoluble fraction after 2 M urea treatment (Fig. 2, lane 3) was solubilized in 7.5 M guanidine hydrochloride and subjected to size exclusion chromatography under denaturing conditions[20,34]. The peak fractions were further subjected to reverse-phase high-performance liquid chromatography (RP-HPLC)[20,34], and analyzed for their mass and N-terminal sequence.

Protein refolding and enzyme assays. A portion of the purified protein after RP-HPLC was dialyzed exhaustively against 25 mM formic acid, pH 2.8. Under these conditions the protein is unfolded and can be stored at 4 °C for several weeks. A 2.5 μl aliquot of protein was then diluted with 47.5 μl of 5 mM acetate, pH 6.0, to attain a final pH of 4.0. Subsequently, 50 μl of 100 mM formate, acetate or phosphate buffers at the appropriate pH were added and the reaction mixture incubated at 25 °C. The time course of the maturation reaction was followed by assaying an aliquot of the reaction mixture for enzymatic activity by using the spectrophotometric assay or by SDS–PAGE followed by staining and densitometry[11,13]. Kinetic parameters were measured using the substrate, Lys-Ala-Arg-Val-Nle-(4-nitrophenylalanine)-Glu-Ala-Nle-NH_2 (California Peptide Research, Napa, California) and the RPB inhibitor, Arg-Val-Leu-(r)Phe-Glu-Ala-Nle-NH_2 (r and Nle denote reduced peptide bond and norleucine, respectively; Bachem Bioscience, King of Prussia, Pennsylvania) in 100 mM sodium acetate, pH 5.0, 1 mM dithiothreitol (DTT), 1 mM EDTA (buffer A) as described[20]. Enzyme concentrations were determined both by active site titration and Bio-Rad protein assay (Bio-Rad Laboratories, Hercules, California) to ensure a renaturation efficiency of >95% (ref. 20). Assays to determine the apparent K_d were performed in buffer A at 25 °C (ref. 20).

Preparation of proteins for NMR studies. Protein (9–10 mg) in 25 mM formic acid was mixed with four volumes of 5 mM acetate buffer, pH 6.0, which contains approximately a five-fold molar excess of a cyclic urea inhibitor (DMP323; ref. 24), and dialyzed twice against an 80-fold excess of 50 mM acetate buffer, pH 5.2. The protein solution was concentrated to yield a final concentration of 0.7–1 mM using 10 kDa cutoff membranes (Amicon, Inc., Beverly, Massachusetts).

NMR. ^{1}H-^{15}N HSQC spectra[35] were recorded at 25 °C on a Bruker DMX600 spectrometer equipped with a triple-resonance x,y,z-shielded gradient probe.

Acknowledgments

We wish to thank J. Hung for technical assistance, N.T. Nashed for discussions, L.K. Pannell for mass spectroscopic analyses, and H.R. Parikh for contract services. DMP323 was a generous gift from N. Hodge, DuPont Merck Pharmaceutical Company. This research was supported by the Intramural AIDS Targeted Program of the Office of the Director of the National Institutes of Health.

Received 4 February, 1999; accepted 19 May, 1999.

