Transient HIV-1 Gag–protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations

Lalit Deshmukh[a], John M. Louis[a], Rodolfo Ghirlando[b], and G. Marius Clore[a,1]

[a]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892; and [b]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by G. Marius Clore, September 13, 2016 (sent for review September 1, 2016; reviewed by Hashim M. Al-Hashimi, Lewis E. Kay, and Michael F. Summers)

Cleavage of the group-specific antigen (Gag) polyprotein by HIV-1 protease represents the critical first step in the conversion of immature noninfectious viral particles to mature infectious virions. Selective pressure exerted by HIV-1 protease inhibitors, a mainstay of current anti–HIV-1 therapies, results in the accumulation of drug resistance mutations in both protease and Gag. Surprisingly, a large number of these mutations (known as secondary or compensatory mutations) occur outside the active site of protease or the cleavage sites of Gag (located within intrinsically disordered linkers connecting the globular domains of Gag to one another), suggesting that transient encounter complexes involving the globular domains of Gag may play a role in guiding and facilitating access of the protease to the Gag cleavage sites. Here, using large fragments of Gag, as well as catalytically inactive and active variants of protease, we probe the nature of such rare encounter complexes using intermolecular paramagnetic relaxation enhancement, a highly sensitive technique for detecting sparsely populated states. We show that Gag-protease encounter complexes are primarily mediated by interactions between protease and the globular domains of Gag and that the sites of transient interactions are correlated with surface exposed regions that exhibit a high propensity to mutate in the presence of HIV-1 protease inhibitors.

HIV-1 Gag | HIV-1 protease | drug resistance mutations | invisible states | paramagnetic relaxation enhancement

The transformation of noninfectious viral particles into mature infectious virions is a hallmark of the retroviral replication cycle. In HIV type 1 (HIV-1), morphological remodeling is initiated on sequential hydrolysis of the group-specific antigen (Gag) polyprotein by the viral homodimeric aspartyl protease, which generates a set of structural proteins (1). Not surprisingly, therefore, HIV-1 protease inhibitors are major components of current anti–HIV-1 therapies (2). The organization of HIV-1 Gag is as follows: matrix (MA)–capsid (CA)–spacer peptide 1 (SP1)–nucleocapsid (NC)–spacer peptide 2 (SP2)–p6, with an HIV-1 protease cleavage site located at each junction (Fig. 1*A*). MA, CA, and NC are globular domains, whereas SP1, SP2, and p6 are intrinsically disordered in solution (3). At neutral pH in vitro, the five cleavage sites in Gag are hydrolyzed by protease in the following order: SP1|NC > SP2|p6 ~ MA|CA > CA|SP1 ~ NC|SP2 (4). The corresponding in vivo cleavage order, deduced from virions with mutated Gag cleavage sites, is consistent with in vitro observations (5). The underlying mechanism governing ordered proteolysis is far from clear. Current understanding of Gag–protease interactions is derived from crystal structures of active protease complexed to nonhydrolysable peptide analogs (6) and of inactive protease (in which the active site Asp25 is replaced by Asn) bound to synthetic peptides comprising the Gag cleavage sites (7). Because all protease-bound peptides are in an extended, asymmetric β-strand conformation, it has been hypothesized that an effective protease substrate comprises six to eight residues of Gag (8) and that protease recognizes the shape of Gag substrates rather than their primary sequence (7). This "substrate envelope" hypothesis predicts variable interactions between side chains of Gag cleavage sites and protease (9), resulting in a unique processing rate for each site. Synthetic peptides, however, are a poor substitute for the Gag polyprotein in terms of long-range intermolecular interactions and conformational changes that can alter the accessibility of Gag cleavage junctions. Additionally both protease and Gag mutate and coevolve in response to protease inhibitors in current clinical use (10–17). Primary mutations (i.e., mutations at the catalytic site of protease) and mutations within the Gag cleavage sites are readily amenable to investigation using peptide analogs. The effects of secondary or compensatory mutations (i.e., mutations outside the protease substrate binding cleft or far away from the Gag cleavage sites), however, cannot be unraveled using synthetic peptides, and their existence hints at a larger role played by the globular Gag domains in Gag–protease interactions.

Here, using state-of-the-art solution NMR methods, and in particular intermolecular paramagnetic relaxation enhancement (PRE) measurements (18), we investigate Gag–protease interactions. We show that globular Gag domains transiently interact with protease to form sparsely populated encounter complexes, and that the patches of surface-exposed residues on the Gag domains and on protease that come into short-lived close contact with one another exhibit a high propensity to mutate in the presence of protease inhibitors.

Significance

Hydrolysis of group-specific antigen (Gag) polyprotein by protease is essential for the formation of infectious HIV-1 virions. In response to treatment with protease inhibitors, drug resistance mutations coevolve in protease and Gag, a significant number of which (known as compensatory mutations) lie outside the active site of protease and the Gag cleavage sites. We show, using paramagnetic NMR spectroscopy, that transient, sparsely populated encounter complexes are predominantly formed between the globular domains of Gag and protease at sites that correlate with the location of secondary drug resistance mutations. These results provide a structural basis for the origins of secondary mutations and suggest that transient encounter complexes play a significant role in guiding protease to the Gag cleavage sites.

Author contributions: L.D. and G.M.C. designed research; L.D. and R.G. performed research; J.M.L. contributed new reagents/analytic tools; L.D. and G.M.C. analyzed data; J.M.L. optimized and contributed new protease O samples used in the study; R.G. collected and analyzed sedimentation velocity data; and L.D. and G.M.C. wrote the paper.

Reviewers: H.M.A.-H., Duke University Medical Center; L.E.K., University of Toronto; and M.F.S., Howard Hughes Medical Institute, University of Maryland Baltimore County.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: mariusc@mail.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615342113/-/DCSupplemental.

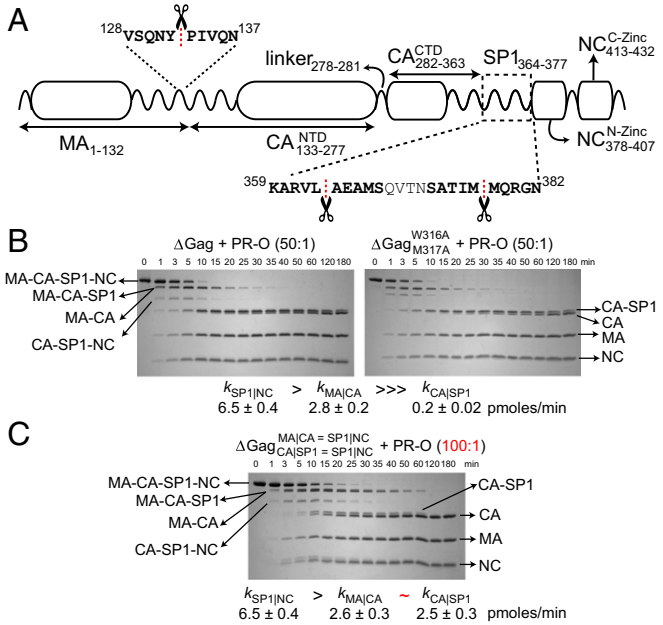

Fig. 1. Cleavage kinetics of ΔGag by PR-O. (*A*) Schematic of ΔGag organization. Sequences of the cleavage sites are shown with the MA|CA, CA|SP1, and SP1|NC junctions indicated by dashed lines and scissors. Note that in full-length Gag, the intrinsically disordered SP2 and p6 domains are located C-terminal to the NC domain, and there are cleavage sites at the NC|SP2 and SP2|p6 junctions. NTD and CTD refer to the N- and C-terminal domains, respectively of CA. Cleavage time course of (*B*) WT ΔGag (*Left*) and the monomeric $\Delta Gag^{W316A}_{M317A}$ mutant (*Right*) and of (*C*) the ΔGag construct, $\Delta Gag^{MA|CA=SP1|NC}_{CA|SP1=SP1|NC}$, in which the MA|CA and the CA|SP1 cleavage sites (highlighted in bold in *A*) have been mutated to the same sequence as that of the SP1|NC cleavage site (SATIM|MQRGN). In all cases, Gag processing by PR-O occurs in the same sequential order. Cleavage rates obtained from analysis of the time courses of the SDS/PAGE band intensities (see Fig. S2 and Table S1) are indicated. Aliquots of the reaction mixture at room temperature were taken at regular time intervals and visualized by PageBlue staining [18% (wt/vol) Tris-glycine gel]. The concentration of PR-O was 1 μM (in subunits), and the ratio of Gag to protease (in subunits) was 50:1 in *B* and 100:1 in *C*. Buffer conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, and 1 mM TCEP.

Results

Gag and Protease Constructs. To elucidate the role of Gag domains in Gag–protease interactions, we made use of a large Gag fragment containing the MA-CA-SP1-NC domains (group M, residues 1–432, hereafter referred to as ΔGag; Fig. 1*A*). ΔGag exhibits a monomer-dimer equilibrium at high ionic strength (K_{dimer} ~35 μM at ≥300 mM NaCl), but at low salt (≤100 mM NaCl) forms immature virus-like particles (3). We also made use of the construct, $\Delta Gag^{W316A}_{M317A}$ (Fig. S1), which carries a double mutation at the CA dimerization interface and is monomeric at high (300 mM NaCl) ionic strength (Fig. S1*A*), and the shorter construct $CA\text{-}SP1\text{-}NC^{W316A}_{M317A}$, which is monomeric at low (50 mM NaCl) ionic strength (Fig. S1*B*). Protease from HIV-1 group O (PR-O, pI ~5.2) was used instead of the more common M group (PR-M, *pI* ~9.5, 70% sequence identity with PR-O), owing to its much better solubility properties at high ionic strength.

Ordered Proteolytic Cleavage of Gag. In vitro cleavage of ΔGag by PR-M proceeds in the following order: NC|SP1 > MA|CA > CA|SP1 (3). Nearly identical cleavage patterns, time courses, and proteolysis rates are observed using PR-O for both ΔGag and $\Delta Gag^{W316A}_{M317A}$ (Fig. 1*B*, Fig. S2, and Table S1), indicating that the order of Gag proteolysis is conserved across HIV-1 groups and is independent of Gag oligomerization state and protease isoelectric point. The order of cleavage is also not dependent on differences in amino acid sequence of the cleavage sites, as a similar sequential order of processing is observed for the construct $\Delta Gag^{MA|CA=SP1|NC}_{CA|SP1=SP1|NC}$, in which the MA|CA and CA|SP1 cleavage junctions (^{128}VSQNY|PIVQN^{137} and ^{359}KARVL|AEAMS^{368}, respectively) were mutated to the same sequence as that of the SP1|NC cleavage site (^{373}SATIM|MQRGN^{383}), although the cleavage rate at the CA|SP1 junction is now enhanced and comparable to that at the MA|CA junction (Fig. 1*C*). These observations lead one to conclude that the globular Gag domains likely play a pivotal role in Gag–protease interactions.

Intermolecular PRE Measurements. To probe transient interactions between Gag and protease, we made use of NMR spectroscopy and specifically intermolecular PRE measurements (18). The intermolecular PRE arises from dipolar interactions between an unpaired electron located on a paramagnetic tag covalently attached to one protein and the protons of the second protein: in our case, backbone amide protons that can be selectively monitored by isotopically labeling the second protein with ^{15}N. When exchange between the unbound major species and the sparsely populated transient encounter complex is fast on the PRE time scale (lifetime < 250–500 μs), the intermolecular PRE observed on the protons of the major species will simply be given by the "true" intermolecular PRE within the complex scaled by the population of the complex (18). Owing to the large magnetic moment of the electron, the PRE effect (which is proportional to the $\langle r^{-6}\rangle$ separation between the unpaired electron and the proton) is very large, and hence transient complexes with occupancies as low as 0.5–1% can be detected (18).

