Published in final edited form as:

J Mol Biol. 2007 October 26; 373(3): 573–586.

Insights From Atomic-Resolution X-Ray Structures Of Chemically-Synthesized Hiv-1 Protease In Complex With Inhibitors

Erik C.B. Johnson^{1,2,‡}, Enrico Malito^{3,‡}, Yuequan Shen^{3,&‡}, Brad Pentelute⁴, Dan Rich⁵, Jan Florián⁶, Wei-Jen Tang³, and Stephen B.H. Kent^{1,2,4}

¹ Department of Biochemistry and Molecular Biology, The University of Chicago, IL 60637, USA

² Institute for Biophysical Dynamics, The University of Chicago, IL 60637, USA

³ Ben-May Department for Cancer Research, The University of Chicago, IL 60637, USA

⁴ Department of Chemistry, The University of Chicago, IL 60637, USA

⁵ School of Pharmacy, The University of Wisconsin-Madison, Madison, WI 53705, USA

⁶ Department of Chemistry, Loyola University Chicago, Chicago, IL 60626, USA

Summary

The HIV-1 protease is an aspartyl protease essential for HIV-1 viral infectivity. HIV-1 protease has one catalytic site formed by the homodimeric enzyme. We have chemically synthesized fully active HIV-1 protease using modern ligation methods. When complexed with the classic substrate-derived inhibitors JG-365 and MVT-101, the synthetic HIV-1 protease formed crystals that diffracted to 1.04 and 1.2 Å resolution, respectively. These atomic resolution structures revealed additional structural details of the HIV-1 protease interactions with its active site ligands. Heptapeptide inhibitor JG-365, which has a hydroxyethylamine moiety in place of the scissile bond, binds in two equivalent antiparallel orientations within the catalytic groove, whereas the reduced isostere hexapeptide MVT-101 binds in a single orientation. When JG-365 was converted into the natural peptide substrate for molecular dynamic simulations, we found putative catalytically competent reactant states for both lytic water and direct nucleophilic attack mechanisms. Moreover, free energy perturbation calculations indicated that the insertion of catalytic water into the catalytic site is an energetically favorable process.

Keywords

HIV-1 protease; chemical protein synthesis; native chemical ligation; MVT-101; JG-365

Introduction

The human immunodeficiency virus 1 (HIV-1) encodes a protease for processing of the gag and gag-pol viral polyprotein precursors, and for the subsequent maturation of infective virion particles¹. Because the HIV-1 protease is essential for viral infectivity, it has been the focus

Address correspondence to: Wei-Jen Tang, PhD, Ben-May Department for Cancer Research, The University of Chicago, Chicago, IL 60637, USA, Tel: 773-702-4331, Fax: 773-702-6260, email: wtang@uchicago.edu.

&Current address: College of Life Science, Nankai University, Tianjin, China 300071

‡Contribute equally

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of a number of successful drug design efforts to inhibit its enzymatic activity², and HIV-1 protease inhibitors are now an essential component of the standard highly active anti-retroviral therapy (HAART) for acquired immunodeficiency syndrome caused by infection with HIV-1. The HIV-1 protease is an aspartyl protease, all of which contain two conserved, catalytically-essential aspartate residues that define this class of enzymes. In contrast to the single chain cell-encoded aspartyl proteases, the HIV-1 protease is a symmetric homodimer^{3, 4} consisting of two 99-amino acid polypeptide chains. In addition to providing one catalytic aspartate residue, each chain also contributes a “flap” that folds over the bound substrate.

Although the HIV-1 protease has been well-studied, the exact mechanism of peptide bond hydrolysis remains to be unambiguously experimentally determined. For example, peptide bond hydrolysis in the HIV-1 protease is believed to be initiated by nucleophilic attack of a water molecule activated by one of the aspartate side chains⁵, but the water molecule nucleophile has never been directly observed in enzyme-inhibitor crystal structures.

As the first step in our efforts to elucidate the exact mechanism of this enzyme, we report here the total chemical synthesis of fully active HIV-1 protease by modern chemical ligation methods. We also report atomic resolution X-ray crystal structures of the synthetic enzyme in complex with two different inhibitors that mimic proposed intermediates formed during catalysis. Finally, we use one of the reported crystal structures as a starting point for classical molecular dynamics (MD) and free energy perturbation (FEP) simulations to examine energetic and structural aspects of the insertion of the catalytic water next to the scissile bond of the substrate bound in the enzyme active site.

Results

Chemical synthesis of HIV-1 protease

The strategy used for the chemical synthesis of the HIV-1 protease is shown in Table 1. Previous syntheses of the HIV-1 protease from ligation of two peptide segments relied on the synthesis of two ~50 residue peptides, one of which contained a C-terminal C^α-thioacid and the other an N-terminal bromoacyl group for the purposes of ligation^{6, 7}. This type of ligation creates a non-native thioester bond at the ligation site. A native amide bond can be created at the ligation site using native chemical ligation (NCL), and therefore this type of ligation chemistry was used. NCL requires an N-terminal cysteine residue for ligation to a peptide-^αthioester. We took advantage of the fact that cysteine could be converted into ‘pseudolysine’ by alkylation with 2-bromoethylamine using tight pH control (Figure 1). Lysines are present at positions 41, 43, 45, and 55 in the protein sequence. Among these lysine residues, the ligation site most favorable for NCL is at position 41, which involves ligation of Cys41-99 to a 1–40 alanine thioester. With the ligation site at residue 41, syntheses of an N-terminal peptide thioester of 40 residues in length and a cysteine peptide of 59 residues in length were required. Both of these peptides were synthesized using the highly-optimized stepwise SPPS protocol previously reported for the synthesis of the 1–99 polypeptide⁸.

Ligation of the peptide segments 1–40^αthioester and Cys41-99 was performed at pH 7.0 using standard protocols, with 200 mM mercaptophenylacetic acid (MPAA) added as a catalyst⁹. Ligation proceeded to completion within one hour. After ligation was complete, piperidine was added to the solution at a concentration of 20% vol/vol to remove formyl protection on tryptophan. The product 1–99 polypeptide was then purified by RP-HPLC.

Alkylation of the introduced cysteine residue with bromoethylamine hydrobromide was performed at pH 7.5 in order to convert it into a pseudolysine residue. After 3 hours, LC-MS analysis showed that alkylation was ~80–90% complete, but that a poorly-separated side product was beginning to emerge on the front side of the peak that was consistent in mass with

a second alkylation event. Therefore, the reaction was quenched after 3.5 hours and the alkylated product purified by RP-HPLC. Overall isolated yield after purification for the synthesis of the 1–99 polypeptide containing the pseudolysine residue was 26%, based on the limiting peptide segment (for synthetic details, see Supplementary Material).