1. Wlodawer, A. & Erickson, J. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585 (1993).
2. Oroszlan, S. & Luftig, R.B. Retroviral proteases. *Curr. Top. Microbiol. Immunol.* **157**, 153–185 (1990).
3. Candotti, D. *et al.* High variability of the gag/pol transframe region among HIV-1 isolates. *C. R. Acad. Sci.* **317**, 183–189 (1994).
4. Vogt, V.M. Proteolytic processing and particle maturation. *Curr. Top. Microbiol. Immunol.* **214**, 95–131 (1996).
5. 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).
6. Condra, J.H. *et al.* In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374**, 569–571 (1995).
7. Louis, J.M., Dyda, F., Nashed, N.T., Kimmel, A.R. & Davies, D.R. Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. *Biochemistry* **37**, 2105–2110 (1998).
8. Beissinger, M. *et al.* Sequence-specific resonance assignments of the 1H-NMR spectra and structural characterization in solution of the HIV-1 transframe protein p6. *Eur. J. Biochem.* **237**, 383–392 (1996).
9. Partin, K. *et al.* Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **88**, 4776–4780 (1991).
10. Zybarth, G. & Carter, C. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. *J. Virol.* **69**, 3878–3884 (1995).
11. 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).
12. 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).
13. 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).
14. Louis, J.M. *et al.* Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in *Escherichia coli*. *Eur. J. Biochem.* **199**, 361–369 (1991).
15. Louis, J.M., Oroszlan, S. & Mora, P.T. Studies of the autoprocessing of the HIV-1 protease using cleavage site mutants. *Adv. Exp. Med. Biol.* **306**, 499–502 (1991).
16. Zybarth, G., Krausslich, H.G., Partin, K. & Carter, C. Proteolytic activity of novel human immunodeficiency virus type 1 proteinase proteins from a precursor with a blocking mutation at the N terminus of the PR domain. *J. Virol.* **68**, 240–250 (1994).
17. Tessmer, U. & Krausslich, 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. Rose, J.R., Salto, R. & Craik, C.S. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J. Biol. Chem.* **268**, 11939–11945 (1993).
19. Mildner, A.M. *et al.* The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry* **33**, 9405–9413 (1994).
20. 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).
21. Lindhofer, H., von der Helm, K. & Nitschko, H. In vivo processing of Pr160gag-pol from human immunodeficiency virus type 1 (HIV) in acutely infected, cultured human T- lymphocytes. *Virology* **214**, 624–627 (1995).
22. Ho, D.D. *et al.* Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *J. Virol.* **68**, 2016–2020 (1994).
23. Weber, I.T. Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases. *J. Biol. Chem.* **265**, 10492–10496 (1990).
24. Lam, P.Y. *et al.* Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380–384 (1994).
25. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. & Sykes, B.D. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**, 67–81 (1995).
26. Yamazaki, T. *et al.* Secondary structure and signal assignments of human-immunodeficiency-virus-1 protease complexed to a novel, structure-based inhibitor. *Eur. J. Biochem.* **219**, 707–712 (1994).
27. Yamazaki, T. *et al.* Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. *Protein Sci.* **5**, 495–506 (1996).
28. 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).
29. al-Janabi, J., Hartsuck, J.A. & Tang, J. Kinetics and mechanism of pepsinogen activation *J. Biol. Chem.* **247**, 4628–4632 (1972).
30. Lin, X.L. *et al.* Enzymic activities of two-chain pepsinogen, two-chain pepsin, and the amino-terminal lobe of pepsinogen. *J. Biol. Chem.* **267**, 17257–17263 (1992).
31. Khan, A.R. & James, M.N. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**, 815–836 (1998).
32. Peters, R.J. *et al.* Pro region C-terminus: protease active site interactions are critical in catalyzing the folding of alpha-lytic protease. *Biochemistry* **37**, 12058–12067 (1998).
33. Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA* **86**, 8247–8251 (1989).
34. Weber, I.T. *et al.* Crystallographic analysis of human immunodeficiency virus 1 protease with an analog of the conserved CA-p2 substrate—interactions with frequently occurring glutamic acid residue at P2′ position of substrates. *Eur. J Biochem.* **249**, 523–300 (1997).
35. Bax, A. & Ikura, M. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly $^{15}N/^{13}C$ enriched proteins. *J. Biomol. NMR* **1**, 99–104 (1991).
36. Wondrak, E.M., Louis, J.M., de Rocquigny, H., Chermann, J.C. & Roques, B.P. The gag precursor contains a specific HIV-1 protease cleavage site between the NC (P7) and P1 proteins. *FEBS Lett.* **333**, 21–24 (1993).
37. Leis, J. *et al.* Standardized and simplified nomenclature for proteins common to all retroviruses. *J.Virol.* **62**, 1808–1809 (1988).