We first carried out PRE experiments with the paramagnetic tag (maleimido-DOTA-Gd^{3+}; see Fig. S3*A*) conjugated to two alternative sites, V82C (PR-O^{D25N}_{V82C}) and L72C (PR-O^{D25N}_{L72C}), located within the catalytic cleft and on the surface, respectively, of an inactive PR-O variant (D25N; Fig. 2*A*). Although the catalytic activity of the V82C variant of PR-O is reduced compared with the wild-type (WT) version, the order of Gag cleavage is preserved (Fig. S4), and the structure of PR-O^{D25N}_{V82C}, determined from backbone amide residual dipolar couplings (RDCs) and backbone chemical shifts using CS-ROSETTA (19, 20), is the same as that of PR-M with the flaps predominantly in the closed conformation (Fig. S5). The intermolecular PRE profiles observed on monomeric ^{15}N/^{13}C/^{2}H-labeled $\Delta Gag^{W316A}_{M317A}$ (high salt) and ^{15}N/^{2}H-labeled $CA\text{-}SP1\text{-}NC^{W316A}_{M317A}$ (low salt) are shown in Fig. 2 *B* and *C*. Very similar intermolecular PRE profiles are also observed for WT ΔGag, although line broadening owing to monomer-dimer exchange precludes the measurement of PREs for much of the capsid domain (Fig. S6). Although the magnitude of the PREs may differ, the general pattern of intermolecular PREs obtained with the two paramagnetically tagged PR-O constructs is rather similar, with significant intermolecular PREs above background ($^{1}H_N$-Γ_2 > 10 s^{-1}) largely confined to specific surface-exposed regions within the three globular domains of Gag, namely MA, CA, and NC (Figs. 2 *B–D*). These intermolecular PREs are specific to the Gag–protease system, as essentially no intermolecular PREs above background are observed in control experiments using either free DOTA-Gd^{3+} or DOTA-Gd^{3+} tagged maltose binding protein (MBP) (Fig. S7). Interestingly, two regions display significant differences in intermolecular PREs between the $\Delta Gag^{W316A}_{M317A}$ and $CA\text{-}SP1\text{-}NC^{W316A}_{M317A}$ constructs. Specifically, larger intermolecular PREs are observed for $CA\text{-}SP1\text{-}NC^{W316A}_{M317A}$ at the N terminus of CA with both paramagnetic PR-O constructs and in the NC domain with PR-O^{D25N}_{L72C}. The former is likely due to the fact that the N terminus of CA undergoes a conformational change from an intrinsically disordered linker to a β-hairpin subsequent to cleavage at the MA|CA junction of Gag (3), while the latter is primarily driven by electrostatic interactions between the positively charged

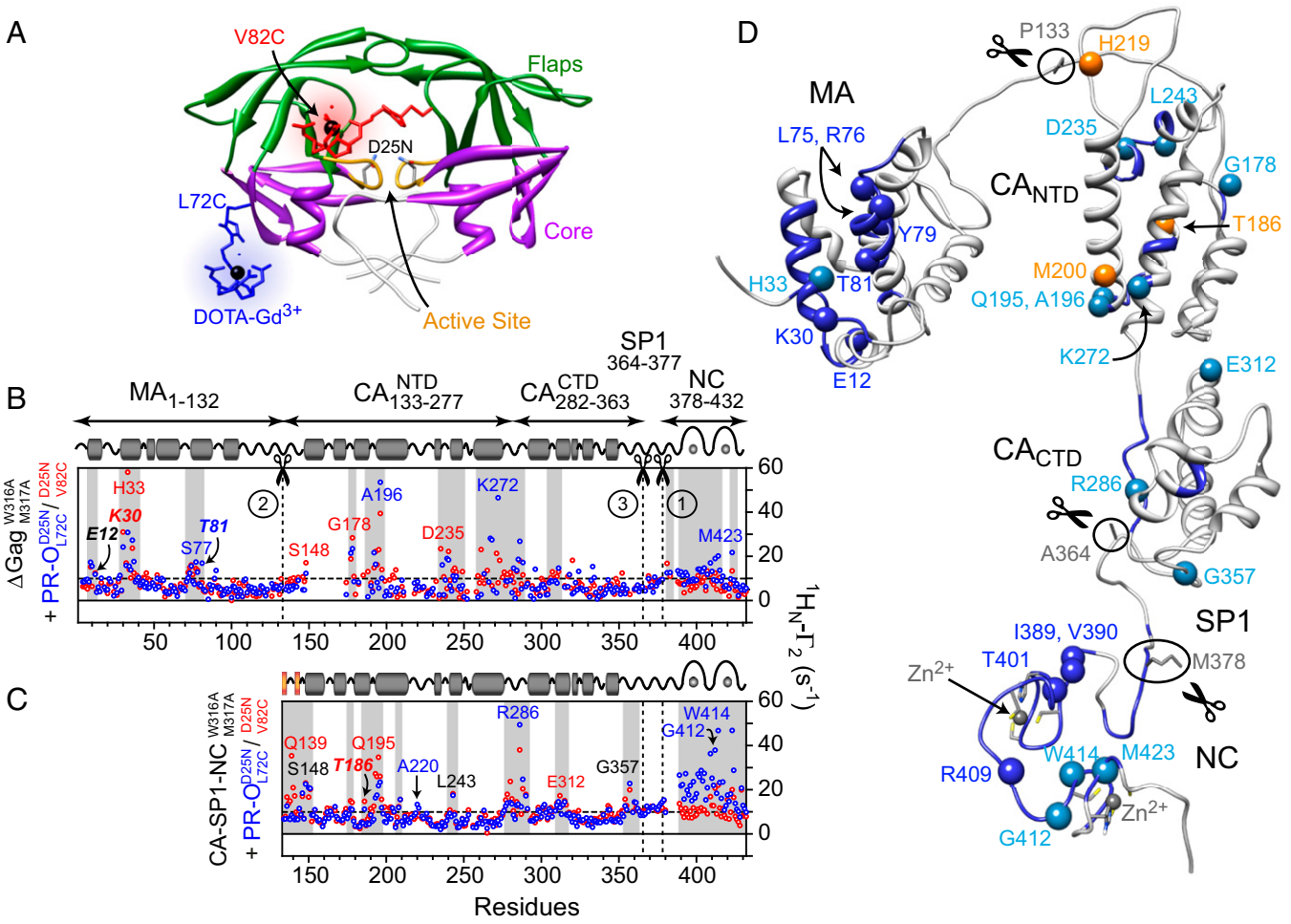

Fig. 2. PRE mapping of transient interactions of Gag with PR-O. (*A*) Ribbon diagram of PR-O with sites of paramagnetic tagging with DOTA-Gd^{3+} indicated (L72C, blue; V82C, red). Intermolecular PRE profiles observed on (*B*) $^{15}N/^{13}C/^{2}H$-labeled $\Delta Gag^{W316A}_{M317A}$ and (*C*) $^{15}N/^{2}H$-labeled CA-SP1-NC^{W316A}_{M317A} monomeric Gag constructs arising from Gd^{3+}-tagged PR-O^{D25N}_{L72C} (blue) and PR-O^{D25N}_{V82C} (red). Regions with intermolecular PREs above background (dashed lines) are indicated by the gray bars; residues exhibiting large PREs and associated with drug resistance (14–16, 28, 29) are indicated by the bold italic labels. The concentration of PR-O (in subunits) is 100 μM with a Gag to PR-O ratio of 3:1 and 2:1 for ΔGag and CA-SP1-NC constructs, respectively. (*D*) Ribbon representation of ΔGag (note the relative orientation of the MA, CA, and NC domains is arbitrary as the globular domains reorient semi-independently of one another in solution) (3). Gag cleavage junctions are highlighted with black circles and scissors. Regions of $\Delta Gag^{W316A}_{M317A}$ exhibiting PREs above background in the presence of paramagnetically tagged PR-O^{D25N}_{V82C} are colored in blue; residues that exhibit large PREs and are associated with drug resistance are depicted as dark blue spheres; those exhibiting large PREs only are shown as light blue spheres. Three residues (Thr186, Met200, and His219) in CA also associated with drug resistance (15, 16) are depicted as orange spheres; two of these (Thr186 and His219) give large PREs at low ionic strength (see *C*), whereas Met200 is close to the patch of CA residues (residues 192–198) that exhibit large PREs at both high and low ionic strengths.

NC domain and the negatively charged protease, which are magnified at low salt.

The unexpected absence of intermolecular PREs in the vicinity of the Gag cleavage sites can be attributed to two factors: (*i*) the occupancy of transient encounter complexes (lifetimes ≤ 250–500 μs probed by PRE) (18) involving the cleavage sites is much lower than that involving the globular domains of Gag; and (*ii*) the slow formation of weak (K_D ~300 μM) productive Gag–protease complexes are rate limited by protease flap opening (21, 22). The flaps of free protease are predominantly closed (23), blocking access to substrate; flap opening is a rare process resulting in slow productive complex formation (lifetime ~100 ms) (22), which is manifested by a significant reduction in $^{1}H_N/^{15}N$ cross-peak intensity for residues at the SP1|NC junction of $\Delta Gag^{W316A}_{M317A}$ on addition of a 3:1 excess of unlabeled protease (Fig. S8*A*). [Note that $^{1}H_N/^{15}N$ cross-peaks for residues at the MA|CA and CA|SP1 junctions are not affected at the concentrations of protease used, indicating that the SP1|NC junction serves as the primary point of association between Gag and protease, consistent with the observed order and rates of proteolytic cleavage (see Fig. 1*B*).] The converse experiment titrating a threefold excess of unlabeled $\Delta Gag^{W316A}_{M317A}$ into $^{15}N/^{2}H$-labeled PR-O^{D25N}_{V82C} results in a decrease in intensity of $^{1}H_N/^{15}N$ cross-peaks within the flaps and catalytic site of protease with no changes in chemical shifts (Fig. S8*B*), as expected for a system on the slow side of intermediate exchange on the chemical shift time scale.

To map the interaction sites for transient encounter complexes on the protease, we carried out intermolecular PRE experiments using $^{15}N/^{2}H$-labeled PR-O^{D25N}_{V82C} and two different paramagnetically(Gd^{3+}) tagged Gag constructs. Site-specific incorporation of a paramagnetic probe on ΔGag presents a challenge owing to the presence of 10 native cysteine residues. For one construct, we therefore used a truncated Gag comprising only MA and the N-terminal domain of CA (MA-CA^{NTD}), in which one of the two native cysteines was mutated to Thr (C57T), whereas the other, Cys87, located close to

