THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 276, No. 52, Issue of December 28, pp. 49110–49116, 2001
Printed in U.S.A.

Folded Monomer of HIV-1 Protease*

Received for publication, August 23, 2001, and in revised form, October 10, 2001
Published, JBC Papers in Press, October 11, 2001, DOI 10.1074/jbc.M108136200

Rieko Ishima‡, Rodolfo Ghirlando§, József Tözsér¶, Angela M. Gronenborn‖, Dennis A. Torchia‡, and John M. Louis‖**

From the ‡Molecular Structural Biology Unit, NIDCR, and the Laboratories of ‖Chemical Physics and §Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-4307 and the ¶Department of Biochemistry and Molecular Biology, Faculty of Medicine, Debrecen University, Debrecen H-4012, Hungary

The mature human immunodeficiency virus type 1 protease rapidly folds into an enzymatically active stable dimer, exhibiting an intricate interplay between structure formation and dimerization. We now show by NMR and sedimentation equilibrium studies that a mutant protease containing the R87K substitution (PR_{R87K}) within the highly conserved Gly^{86}-Arg^{87}-Asn^{88} sequence forms a monomer with a fold similar to a single subunit of the dimer. However, binding of the inhibitor DMP323 to PR_{R87K} produces a stable dimer complex. Based on the crystal structure and our NMR results, we postulate that loss of specific interactions involving the side chain of Arg^{87} destabilizes PR_{R87K} by perturbing the inner C-terminal β-sheet (residues 96–99 from each monomer), a region that is sandwiched between the two β-strands formed by the N-terminal residues (residues 1–4) in the mature protease. We systematically examined the folding, dimerization, and catalytic activities of mutant proteases comprising deletions of either one of the terminal regions (residues 1–4 or 96–99) or both. Although both N- and C-terminal β-strands were found to contribute to dimer stability, our results indicate that the inner C-terminal strands are absolutely essential for dimer formation. Knowledge of the monomer fold and regions critical for dimerization may aid in the rational design of novel inhibitors of the protease to overcome the problem of drug resistance.

Human immunodeficiency virus type 1 (HIV-1)[1] is a 99-residue aspartic acid protease and is active as a homodimer. The interface of the free protease dimer is stabilized through interactions between the two subunits at the active site and at the termini (1, 2). The protease catalyzes its own release from the Gag-Pol polyprotein in addition to the maturation of the virally encoded structural proteins and replication enzymes required for the assembly and production of viable virions (3). Thus, the protease has served as one of the primary targets for the development of drugs against AIDS. Structure-based design of drugs targeted against the wild-type mature protease has aided in the development of several potent inhibitors that are specific for binding to the active site (4). Although several of these drugs are in clinical use and have curtailed the progression of the disease, the effectiveness of long-term treatment has been restricted due to naturally selected protease variants exhibiting lower affinity for the drugs than the wild-type enzyme. Various drug-resistant mutants of the protease have been identified (5). Different resistance mechanisms based on the observed structural changes in drug-resistant mutants have been proposed. In general, the mutants modulate structure and interactions within the active site as well as inter- and intrasubunit flexibility (6–8).

Because of the difficulty in designing active-site inhibitors that avoid the problem of drug resistance, the target region for drug development has been extended to areas in which no drug-selected mutations have occurred to date (9, 10). These regions are generally well conserved and critical to the structure and function of the protease. Sequence alignment of retroviral proteases shows that the most conserved regions correspond to (i) the active site (amino acids 22–34), (ii) the flap (amino acids 47–52), and (iii) amino acids 84–94 encompassing a single α-helix (11). Although the active-site triad Asp^{25}-Thr^{26}-Gly^{27} is common to all aspartic acid proteases, Gly^{86}-Arg^{87}-(Asn/Asp^{88}) in the α-helix is unique to retroviral proteases (12), and its structural significance is not fully understood. We showed in earlier studies that a conservative substitution of Arg^{87} with Lys results in total loss of proteolytic activity, although binding to pepstatin A, a well known inhibitor of aspartic acid proteases, is not curtailed (13). However, detailed structural investigations of this mutant were not feasible earlier due to lack of an expression system for producing large amounts of this protein and limited knowledge of optimized conditions for studies by solution NMR.

Although NMR solution studies of free wild-type PR are not possible due to rapid self-degradation, they have become feasible using a stabilized protease variant that carries the following five mutations, Q7K, L33I, and L63I to restrict self-degradation and C65A and C95A to avoid cysteine-thiol oxidation (14). This protease variant (termed PR herein) exhibits essentially the same kinetic parameters and stability as the wild-type protein (15).

To investigate the importance of Arg^{87} for protease structure and function, we introduced the R87K mutation into PR (termed PR_{R87K}). NMR and sedimentation equilibrium experiments show that PR_{R87K} forms a semi-stable folded monomer in solution. Our results further suggest that the R87K mutation exerts an effect on dimer formation of the protease, interfering with the correct positioning of the C-terminal β-strands (residues 96–99) at the interface. The contribution of the terminal β-sheet to folding and dimer formation was systematically analyzed using proteins that lack either the N-terminal (PR-(5–99)) or C-terminal (PR-(1–95)) 4 residues or both (PR-

* This work was supported by the Intramural AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed: Laboratory of Chemical Physics, Rm. 411, Bldg. 5, NIDDK, NIH, Bethesda, MD 20892. Tel.: 301-594-3122; Fax: 301-480-4001; E-mail: jmlouis@helix.nih.gov.

[1] The abbreviations used are: HIV-1, human immunodeficiency virus type 1; PR, HIV-1 protease; Nle, norleucine; HSQC, heteronuclear single quantum coherence spectroscopy; NOE, nuclear Overhauser effect.

