Supporting Information

Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease

Seema Mittal, [†] Yufeng Cai, [†] Madhavi N. L. Nalam, [†] Daniel N. A. Bolon[†] and Celia A. Schiffer*[†]

[†]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605. Email: Seema.mittal@umassmed.edu; Yufeng.Cai@umassmed.edu; Madhavi.Nalam@umassmed.edu; Dan.Bolon@umassmed.edu; and *Celia.Schiffer@umassmed.edu.

METHODS

Nomenclature for structural and molecular dynamic simulation studies: We determined six crystal structures and developed the following nomenclature to serve as concise references. Inhibitors when present are indicated with a subscript prefix followed by the cysteine mutations and a subscript suffix indicates the oxidation state of each protease monomer. Thus, $_{DRV}(G16C/L38C)_{rr}$ refers to DRV complexed with the G16C/L38C variant where the introduced cysteines are in their reduced state in both monomers. The rr could be replaced with oo for crosslinked/oxidized or by or for mixed population of reduced and oxidized forms of the two monomers in the protease.

Detection of Cross-links: Cross-linking of cysteines was monitored by quantifying free sulfhydryl content by Ellman's test[1,2] and further verified by alkylation of cysteines and mass spectrometry[3].

For Ellman's test, under reducing conditions, detection of four molecules of free sulfhydryls per protease dimer in both variants (G16C/L38C and R14C/E65C) confirmed the absence of disulfide bonds. Oxidizing conditions, however, yielded no Ellman's reaction, indicating the absence of reduced sulfhydryls in the two oxidized samples, confirming complete cross-linking. As positive and negative controls for quantifying sulfhydryls, bovine serum albumin (BSA, containing one free cysteine) and the G16A/L38A variant (containing no cysteine), respectively, were assayed by Ellman's test. Determination of the crystal structures of the designed variants further verified these results.

The presence of disulfide bond in $(G16C/L38C)_{oo}$ protein preparation was further confirmed by using maleimides; NEM (N-ethyl maleimide) and NMM (N-methyl maleimide) as thiol modifying agents followed by molecular weight detemination using mass spectrometry. Maleimides readily react with the thiol group found on cysteine to form a stable carbon-sulfur bond. DTT (Dithiothreitol) was used as a reducing agent. Protein samples for WT and $(G16C/L38C)_{oo}$ were prepared in 0.1M sodium phosphate buffer pH 7.0. All reagents including maleimides and DTT were prepared in 0.1M PBS pH7.0. 10μg of protein (38μl) was diluted with 0.1M PBS (19μl) followed by 5μl of 63mM NEM (5mM final concentration). The reaction was left at RT for 1.5hr to ensure complete alkylation. Only those cysteines that are not involved in disulfide bond will be available for alkylation. 5μl of 205.5mM DTT (15mM final concentration) was then added to this reaction mixture and the reaction was heated at 60°C for 30min. Finally, in the second step of alkylation, 20μl of 199.1mM (45mM final concentration) NEM was added and the reaction was incubated at RT for 1.5hr to alkylate the cysteines made available upon reduction by DTT. The reverse procedure (NMM followed by DTT followed by NEM) was also performed at the same concentrations mentioned above. 5μl of this reaction was injected in PLRPS column and washed at 5μl/minute for 10 minutes in 95% of 0.1% Formic acid. The column was further washed at 3 μl/minute with a gradient of 0.1% formic acid in acetonitrile at the following concentrations: 5%(1 mint) to 60% (2mints) to 95% (2.5mints) and finally 5% (2.5 mints.) and resulting spectra analyzed. The MS method was to scan from m/z 400-2000 for 10min in + V mode with 1 sec scans and a lock-mass scan every 10seconds. The data was deconvoluted with the Maxent software[4].