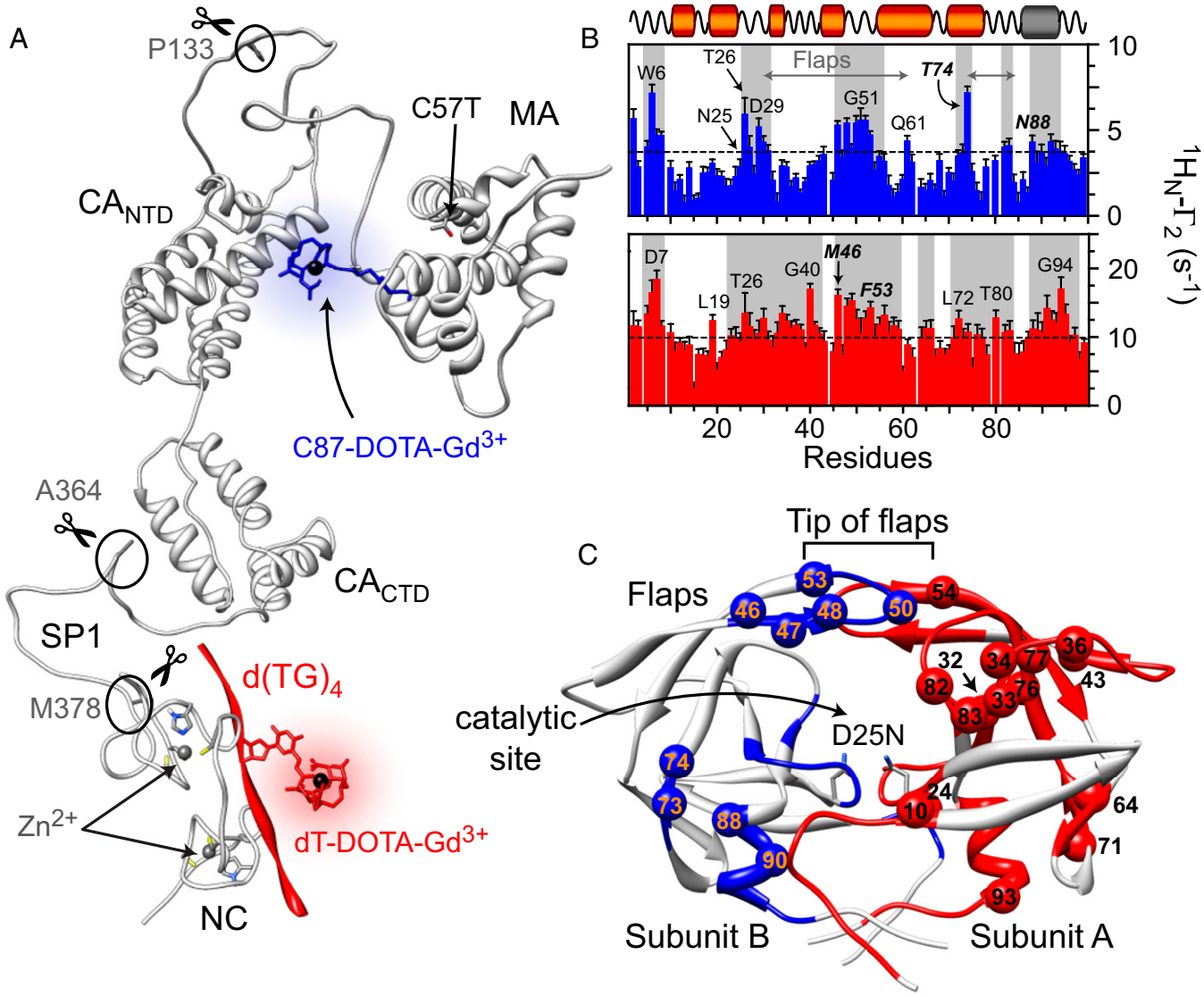

Fig. 3. PRE mapping of transient PR-O interactions with Gag. (*A*) Ribbon diagram of ΔGag with the location of the Gd^{3+} paramagnetically tagged sites indicated. Site-specific incorporation of a paramagnetic probe on ΔGag using traditional cysteine chemistry is difficult due to the presence of 10 native cysteine residues. Two constructs were therefore used. The first comprises MA-$CA^{C57T,C87}_{NTD}$ in which Cys57 is mutated to Thr and Cys87 is paramagnetically tagged. The second comprises $\Delta Gag^{W316A}_{M317A}$ complexed via the NC domain to the single stranded oligonucleotide d(TGT*GTGTG), where T* is DOTA-Gd^{3+} derivatized deoxythymidine. (*B*) Intermolecular PRE profiles observed on $^{15}N/^{2}H$-labeled PR-O^{D25N}_{V82C} in the presence of the paramagnetically tagged Gag variants MA-$CA^{C57T,C87}_{NTD}$ (blue; *Upper*) and $\Delta Gag^{W316A}_{M317A}$-$d(TG)_4$ (red; *Lower*). Regions with intermolecular PREs above background (dashed lines) are indicated by the gray bars. Residues exhibiting large PREs and associated with drug resistance (17) are labeled in bold italics. The concentration of PR-O was ~150 μM (in subunits), with PR-O to Gag molar ratios of 5:1 for MA-$CA^{C57T,C87}_{NTD}$ and 3:1.2 for $\Delta Gag^{W316A}_{M317A}$ with 50 μM ssDNA. (*C*) Ribbon representation of the PR-O dimer. Residues exhibiting large intermolecular PREs from paramagnetically tagged MA-$CA^{C57T,C87}_{NTD}$ and $\Delta Gag^{W316A}_{M317A}$-$d(TG)_4$ are shown in blue (on subunit A) and red (on subunit B), respectively. Residues that show large PREs and are associated with drug resistance (17) are depicted as spheres.

the MA|CA cleavage junction, was conjugated with DOTA-Gd^{3+} (MA-$CA^{C57T,C87}_{NTD}$) (Fig. 3*A*). In the second, monomeric $\Delta Gag^{W316A}_{M317A}$ was complexed to a single-stranded (ss)DNA 8mer, $d(TG)_4$ containing a single DOTA-Gd^{3+} derivatized deoxythymidine (Fig. 3*A* and Fig. S3*B*). A slight molar excess of $\Delta Gag^{W316A}_{M317A}$ was used, ensuring that all ssDNA was bound to NC, as the interaction of MA with ssDNA is very weak in comparison (3, 24). With paramagnetically tagged MA-$CA^{C57T,C87}_{NTD}$, regions of protease exhibiting PREs above background ($^{1}H_N$-$\Gamma_2 > 4\ s^{-1}$) comprise residues near or in the catalytic site (residues 6–8 and 26–31), residues in the flaps (residues 46–53 and 82–83), and a small stretch of core residues (residues 88–94; Fig. 3*B*, *Upper*). With paramagnetically tagged $\Delta Gag^{W316A}_{M317A}$-$d(TG)_4$, regions of protease exhibiting significant PREs include residues near or in the catalytic site (residues 5–10 and 23–29), the flaps (residues 30–43, 46–59, 71–77 and 80–83), and part of the core (residues 64–66 and 87–97; Fig. 3*B*, *Lower*). Essentially no intermolecular PREs above background are observed on protease in control experiments using DOTA-Gd^{3+} or Gd^{3+}-tagged MBP (Fig. S9).

Discussion

Role of Gag–Protease Encounter Complexes. The observed pattern of intermolecular PRE profiles are indicative of the formation of transient encounter complexes involving the globular domains of Gag, which serve to guide protease in the vicinity of the Gag cleavage sites such that productive complex formation can occur efficiently on rare protease flap opening (Fig. 4). These encounter complexes are likely to have a significant impact on the

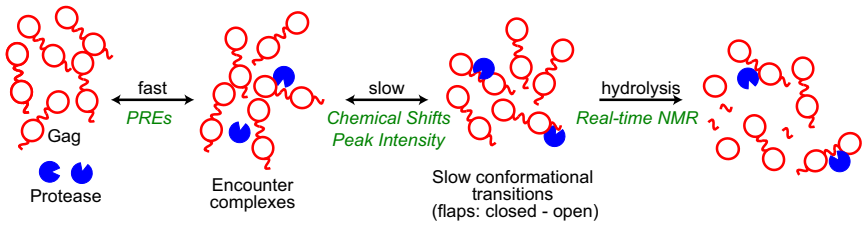

Fig. 4. Schematic of the interactions of HIV-1 protease and Gag. Each illustrated step occurs at successively slower time scales that are accessible to different types of NMR measurements (as indicated in green). Transient encounter complexes, in fast exchange with the unbound species, are detected by PRE measurements (see Figs. 2 and 3). These ultra-weak complexes, involving quite extensive interactions surfaces, are initially formed and serve to guide the formation of productive complexes, thereby dictating the sequential order of Gag cleavage (i.e., we speculate that these transient encounter complexes accelerate the formation of productive complexes). The formation of productive complexes at the Gag cleavage sites are associated with rare conformational transitions involving flap opening (22, 23) that are detected by changes in cross-peak intensities and chemical shifts (intermediate exchange; see Fig. S8). Finally, cleavage/liberation of individual MA, CA, and NC domains is a slow process that can be monitored by real-time NMR by observing site specific build-up of cross-peaks arising from the cleavage products (3).

cleavage rates at the different Gag junctions. (It should be noted, however, that functional testing of the role of encounter complexes by in vitro mutagenesis is likely to be very difficult as the interfaces are relatively large and the contribution of any individual intermolecular interaction is likely to be very small. Hence, a large number of cumulative mutations would likely be required.) Other lines of biochemical evidence support this picture: the presence of single-stranded nucleic acids significantly accelerates hydrolysis at the SP1|NC junction while leaving the cleavage rate at the MA|CA and CA|SP1 unaltered (3). The two zinc knuckles of NC sample a large region of conformational space relative to one another in the absence of nucleic acids but behave as a single globular entity when bound to nucleic acids (24, 25), which would be predicted to enhance the formation of transient encounter complexes with protease. The two slowest cleavage sites (CA|SP1 and NC|SP2) occur at junctions that are not flanked by globular domains, suggesting that transient confinement of protease between Gag domains is important in sequential processing. A recent report (26) suggested that the six-helix bundle formed by SP1 (27) in immature Gag assemblies may block access to the CA|SP1 junction, thereby slowing down the proteolysis at this site. However, SP1 and neighboring residues are intrinsically disordered and fully accessible in solution in the context of ΔGag, as evidenced by NMR chemical shifts, relaxation parameters, and backbone amide RDCs (3, 24), and yet hydrolysis at the CA|SP1 junction is remarkably slow in vitro (compare Fig. 1*B*), even when mutated to the sequence of the SP1|NC junction (see Fig. 1*C*). Moreover, of the five Gag cleavage sites, only the CA|SP1 and SP1|NC junctions, located a mere 15 residues apart, are interdependent on one another such that blocking hydrolysis at the SP1|NC junction by site-directed mutagenesis accelerates hydrolysis of the upstream CA|SP1 junction (4). These observations suggest that protease, guided by the CA and NC domains, can competitively interact with both cleavage junctions on either side of SP1, with an intrinsic preference for the SP1|NC junction. Based on these observations, we conclude that sequential in vitro hydrolysis of Gag is governed by several factors: transient interaction of Gag domains with protease, accessibility of Gag cleavage junctions, and the primary sequence of Gag cleavage sites. Moreover, because the sequential order of proteolytic processing is the same in vitro and in vivo, we speculate that these factors also dictate Gag hydrolysis in vivo. Additionally, steric hindrance due to the assembled Gag lattice and the protonation states of cleavage site residues are likely to also contribute to Gag processing in vivo.

Correlation of Sites of Encounter Complexes with Compensatory Drug Resistance Mutations. A comparison of regions involved in transient Gag–protease encounter complexes with residues associated with drug-resistance mutations (14–17, 28) in Gag and protease is provided in Figs. 2*D* and 3*C*, respectively (also see Table S2 for additional details). The emergence of mutations resistant to HIV-1 protease inhibitors usually involves a stepwise process whereby primary mutations in protease alter the substrate-binding pocket leading to a reduction in inhibitor binding and concomitantly productive Gag–protease complex formation, whereas secondary or compensatory mutations in both protease and Gag may help partially restore the loss of viral fitness caused by the primary mutations (10, 15, 16). The regions of Gag and protease that undergo compensatory mutations are strikingly similar to those exhibiting significant intermolecular PREs, thereby providing a plausible underlying structural basis behind these mutations. For example, conservative noncleavage site mutations in MA, specifically K30R, R76K, Y79F, and T81A, restore the fitness deficit arising from drug-resistant protease by improving replication capacity and generating a significant reduction in susceptibility to protease inhibitors in vivo (14, 29). These residues along with two other drug-resistance mutations, E12K and L75R, reside in regions of MA that give large intermolecular PREs. Similarly, among the few compensatory mutations in CA that have been reported, T186M and H219Q display large intermolecular PREs, whereas the third, M200I, is close to a region of CA that exhibits large PREs (residues 192–198). Further, several residues of NC mutate on exposure to protease inhibitors, and every single one displays large intermolecular PREs. It should also be noted that some of these residues are involved in other well-established functions [e.g., Lys30 and Arg76 of MA are involved in Gag–membrane interactions (30, 31), and His219 is located in an exposed loop within the N-terminal domain of CA loop that binds to cyclophilin-A (32)], implying multiple functions for these residues. The primary and secondary mutations in protease also show a close correlation with residues that exhibit significant intermolecular PREs, including residues in the core (residues 10, 71, 73, 88, 90, and 93) and flaps (residues 32–34, 36, 46, 47, 48, 50, 53, 54, and 82). Based on the above correlations, we conclude that HIV-1 Gag–protease interactions have evolved a conserved mechanism whereby the globular domains of Gag act as guideposts for the formation of transient encounter complexes that facilitate access of the protease to the cleavage junctions, thereby enhancing cleavage rates and modulating the sequential order of Gag cleavage.