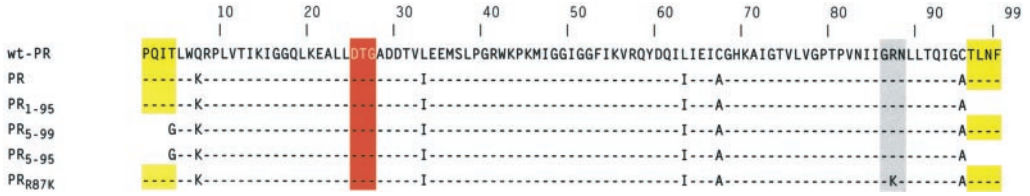

FIG. 1. **Amino acid sequences of mature HIV-1 protease constructs.** *Red* and *gray* indicate the two highly conserved regions, the active-site (DTG) and C-terminal (GR(N/D)) triads, in retroviral proteases. All of the constructs bear five mutations: three mutations (Q7K, L33I, and L63I) to restrict degradation and two mutations (C65A and C95A) to avoid Cys-thiol oxidation. The N- and C-terminal residues involved in forming the interface β-sheet (see Fig. 4*b*) are highlighted in *yellow*. *wt-PR*, wild-type PR.

(5–95)). The enzymatic activities of these mutant proteins are reported and compared with those of PR and wild-type PR.

EXPERIMENTAL PROCEDURES

Sample Preparation—The R87K mutation and a stop codon were introduced into the PR template (which encodes the five mutations Q7K, L33I, L63I, C67A, and C95A) (15) to generate the PR_{R87K} and PR-(1–95) constructs, respectively, using the QuickChange mutagenesis protocol (Stratagene, La Jolla, CA). The PR-(5–99) construct (15) was used to generate the PR-(5–95) clone using the same mutagenesis protocol and the primers used for creating PR-(1–95). The mutant proteases used in the study are listed in Fig. 1.

PR and its derived mutants were expressed in *Escherichia coli* BL21(DE3) in LB or minimal medium containing [^{15}N]ammonium chloride and/or [^{13}C]glucose as the sole nitrogen and carbon sources at 37 °C (16). All proteins were isolated using a common protocol, which comprises isolation of inclusion bodies followed by fractionation of the protease by size-exclusion and reversed-phase high-pressure liquid chromatography. Proteins (2–3 mg) at a concentration of ≤0.33 mg/ml in 35% acetonitrile/water containing 0.05% trifluoroacetic acid were folded by dialysis against 2 liters of 30 mM formic acid (pH 2.8) for 1.5 h, followed by dilution of the protein with a 5-fold excess of 10 mM acetate buffer (pH 6.0) to shift the pH to 4, either with or without a saturating amount of DMP323 (K_i < 10 nM) (17), and dialysis against 4 liters of 20 mM sodium phosphate (pH 5.8) for another 1.5 h. Proteins were concentrated to achieve a final concentration of ~0.4 mM for NMR analysis and to a concentration of ~0.05 mM for sedimentation equilibrium studies and enzyme kinetics. Concentrations of PR and its mutants are expressed for a dimer, unless otherwise noted. Although the amount of DMP323 added is equivalent to a 5 mM final concentration, the actual concentration is unknown due to the poor solubility of DMP323 in aqueous buffers.

Protease Assays—Protease assays were initiated by mixing 5 μl of freshly folded PR_{R87K} (59 μM) or PR-(1–95) (52 μM), 10 μl of 2× incubation buffer (200 mM potassium phosphate buffer (pH 5.6) and 0.2 M (buffer A) or 2 M (buffer B) NaCl), and 5 μl of 0.2 mM substrate. PR-(5–99) was assayed in buffer B at a final enzyme concentration of 1 μM. The reaction mixture was incubated at 37 °C for 1 h, and the reaction was terminated by the addition of guanidine HCl to a final concentration of 6 M. The solution was acidified with trifluoroacetic acid, and an aliquot was injected onto a Nova-Pak C_{18} reversed-phase chromatography column (3.9 × 150 mm; Waters Associates, Inc., Milford, MD) using an automatic injector. Substrates and the cleavage products were separated using a water/acetonitrile gradient (0–100%) in the presence of 0.05% trifluoroacetic acid. PR-(5–99) was assayed using the substrates listed in Table I. Substrates corresponding to the cleavage sites CA/p2, $p6^{pol}$/PR, PR/RT, RT/IN (for the oligopeptide sequences, see Table I), MA/CA (VSQNY↓PIVQ), and p2/NC (TATIM↓MQRG) were used to assay PR_{R87K} and PR-(1–95).

The mutant enzymes PR_{R87K} and PR-(1–95) were also assayed using a spectrophotometric substrate, Lys-Ala-Arg-Val-Nle-(4-nitrophenylalanine)-Glu-Ala-Nle-NH_2, in 100 mM sodium acetate (pH 5.0) at 25 °C in a final concentration of 0.96 μM PR-(1–95) or 2.9 μM PR_{R87K} and 460 μM substrate as described previously (15). The kinetic parameters for PR-(5–99)-catalyzed hydrolysis of the identical spectrophotometric substrate were reported previously (15).

Sedimentation Equilibrium Analyses—Sedimentation equilibrium experiments were carried out at 20 °C using four different rotor speeds (10,000, 12,000, 14,000, and 30,000 rpm) on a Beckman Optima XL-A analytical ultracentrifuge. Data were acquired as an average of eight absorbance measurements at a nominal wavelength of 280 nm and a radial spacing of 0.001 cm. Equilibrium was achieved within 24 h. Protein samples (PR, PR-(1–95), PR-(5–99), and PR_{R87K}), in the absence or presence of an ~10-fold excess of the potent inhibitor DMP323 over the enzyme concentration, were prepared in 20 mM sodium phosphate (pH 5.8) and loaded into the ultracentrifuge cells at nominal loading concentrations of 0.80 A_{280} units.

Data were analyzed in terms of a single ideal solute by fitting to the following equation: $A_r = A_{o,r}\cdot\exp(HM(r^2 - r^2_o)) + E$, where $A_{o,r}$ is the absorbance of the solute at a reference radius (r_o) and A_r is the absorbance at a given radial position (r). H represents $\omega^2/2RT$, where ω is the angular speed in radians/s, R is the gas constant, and T is the absolute temperature. The value of M represents the experimentally returned value of the buoyant molecular mass, $M(1 - v\rho)$. E is a small base-line correction determined experimentally at 30,000 rpm.