Characterization of synthetic HIV-1 protease

The pure 1–99 polypeptide was then folded by dialysis. The presence of correctly folded protein was determined by the ability of the enzyme to cleave the p24/p15 substrate peptide derived from the gag-pol polyprotein (Figure 2a), as well as a fluorogenic peptide substrate (Figure 2b)¹⁰. Measured at 50 mM sodium acetate (pH 5.6), the K_m and k_{cat} values for the enzyme were 27.5 μM and 26.5 s^{-1} , respectively, which is in agreement with previously reported values¹¹. Folding yield was 54%, for an overall synthetic yield of 14% based on the limiting peptide segment.

In order to determine the molecular structure of the folded HIV-1 protease, we crystallized the synthetic protein with two different inhibitors, JG-365 and MVT-101. Crystals of both enzyme-inhibitor complexes grew within 3–5 days, and diffracted to 1.04 and 1.2 Å for HIV1-PR•JG-365 and HIV1-PR•MVT-101 complexes, respectively. The structures were solved by molecular replacement, and refined by least squares methods in SHELX-97¹² (statistics for data collection and refinement are shown in Table 2). The atomic resolution of the diffraction data allowed refinement of individual anisotropic displacement parameters for all atoms in the models (Table 3). The excellent quality of the electron density maps permitted the placement of several residues into alternative conformations and the introduction of several partially occupied water sites. However, two regions in the structures of both complexes could not be unambiguously modeled: residues 16–18 and residues 34–37 of chain B were disordered. These residues are positioned on the surface of the protein and are involved in contacts with symmetry related molecules. Final R_{fac} and R_{free} values of 12.70 and 16.88% were obtained for the HIV1-PR•JG-365 complex, and 13.85 and 19.02% for the HIV1-PR•MVT-101 complex.

The presence of the non-natural amino acids in the synthetic enzyme was readily apparent. Methionine and cysteine residues were modified to include carbon atoms instead of sulfur (L-norleucine and L- α -amino-n-butyric acid (Abu), respectively). In addition, the CG atom of Lys41 was replaced by a sulfur atom (Slys). The primary structure of the synthetic protease utilized in these studies contained several amino acid substitutions: Gln7 and Leu33 were mutated to Lys and Ile, respectively, giving rise to an autolysis-resistant mutant¹³. All of these residues were identified in the electron density map, confirming the fidelity of our chemically synthesized HIV-1 protease.

The interaction of synthetic HIV-1 protease with JG-365

We chose JG-365 for co-crystallization because this heptapeptide analog has high affinity (240 pM) to HIV-1 protease and it contains a non-cleavable hydroxyethylamine isostere of the scissile amide bond (Figure 3a)¹⁴. JG-365 is posited to mimic the tetrahedral intermediate formed by the nucleophilic attack of water on the scissile amide bond^{14, 15}.

The synthetic enzyme•JG-365 complex crystallized in the $P2_12_12_1$ space group, the same as previously observed¹⁶. The conformation of the protease in complex with JG-365 was compared with the previously published 2.4 Å model¹⁶ (pdb ID: 7HVP) by using *LSQKAB* as implemented in CCP4¹⁷. This comparison revealed close agreement with the previous model, with a root mean square deviation (rmsd) of 0.44 Å for 197 common α -carbons in both subunits. During refinement, inspection of both σ_A -weighted $2F_o - F_c$ and $F_o - F_c$ difference electron density maps revealed that the JG-365 inhibitor could be best modeled as occupying two alternate orientations in the binding groove of the enzyme with occupancy of 50% (Figure 3).

Since the HIV-1 protease is a symmetric homodimer, it is conceivable that a peptide inhibitor could occupy the binding groove in two different orientations, as has previously been reported for other inhibitors^{18, 19}. Interestingly, the previously reported 2.4 Å resolution structure of the HIV1-PR•JG-365 complex showed JG-365 bound in only one orientation in the catalytic groove¹⁶.

Related to the alternate orientations of JG-365 in the active site of HIV-1 protease, the conserved water 301 can be seen in two distinct positions (HOH/27 and HOH/28 chain E in our structure), and has been modeled assigning partial occupancies to each (Figure 3b and 3c). The H-bonding interactions made by the water molecule with the backbone carbonyls between P1 and P2 and between P1' and P2' on JG-365, and the backbone amide group of the flaps (residues Leu50/A and Leu50/B) (Figure 3b and 3c) create approximately tetrahedral geometry. These observations confirm the role of water 301 that contributes in inducing the fit of the flaps over the inhibitor¹⁶. Similarly, in the JG-365 complex the carbonyl group of the peptide bond of the flaps (between residues Ile50 and Gly51 in both chain A and B) can be modeled in two opposite directions, flipped by 180 with respect to each other (Figure 3b and 3c). The carboxylate oxygen atoms of the catalytic aspartates 25 and 25' are positioned as previously reported¹⁶. They are arranged in almost the same plane and reside within a 2.4–2.96 Å distance from the hydroxyl group on the tetrahedral carbon of JG-365. With respect to a true isosteric analog, JG-365 contains one extra backbone atom, the methylene adjacent to the CH(OH) group. Interestingly, the catalytic aspartates form a hydrogen bond to the hydroxyl group of the inhibitor, which presumably mimics the putative water molecule nucleophile attacking the carbonyl carbon of the scissile bond substrate²⁰.

The interaction of synthetic HIV-1 protease with MVT-101

We also used MVT-101 to determine the structure of the protease-inhibitor complex. MVT-101 contains a reduced peptide bond isostere ([CH₂-NH]) in place of the scissile amide⁴. HIV1-PR•MVT-101 crystallized in the P₂₁2₁2₁ space group, the same as previously reported⁴. The overall conformation of the protease is similar to the previously published structure; superposition of our model was performed with the algorithm *LSQKAB*²¹ using the previously reported 2.3 Å protease-MVT-101 complex structure (pdb ID: 4HVP). This comparison gave an rmsd of 0.43 Å for 197 common α-carbons. In contrast with JG-365, the MVT-101 inhibitor was best modeled as occupying only one orientation (Figure 4a). Previously reported structures of the protease in complex with MVT-101 showed the inhibitor bound either in the orientation opposite to that of our current model⁴ or in a ratio of two opposite orientations²².