Materials and Methods

Protein Expression and Purification. Full details on cloning, expression, site-directed mutagenesis, isotope ($^{2}H/^{15}N/^{13}C$ and $^{2}H/^{15}N$) labeling, purification, paramagnetic tagging with Gd^{3+}, and analytical ultracentrifugation experiments are provided in *SI Materials and Methods*. Samples for NMR were prepared in a buffer containing 20 mM sodium phosphate, pH 6.5, 0.1 mM $ZnCl_2$, and 1 mm Tris(2-carboxyethyl)phosphine (TCEP); additionally, the

buffer contained 300 and 50 mM NaCl in the case of ΔGag and CA-SP1-NC constructs, respectively.

NMR Spectroscopy. All heteronuclear NMR experiments were carried out at 30 °C on Bruker 500-, 600-, and 800-MHz spectrometers equipped with *z*-gradient triple resonance cryoprobes. Full details of the sequential assignment, PRE and RDC experiments, and NMR spectral processing and analysis are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank M. Bayro, J. Werner-Allen, C. D. Schwieters, D. Appella, and V. Tugarinov for useful discussions; Y. Shen for help with CS-ROSETTA calculations; Y. Kim for help with DOTA-Gd^{3+} purification; A. Aniana for technical assistance; and J. Lloyd [National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Advanced Mass Spectrometry Core] for technical support. This work was supported by the Intramural Program of NIDDK/NIH and the AIDS Targeted Antiviral Program of the Office of the NIH Director (to G.M.C.).

1. Freed EO (2015) HIV-1 assembly, release and maturation. *Nat Rev Microbiol* 13(8): 484–496.
2. Arts EJ, Hazuda DJ (2012) HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* 2(4):a007161.
3. Deshmukh L, Ghirlando R, Clore GM (2015) Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proc Natl Acad Sci USA* 112(11):3374–3379.
4. Pettit SC, et al. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 68(12):8017–8027.
5. Wiegers K, et al. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* 72(4):2846–2854.
6. Mahalingam B, Louis JM, Hung J, Harrison RW, Weber IT (2001) Structural implications of drug-resistant mutants of HIV-1 protease: High-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins* 43(4):455–464.
7. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* 10(3):369–381.
8. Chou KC (1996) Prediction of human immunodeficiency virus protease cleavage sites in proteins. *Anal Biochem* 233(1):1–14.
9. Ozen A, Haliloğlu T, Schiffer CA (2011) Dynamics of preferential substrate recognition in HIV-1 protease: Redefining the substrate envelope. *J Mol Biol* 410(4): 726–744.
10. Doyon L, et al. (1996) Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* 70(6):3763–3769.
11. Zhang YM, et al. (1997) Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* 71(9):6662–6670.
12. Mammano F, Petit C, Clavel F (1998) Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: Phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* 72(9):7632–7637.
13. Kolli M, Lastere S, Schiffer CA (2006) Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* 347(2):405–409.
14. Parry CM, et al. (2011) Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrob Agents Chemother* 55(3):1106–1113.
15. Garcia Diaz A (2012) An investigation of the role of HIV-1 Gag mutations in failure of protease inhibitors. PhD thesis (University College London, London).
16. Fun A, Wensing AM, Verheyen J, Nijhuis M (2012) Human immunodeficiency virus Gag and protease: Partners in resistance. *Retrovirology* 9:63.
17. Wensing AM, et al. (2015) 2015 Update of the drug resistance mutations in HIV-1. *Top Antivir Med* 23(4):132–141.
18. Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* 109(9):4108–4139.
19. Shen Y, et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105(12):4685–4690.
20. Das R, et al. (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci USA* 106(45):18978–18983.
21. Furfine ES, et al. (1992) Two-step binding mechanism for HIV protease inhibitors. *Biochemistry* 31(34):7886–7891.
22. Katoh E, et al. (2003) A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex. *Protein Sci* 12(7):1376–1385.
23. Roche J, Louis JM, Bax A (2015) Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. *ChemBioChem* 16(2):214–218.
24. Deshmukh L, Ghirlando R, Clore GM (2014) Investigation of the structure and dynamics of the capsid-spacer peptide 1-nucleocapsid fragment of the HIV-1 gag polyprotein by solution NMR spectroscopy. *Angew Chem Int Ed Engl* 53(4):1025–1028.
25. Deshmukh L, Schwieters CD, Grishaev A, Clore GM (2016) Quantitative characterization of configurational space sampled by HIV-1 nucleocapsid using solution NMR, X-ray scattering and protein engineering. *ChemPhysChem* 17(11):1548–1552.
26. Wagner JM, et al. (2016) Crystal structure of an HIV assembly and maturation switch. *eLife* 5:e17063.
27. Schur FK, et al. (2016) An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* 353(6298):506–508.
28. Gatanaga H, et al. (2002) Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem* 277(8):5952–5961.
29. Sutherland KA, Mbisa JL, Cane PA, Pillay D, Parry CM (2014) Contribution of Gag and protease to variation in susceptibility to protease inhibitors between different strains of subtype B human immunodeficiency virus type 1. *J Gen Virol* 95(Pt 1):190–200.
30. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI (1996) Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. *Proc Natl Acad Sci USA* 93(7): 3099–3104.
31. Saad JS, et al. (2006) Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc Natl Acad Sci USA* 103(30):11364–11369.
32. Gamble TR, et al. (1996) Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 87(7):1285–1294.
33. Zhao H, Brautigam CA, Ghirlando R, Schuck P (2013) Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. *Curr Protoc Protein Sci* Chap 20, Unit 20:12.
34. Ghirlando R, et al. (2013) Improving the thermal, radial, and temporal accuracy of the analytical ultracentrifuge through external references. *Anal Biochem* 440(1):81–95.
35. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606–1619.
36. Cole JL, Lary JW, P Moody T, Laue TM (2008) Analytical ultracentrifugation: Sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* 84:143–179.
37. Clore GM, Starich MR, Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120(40):10571–10572.
38. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
39. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687–696.
40. Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* 16(1):22–34.
41. Fitzkee NC, Bax A (2010) Facile measurement of ^{1}H-^{15}N residual dipolar couplings in larger perdeuterated proteins. *J Biomol NMR* 48(2):65–70.
42. Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* 48(1):47–62.
43. Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against ^{1}H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126(18): 5879–5896.
44. Hu K, Doucleff M, Clore GM (2009) Using multiple quantum coherence to increase the ^{15}N resolution in a three-dimensional TROSY HNCO experiment for accurate PRE and RDC measurements. *J Magn Reson* 200(2):173–177.
45. Kuzmic P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal Biochem* 237(2):260–273.
46. Sayer JM, Liu F, Ishima R, Weber IT, Louis JM (2008) Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. *J Biol Chem* 283(19):13459–13470.
47. Heaslet H, et al. (2007) Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr D Biol Crystallogr* 63(Pt 8):866–875.
48. Martin P, et al. (2005) "Wide-open" 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target. *Structure* 13(12):1887–1895.
49. Deshmukh L, et al. (2013) Structure and dynamics of full-length HIV-1 capsid protein in solution. *J Am Chem Soc* 135(43):16133–16147.
50. Clore GM, Garrett D (1999) R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *J Am Chem Soc* 121(39):9008–9012.

Supporting Information

Deshmukh et al. 10.1073/pnas.1615342113

SI Materials and Methods

Materials. Single-stranded (ss)DNA, 5′-d$(TG)_4$, containing a single 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-Gd^{3+} derivatized deoxythymidine, was purchased from Bio-Synthesis. ssDNA was dissolved in deionized water and dialyzed overnight (Spectra/Pro, Micro Float-A-Lyzer Dialysis Device; 100- to 500-Da cutoff, catalog no. F235049) in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, and 1 mM TCEP. Soluble paramagnetic cosolute, DOTA-Gd^{3+} (also known as gadoteric acid) was purchased from Macrocyclics (catalog no. M-147) and dissolved in a buffer containing 50 mM sodium acetate, pH 6, to a final concentration of ∼10 mg/mL.

Protein Expression and Purification. WT and mutant variants of HIV-1 Gag polyprotein, ΔGag (MA-CA-SP1-NC, residues 1–432, strain HXB2, group M), CA-SP1-NC (residues 133–432, strain pLN4-3, group M), MA-CA_{NTD} (residues 1–278, strain HXB2, group M), and HIV-1 protease (residues 1–99, group O), as well as apo-maltose binding protein (MBP), were subcloned into pET-11a vectors (Novagen; EMD Millipore) and expressed in BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies). Site-directed mutagenesis was performed using the QuikChange kit (Agilent Technologies).

ΔGag constructs were expressed at 19 °C, whereas CA-SP1-NC and MA-CA_{NTD} constructs were expressed at 37 °C using our previously published protocols (3, 24). Briefly, cells were grown at 37 °C in 1 L Luria-Bertani (LB) medium at natural isotopic abundance or minimal M9 medium for isotopic labeling. The latter contained 0.3 g/L $^2H/^{15}N/^{13}C$ Isogro (Sigma-Aldrich), 99.9% (vol/vol) D_2O, 1 g/L $^{15}NH_4Cl$, and 3 g/L $^2H_7,^{13}C_6$-D-glucose for $^2H/^{15}N/^{13}C$ labeling and 0.3 g/L $^2H/^{15}N$ Isogro (Sigma-Aldrich), 99.9% (vol/vol) D_2O, 1g/L $^{15}NH_4Cl$, and 3g/L $^2H_7,^{12}C_6$-D-glucose for $^2H/^{15}N$ labeling. For ΔGag constructs, ∼30 min before induction, the temperature was reduced to 19 °C. Cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density of A_{600} ∼0.8. For CA-SP1-NC and MA-CA_{NTD} constructs, the cells were harvested 8 h after induction, whereas for ΔGag constructs, the cells were harvested after 24 h. For HIV-1 protease, cells were grown at 37 °C in 1 L of either LB or minimal M9 medium and were harvested 3 h after induction with IPTG. In the case of MBP, cells were grown at 37 °C in 1 L LB medium and harvested 3 h after induction with IPTG.