Values for the experimental molecular mass were determined using tabulated values for the density (ρ), whereas the partial specific volume (v) was calculated based on the amino acid composition using the consensus data for the partial specific molar volumes of amino acids published by Perkins (18). This value was corrected to account for the isotopic composition of the proteins studied. For PR_{R87K}, PR-(1–95), and PR-(5–99), uniformly ^{15}N-labeled proteins were used; and for PR, 10% ^{13}C-labeled and uniformly ^{15}N-labeled proteins were used.

NMR Experiments—NMR experiments were carried out using protein concentrations of 0.4–0.5 mM in 20 mM phosphate buffer in 95% H_2O and 5% D_2O and a sample volume of ~280 μl in a 5-mm Shigemi tube (Shigemi, Inc., Allison Park, PA). NMR spectra were acquired on a DMX500 spectrometer (Bruker Instruments, Billerica, MA). All experiments were conducted at 20 °C unless noted otherwise. Data were processed and analyzed using nmrPipe, nmrDraw, and PIPP software (19, 20).

^{1}H-^{15}N HSQC spectra of PR_{R87K}, PR-(1–95), and PR-(5–99) were recorded at pH 5.8 with and without DMP323. Backbone signal assignments of PR and PR_{R87K}, in the presence and absence of DMP323, were made using HNCA and CBCA(CO)NH experiments (21, 22). Three-dimensional experiments for the assignment of free PR_{R87K} were carried out at pH 4.5 to slow aggregation of the sample. ^{15}N longitudinal relaxation times (T_1), ^{15}N transverse relaxation times (T_2), and ^{15}N-{H} NOEs for free PR_{R87K} were measured in an interleaved manner (23) with repetition delays of 2 s for T_1 and T_2 determination and 3 s for NOE determination. Relaxation delays were 0.0016, 0.16, 0.32, 0.48, 0.64, 0.8, and 0.96 s for T_1 measurements and 6, 12, 24, 42, 60, 84, and 96 ms for T_2 measurements. Diffusion experiments were carried out in D_2O phosphate buffer using bipolar gradient pulses in a longitudinal-eddy-current delay pulse sequence (24) for PR_{R87K}, with and without DMP323, and for ubiquitin.

RESULTS AND DISCUSSION

Structural Features of PR_{R87K} Derived from Chemical Shifts—PR folds from a denatured state into an enzymatically active stable dimer similar to wild-type PR when dialyzed from pH 2.8 to 4.2–5.8. The ^{1}H-^{15}N correlation spectrum recorded on a freshly prepared sample of PR_{R87K}, folded under identical conditions as PR, displayed a set of well dispersed signals, indicating a folded conformation of the protein (Fig. 2*a*). However, in contrast with PR, PR_{R87K} aggregated at a concentration of 0.4 mM, resulting in an ~50% loss in signal intensity within 1 day. The addition of the potent inhibitor DMP323 to PR_{R87K} resulted in two sets of signals (Fig. 2*b*), with the minor set belonging to free PR_{R87K}. The C-α chemical shifts of the major set for PR_{R87K} in the presence of DMP323 (Fig. 3*b*) were nearly identical to those of the PR-DMP323 complex, suggesting that both dimer complexes exhibit very similar structures.

Even in the absence of DMP323, the overall backbone C-α chemical shifts of PR_{R87K} residues were similar to those of PR

TABLE I

Kinetic parameters for PR and PR-(5–99) protease-catalyzed hydrolysis of Gag-Pol substrates

Assays were performed in 250 mM sodium phosphate (pH 5.6) and 2 M NaCl at 37 °C in a final enzyme concentration of 0.01–0.1 μM PR and 1 μM PR-(5–99). ND, not determined. MA, CA, NC, RT, and IN denote matrix, capsid, nucleocapsid, reverse transcriptase, and integrase proteins, respectively.

Cleavage site	Substrate	Enzyme	K_m	k_{cat}	k_{cat}/K_m
			mM	s^{-1}	$mM^{-1}\ s^{-1}$
		PR	0.021 ± 0.004	0.89 ± 0.03	42.4
CA/p2	KARVL ↓ AEAMS	PR-(5–99)	0.192 ± 0.02	0.11 ± 0.02	0.6
		PR-(5–99)	<0.01	0.09	ND
$p6^{pol}$/PR	VSFNF ↓ PQITL	PR-(5–99)	<0.02	0.002	ND
		PR	0.044 ± 0.008	1.72 ± 0.08	39.1
PR/RT	CTLNF ↓ PISP	PR-(5–99)	0.087 ± 0.014	0.05 ± 0.01	0.6
		PR	0.032 ± 0.006	0.41 ± 0.02	12.8
In RT	AETF ↓ YVDGAA	PR-(5–99)	0.041 ± 0.01	0.008 ± 0.001	0.2
		PR	0.016 ± 0.005	1.20 ± 0.06	75.0
RT/IN	IRKIL ↓ FLDG	PR-(5–99)	0.046 ± 0.006	0.08 ± 0.01	1.7