The carbonyl groups on the inhibitor molecule sites P1 and P1' are positioned within H-bond distances from the conserved water molecule 301 (named HOH/7 chain E in this structure) (Figure 4b). As previously shown, this water molecule is tetrahedrally coordinated with the backbone amino group of the residues Leu50/A and Leu50/B of the flaps. In contrast with the JG-365 complex, in the MVT-101 complex there is no evidence of disorder either in the position of the conserved water molecule or in the conformation of the flaps. This is consistent with the single orientation of MVT-101 in the binding groove. The additional important interactions made by MVT-101 with the enzyme are between the catalytic aspartates 25 and 25' and the reduced bond between the sites P1 and P1'. Given our high resolution data, hydrogen atoms were included in the model at their theoretically calculated positions using SHELXL¹². The putative water nucleophile is not observed.

Computer simulations to address the absence of lytic water—To provide structural and energetic information that complement our X-ray diffraction data, we carried out classical molecular dynamics (MD) and free energy perturbation (FEP) simulations of several reactant state conformations for the complex with the peptide substrate rather than inhibitor bound in

the active site. These conformations differed by the presence and absence of putative water nucleophile (lytic water) and the topology of H-bonds in the active site. We used both JG-365 and MVT-101 bound HIV-1 protease structures as the initial model. Here, we only reported the calculation using the former model since the conformational state derived from the latter model is unstable in our nanosecond time-scale simulation. We replaced JG-365 with the corresponding peptide substrate, in the absence of lytic water. From two orientation of JG-365 at the catalytic site, we obtained two structures that differ in relative positions of the two catalytic aspartates and the substrate. Of those, one is stable in our nanosecond time-scale simulation. The positions of substrate and important water molecules in relation to the protease are shown in Figure 5.

The conformation depicted as state 1A in Figure 6 is the most reminiscent of the geometry of the two catalytic Asp residues in the crystal structure with JG-365 (Figure 6, yellow sticks) in the sense that it contains the OD1H $\cdots$ OD1 H-bond. Using a different set of initial simulation conditions we obtained state 1B, in which Asp25' is perfectly positioned for the direct nucleophilic attack on the carbon of the peptidyl group. Both 1A and 1B conformations are stable on the nanosecond time scale. The rearrangement of hydrogen bonds to form the Asp25'–OD2H $\cdots$ O-substrate hydrogen bond (state 1C in Figure 6) is associated with a large unfavorable free energy difference of 4.9 ± 1.6 kcal/mol (Table 4).

However, based on the calculated standard free energy of -3.5 ± 2.1 kcal/mol for the insertion of lytic water (Figure 6 and Table 4), neither of the configurations 1A – 1C is likely to be significantly populated in the substrate-bound HIV-1 protease in aqueous solution. The most stable conformation containing lytic water (state 1D in Figure 6) has the two Asp residues interconnected by an eight-membered ring, which includes H-bonds formed by both hydrogen atoms of lytic water and an H-bond between the OD1 and OD2 atoms of the Asp25' and Asp25 residues, respectively. This OD1 $\cdots$ H $\cdots$ OD2 H-bond is in fast equilibrium with a less stable OD1–H $\cdots$ OD1 H-bond, which is present in about 5% of all configurations sampled along the simulated 1 ns trajectory.

The rearrangement of these H-bonds to form a pre-catalytic configuration that features the Asp25'–OD2H $\cdots$ O-substrate H-bond (Figure 6, state 1E) is an energetically unfavorable process. This process can be quantitatively characterized by a free energy difference of 3.7 ± 0.9 kcal/mol (Table 4). In the conformation 1E, the nucleophilic oxygen is located on average 3.0 Å from the carbon of the peptidyl group. This proximity and the accompanying hydrogen bond network make conformation 1E an ideal starting point for the catalysis initiated by the nucleophilic attack of the lytic water.

Discussion

Our goal is to experimentally determine the exact catalytic mechanism of the HIV-1 protease. We reasoned that this task could be simplified with atom-by-atom control over the enzyme using a chemically-synthesized protease, access to atomic-resolution structural information on such a synthetic enzyme, and energetic data on the catalytic apparatus on which to guide our future experiments. Here, we report results regarding these three aims.

In order to elucidate the mechanistic details of peptide bond hydrolysis catalyzed by the HIV-1 protease, it is highly desirable to have atom-by-atom control over the enzyme in order to perturb the protein in a highly precise and controlled fashion. Total chemical protein synthesis using native chemical ligation allows such control over the covalent structure of the protein while maintaining a fully-native polypeptide backbone. This level of experimental control allows modifications to the catalytic aspartates and flaps that are not currently possible using recombinant molecular biology. As demonstrated by multiple experimental techniques

including reversed-phase HPLC, mass spectrometry, and X-ray crystallography, the synthetic HIV-1 protease prepared in this study was of very high purity and contained full catalytic activity.

Crystals of the synthetic protease in complex with two classical inhibitors, JG-365 and MVT-101, diffracted to atomic resolution. These structures revealed differences from previously published structures with respect to inhibitor binding within the catalytic groove of the enzyme. In the case of JG-365, our model unambiguously showed the inhibitor binding in opposite orientations within the groove, in a ratio of approximately 50%. With MVT-101, however, the inhibitor was observed to bind in only one orientation. Because the HIV-1 protease active site is formed from the dimerization of two identical subunits, it is certainly possible that a substrate could bind in either orientation within the binding groove as observed with JG-365. Why only one orientation is preferred for MVT-101 while both orientations are observed for JG-365 is not clear. Differences with previously published models of the HIV-1 protease in complex with these inhibitors may be due to the introduction of mutations from the autolytic mutant, differences in the crystallization conditions, crystal variability, and/or the resolution limit for the model building.

Energetic information gained from classical molecular dynamics and free energy perturbation simulations allow us to contribute new insights into the aspartyl protease catalytic mechanism, and provide a theoretical basis for future experiments. Although several catalytic mechanisms have been proposed for this family of proteases, the most widely accepted mechanism involves general acid/general base catalysis of water attack on the scissile amide bond of the substrate⁵. In this model, one active site aspartate acts as a general base in activating a water molecule, postulated to reside between the catalytic aspartic acid side chain carboxyls, that consequently acts as a nucleophile attacking the carbonyl carbon of the scissile amide bond. The other aspartic acid side chain is postulated to act as a general acid to protonate the carbonyl oxygen of the scissile amide bond.