For ΔGag and MA-CA_{NTD} constructs, the cells were resuspended in a lysis buffer containing 100 mM Tris, pH 8.0, 500 mM NaCl, 5 mM β-mercaptoethanol (BME), and 1 cOmplete Protease Inhibitor mixture tablet (Roche Applied Science). For CA-SP1-NC constructs, the lysis buffer contained 100 mM Tris, pH 8.0, 50 mM NaCl, 0.1 mM $ZnCl_2$, 5 mM BME, and 1 cOmplete tablet. Cells were lysed using a microfluidizer and cleared by centrifugation. Additionally, in the case of CA-SP1-NC constructs, nucleic acids were precipitated by adding 4% (wt/vol) polyethyleneimine, pH 8.0 (Sigma-Aldrich), to the cell lysate to a final concentration of ∼0.4% (wt/vol). All constructs were purified by a combination of ion exchange and size exclusion chromatography. For ΔGag and MA-CA_{NTD} variants, the cell lysate was loaded onto a HiPrep 16/10 Q FF column (GE Healthcare) with a 0.5–1 M NaCl gradient in a buffer containing 100 mM Tris, pH 8.0, 500 mM NaCl, and 5 mM BME. For CA-SP1-NC constructs, a 0–1 M NaCl gradient was used with a running buffer containing 100 mM Tris, pH 8.0, 0.1 mM $ZnCl_2$, and 5 mM BME. In the case of ΔGag and MA-CA_{NTD} constructs, relevant flow-through fractions were diluted in a buffer containing 100 mM Tris, pH 8.0, and 5 mM BME (1:1 dilution), whereas for CA-SP1-NC constructs, the flow-through fractions were used without any further dilution. These fractions were loaded onto a HiLoad 16/10 SP Sepharose HP column (GE Healthcare) with a 0–1 M NaCl gradient containing 100 mM Tris, pH 8.0, and 5 mM BME. The eluted proteins were concentrated (Amicon ultra-15, 10-kDa cutoff for MA-CA_{NTD} and CA-SP1-NC constructs, and 30-kDa cutoff for ΔGag constructs) and loaded onto a HiLoad 26/60 Superdex 200 column (GE Healthcare) pre-equilibrated with 100 mM Tris, pH 8, 500 mM NaCl, and 5 mM BME. For CA-SP1-NC and MA-CA_{NTD} constructs, a HiLoad 26/60 Superdex 75 column (GE Healthcare) was used pre-equilibrated with 100 mM Tris, pH 8, 300 mM NaCl, and 5 mM BME. Relevant fractions were pooled and diluted in a buffer containing 20 mM Tris, pH 8.0, and 5 mM BME (1:1 dilution), and further purified using a Mono S 10/100 GL column (GE Healthcare) with a 0–1 M NaCl gradient. Note that for ΔGag and CA-SP1-NC constructs, every buffer solution was supplemented with 0.1 mM $ZnCl_2$.

For WT active protease-O (PR-O), the cells were resuspended in a bacterial protein extraction reagent (ThermoFisher; catalog no. 78248) supplemented with 5 mM benzamidine. Cells were lysed using a sonicator and cleared by centrifugation. Ammonium sulfate was added to the supernatant to a final concentration ∼20% (wt/vol) and incubated at room temperature followed by centrifugation. The pellet was resuspended in buffer containing 25 mM Tris, pH 7.5, 50 mM NaCl and spun, and the supernatant subjected to fractionation on a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 25 mM Tris, pH 7.5, and 50 mM NaCl. Peak fractions corresponding to active PR-O were pooled, concentrated (Amicon ultra-15, 10-kDa cutoff), and stored at −70 °C.

In the case of the two PR-O variants, PR-O^{D25N}_{L72C} and PR-O^{D25N}_{V82C}, the proteins were purified using the following two schemes. For unlabeled proteins, the cells were resuspended in buffer containing 50 mM Tris, pH 8, 10 mM EDTA, 10 mM DTT, and ∼100 μg/mL lysozyme. The insoluble recombinant protein was washed using a buffer containing 50 mM Tris, pH 8, 10 mM EDTA, 10 mM DTT, 2 M urea, and 1% (vol/vol) Triton X-100. The insoluble fraction was pelleted by centrifugation and solubilized in 50 mM Tris, pH 8.0, 7.5 M guanidine hydrochloride, 5 mM EDTA, and 10 mM DTT. The solubilized fraction was injected on a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated in 50 mM Tris, pH 8, 4 M guanidine hydrochloride, 5 mM EDTA, and 1 mM DTT. Eluted fractions were pooled and subjected to reverse-phase HPLC on a POROS R2 20-μm resin (ThermoFisher; catalog no. 1112906). Peak fractions were pooled and stored at −70 °C. For the $^2H/^{15}N$-labeled PR-O variants, a similar protocol was used except that the supernatant fraction, instead of the insoluble fraction, was stirred for 1 h at room temperature in the presence of 20% (wt/vol) ammonium sulfate, followed by centrifugation. The pellet was resuspended in 25 mM Tris, pH 7.5, and 50 mM NaCl and spun. The supernatant was adjusted to 2 M guanidine hydrochloride and concentrated (Amicon ultra-15, 10-kDa cutoff). Proteins were fractionated using size-exclusion chromatography and reverse-phase HPLC as described above.

To fold PR-O samples, aliquots were lyophilized and later redissolved in 4 M guanidine hydrochloride, 50 mM Tris, pH 7.5, and 2 mM DTT to a final concentration of ∼5 mg/mL, followed by extensive dialysis in aqueous buffers without guanidine hydrochloride.

For apo-maltose binding protein, the cells were resuspended in a lysis buffer containing 100 mM Tris, pH 8.0, 500 mM NaCl, 1 mM EDTA, and 5 mM BME. Cells were lysed using a microfluidizer, cleared by centrifugation, and then loaded onto an Amylose resin column (New England BioLabs) with a 0–10 mM maltose gradient containing 100 mM Tris, pH 8.0, 500 mM NaCl, 1 mM EDTA, and 5 mM BME. The eluted protein was concentrated (Amicon ultra-15, 10-kDa cutoff) and loaded onto a HiLoad 26/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 100 mM Tris, pH 8, and 500 mM NaCl.

All protein constructs were verified by DNA sequencing and mass spectrometry.

Site-Specific Spin Labeling with Gd^{3+}. The commercially available paramagnetic probe, maleimido-monoamide-DOTA (Macrocyclics: catalog no. B-272; 1,4,7,10-tetraazacyclododecane-1,4,7-tris(t-butyl-acetate)-10-(aminoethylacetamide)), was mixed with Gd(III) chloride hexahydrate (Sigma-Aldrich; catalog no. G7532) at a molar ratio 1:2 in 50 mM sodium acetate buffer, pH 5.5. The mixture was incubated overnight at room temperature, and the reaction was verified by mass spectrometry. The resultant complex was purified by reverse-phase HPLC (Beckman Ultrasphere ODS column, MAC-MOD Analytical; catalog no. 235328). Relevant fractions were pooled, lyophilized, and later redissolved in 50 mM sodium acetate buffer, pH 6.0. For site-specific paramagnetic spin labeling, a DOTA-Gd^{3+} stock solution (~8 mg/mL) was incubated for 1 h at room temperature with the protein of interest (Gag, MBP, or PR-O variants containing a single surface-exposed cysteine residue) at a molar ratio of protein to paramagnetic label of ~1:1.2 in 100 mM Tris, pH 8. The conjugation reaction was tested for completion by mass spectrometry, and excess unreacted paramagnetic spin-label was removed by dialysis (Slide-A-Lyzer G2 Dialysis Cassettes; ThermoFisher Scientific). All buffers were treated with chelex-100 (Sigma-Aldrich; catalog no. 88593) to remove any potential trace metal contamination.

Sedimentation Velocity Analytical Ultracentrifugation. Sedimentation velocity experiments were conducted at 50,000 rpm using an An50-Ti rotor (Beckman Coulter) on a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge following protocols described previously (33). Samples of $\Delta Gag^{W316A}_{M317A}$ were studied at loading concentrations ranging from 2.5 to 66 μM in 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 1mM TCEP, and 0.1 mM $ZnCl_2$ and 30 °C. Samples of CA-SP1-NC^{W316A}_{M317A} were studied at 20 °C and loading concentrations of 10–84 μM in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 1 mM TCEP, and 0.1 mM $ZnCl_2$. Samples were loaded in two-channel centerpiece cells, and data were collected using the absorbance (280 nm) and Rayleigh interference (655 nm) optical detection systems. In both cases, the highest concentration samples were loaded in 3-mm path-length cells, whereas all other samples were loaded in standard 12-mm path-length cells. Sedimentation data were time-corrected (34) and analyzed in SEDFIT15.01c (35) in terms of a continuous c(s) distribution of Lamm equation solutions with a resolution of 0.05 S and a maximum entropy regularization confidence level of 0.68. Excellent data fits were observed, with RMSD values ranging from 0.0037 to 0.0079 absorbance units and 0.0060 to 0.011 fringes. The solution densities and viscosities were determined based on the buffer composition in SEDNTERP (36) (sednterp.unh.edu). Protein partial specific volumes were calculated based on the amino acid composition in SEDNTERP, and sedimentation coefficients s were corrected to $s_{20,w}$ values at standard conditions.

NMR Sample Preparation. All heteronuclear NMR experiments were performed on uniformly $^{15}N/^{13}C/^{2}H$- or $^{15}N/^{2}H$-labeled Gag or protease samples (unless stated otherwise). ΔGag constructs were prepared in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, 93% (vol/vol) H_2O/7% (vol/vol) D_2O, and 1 mM TCEP. For CA-SP1-NC constructs, an identical buffer was used but with the concentration of NaCl reduced to 50 mM. Aligned PR-O^{D25N}_{V82C} samples were prepared using ~11 mg/mL phage pf1 (ASLA Biotech) (37) in the same buffer as that used for ΔGag constructs. Backbone amide ($^1D_{NH}$) RDC data were measured on samples containing 0.1 mM PR-O^{D25N}_{V82C}. For PRE experiments, the following sample concentrations were used: 0.3 mM $^{15}N/^{13}C/^{2}H$-labeled ΔGag or 0.2 mM $^{15}N/^{2}H$-labeled CA-SP1-NC constructs in the presence of 0.1 mM paramagnetically tagged PR-O variants at natural isotopic abundance and 0.15 mM $^{15}N/^{2}H$-labeled PR-O^{D25N}_{V82C} in the presence of either 0.03 mM paramagnetically tagged MA-$CA^{C57T,C87S}_{NTD}$ or a mixture of 0.06 mM monomeric $\Delta Gag^{W316A}_{M317A}$ and 0.05 mM paramagnetically tagged $d(TG)_4$ at natural isotopic abundance.

NMR Spectroscopy. All heteronuclear NMR experiments were carried out at 30 °C on Bruker 500-, 600-, and 800-MHz spectrometers equipped with z-gradient triple resonance cryoprobes. Spectra were processed using NMRPipe (38) and analyzed using the CCPN software suite (39). Sequential ^1H, ^{15}N, and ^{13}C backbone resonance assignments were performed using conventional transverse relaxation optimized (TROSY)-based through-bond 3D triple resonance experiments (40). $^1D_{NH}$ RDCs (given by the difference in $^1J_{NH}$ coupling constants in aligned and isotropic media) were measured using the TROSY-based ARTSY pulse sequence (41) and analyzed with Xplor-NIH (42). Transverse ^1H_N-Γ_2 PRE rates were obtained from the differences in the transverse ^1H_N-R_2 relaxation rates between the paramagnetic and diamagnetic samples (18, 43). Two time points (separated by 15 ms) were used for the measurements of ^1H_N-R_2 rates, and the errors in the ^1H_N-Γ_2 PRE rates were calculated as described previously (18, 43). The transverse ^1H_N-R_2 relaxation rates were measured using 3D HNCO-based (44) and 2D (18) pulse schemes with TROSY readout for ΔGag and CA-SP1-NC/protease constructs, respectively. TROSY ^1H-^{15}N correlation spectra of the proteins were recorded before and after the PRE experiments to verify that no spurious intermolecular disulfides were formed during the course of the measurements.

ΔGag Cleavage Assay. ΔGag hydrolysis was carried out using our previously published protocol (3). Briefly, ΔGag constructs were incubated for 3 h at room temperature with 1 μM PR-O variants in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, and 1 mM TCEP. Aliquots (5 μL each) were taken at regular time intervals, mixed with SDS protein gel loading solution (Quality Biological; catalog no. 351-082-661), boiled at 99 °C for 2 min, and loaded onto a SDS/PAGE gel [18% (wt/vol) Tris-glycine gel; Life Technologies; catalog no. EC65055BOX]. Cleavage products were visualized by PageBlue staining (ThermoFisher Scientific; catalog no. 24620).