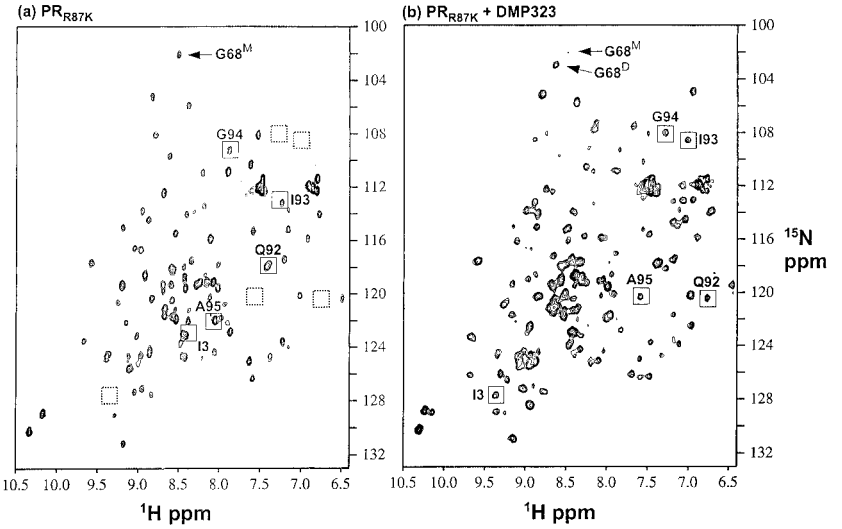

FIG. 2. **Amide ^{1}H-^{15}N HSQC spectra of freshly prepared PR$_{R87K}$ (*a*) and PR$_{R87K}$ in the presence of DMP323 (*b*) measured at 20 °C.** *Boxes* in *a* and *b* delineate the locations of peaks that exhibited significant changes due to dimer formation. Peaks of Gly^{68} for the dimer and monomer forms of PR$_{R87K}$ are labeled *G68^D* and *G68^M*, respectively.

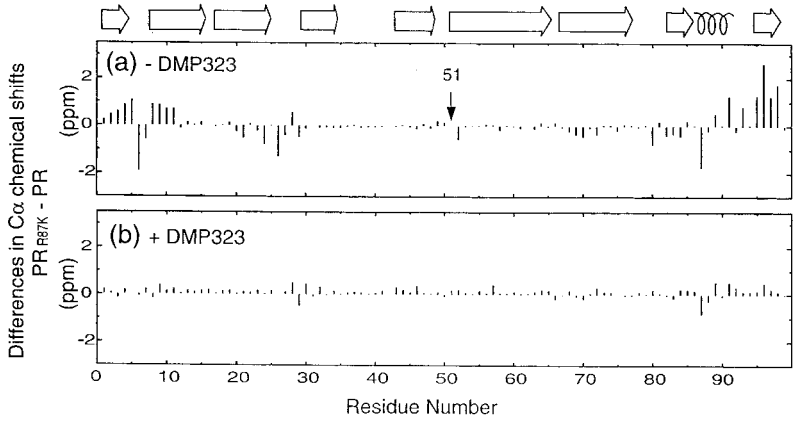

FIG. 3. **Differences in backbone C-α chemical shifts between PR and PR$_{R87K}$ in the absence (*a*) and presence (*b*) of DMP323 together with the secondary structure of PR.** The single α-helix and the β-strands are depicted as a *coil* and *white arrows*, respectively. Residue 51, whose C-α was not assigned due to broadening of the signal in PR$_{R87K}$, is indicated.

(Fig. 3*a*). However, unlike the DMP323-bound forms of PR and PR$_{R87K}$, significant differences in the chemical shifts of the free forms of PR and PR$_{R87K}$ were noted for residues at the dimer interface, *i.e.* near the active site (residues 24–29) and the N- and C-terminal regions (residues 1–10 and 90–99, respectively). In particular, the peaks for Ile^3, Gln^{92}, Ile^{93}, Gly^{94}, and Ala^{95} that significantly shifted in the dimer due to intermonomer and not DMP323 interaction were not observed in the corresponding positions in the free PR$_{R87K}$ spectrum (Figs. 2*a* and 3*a*). Since Arg^{87} is located far from the N- and C-terminal residues (>15 Å), the observed chemical shift differences between PR$_{R87K}$ and PR corresponding to the terminal residues were not caused by changes in the local magnetic environment upon mutation, but rather by the structural changes of the protein. Overall, the similarity in chemical shifts for uncomplexed free PR$_{R87K}$ and PR, except for the dimer interface regions, suggests that the fold of PR$_{R87K}$ is similar to that of PR.

PR$_{R87K}$ Forms a Semi-stable Monomer—The dynamics of PR$_{R87K}$ in the absence of DMP323 were probed by carrying out ^{15}N relaxation experiments (see Fig. 5). For the most part, the T_1 and T_2 values of PR$_{R87K}$ were quite uniform (0.5–0.6 s and 90–100 ms, respectively), although different from those of PR (0.7–0.9 s and 60–70 ms, respectively).[2] The rotational correlation time estimated from the T_1/T_2 ratios was ~7.5 ns. This

[2] Freedburg, D. I., Ishima, R., Jocob, J., Wang, Y.-X., Kustanovich, I., Louis, J. M., and Torchia, D. A. (2001) *Protein Sci.,* in press.

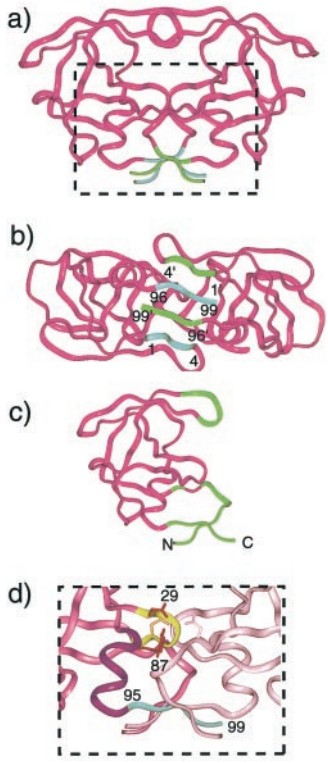

FIG. 4. **Ribbon representations of the structure of mature HIV-1 protease (*a*), the terminal β-sheet interface (*b*), regions that exhibit pico/nanosecond time scale motions (NOE < 0.5) in the PR_{R87K} monomer (shown in *green*) (*c*), and details of the structure surrounding Arg^{87} (*d*).** In *a* and *b*, the terminal β-strands (residues 1–4 and 96–99) of one subunit are colored in *green* and the other *cyan*. In *d*, the side chains of Asp^{29} and Arg^{87} are depicted in *red*, and the side chains of the active-site Asp^{25} residues are colored in the same color as the corresponding subunit backbone. The region (residues 87–95) encompassing the α-helix is shown in *purple*, and the adjoining C-terminal β-strand is shown in *cyan* in *d*. The structures were drawn using the coordinates of Protein Data Bank code 1A8K (www.rcsv.org).