The structures obtained in the current work, by virtue of their high resolution, provide a useful starting point for the examination of the energetics of water binding in the catalytic site. The FEP calculations used a protocol in which the equilibrated lytic water was slowly mutated to 'nothing' and the free energy cost of this water annihilation process was subtracted from the energy of the corresponding process occurring in aqueous solution. In these calculations, we utilized several variants of the flat-bottom harmonic constraints, the purpose of which was to prevent the annihilated water to leave the active site and to provide a reliable measure of the precision (standard deviation) of our results (Table 4). The calculated annihilation free energy in water of 6.9 ± 0.6 kcal/mol (Table 4) agrees with the negative of the observed solvation free energy of water in water of -6.4 kcal/mol²³. To relate the results of our calculations to the situation where the water molecules are actually known to bind, we used the identical computational methodology to calculate binding free energies of the conserved structural water W301, W3 and W4 (Figure 5). Considering that the concentration of water molecules in aqueous solution is 55.5 M, the calculated standard binding free energies of $+0.1 \pm 1.8$, -0.4 ± 1.9 , and -0.3 ± 0.9 kcal/mol for W301, W3 and W4, respectively, correspond to over 97% occupancy of each protein site by the structural water molecule. This result is consistent with the observation of these water molecules in the crystal structures, and with results of previous calculations²⁴. Based on a reasonable accuracy of the calculated free energies for structural waters we can consider the two calculated binding free energies for the lytic water of -3.5 ± 2.1 and -4.3 ± 2.5 kcal/mol (Figure 6 and Table 4) to be plausible.

Considering that the lytic water has not been observed in crystal structures of inhibitor-bound HIV-1 protease structures, it is surprising that the calculated affinity of the protein with bound substrate for the lytic water molecule is much larger than for any of the structural waters. Such

a high affinity is due to the fact that lytic water forms a strong hydrogen bond with a negatively charged Asp25 and a medium-strength hydrogen bond with neutral Asp25', and that these hydrogen bonds are not replaced when lytic water is removed. Of course, the lytic water would not be observed if significant steric repulsion from non-peptide inhibitors increased its binding free energy, which may be the case in all current crystal structures.

The calculated relative free energies (Figure 6 and Table 4) indicate that the catalytically competent reactant state for the lytic water mechanism (state 1E, Figure 6) is likely not a true ground state for the reaction but rather the reaction intermediate. Thus, although the topology of this conformation appears to be a perfect starting point for the reaction, and the corresponding activation barrier is consistent with the observed rate constants²⁴, the activation free energies calculated using the structure 1E as the ground state need to be increased by 3.7 kcal/mol to account for the proton relocation free energy (Figure 6)¹. However, this result does not make the alternative model of direct nucleophilic attack by Asp more viable because approximately the same energy (3.5 kcal/mol) is required to remove the lytic water molecule hindering the direct attack (Figure 6). In addition, the direct attack mechanism appears to have a 'handicap' of lacking the hydrogen bonding interaction capable of stabilizing developing negative charge on the oxygen of the carbonyl group in the transition state. Of course, proton translocation penalties in each of the two mechanisms could be lowered if the proton translocation via the rotation about the C_γ-C_δ single bond in Asp25' would occur in a concerted way together with the nucleophilic attack. Thus, it is quite possible that both mechanisms and/or their stepwise variants are operational in HIV-1 protease.

Using these insights into the catalytic mechanism of HIV-1 protease, in addition to baseline structural data at atomic resolution of the protease in complex with inhibitors that mimic two proposed catalytic intermediates, we are now in a position to rationally perturb the enzyme in a highly precise fashion using chemical synthesis in order to fully characterize the catalytic mechanism. Information gained about the mechanism of the HIV-1 protease will hopefully lead to improved inhibitors that can more effectively and specifically target this enzyme.

Materials and Methods

Materials

2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *S*-trityl-mercaptopropionic acid, *p*-methylbenzhydrylamine (MBHA) resin, Boc-Lys(2ClZ)-OCH₂-PAM-resin, and Boc-amino acids were obtained from Peptides International, Kentucky. Boc-Phe-OCH₂-PAM-resin was purchased from Applied Biosystems, Foster City. Side chain protecting groups used were: Arg(Tos), Cys(4MeBzl), Asp(cHex), Asn(Xan), Glu(cHex), His(Bom), Lys(2ClZ), Thr(Bzl), Trp(CHO), Tyr(BrZ). *N,N*-diisopropylethylamine (DIEA) was obtained from Applied Biosystems. *N,N*-Dimethylformamide (DMF), dichloromethane, diethyl ether, and HPLC grade acetonitrile were purchased from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon Products, New Jersey. HF was purchased from Matheson. All other reagents were purchased from Sigma-Aldrich. 4-mercaptophenylacetic acid (MPAA) was repurified by HPLC.

Peptide Synthesis

The 1-40^αCOSCH₂CH₂COArg thioester peptide was synthesized as described previously²⁵. The 1-40^αCOSR and fluorogenic substrate peptides were synthesized on MBHA-resin, the p24/

¹Although not explicitly shown in reference ²⁴, the reaction simulations starting from the tetrahedral intermediate relax via 1E to the shared proton situation (1D). Hence the 3.7 kcal/mol difference between the two structures is already included in the free energy profiles reported in reference ²⁴ (J. Åqvist, personal communication).

p15 substrate peptide was synthesized on Boc-Lys(2ClZ)-OCH₂-Pam-resin, and the Cys41-99 peptide was synthesized on a Boc-Phe-OCH₂-Pam-resin. Peptides were synthesized on a 0.4 mmol scale using the manual Boc chemistry *in situ* neutralization/HBTU protocol described previously²⁶, with slight modifications. Briefly, N^α-Boc removal was achieved with neat TFA, using one rapid wash followed by a 2 minute batch treatment. The N^α-deprotected peptide-resin was then subjected to a single flow wash with DMF, and the solvent drained from the resin. Amino acids (2.2 millimole) were dissolved in 4mL of 0.5M HBTU (i.e. 2.0 millimole) and were activated by addition of DIEA (1mL, 6.6 millimole) for 1 minute, and then added to the peptide-resin; coupling was carried out for 12 minutes, followed by a single DMF flow wash to remove excess activated amino acid and soluble co-products, and the solvent drained from the resin. The cycle of deprotection and coupling was repeated. After chain assembly was complete, the N^α-Boc group was removed as described above, and the peptide-resin was washed with dichloromethane and dried by aspiration. The peptide was cleaved from the resin and the side-chain protecting groups were simultaneously removed by treatment with anhydrous HF at 0°C for 1 hour, with 10% p-cresol added as a scavenger. After removal of HF by evaporation, the peptide was precipitated and washed with cold diethyl ether and dissolved in aqueous acetonitrile + 0.1% TFA, then lyophilized.

LC analysis

Analytical reversed-phase HPLC was performed on an Agilent 1100 system with a Microsorb C-4 (5 μm 2.1 × 50 mm) silica column packed in-house, at a flow rate of 0.5 mL/min. Peptides were eluted from the column using a gradient of 5–65% acetonitrile/0.08% TFA versus water/0.1% TFA. Preparative HPLC of crude peptides after SPPS was performed on a Vydac C-18 (50 × 250 mm) column, and preparative HPLC of 1–99 was performed on an Agilent C-3 (22 × 250 mm) column. Peptides were eluted from the column using an appropriate shallow gradient of acetonitrile/0.08% TFA versus water/0.1% TFA. Fractions containing the desired purified peptide product were identified by analytical LC-MS, then combined and lyophilized.