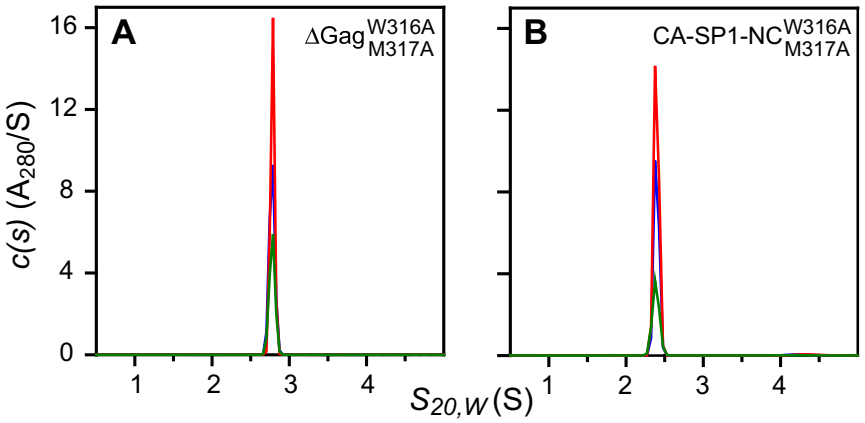

Fig. S1. Sedimentation velocity absorbance c(*s*) profiles for monomeric Gag constructs. (*A*) $\Delta Gag^{W316A}_{M317A}$ at loading concentrations of 66 (blue), 20 (red), and 10 (green) μM and (*B*) CA-SP1-NC^{W316A}_{M317A} at loading concentrations of 84 (blue), 31 (red), and 10 (green) μM. The presence of a single species at 2.78 S with a mass of ~47 kDa for $\Delta Gag^{W316A}_{M317A}$ and at 2.40 S with a mass of ~33 kDa for CA-SP1-NC^{W316A}_{M317A} confirm that these constructs are entirely monomeric. Note that the highest concentration data (blue curves) were collected using 3-mm path-length cells. Identical profiles were obtained with the interference optical detection system. The buffer used comprised 20 mM sodium phosphate, pH 6.5, 1 mM TCEP, and 0.1 mM $ZnCl_2$; the $\Delta Gag^{W316A}_{M317A}$ sample also contained 300 mM NaCl, whereas the CA-SP1-NC^{W316A}_{M317A} sample contained 50 mM NaCl.

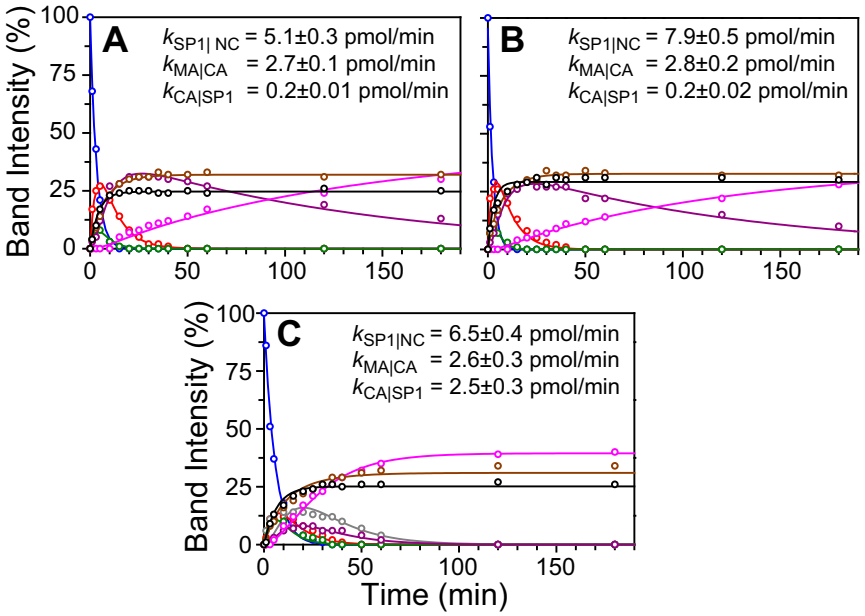

Fig. S2. Quantitative analysis of ΔGag cleavage by PR-O. (*A*) WT ΔGag, (*B*) monomeric $\Delta Gag^{W316A}_{M317A}$, and (*C*) $\Delta Gag^{MA|CA=SP1|NC}_{CA|SP1=SP1|NC}$ at molar ratios (in subunits) of 50:1, 50:1, and 100:1, respectively, Gag to protease. The concentration of PR-O was 1 μM (in subunits). The apparent rate constants, $k_{SP1|NC}$, $k_{MA|CA}$, and $k_{CA|SP1}$, are the rates of proteolysis at the SP1|NC, MA|CA, and CA|SP1 cleavage junctions, respectively. Experimental data are shown as circles, whereas the best-fit curves are shown by the continuous lines. The 1D gel analysis module from ImageQuant TL (GE Healthcare) was used to determine band intensities in SDS/PAGE gels (see Fig. 1 *B* and *C*), and the cleavage rates were obtained by nonlinear least squares fitting and solving the appropriate simultaneous first-order ordinary differential equations using the program DyanaFit (45). The color scheme is as follows: ΔGag, blue; intermediates, MA-CA-SP1 + MA-CA, red, CA-SP1-NC, green, CA-SP1, magenta; final products: CA, lilac, MA, brown, and NC, black. In the case of the $\Delta Gag^{MA|CA=SP1|NC}_{CA|SP1=SP1|NC}$ variant, the SDS/PAGE bands for the intermediates, MA-CA-SP1 and MA-CA, can be resolved (see Fig. 1*C*) and are depicted in red and gray, respectively.

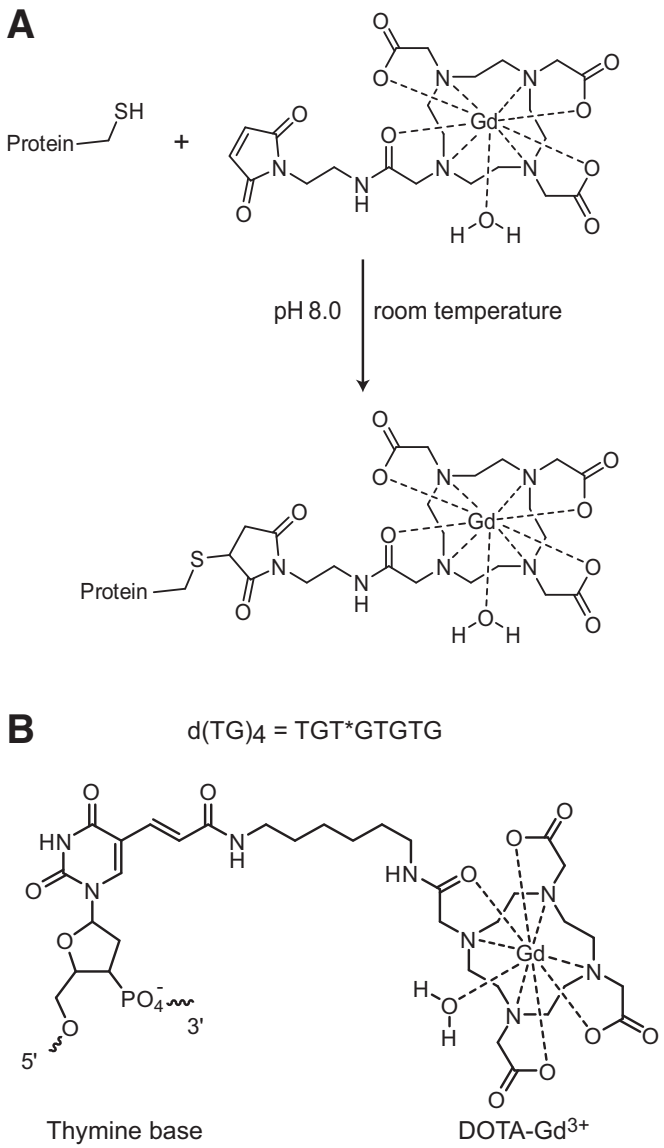

Fig. S3. Site-specific spin labeling with Gd(III). (*A*) Protein labeling. The paramagnetic probe, maleimido-DOTA-Gd^{3+}, is covalently attached to the protein of interest via a single surface-exposed cysteine residue using a chemo-specific reaction involving the formation of a S-C bond between the maleimide moiety of the probe and the thiol functional group of the cysteine. A DOTA-Gd^{3+} stock solution was incubated with the protein of interest for 1 h at room temperature at a protein to tag molar ratio of ~1:1.2. The buffer conditions were as follows: 100 mM Tris, pH 8. See *SI Materials and Methods* for additional details. (*B*) DNA labeling. The DOTA-Gd^{3+} paramagnetic tag is covalently attached via a linker to a thymine base by a C=C double bond involving the methyl carbon. In this instance the paramagnetic label was attached to the third base of the 8mer single stranded $(TG)_4$ DNA (the asterisk indicates the site of labeling) purchased from Biosynthesis and purified as described in *SI Materials and Methods*.

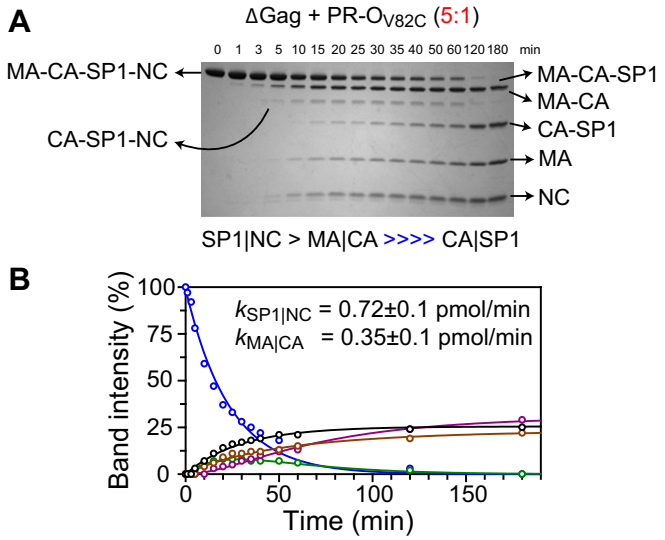

Fig. S4. ΔGag cleavage assay using PR-O$_{V82C}$. (*A*) WT ΔGag was incubated with PR-O$_{V82C}$ for 3 h at room temperature. The concentration of PR-O$_{V82C}$ was 10 μM (in subunits) with a Gag:protease molar ratio of 5:1. Although a large excess of PR-O$_{V82C}$ was used compared with the ΔGag cleavage assay with the WT PR-O (10 vs. 1 μM; see Fig.1*B*), Gag processing is significantly slower with PR-O$_{V82C}$, as the cleavage pattern observed with PR-O$_{V82C}$ after 3 h resembles that after ~10–15 min with WT PR-O. Note that despite the 3-h incubation of ΔGag with PR-O$_{V82C}$, the primary cleavage product from the hydrolysis at the CA|SP1 junction, namely CA, is not observed, making it difficult to quantitate the rate of CA|SP1 cleavage, other than to note that it is extremely slow compared with the other two cleavages (marked in blue). Buffer conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, and 1 mM TCEP. (*B*) Quantitative analysis of ΔGag cleavage by PR-O$_{V82C}$. The apparent rate constants $k_{SP1|NC}$ and $k_{MA|CA}$ are Gag proteolysis rates (pmol/min) at the SP1|NC and MA|CA cleavage junctions, respectively. Experimental data are depicted as circles and the best-fit curves by continuous lines (see Fig. S2 and Table S1 for further details). The color scheme is as follows: ΔGag, blue; intermediates: CA-SP1-NC, green, and CA-SP1, magenta; final products: MA, brown, and NC, black.