correlation time is significantly smaller than that observed for PR (12.9 ns) under similar conditions (0.3 mM protein in 20 mM sodium phosphate (pH 5.8) at 20 °C) (14) despite the noted tendency of PR_{R87K} to aggregate. Using an NMR approach (24), we determined the relative translational diffusion coefficient of PR_{R87K} to be 1.5×10^{-10} m^2/s. This value lies between those measured for the PR-DMP323 complex (1.0×10^{-10} m^2/s) and ubiquitin (2.1×10^{-10} m^2/s), proteins with molecular masses of 22 and 8.5 kDa, respectively. This suggests that PR_{R87K} has a mass close to that expected for the protease monomer (Fig. 4*c*). Sedimentation equilibrium data for the PR_{R87K} mutant confirmed the monomeric state. The data fit well to a single solute model with a molecular mass equal to that of the monomer (Table II). All of the above results together suggest that PR_{R87K} is a folded monomer in solution. The tendency for the PR_{R87K} monomer to aggregate at a concentration of ~0.5 mM suggests that it is less stable, unlike dimeric PR, which exhibits a stable fold and does not aggregate even at 1 mM.

Analysis of the ^{15}N relaxation experiments revealed significant motions on the subnanosecond time scale for both N- and C-terminal residues 2–10 and 93–99, respectively; the elbow residues 37–42, and the flap residues 48–53 in PR_{R87K} as shown by the elevated T_1 and T_2 values and reduced NOE values (Fig. 5). Although the mobilities of the flap and elbow regions have been shown for free PR, increased motion of the termini had not been observed (25, 26). Inspection of a model of monomeric protease (Fig. 4*c*) revealed that the areas identified as mobile correspond to solvent-exposed regions, assuming that the monomeric structure remains essentially identical to a single subunit of the dimer.

Analysis of protease crystal structures revealed that the N-terminal (residues 1–4) and C-terminal (residues 96–99) residues contribute ~50% of the dimer interface, critical for PR stability (1). The four-stranded antiparallel β-sheet is organized such that the two C-terminal β-strands are sandwiched between the two N-terminal β-strands at the bottom of the protease structure (Fig. 4*b*). The highly conserved Arg^{87} residue, residing on the sole α-helix (residues 87–91), forms a hydrogen bond with Asp^{29}, a residue that is located near the active site (Fig. 4*d*) (27). In PR_{R87K}, chemical shift perturbations and subnanosecond motions detected for the N-terminal residues of the α-helix increased further in the C-terminal β-strand. Taking into account the above NMR results together with information from crystal structures, we propose that disruption of the specific interactions involving the Arg^{87} side chain could destabilize the terminal β-sheet and induce enhanced mobility at the dimer interface.

Inhibitor Binding Stabilizes the PR_{R87K} Dimer—As shown above, inhibitor DMP323 binding could induce dimerization of PR_{R87K}, although a small fraction of the PR_{R87K} monomer was still present as evidenced by the HSQC spectrum (Fig. 2*b*). Since DMP323 has very poor solubility in aqueous buffers (see "Experimental Procedures"), complete saturation of PR_{R87K} with inhibitor could not be achieved. The above result is consistent with the monomer/dimer-DMP323 complex ratio determined by sedimentation equilibrium studies (Table II). Under the same conditions of inhibitor saturation, we never observed free PR, indicating that the affinity of DMP323 for PR has to be significantly higher. It has been shown that the highly hydrophobic DMP323 interacts tightly with the methyl groups of aliphatic residues that line the substrate-binding pocket (28). This binding site is composed of residues residing in both monomeric subunits; thus, binding of DMP323 would favor the dimer over the monomer structure, *i.e.* binding induces dimerization. Indeed, DMP323 could be viewed as a staple keeping the monomers together, the molecular interactions alone of which are not sufficient for this purpose, as in the case of the PR_{R87K} mutant. However, substrates are not expected to stabilize the PR_{R87K} dimer to the same extent as DMP323, as the binding affinity of substrates is much lower than that of DMP323. This interpretation is consistent with our observation that a substrate analog inhibitor of PR (Arg-Val-Leu-®-Phe-Glu-Ala-Nle-NH_2, where ® denotes a reduced peptide bond; K_i = 89 nM) (15) and substrate (Lys-Ala-Arg-Val-Nle-(4-nitrophenylalanine)-Glu-Ala-Nle-NH_2; K_m = 177 μM) (15) give spectra similar to those shown in Fig. 2*a* (data not shown), implying that substrates do not stabilize the dimeric structure of PR_{R87K}.

Loss of the Catalytic Activity of PR_{R87K}—The fact that PR_{R87K} is a monomer is consistent with the observed loss of its catalytic activity. PR_{R87K} was assayed against substrates representing cleavage sites between the major Gag and Gag-Pol domains. No sign of cleavage was observed using PR_{R87K}, except for the MA/CA and $p6^{pol}$/PR substrates, for which very low levels of cleavage were observed. Based on the specific activity values, PR_{R87K} was calculated to have 6700- and 1800-fold lower activity than PR in buffers A and B, respectively (Table III). PR_{R87K} exhibited a 4600-fold lower specific activity compared with PR as measured by the spectrophotometric assay.

Influence of Terminal Deletions on Folding, Dimerization, and Catalytic Activity—The role of the interfacial β-sheet comprising the N- and C-terminal strands in protease structure and function was examined using deletions mutants of either one of the terminal β-strands (residues 1–4 or 96–99) or both.