MS analysis

Peptide masses were obtained using LC-MS with on-line electrospray detection with an Agilent 1100 LC/MSD ion trap. MALDI MS analysis of the folded protein was performed by mixing a sample of the protein and a sample of sinapinic acid (saturated in aqueous acetonitrile + 0.1% TFA) in a 1:1 mixture. MALDI spectra were acquired on a Perseptive Biosystems Voyager-DE in linear mode.

Ligation

Ligation was performed on a 9.3 μmol scale (2 mM), in 6M GdmHCl phosphate buffer (0.2 M) at pH 7.0, with 200 mM MPAA added as a catalyst and 20 mM TCEP added as a reducing agent. After the ligation was complete, formyl protection was removed by addition of 1M cysteine and piperidine (20% by volume). After 20 minutes, the solution was acidified with HCl and purified by preparative RP-HPLC. Fractions containing the desired purified peptide product were identified by analytical LC-MS, then combined and lyophilized. After lyophilization, the peptide was redissolved in 6M GdmHCl phosphate buffer (0.2 M) with 20 mM TCEP. Bromoethylamine hydrobromide was added to a concentration of 200 mM, and the pH adjusted to 7.5. After 3.5 hours of reaction at pH 7.5, the solution was acidified with HCl and purified by preparative RP-HPLC. Fractions containing the desired purified peptide product were identified by analytical LC-MS, then combined and lyophilized.

Folding

13 mg of 1–99 polypeptide was dissolved in 12 mL 6M guanidine HCl phosphate buffer (0.2M, pH 7.4) and folded by a three-step dialysis procedure against sodium acetate buffer (pH 5.6)

using a 3500 MW cutoff membrane. The first step was against 2L of 50 mM sodium acetate buffer (pH 5.6) for 3 hours. The second step was against 2L of 25 mM sodium acetate buffer (pH 5.6) for 2 hours. The third step was against 2L of 10 mM sodium acetate buffer (pH 5.6) overnight at 4°C. Final conditions were 10 mM sodium acetate pH 5.6. The final concentration of folded protein was ~16 μ M, determined by absorbance at 280 nm using a calculated dimer extinction coefficient of 25120 M⁻¹ cm⁻¹. Folded protein was separated from precipitated protein material by centrifugation, and was used directly for further studies without purification.

Crystallization

Crystals were grown at room temperature by the hanging drop vapor diffusion method from a well solution consisting of 0.1M citrate, 0.2M sodium phosphate, 30% saturated ammonium sulfate, 10% DMSO, pH 6.0. Protein solution, concentrated to ~0.4 mM enzyme using a Centricon YM-3 3,000 molecular weight cutoff membrane, was preincubated with a 20-fold molar excess of either (S)-JG-365 or MVT-101, and was then mixed in a 2:1 (v/v) ratio with well solution. Crystals grew within 3–5 days and were frozen in liquid nitrogen using a cryoprotectant of 30% glycerol.

Structure determination and refinement

Data was collected at 14-BM-C and 19-BM stations in the Advanced Photon Source (APS) at Argonne National Laboratory and processed using HKL2000²⁷. PHASER²⁸ was used to obtain the initial phases for HIV1-PR•MVT-101 and HIV1-PR•JG-365 using the model of HIV-1 protease in complex with MVT-101 (pdb code: 4HVP)⁴. The starting models were initially refined by the CNS program²⁹ using data between 50–1.5 Å. After free R factor convergence³⁰, the models were fed into SHELX97¹² for full anisotropic temperature factor refinement, using data to the highest resolution shell and with low-resolution cutoff at 10 Å (Table 2 and Table 3). Iterative model building and refinement were performed using COOT³¹, and SHELXL¹² until the final $R_{\text{cryst}}/R_{\text{free}}$ values converged to 12.70/16.88% for HIV1-PR•JG-365 and 13.85/19.02% for HIV1-PR•MVT-101. Coordinates for HIV-1 protease in complex with JG-365 and MVT-101 have been deposited in the Protein Data Bank with accession codes 2j9j and 2j9k, respectively. Model analysis was carried out using programs of the CCP4 package¹⁷. Graphics were generated by Pymol².

Computer simulations

Starting structures for MD simulations were generated from the crystal structures of HIV-1 protease in complex with JG-365. Of the two complexes, the JG-365 complex was used for the calculations because the observation of both binding modes of the substrate indicates that the effect of the crystal field is negligible. Furthermore, the JG-365 sequence has proline as the N-terminal residue of the leaving group, which is identical with the sequence used for key isotope exchange experiments³². To mimic the substrate, the hydroxyethylamine moiety of JG-365 was altered to form regular peptide bonds. A water molecule was also added to the structures between the two catalytic aspartates. All crystallographic water molecules were preserved and modeled by the TIP3P model, and all residues farther than 21 Å from the carbonyl group of the substrate were kept in their electroneutral forms. New TIP3P water molecules were added by immersing the simulation sphere into the sphere of bulk water molecules.

MD simulations were carried out in a simulation sphere (24 Å radius) centered on the oxygen atom of the substrate carbonyl group using the program Q³³ and Amber 95 force field³⁴. These simulations sampled the reactant state equilibria prior to the catalytic reaction of HIV-1

²<http://pymol.sourceforge.net>

protease. Following a 100 ps equilibration period, MD simulations were carried out at 298K using the step size of 1 fs and the SHAKE algorithm³⁵. The trajectory analysis was carried out using the program VMD 1.8.4³⁶.

Relative free energies for the proton relocation were calculated from energies sampled on MD trajectories using the free energy perturbation (FEP)/alchemistic mutation approach³⁷. The proton relocation free energy was determined by slowly annihilating hydrogen in the hydrogen bond between Asp25' and Asp25 residues and concomitantly creating hydrogen in a hydrogen bond between Asp25' and the oxygen of the carbonyl group of the substrate (Figure 6). This mutation pathway was calculated in the presence of distance or torsion restraints to prevent the interchange of the initial and final states along the pathway (Table 4). The energetic contribution of these restraints was removed from the calculated free energies. In each case, the magnitude of the removed restraint contribution was smaller than the statistical error (Table 4). The atom charges and van der Waals radii were changed linearly between the end states of each mutation pathway. The parameters of these end states were the same as the standard Amber 95 parameters for the neutral Asp residue. The only difference between the two end states was in the location of the protonation site (i.e. OD1 or OD2). The binding free energy of the lytic water and structural waters (Table 4) was calculated using the thermodynamic cycle shown in Figure 4, Supplementary Material. Two legs of this cycle were obtained by FEP calculations of the free energy associated with the alchemistic mutation of the water molecule to nothing in the protein active site and in aqueous solution. The atom charges and van der Waals well-depths (ϵ) of the TIP3P water model were changed linearly and concomitantly between the end states of each annihilation pathway. Pathways of all FEP calculations were subdivided into 101 simulation windows, with sampling time of 10 ps per window. Energies were sampled every 10 fs and the first 10 energies in each window were excluded from the free energy averages.