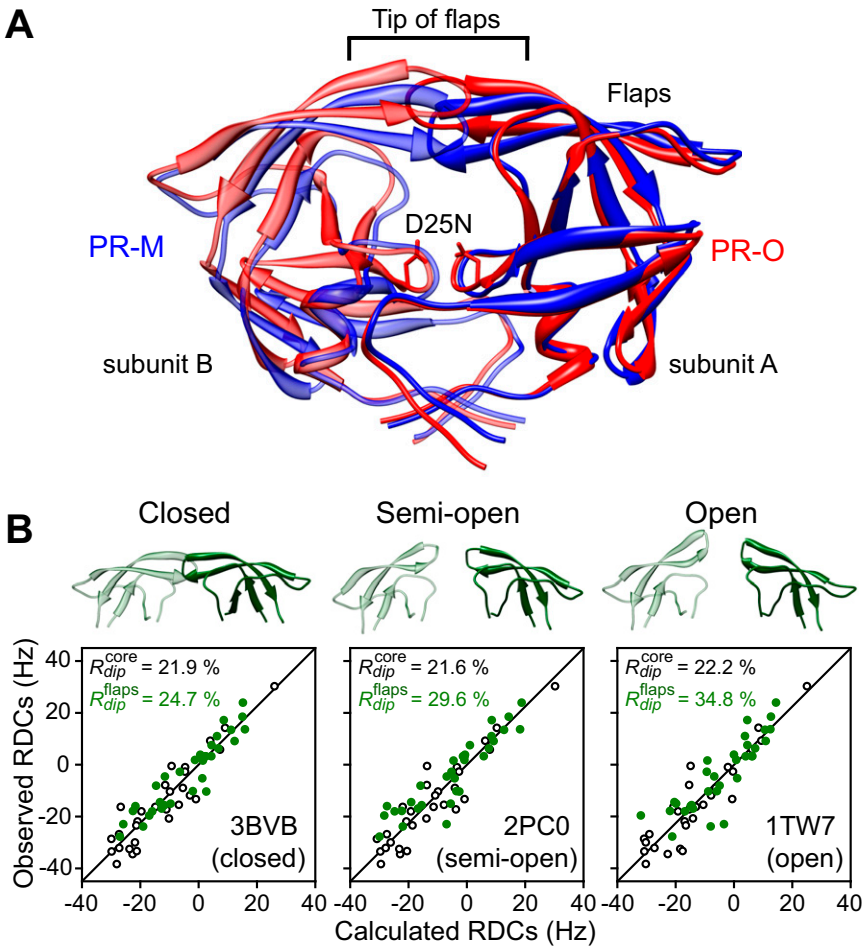

Fig. S5. Comparison of the structures of HIV-1 proteases: PR-M and PR-O. (*A*) Backbone superposition of catalytically inactive PR-M [Protein Data Bank (PDB) ID code 3BVB (46), blue] and PR-O_{V82C}^{D25N} (red), with subunits A and B depicted in solid and semitransparent ribbons, respectively. The core region of subunit A (residues 10–23, 62–73, and 87–93) was used for the superposition (Cα RMSD = 0.92 Å). The model of PR-O_{V82C}^{D25N} was calculated from NMR backbone chemical shifts (^{15}N, $^{1}H_N$, $^{13}C_\alpha$, $^{13}C_\beta$, and $^{13}C'$) and backbone amide RDCs using the program CS Rosetta (19, 20). (*B*) Backbone RDC analysis of PR-O_{V82C}^{D25N}. Panels show best-fit agreements obtained by singular value decomposition (SVD) between $^{1}D_{NH}$ RDCs measured in phage pf1 (~11 mg/mL) and those calculated from coordinates of three high-resolution (≤1.4 Å) X-ray structures of PR-M, each representing a unique flap conformation, closed [PDB ID code 3BVB (46)], semiopen [PDB ID code 2PC0 (47)], and open [PDB ID code 1TW7 (48)]. (Ribbon diagrams of the three flap conformations are shown above the corresponding panels.) The fits for the core region (residues 10–23, 62–73, and 87–93) are shown as black circles. The agreement between observed and calculated RDCs for the flaps (residues 30–61 and 74–84) are depicted by filled-in green circles, with predicted values calculated from the alignment tensor parameters obtained from the SVD fits to the core region. Based on RDC *R*-factors, the upper limit for the population of the open-flap conformation is <10% (49). The RDC *R*-factor, R_{dip}, is given by $\{<(D_{obs} - D_{calc})^2>/(2 < D_{obs}^2>)\}^{1/2}$, where D_{obs} and D_{calc} are the observed and calculated RDC values, respectively (50). Buffer and experimental conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM TCEP, and 30 °C.

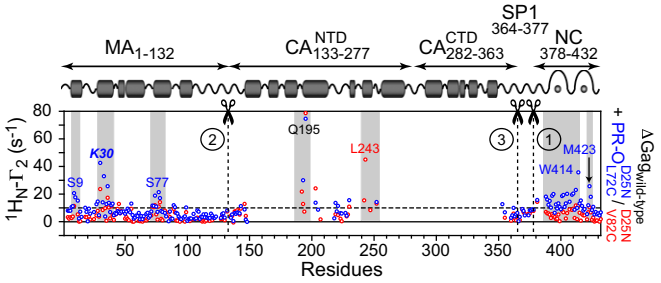

Fig. S6. Overview of fast transient WT ΔGag–protease interactions. Intermolecular PRE profiles observed between paramagnetically tagged PR-O variants and $^{15}N/^{13}C/^{2}H$-labeled WT ΔGag. Secondary structure elements and Gag domain organization are indicated above the panels. PREs arising from addition of paramagnetically(Gd^{3+}) tagged PR-O^{D25N}_{L72C} and PR-O^{D25N}_{V82C} are displayed as blue and red circles, respectively. Gag cleavage sites are indicated by vertical dashed lines and scissors. The numbers in circles indicate the order in which each site is cleaved, with the SP1|NC site being the first cleaved. ΔGag regions that come into close transient contact with protease ($^{1}H_N$-$\Gamma_2 \geq 10\ s^{-1}$) are delineated by transparent gray bars. A few of the residues that exhibit large PREs and undergo drug-resistant mutations are labeled with the latter in bold italics. Buffer and experimental conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, and 30 °C. The concentration of protease was 100 μM (in subunits) with a 3:1 molar ratio of Gag to protease. Severe resonance line broadening of resonances as a consequence of monomer-dimer exchange of the CA_{CTD} results in far fewer PRE data points within the capsid region of WT ΔGag compared with the monomeric variant $\Delta Gag^{W316A}_{M317A}$.

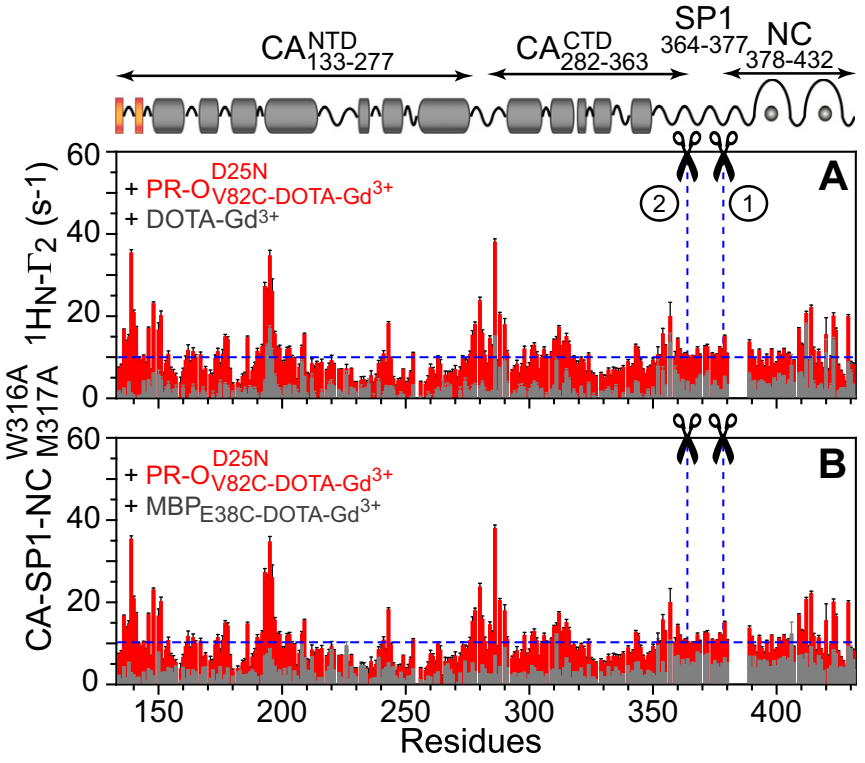

Fig. S7. Intermolecular PRE controls for transient interactions between Gag and paramagnetically labeled protease. Negative PRE controls (gray bars) are provided by the intermolecular PRE profiles observed on $^{15}N/^{2}H$-labeled CA-SP1-NC^{W316A}_{M317A} on addition of (*A*) soluble DOTA-Gd^{3+} and (*B*) apo-maltose binding protein (MBP_{E38C}) conjugated to a paramagnetic maleimido-DOTA-Gd^{3+} probe. (Note that residue E38C is located on the surface of MBP.) For comparison, the intermolecular PREs observed on addition of paramagnetically(Gd^{3+}) tagged PR-O^{D25N}_{V82C} are shown as red bars. A few small solvent PREs with $^{1}H_N$-Γ_2 values in the range of 10–17 s^{-1} (at residues^{194}HQA^{196}, R286, G357, G412) are observed with DOTA-Gd^{3+} owing to random collisions at solvent-exposed, surface accessible sites; these solvent PREs are much reduced ($^{1}H_N$-$\Gamma_2 \leq 10\ s^{-1}$) with MBP_{E38C}-DOTA-Gd^{3+}, as the larger size of MBP (~40 kDa) precludes preferential access to these sites. All PREs observed with either DOTA-Gd^{3+} or MBP_{E38C}-DOTA-Gd^{3+} (gray bars) are much smaller than those observed with PR-O^{D25N}_{V82C}-DOTA-Gd^{3+} (red bars). Gag cleavage sites are marked by vertical dashed lines and scissors. The numbers in circles indicate the order in which each site is cleaved with the SP1|NC site being the first cleaved. Buffer and experimental conditions were as follows: 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 0.1 mM $ZnCl_2$, and 30 °C. The concentration of CA-SP1-NC^{W316A}_{M317A} was 200 μM with 2:1 molar ratios of CA-SP1-NC^{W316A}_{M317A} to protease-DOTA-Gd^{3+}, MBP_{E38C}-DOTA-Gd^{3+}, and DOTA-Gd^{3+}.