TABLE II
Estimated molecular masses and major folded species of HIV-1 protease constructs

Construct[a]	Inhibitor DMP323[b]	Molecular masses estimated by sedimentation equilibrium analysis			Major folded species[c]
		$M_{experimental}$	$M_{calc(monomer)}$	$M_{experimental}/M_{calc(monomer)}$	
		g/mol	*g/mol*		
PR	−	20,370 ± 240	10,906.6	1.87 ± 0.02	Dimer
	+	21,980 ± 560		1.96 ± 0.01	Dimer-DMP323 complex
PR_{R87K}	−	10,620 ± 530	10,827.9	0.98 ± 0.05	Monomer[d]
	+	17,420 ± 940[e]		1.57 ± 0.09	Monomer[d] + dimer-DMP323 complex
PR-(1–95)	−	10,440 ± 370	10,377.4	1.01 ± 0.04	Monomer[d]
	+	11,790 ± 1150[e]		1.13 ± 0.11	Monomer[d]
PR-(5–99)	−	10,860 ± 120	10,471.5	1.04 ± 0.01	Monomer[d] + dimer
	+	20,080 ± 690		1.87 ± 0.06	Dimer-DMP323 complex

[a] PR-(5–95) precipitated completely with no observable NMR signal.
[b] Calculated molecular mass of DMP323 = 567 Da.
[c] Data were determined by NMR and sedimentation equilibrium studies.
[d] These samples have a tendency to aggregate at 0.4 mM (as a dimer).
[e] Data were forced to fit to a single exponential decay function. To prevent underestimate of the errors, the largest fitting errors among the data measured at various rotor speeds are shown for these two cases. Otherwise, errors are standard deviations of the data measured at various rotor speeds and are compatible with or slightly larger than the fitting error.

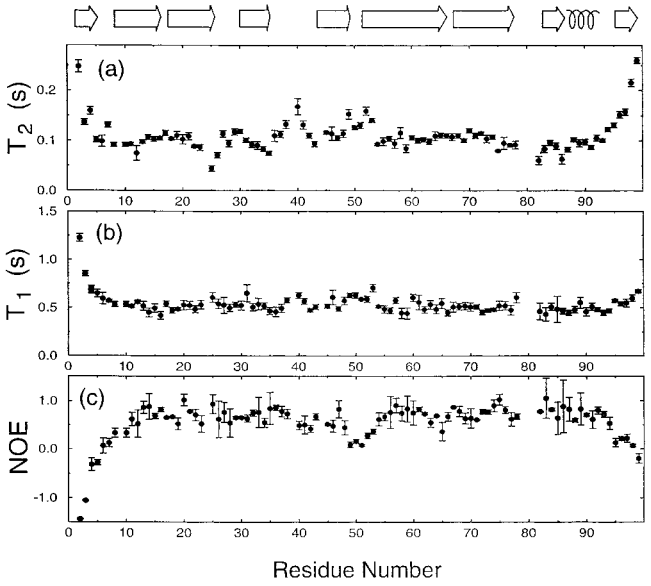

FIG. 5. **^{15}N T_2 (*a*), ^{15}N T_1 (*b*), and $^{15}N\{H\}$ NOE (*c*) values of backbone amides of PR_{R87K}.** The secondary structure of PR is depicted at the top.

TABLE III
Comparison of specific activities of PR and PR_{R87K}

Assays were performed in 100 mM potassium phosphate (pH 5.6) either in 0.1 M (buffer A) or 1 M (buffer B) NaCl at a final concentration of 200 μM $p6^{pol}$/PR substrate and 0.02 μM PR or 15 μM PR_{R87K}.

Enzyme	Specific activity	
	Buffer A	Buffer B
	nmol products s^{-1} $nmol^{-1}$ enzyme	
PR	0.06	0.09
PR_{R87K}	0.9×10^{-5}	5×10^{-5}

The truncated proteins PR-(1–95), PR-(5–99), and PR-(5–95) (Fig. 2) were purified and refolded essentially as described for PR and PR_{R87K} (see "Experimental Procedures"). At low concentration (~50 μM), PR-(1–95) and PR-(5–99) were soluble in the pH range of 4–5.8, similar to PR and PR_{R87K}. However, at higher concentrations, in either the absence or presence of DMP323, PR-(5–95) displayed severe aggregation, precluding any structural analysis by NMR. Apparently at least one of the terminal β-strands is required to maintain structural stability and reasonable solubility of the protease.

Although the three mutations Q7K, L33I, and L63I in the protease reduce self-degradation at pH 5.0, cleavage still occurs, albeit at a much slower rate compared with wild-type PR, permitting NMR studies (26, 29). Degradation of PR can be further reduced by conducting the experiments at a slightly higher pH and lower temperature, such as 20 mM sodium phosphate (pH 5.8) and 20 °C (14). The truncated proteins PR-(1–95) and PR-(5–99) behaved differently under these conditions. PR-(1–95) did not show any degradation products (similar to PR_{R87K}), whereas PR-(5–99) clearly was proteolyzed, as evidenced by the appearance of additional resonances in the ^1H-^{15}N HSQC spectrum (Fig. 6*a*, *dotted enclosed area*). The degradation products of PR-(5–99) most likely arise from cleavage at the dipeptide sequences Ile^{33}-Glu^{34} and Ile^{63}-Ile^{64} (29). Based on these observations, we suggest that PR-(5–99) forms an active dimeric protein that is capable of autoproteolysis at higher protein concentrations. The kinetic parameters for PR- and PR-(5–99)-catalyzed hydrolysis of substrates under identical conditions are summarized in Table I. The catalytic activity of PR is similar to that of wild-type PR (30). The low k_{cat}/K_m observed for PR-(5–99) using five different substrates suggests that only a small fraction of the protein contributes to the observed activity. The present results are consistent with our earlier study, in which we observed an ~50-fold lower k_{cat}/K_m for PR-(5–99) compared with PR assayed using a spectrophotometric substrate (15). PR-(1–95) did not exhibit detectable cleavage of any of the substrates tested.