Protein Data Bank accession codes

Coordinates for HIV-1 protease in complex with JG-365 and MVT-101 have been deposited in the Protein Data Bank with accession codes 2j9j and 2j9k, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Hookang Im and other members of the Tang Lab for assistance with the X-ray data collection, and Karl Jablonowski for his computer support and insightful advises. E.C.B.J. is funded by the MD/PhD Graduate Training in Growth and Development program at the University of Chicago (NIH T32 HD007009). Y.S. is funded by the AHA fellowship (0520123Z), and E.M. is funded by AHAF. This research was supported by NIH GM62548 to W.-J. Tang. We also gratefully acknowledge funding from the Department of Energy Genomes to Life Genomics Program (Grant DE-FG02-04ER63786).

References

1. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 1998;67:1–25. [PubMed: 9759480]
2. Swanstrom R, Erona J. Human immunodeficiency virus type-1 protease inhibitors: therapeutic successes and failures, suppression and resistance. *Pharmacol Ther* 2000;86:145–70. [PubMed: 10799712]
3. Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 1989;337:615–20. [PubMed: 2645523]

4. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 1989;246:1149–52. [PubMed: 2686029]
5. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Freeman; New York: 1999.
6. Schnolzer M, Kent SB. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* 1992;256:221–5. [PubMed: 1566069]
7. Baca M, Muir TW, Schnolzer M, Kent SBH. Chemical Ligation of Cysteine-Containing Peptides - Synthesis of a 22-Kda Tethered Dimer of Hiv-1 Protease. *Journal of the American Chemical Society* 1995;117:1881–1887.
8. Kent, SBH.; Alewood, D.; Alewood, P.; Baca, M.; Jones, A.; Schnolzer, M. *Innovations and Perspectives in Solid Phase Synthesis*. Epton, R., editor. Intercept Ltd; Andover, UK: 1992. p. 1. references cited therein
9. Johnson EC, Kent SB. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 2006;128:6640–6. [PubMed: 16704265]
10. Toth MV, Marshall GR. A Simple, Continuous Fluorometric Assay for Hiv Protease. *International Journal of Peptide and Protein Research* 1990;36:544–550. [PubMed: 2090647]
11. Szeltner Z, Polgar L. Rate-determining steps in HIV-1 protease catalysis. The hydrolysis of the most specific substrate. *J Biol Chem* 1996;271:32180–4. [PubMed: 8943273]
12. Sheldrick GM, Schneider TR. SHELXL: High-resolution refinement. *Macromolecular Crystallography, Pt B* 1997;277:319–343.
13. Mildner AM, Rothrock DJ, Leone JW, Bannow CA, Lull JM, Reardon IM, Sarcich JL, Howe WJ, Tomich CS, Smith CW, et al. The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry* 1994;33:9405–13. [PubMed: 8068616]
14. Rich DH, Green J, Toth MV, Marshall GR, Kent SBH. Hydroxyethylamine Analogs of the P17/P24 Substrate Cleavage Site Are Tight-Binding Inhibitors of Hiv Protease. *Journal of Medicinal Chemistry* 1990;33:1285–1288. [PubMed: 2184237]
15. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Krohn A, et al. Rational design of peptide-based HIV proteinase inhibitors. *Science* 1990;248:358–61. [PubMed: 2183354]
16. Swain AL, Miller MM, Green J, Rich DH, Schneider J, Kent SB, Wlodawer A. X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc Natl Acad Sci U S A* 1990;87:8805–9. [PubMed: 2247451]
17. Collaborative Computational Project N. The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* 1994;50:760–763.
18. Fitzgerald PM, McKeever BM, VanMiddlesworth JF, Springer JP, Heimbach JC, Leu CT, Herber WK, Dixon RA, Darke PL. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-Å resolution. *J Biol Chem* 1990;265:14209–19. [PubMed: 2201682]
19. Dreyer GB, Lambert DM, Meek TD, Carr TJ, Tomaszek TA Jr, Fernandez AV, Bartus H, Cacciavillani E, Hassell AM, Minnich M, et al. Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease: structure-activity analysis using enzyme kinetics, X-ray crystallography, and infected T-cell assays. *Biochemistry* 1992;31:6646–59. [PubMed: 1637805]
20. Suguna K, Padlan EA, Smith CW, Carlson WD, Davies DR. Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase from *Rhizopus-Chinensis* - Implications for a Mechanism of Action. *Proceedings of the National Academy of Sciences of the United States of America* 1987;84:7009–7013. [PubMed: 3313384]
21. Bailey S. The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* 1994;50:760–763.
22. Miller M, Geller M, Gribskov M, Kent SB. Analysis of the structure of chemically synthesized HIV-1 protease complexed with a hexapeptide inhibitor. Part I: Crystallographic refinement of 2Å data. *Proteins* 1997;27:184–94. [PubMed: 9061782]

23. Ben-Naim A, Marcus Y. Solvation Thermodynamics of Nonionic Solutes. *Journal of Chemical Physics* 1984;81:2016–2027.
24. Bjelic S, Aqvist J. Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry* 2006;45:7709–23. [PubMed: 16784222]
25. Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* 1999;96:10068–73. [PubMed: 10468563]
26. Schnolzer M, Alewood P, Jones A, Alewood D, Kent SB. In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* 1992;40:180–93. [PubMed: 1478777]
27. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* 1997;276:307–326.
28. Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D-Biological Crystallography* 2001;57:1373–1382.
29. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* 1998;54:905–921.
30. Brunger AT. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* 1992;355:472–475.
31. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* 2004;60:2126–2132.
32. Hyland LJ, Tomaszek TA Jr, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD, Culp JS, et al. Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by ¹⁸O isotope exchange. *Biochemistry* 1991;30:8441–53. [PubMed: 1883830]
33. Marelus J, Kolmodin K, Feierberg I, Aqvist J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J Mol Graph Model* 1998;16:213–25. 261. [PubMed: 10522241]
34. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* 1995;117:5179–5197.
35. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* 1977;23:327–341.
36. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996;14:33. [PubMed: 8744570]
37. Kollman P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* 1993;93:2395–2417.
38. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* 1993;26:283–291.