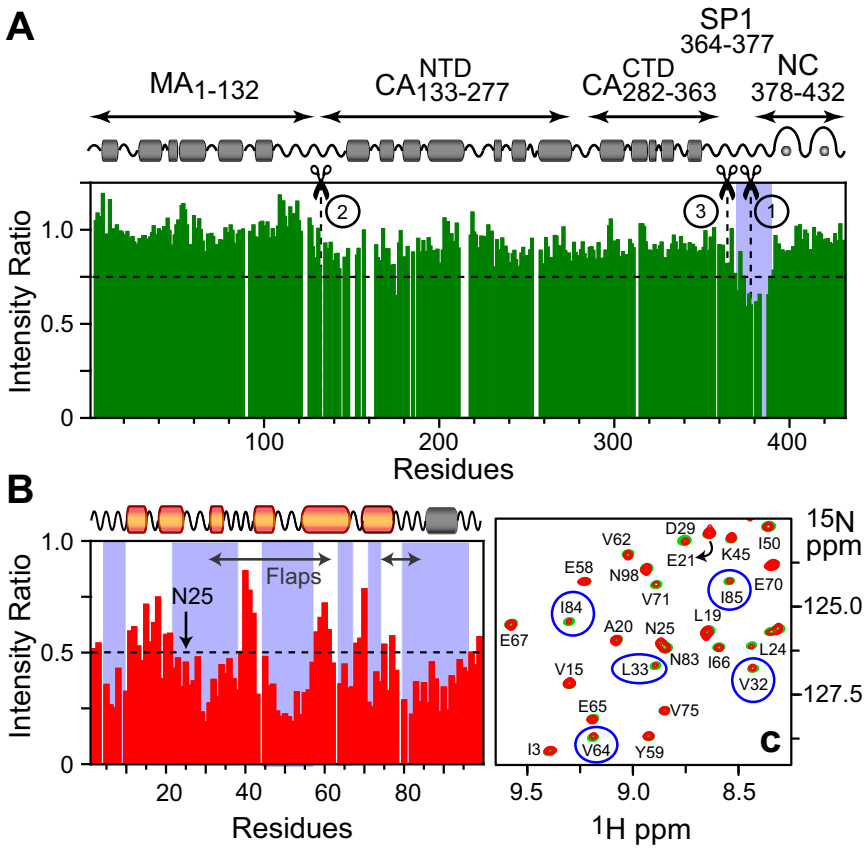

Fig. S8. Changes in $^{1}H_{N}/^{15}N$ cross-peak intensities and chemical shifts on protease–Gag association at the Gag cleavage sites. The reduction in cross-peak intensities of (*A*) $^{15}N/^{2}H$-labeled $\Delta Gag^{W316A}_{M317A}$ on addition of excess of PR-O^{D25N}_{V82C}, and (*B*) PR-O^{D25N}_{V82C} on addition of excess of $\Delta Gag^{W316A}_{M317A}$, is indicative of intermediate exchange on the chemical shift time scale. Secondary structure elements, Gag domain organization, location of protease flaps and active site asparagine (D25N) are marked. Gag cleavage sites are delineated by vertical dashed lines and scissors. The numbers in circles indicate the order in which each site is cleaved with the SP1|NC site being cleaved first. Regions of Gag and protease that exhibit significant reduction in cross-peak intensities are highlighted by the semitransparent blue bars. (*C*) Overlay of an expanded region of the ^{1}H-^{15}N TROSY correlation spectra of $^{15}N/^{2}H$-labeled PR-O^{D25N}_{V82C} in the absence (green) and presence (red) of $\Delta Gag^{W316A}_{M317A}$ at natural isotopic abundance. No significant chemical shift perturbations were observed in PR-O signals on addition of excess monomeric $\Delta Gag^{W316A}_{M317A}$. PR-O residues that exhibit a large reduction in cross-peak intensity on addition of $\Delta Gag^{W316A}_{M317A}$ are marked with blue circles. Buffer and experimental conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM TCEP, and 30 °C. The concentration of isotopically labeled protein was 50 μM with a molar ratio of unlabeled to isotopically labeled protein of 3:1 (in subunits).

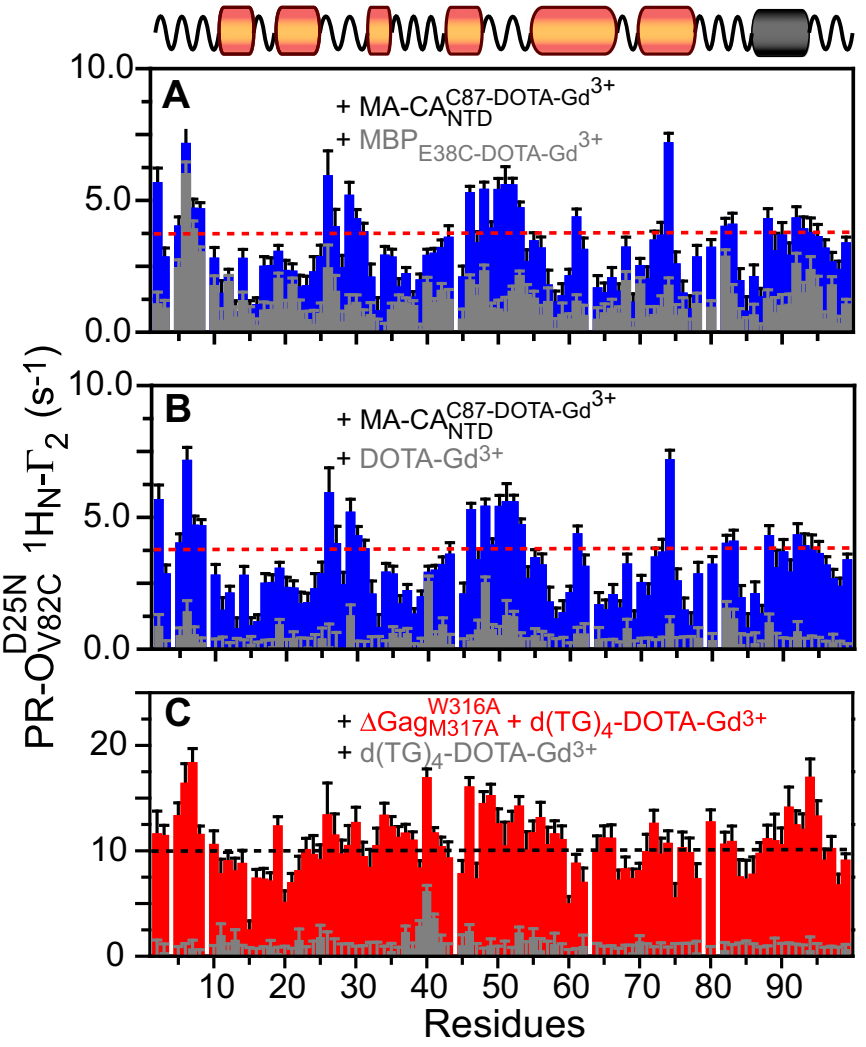

Fig. S9. Intermolecular PRE controls for the transient interactions between protease and paramagnetically labeled Gag. Negative PRE controls (gray bars) are provided by the intermolecular PRE profiles observed on $^{15}N/^{2}H$-labeled PR-O$^{D25N}_{V82C}$ on addition of (*A*) MBP$_{E38C}$-DOTA-Gd^{3+}, (*B*) soluble DOTA-Gd^{3+}, and (*C*) d(TG)$_4$-DOTA-Gd^{3+}. For comparison, intermolecular PRE profiles observed on addition of the paramagnetically tagged Gag variants MA-CA$^{C57T,C87}_{NTD}$-DOTA-Gd^{3+} and ΔGag$^{W316A}_{M317A}$ + d(TG)$_4$-DOTA-Gd^{3+} are shown as blue and red bars, respectively. With MBP$_{E38C}$-DOTA-Gd^{3+} (*A*), only residues 5–8 of PR-O show any significant intermolecular PREs (^1H_N-Γ$_2$ ~4–7 s^{-1}); the remaining regions of PR-O exhibit PREs that are barely above the background (^1H_N-Γ$_2$ ≤ 2.5 s^{-1}). Negligible intermolecular PREs are observed with soluble DOTA-Gd^{3+} with the exception of Gly40, which is highly solvent exposed (*B*). The same region (residues 39–41) also exhibits a noticeable intermolecular PRE effect in the presence of d(TG)$_4$-DOTA-Gd^{3+} (*C*). Based on the fact that the PREs observed with the paramagnetically tagged Gag variants are much larger than and distinct from those for the negative controls, we conclude that the former arise from transient encounter complexes between PR-O and Gag polyproteins. Buffer and experimental conditions were as follows: 20 mM sodium phosphate, pH 6.5, 0.1 mM $ZnCl_2$, and 30 °C; 50 mM NaCl was used in *A* and *B* and 300 mM in *C*. The concentration of $^{15}N/^{2}H$-labeled PR-O$^{D25N}_{V82C}$ was 150 μM (in subunits). The molar ratios of PR-O to MA-CA$^{C57T,C87}_{NTD}$-DOTA-Gd^{3+} (blue) was 5:1; to ΔGag$^{W316A}_{M317A}$ and d(TG)$_4$-DOTA-Gd^{3+} (red), 3:1.2 and 3:1, respectively; to MBP$_{E38C}$-DOTA-Gd^{3+} (gray), 3:1; and to d(TG)$_4$-DOTA-Gd^{3+} (gray), 3:1.

Table S1. Proteolysis rates (pmol/min) of ΔGag variants by HIV-1 protease

Apparent rate constant (pmol/min)	ΔGag +PR-M (50:1)*	ΔGag +PR-O (50:1)	$\Delta Gag^{W316A}_{M317A}$ +PR-O (50:1)	$\Delta Gag^{MA\|CA=SP1\|NC}_{CA\|SP1=SP1\|NC}$ +PR-O (100:1)	ΔGag +PR-O_{V82C} (5:1)†
$k_{SP1\|NC}$	6.1 ± 0.6	5.1 ± 0.3	7.9 ± 0.5	6.5 ± 0.4	0.72 ± 0.1
$k_{MA\|CA}$	2.2 ± 0.3	2.7 ± 0.1	2.8 ± 0.2	2.6 ± 0.3	0.35 ± 0.1
$k_{CA\|SP1}$	0.4 ± 0.06	0.2 ± 0.01	0.2 ± 0.02	2.5 ± 0.3	—

The apparent rate constants for ΔGag cleavage, $k_{SP1|NC}$, $k_{MA|CA}$, and $k_{CA|SP1}$, represent Gag proteolysis rates at the SP1|NC, MA|CA, and CA|SP1 cleavage junctions, respectively. The 1D gel analysis module from ImageQuant TL (GE Healthcare) was used to determine band intensities in SDS/PAGE gels (see Fig. 1 *B* and *C* and Fig. S3). Cleavage rates were obtained by nonlinear least-squares fitting and solving the appropriate simultaneous first-order ordinary differential equations using the program DyanaFit (45). See Fig. S2 for additional details. ΔGag constructs were incubated with PR-M and PR-O variants for 3 h at room temperature. Gag to protease molar ratios are noted in parentheses. The concentration of WT PR-M and PR-O was 1 μM (in subunits), whereas the concentration of PR-O_{V82C} was 10 μM (in subunits). Buffer conditions were as follows: 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM $ZnCl_2$, and 1 mM TCEP.

*Raw data used to derive Gag hydrolysis rates with PR-M were taken from Deshmukh et al. (3).

†The apparent rate constant $k_{CA|SP1}$ could not be determined in the case of PR-O_{V82C} due to lack of formation of the cleavage product, CA, within the 3-h timeframe of the experiment, owing to the lower catalytic efficiency of the mutated protease (compare Fig. S3).

Table S2. Mutations in protease and Gag associated with exposure and/or resistance to protease inhibitors

Protease*	Active site	Core	Flaps
	L24I	L10I (/F/R/V/C)	V32I
		A71V (/I/T/L)	L33I(/F)
		G73C (/S)	E34Q
		N88S (/D)	M36I(/L/V)
		L90M	M46I(/L)
		I93L(/M)	I47V(/A)
			G48V
			I50L(/V)
			F53L(/Y)
			I54L(/V/M/T/A)
			V82A(/T/F/I)
ΔGag†	Matrix	Capsid	Nucleocapsid
	E12K	T186M	I389T
	K30R	M200I	V390A(/D)
	L75R	H219Q(/P)	T401V
	R76K		R409K
	Y79F		
	T81A		

The primary literature sources for the mutations are Wensing et al. (17), Fun et al. (16), Gatanaga et al. (28), Garcia-Diaz (15), Parry et al. (14), and Sutherland et al. (29). All listed residues exhibit large intermolecular PREs, with the exception of Met200 of ΔGag, which is close to the patch of Gag residues (residues 192–198) that exhibit large PREs (see Figs. 2 and 3).

*Protease regions are defined as follows (23): active site (residues 24–29), core (residues 10–23, 62–73 and 87–93), and flaps (residues 30–61 and 74–84), and are colored in orange, magenta and green ribbons, respectively, in Fig. 2*A*.

†ΔGag domain organization is as follows: matrix (residues 1–132), capsid (residues 133–363), spacer peptide 1 (residues 364–377), and nucleocapsid (residues 378–432) (see Fig. 1*A*).