The ^1H-^{15}N HSQC spectra of freshly prepared PR-(1–95) and PR-(5–99) recorded in the absence of DMP323 were similar to that of the PR_{R87K} monomer spectrum. Peaks that were unique to the dimer were absent in both cases (Fig. 6, *dotted boxes*). The addition of excess DMP323 to the proteins shifted the equilibrium for PR-(5–99) to a dimer (similar to the spectrum shown in Ref. 15), but not that for PR-(1–95), which still exhibited the spectrum of a monomer (similar to Fig. 6*b*). The molecular masses estimated for PR-(1–95) and PR-(5–99) by sedimentation equilibrium analysis and the consensus derived from both NMR and sedimentation equilibrium analyses regarding the observed folded dimer and monomer forms of the different constructs used in this study are shown in Table II.

In earlier studies, we had shown that after 2 days at a concentration of ~1.2 mM, PR-(5–99) exhibited a spectrum typical of a random-coil polypeptide; in addition, weak intensity dispersed signals corresponding to a folded conformation were also observed (15). We reexamined freshly prepared samples of PR-(1–95) and PR-(5–99) at a concentration of ~1 mM to determine the effect of protein concentration on dimerization. The spectrum of PR-(1–95) essentially remained that of a monomer (Fig. 6*b*), whereas the spectrum of PR-(5–99) showed a significant fraction of the signals corresponding to a dimer. The

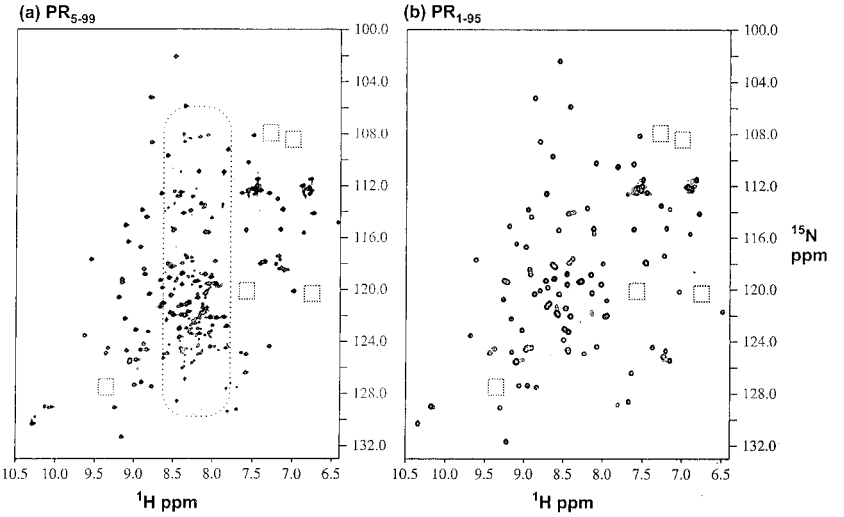

FIG. 6. **Amide ^{1}H-^{15}N HSQC spectra of freshly prepared PR-(5–99) (*a*) and PR-(1–95) (*b*) at 20 °C.** The locations of peaks unique to the dimer are *boxed* (see legend to Fig. 2). Additional peaks within the large oval in *a* result from self-degradation of PR-(5–99).

propensity of PR-(5–99) to dimerize with increasing protein concentration was similar to that observed for an analogous deletion construct of the Rous sarcoma virus protease (31).

Three major conclusions can be drawn from the analysis of the protease constructs that have either one or both of the terminal β-stands deleted. 1) The presence of at least one of the terminal β-strands (residues 1–4 or 96–99) is required for monomer folding; 2) both terminal strands contribute to dimer stability; and 3) the interaction between the C-terminal β-strand residues is pivotal for dimerization.

Relationship between Protease Dimer Stability and Function—In HIV, a single copy of the protease is synthesized as part of the Gag-Pol polyprotein, flanked by the $p6^{pol}$ and reverse transcriptase domains at the N and C termini, respectively. Previous studies indicated that the protease precursor is mainly in an unfolded form (15). Transient dimer formation leads to intramolecular autocatalytic cleavage first at the N terminus of the protease ($p6^{pol}$/PR site) concomitant with stable structure formation and the appearance of enzymatic activity (15, 32, 33). Subsequent cleavage at the C terminus (PR/RT site) of the protease occurs via an intermolecular process (34). Our present results indicate that the β-sheet formed by the central C-terminal β-strands of the four-stranded β-sheet contributes critically to the stability of the transient dimeric structure of the protease precursor such that the flexible N-terminal cleavage site sequence ($p6^{pol}$/PR) is accessible for intramolecular cleavage. Binding of inhibitor or possibly substrate to the active site promotes dimer formation and thereby could compensate for the loss of dimer stability, due to displacement of the outer β-strands (residues 1–4). Similarly, binding of the N-terminal polypeptide comprising the $p6^{pol}$/PR cleavage site to the active site may also contribute to the stability of the transient dimer formation of the precursor, leading to the hydrolysis of the scissile peptide bond and formation of the stable dimeric structure.

The highly conserved Arg^{87} residue in the conserved triad Gly^{86}-Arg^{87}-(Asn/Asp^{88}) of retroviral proteases plays a crucial role in the stability of the dimer. Loss of specific interactions involving Arg^{87}, *e.g.* the hydrogen bond with Asp^{29}, results in the destabilization of the dimer interfaces, particularly between the C-terminal β-stands as discussed above. The crystal structures of proteases of related viruses such as HIV-2, equine infectious anemia virus, feline immunodeficiency virus, Rous sarcoma virus, and simian immunodeficiency virus also demonstrate close proximity between the side chains of Arg^{87} and Asp^{29}. It will be important to investigate whether substitution of the conserved Arg^{87} residue in these related retroviral proteases also destabilizes the corresponding dimer.