Abbreviations

Boc	tert-butoxycarbonyl
COSY	C-terminal thioester
DIEA	<i>N,N</i> -diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide

GdmHCl	guanidine hydrochloride
GBTU	2-(1 <i>H</i> - benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HF	hydrogen fluoride
HIV-1	human immunodeficiency virus 1
MALDI-TOF	matrix-assisted laser desorption ionization time of flight
MBHA	4-methylbenzhydramine
MPAA	(4-carboxymethyl)thiophenol
NCL	native chemical ligation
RP-HPLC	reversed-phase high pressure liquid chromatography
SPPS	solid phase peptide synthesis
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TFA	trifluoroacetic acid

Figure 1.

Synthesis of HIV-1 protease from two peptide segments. Two peptides were ligated using native chemical ligation, and the formyl protecting groups on tryptophan were removed under basic conditions. Cys⁴¹ was subsequently converted into pseudolysine by alkylation with bromoethylamine hydrobromide. The polypeptide was then folded into the active protease, and analyzed by RP-HPLC (inset, left) and MALDI-TOF mass spectrometry (inset, right) (expected mass 10735.5 Da average isotopes, observed mass 10732.5 ± 5.4 Da).

Figure 2.

Characterization of refolded HIV-1 protease. a) Cleavage of the p24/p15 substrate peptide by synthetic HIV-1 protease. p24/p15 peptide (1.8 mg, 0.5 mM) was dissolved in 1 mL acetate buffer (50 mM, pH 5.6), and 16 pmol of enzyme was added to the peptide solution. The solution was incubated for 9 hours at 37°C, and then analyzed by MALDI-TOF MS using α -cyano-4-hydroxy-cinnamic acid as the matrix (GHKARVL expected mass 779.9 Da average isotopes, observed mass 779.3 ± 0.4 Da; MQRGNPRNQRK expected mass 1384.6 Da, observed mass 1385.1 ± 0.7 Da; GHKARVLA EAMSQVTNSATIM expected mass 2215.5 Da, observed mass 2217.2 ± 1.1 Da; AEAMSQVTNSATIMMQRGNPRNQRK expected mass 2820.2 Da, observed mass 2824.0 ± 1.4 Da). The small peaks to the right of the principal peaks correspond to sodium adducts. b) Steady-state kinetics of synthetic HIV-1 protease. The fluorogenic substrate Abz-TI(Nle)F(NO₂)QR-CONH, where Abz = 2-aminobenzoyl, Nle = norleucine,

and F(NO₂) = *p*-nitro phenylalanine, was incubated at various concentrations with the enzyme, and initial velocities were determined by the initial rate of increase in fluorescence after substrate cleavage. Assays at each concentration were performed in triplicate. Final assay conditions were 50 mM sodium acetate buffer, pH 5.6, 1% DMSO, 37°C. k_{cat} and K_{m} were determined after fitting the data points to the Michaelis-Menten equation using a nonlinear least squares fitting program in Origin 7.0.

Figure 3.

Heptapeptide analog JG-365 and its interaction with the catalytic site of HIV-1 protease. a) Stick representation of the two alternate orientations of JG-365. Residues are labeled color-coded accordingly with the shown orientations (P3'-P4 green, P4-P3' yellow). The hydroxyethylamine linkage between sites P1-P1' is shown by HEA. Final σ_A -weighted $2F_o - F_c$ electron density map, contoured at a level of 1σ , is shown in blue. b) and c) Stereo diagrams of alternate orientations for JG-365 and its interactions with flaps (residues Leu50/B and Leu50/A). Consistent with the colors used in a), JG-365 is depicted as green sticks in b) and as yellow sticks in c). Protein residues are represented as grey sticks in both figures, and oxygen and nitrogen atoms are depicted in red and blue, respectively. σ_A -weighted $2F_o - F_c$ map is shown as blue mesh and contoured at 1σ . The conserved water molecule (commonly identified as 301) is shown as red sphere where involved in bonding, and as grey sphere where not involved.

Figure 4.

Model of the inhibitor MVT-101. a) Final σ_A -weighted $2F_o - F_c$ electron density map, contoured at a level of 1σ , around MVT-101. Residues labeled (P3-P3') from the amino to the carboxyl termini. Location of the reduced peptide bond is shown by CH2. b) Stereo diagram of the hydrogen bond interactions among flaps, MVT-101, water 301, and catalytic aspartates. Hydrogen bonds are shown with dashed lines. Carbon, oxygen, nitrogen and hydrogen atoms are shown in grey, red, blue and cyan, respectively.

Figure 5. Schematic representation of the location of structural and catalytic water molecules in HIV-1 protease. Water molecules are shown as CPK. Structural waters are indicated by labels W1–W4, catalytic water is indicated by LW, and the water molecule interacting with the flaps is indicated as W301 (corresponding to residues HOH/27 and HOH/28 chain E in 2j9j, and HOH/7 chain E in 2j9k). The positions of the natural peptide substrate used in MD calculations (green sticks) and of the catalytic aspartates (D25 and D25', grey sticks) are also shown.

Figure 6.

Average geometry (Å) and relative free energies (kcal/mol) of studied reactant state conformations for the reaction catalyzed by HIV-1 protease. Proton translocation and water insertion free energies are indicated above the arrows connecting individual conformers. Crystal structure geometry is drawn in yellow sticks. The measured average interatomic distances are indicated by dashed lines.

Table 1

Strategy for the chemical synthesis of the HIV-1 protease

Name	Peptide Sequence
1-40 ^a COSR	POITLWKRPL ¹⁰ VTIRIGGOLK ²⁰ EALLDTGADD ³⁰ TVIEE(Nle)NLPG ⁴⁰ -COSR-Arg
Cys41-99	(K→C)WKPK(Nle)IGGI ⁵⁰ GGFIKVRQYD ⁶⁰ QIPVEI(Abu)GHK ⁷⁰ AIGTVLVGPT ⁸⁰ PVNIIGRNLI ⁹⁰ TOIG(Abu)TLNE ⁹⁹

Abu stands for L- α -amino-n-butyric acid, Nle stands for L-norleucine, and COSR stands for the mercaptopropionyl thioester. The protein sequence is from the SF2 isolate of the virus, with mutations Q7K and L33I to increase resistance to autolysis¹³. The catalytic Asp²⁵ residue is in bold.