The results demonstrate that no alkylation was detected after the first alkylation reaction (*Figure SI2*). However, the second alkylation following reduction of disulfides upon treatment with DTT resulted in mass spectra consistent with the additive molecular weight of protease monomer alkylated with two molecules of NMM. These results confirm that the engineered cysteines in $(G1C/L38C)_{oo}$ protein sample were completely oxidized and therefore, unavailable for reaction with NEM. Treatment with DTT resulted in reduction of disulfide bond, thereby making two cysteines in each monomer available for reaction with NMM. A small fraction of N-terminal alkylation was also detected in $(G1C/L38C)_{oo}$ sample. These values were consistent with the N terminal alkylation observed for WT sample.

Mutagenesis, protein purification, and crystallization: HIV-1 protease variants were expressed from a Pet11A E. coli plasmid vector, as described[5]. Native cysteines at positions 67 and 95 were both mutated to alanine. The variants for this study were generated using QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) to introduce the G16C, L38C, R14C, E65C, G16A and L38A substitutions. HIV-1 protease variants were expressed in 6 liter cultures of E. coli BL21(DE3) cells at 37˚C. Protein expression was initiated at an OD600 of 0.6 by adding 2mM IPTG. Cells were harvested after 3 hr and lysed using a cell disruptor. The inclusion bodies were isolated and resuspended in 7.5 M guanidinium. The soluble fraction was isolated after centrifugation and the supernate was purified using superdex75 gravity gel filtration chromatography in 10 mM Tris pH 7.5 and 7M guanidine HCl at

room temperature. The purified protease was then refolded by rapid 10-fold dilution into the refolding buffer, containing 0.05 M sodium acetate (pH 5.5), 5% ethylene glycol, 10% glycerol, and 5 mM DTT (for the non-cross-linked protease) and in 100mM sodium phosphate buffer (pH 7-7.5) (for the disulfide-bonded protease), both methods performed over ice. The refolded and diluted sample was concentrated and dialyzed against respective buffers to remove residual guanidine. For the disulfide cross-linked protease, the dialysis was carried out against 100mM sodium phosphate buffer with stepwise pH change from pH 7.5 to pH 6.0 and the purified cross-linked protease was stored in 100mM sodium phosphate buffer (pH 6.0). Protease intended for crystallization was further purified on a Pharmacia Superdex 75 FPLC column equilibrated with respective refolding/storage buffer. Crystals were grown by the hanging drop vapor diffusion method at room temperature. The protein concentrations were 1.5–2 mg/ml. The inhibitor was crystallized in 5–10 molar excess. Inhibitor stock was made in dimethyl sulfoxide (DMSO). Small crystals started appearing after 6–9 days (for the non-crosslinked protease complexes) and after 14-20 days (for the cross-linked protease complex) and grew at room temperature. For apo protease, however, crystals appeared in 4 to 5 months at room temperature.

Collection and processing of crystallographic data: Protein crystals were harvested and flash frozen by dipping in liquid nitrogen. All data was collected under cryo conditions from various sources, including an in-house Rigaku X-ray generator with an R-axis IV image plate, and a synchrotron radiation source at Argonne National Laboratory (APS, Chicago, IL) BioCARS 14-BMC. Diffraction images were indexed and scaled using HKL2000[6].

Crystallographic refinement and structural analysis: Crystal structures were solved and refined using programs within the CCP4I suite[7]. Structures were solved by molecular replacement using PHASER[8]. Molecular replacement phases were improved by building solvent molecules using ARP/warp. Structures were refined with REFMAC 5 in the CCP4 suite using a combination of restrained and TLS[9] (translation, libration, screw rotation) refinement. The possibility of over refinement was limited by using a free R value with 5% of the data. Electron density was viewed and interactive models were built using the program COOT[10].