Our results presented here reveal certain critical regions essential to the structure, stability, and catalytic activity of the mature HIV-1 protease. As pointed out in the Introduction, anti-HIV drugs based on active-site inhibitors contribute to the emergence of drug-resistant strains, a major problem in current treatment of AIDS. Based on our results, it may be possible and desirable to devise newer strategies for the rationale design of drugs targeting the dimer interface. Two critical interactions can be disrupted: 1) the inner two C-terminal β-strands and 2) the pivotal interaction between the Arg^{87} and Asp^{29} side chains. On the other hand, the protease monomer may aid in the screening of novel compounds that are specific for binding to the exposed interface.

Acknowledgments—We thank F. Delaglio and D. Garrett for data processing software, I. Neshiwat for expert technical assistance, and DuPont Pharmaceuticals for DMP323. We are grateful to S. Oroszlan for many thoughtful and inspiring discussions.

REFERENCES

1. Weber, I. T. (1990) *J. Biol. Chem.* **265,** 10492–10496
2. Strisovsky, K., Tessmer, U., Langner, J., Konvalinka, J., and Krausslich, H. G. (2000) *Protein Sci.* **9,** 1631–1641
3. Oroszlan, S., and Luftig, R. B. (1990) *Curr. Top. Microbiol. Immunol.* **157,** 153–185
4. Erickson, J. W., and Burt, S. K. (1996) *Annu. Rev. Pharmacol. Toxicol.* **36,** 545–571
5. Parikh, U., Hammond, J., Calef, C., Larder, B., Schinazi, R., and Mellors, J. W. (2000) in *The 2000 HIV Sequence Compendium: Mutations in Retroviral Genes Associated with Drug Resistance*, pp. 125–138, Los Alamos, NM (www.hiv.lanl.gov)
6. Rose, R. B., Craik, C. S., and Stroud, R. M. (1998) *Biochemistry* **37,** 2607–2621
7. Erickson, J. W., Gulnik, S. V., and Markowitz, M. (1999) *AIDS* **13,** S189–S204
8. Mahalingam, B., Louis, J. M., Reed, C. C., Adomat, J. M., Krouse, J., Wang, Y. F., Harrison, R. W., and Weber, I. T. (1999) *Eur. J. Biochem.* **263,** 238–245
9. Shultz, M. D., and Chmielewski, J. (1999) *Bioorg. Med. Chem. Lett.* **9,** 2431–2436
10. Zutshi, R., and Chmielewski, J. (2000) *Bioorg. Med. Chem. Lett.* **10,** 1901–1903
11. Rao, J. K., Erickson, J. W., and Wlodawer, A. (1991) *Biochemistry* **30,** 4663–4671
12. Pearl, L. H., and Taylor, W. R. (1987) *Nature* **329,** 351–354
13. Louis, J. M., Wondrak, E. M., Copeland, T. D., Smith, C. A., Mora, P. T., and Oroszlan, S. (1989) *Biochem. Biophys. Res. Commun.* **159,** 87–94
14. Ishima, R., Louis, J. M., and Torchia, D. A. (2001) *J. Mol. Biol.* **305,** 515–521
15. Louis, J. M., Clore, G. M., and Gronenborn, A. M. (1999) *Nat. Struct. Biol.* **6,** 868–875
16. Yamazaki, T., Hinck, A. P., Wang, Y. X., Nicholson, L. K., Torchia, D. A., Wingfield, P., Stahl, S. J., Kaufman, J. D., Chang, C. H., Domaille, P. J., and Lam, P. Y. (1996) *Protein Sci.* **5,** 495–506
17. Lam, P. Y., Ru, Y., Jadhav, P. K., Aldrich, P. E., DeLucca, G. V., Eyermann, C. J., Chang, C. H., Emmett, G., Holler, E. R., Daneker, W. F., Li, L., Confalone, P. N., McHugh, R. J., Han, Q., Li, R., Markwalder, J. A., Seitz, S. P., Sharpe, T. R., Bacheler, L. T., Rayner, M. M., Klabe, R. M., Shum, L.,

Winslow, D. L., Kornhauser, D. M., and Hodge, C. N. (1996) *J. Med. Chem.* **39,** 3514–3525
18. Perkins, S. J. (1986) *Eur. J. Biochem.* **157,** 169–180
19. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson.* **95,** 214–220
20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
21. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson.* **96,** 432–440
22. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc.* **114,** 6291–6293
23. Tjandra, N., Szabo, A., and Bax, A. (1996) *J. Am. Chem. Soc.* **118,** 6986–6991
24. Wu, D. H., Chen, A. D., and Johnson, C. S. (1995) *J. Magn. Reson.* **115,** 260–264
25. Todd, M. J., Semo, N., and Freire, E. (1998) *J. Mol. Biol.* **283,** 475–488
26. Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M., and Torchia, D. A. (1999) *Structure* **7,** 1047–1055
27. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. (1989) *Science* **245,** 616–621
28. Ala, P. J., Huston, E. E., Klabe, R. M., McCabe, D. D., Duke, J. L., Rizzo, C. J., Korant, B. D., DeLoskey, R. J., Lam, P. Y., Hodge, C. N., and Chang, C. H. (1997) *Biochemistry* **36,** 1573–1580
29. Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J., Tomich, C. S., Smith, C. W., Heinrickson, R. L., and Tomasselli, A. G. (1994) *Biochemistry* **33,** 9405–9413
30. Tozser, J., Blaha, I., Copeland, T. D., Wondrak, E. M., and Oroszlan, S. (1991) *FEBS Let.* **281,** 77–80
31. Schatz, G. W., Reinking, J., Zippin, J., Nicholson, L. K., and Vogt, V. M. (2000) *J. Virol.* **75,** 4761–4770
32. Louis, J. M., Nashed, N. T., Parris, K. D., Kimmel, A. R., and Jerina, D. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 7970–7974
33. Louis, J. M., Wondrak, E. M., Kimmel, A. R., Wingfield, P. T., and Nashed, N. T. (1999) *J. Biol. Chem.* **274,** 23437–23442
34. Wondrak, E. M., Nashed, N. T., Haber, M. T., Jerina, D. M., and Louis, J. M. (1996) *J. Biol. Chem.* **271,** 4477–4481