Table 2

Data collection and refinement statistics

Crystal name	HIV-1 PR JG365	HIV-1 PR MVT101
Beamline	APS, 19-BM	APS, 14-BM-C
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)		
<i>a</i>	50.55	50.57
<i>b</i>	58.81	58.77
<i>c</i>	60.90	61.27
Resolution (Å)	50-1.04	50-1.2
Unique reflections ^a	82,488 (7,137)	56,850 (5,509)
Completeness (%) ^a	93.8 (82.2)	98.9 (97.9)
Redundancy ^a	10.1 (3.2)	10.4 (6.8)
$R_{\text{merge}}(\%)^{a,b}$	6.2 (41.7)	10.0 (58.4)
I/σ^a	33.0 (2.0)	33.7 (3.0)
Data range for refinement (Å)	10–1.04	10–1.20
R-work (R_{free}) (%) ^c	12.70 (16.88)	13.85 (19.02)
No. of non-hydrogen atoms		
Protein	1608	1561
Inhibitor	127	130
Solvent	277	158
No. of residues modeled in alternate conformations	11	9
Deviation from ideality		
Bond lengths (Å)	0.019	0.015
Angle distances (Å)	0.041	0.034
Ramachandran plot (%) ^d		
Most favored regions	94.4	95.8
Additional allowed	5.6	4.2
Generously allowed	0.0	0.0
Disallowed	0.0	0.0
PDB entry	2j9j	2j9k
Average anisotropy for all protein atoms	0.424	0.499

^aData for the highest resolution shell are given in parenthesis

^b $R_{\text{merge}} = \Sigma (I - \langle I \rangle) / \Sigma \langle I \rangle$

^c $R_{\text{factor}} = \Sigma ||F_{\text{obs}}| - k |F_{\text{calc}}|| / \Sigma |F_{\text{obs}}|$. R_{free} is the R_{factor} value for 5% of the reflections excluded from the refinement.

^dFigures from PROCHECK³⁸.

Table 3Refinement Progression ($R_{\text{work}}/R_{\text{free}}$)^a

	HIV-1 PR JG365	HIV-1 PR MVT101
CNS refinement		
rigid body refinement	32.92/32.82	31.52/34.51
simulated annealing	28.01/31.47	27.38/31.65
final model	23.90/24.25	27.22/28.07
SHELXL97 refinement		
ADPs added	18.88/22.30	17.71/22.16
riding hydrogens added	14.08/18.37	14.14/18.93
final model ^b	12.70/16.88	13.85/19.02

^a $R_{\text{factor}} = \frac{\sum ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$. R_{free} is the R_{factor} value for 5% of the reflections excluded from the refinement.

^b Weighting changed to W=0.2 in the last stage of SHELXL refinement¹².

Table 4

Calculated a) proton relocation and b) water annihilation free energies in HIV-1 PR

<i>Initial state^a</i>	<i>Restraint force constant (kcal/ mol-Å²)</i>	<i>ΔG (kcal/mol)^c</i>	<i>ΔG_{binding} (kcal/mol)^d</i>
a)			
OD1H---OD2 + w ^e	1.5 ^h	4.3	
OD1H---OD2 + w ^e	10 ⁱ	2.7	
OD1H---OD2 + w ^e	10 ^j	4.0	
		<hr/>	
		3.7 ± 0.9	
OD1H---OD1 ^f	10 ^j	3.1	
OD1H---OD2 ^g	10 ^j	6.3	
OD1H---OD2 ^g	20 ^j	5.3	
		<hr/>	
		4.9 ± 1.6	
b)			
WAT aqueous solution	3 ^m	7.4	
WAT aqueous solution	1 ^m	6.2	
WAT aqueous solution	2 ^m	7.0	
		<hr/>	
		6.9 ± 0.6	
WAT_3 ⁿ	1 ^m	8.2	
WAT_3 ⁿ	2 ^m	6.4	
		<hr/>	
		7.3 ± 1.3	
WAT_4 ^o	1 ^m	7.4	
WAT_4 ^o	2 ^m	7.0	
		<hr/>	
		7.2 ± 0.3	
WAT_301 ^p	1 ^m	7.6	
WAT_301 ^p	2 ^m	5.9	
		<hr/>	
		6.8 ± 1.2	
LW, OD1---OD1H	1 ^m	10.0	
LW, OD1H---OD1	1 ^m	9.0	
LW, OD1---OD1H	2 ^m	9.6	
LW, OD1H---OD1	2 ^m	11.2	
LW, OD1---OD1H	1, ^{mh}	8.4	
LW, OD1H---OD1	1, ^{mh}	12.4	
LW, OD1---OD1H	2, ^{mh}	10.5	
LW, OD1H---OD1	2, ^{mh}	12.1	
		<hr/>	
		10.4 ± 1.5	
LW, OD2H---OD2	1, 1.5 ^m	14.0	
LW, OD2H---OD2	1, 1.5 ^m	13.1	
LW, OD2H---OD2	2, 1.5 ^m	9.2	
LW, OD2H---OD1	2, 1.5 ^m	8.7	
LW, OD2H---OD2	1, ^{mh}	11.1	
LW, OD2H---OD2	2, ^{mh}	11.4	
		<hr/>	
		11.2 ± 1.9	
			-0.4 ± 1.9
			-0.3 ± 0.9
			+0.1 ± 1.8
			-3.5 ± 2.1
			-4.3 ± 2.5

^a Asp25' and Asp25 are shown in the order^b Forward (f) or reverse (r) trajectory was used to generate sampled configurations in the FEP calculations.^c Free energy difference at 298K, ΔG = G_{final} - G_{initial}^d Standard binding energy, see thermodynamic cycle in Figure 4, Supplementary Material.^e Initial and final states correspond to configurations depicted in Figure 6, states 1D and 1E, respectively.^f Initial and final states correspond to configurations depicted in Figure 6, states 1A.^g Initial and final states correspond to configurations depicted in Figure 6, states 1B and 1C, respectively.^h Additional two-fold torsional potential with minima at 0° and 180° was added for the Cα-Cβ-Cγ-OD1 torsion in the Asp25' residue.

ⁱ A single harmonic flat-bottom distance restraint between OD2(Asp25') and O(carbonyl substrate) atoms was used. The potential was zero for distances between 2.5 and 4.0 Å and was active in both initial and final state and along the whole pathway connecting these two states.

^l Three harmonic flat-bottom distance restraints were used

1. OD2(Asp25')---O(carbonyl substrate), zero between 2.5 Å and 4.0 Å
2. OD2(Asp25')---OD2(Asp25), zero between 4.0 Å and infinity
3. OD2(Asp25') ---OD1(Asp25), zero between 4.0 Å and infinity

All three restraints were inactive in the initial state and their magnitude varied linearly between the initial and final state. The force constant for the final state is reported.

^m Positional sequence restraints were applied to the water molecule only, with the shown force constant.

^{n, o, p} Initial state corresponds to configurations shown in Figure 5.