Graphical visualization and analysis was carried out using PyMOL (Delano Scientific LLC). All figures were also generated in PyMOL. The disulfide bonds were analyzed using Disulfide Bond Analysis software[11]. The electron density analysis revealed that cysteine substitutions were well tolerated at all the positions (G16, L38, R14 and E65). However, disordered electron density was observed in the loops containing G16A and L38A substitutions in the DRV complex of alanine variant (G16A/L38A). Residues 36 and 37 were therefore, not modeled in one of the monomers of $_{DRV}$(G16A/L38A) structure. Additionally, while the cross-linked $(R14C/E65C)_{oo}$ protease did not crystallize, the formation of cross-links was verified by Ellman's test and mass spectrometry analysis as mentioned above.

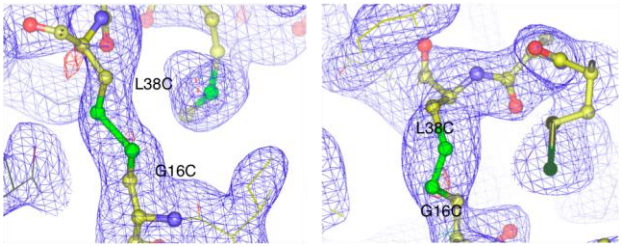

Figure SI1: Electron density maps contoured at 1σ showing the disulfide bonds in both monomers of $_{apo}(G16C/L38C)_{oo}$

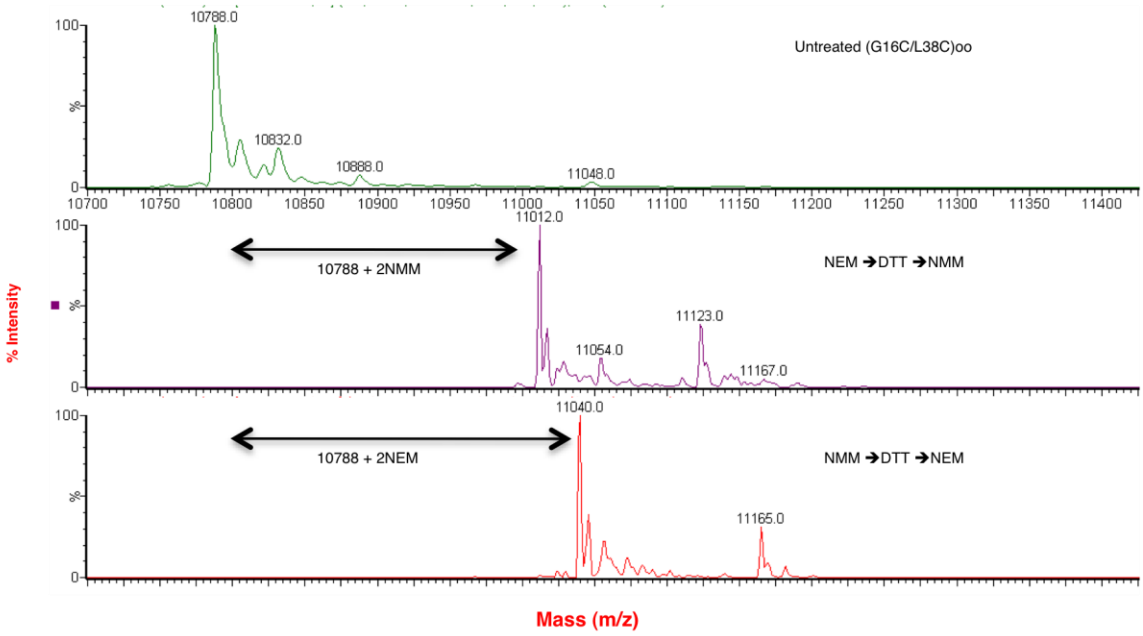

***Figure SI2*:** Confirmation of cysteine cross-linking in $(G16C/L38C)_{oo}$ protease by mass spectrometry. $(G16C/L38C)_{oo}$ protein sample was treated with NEM, DTT and NMM in successive steps and spectra analyzed. The process was also repeated in the reverse order. No alkylation with the first alkylating reagent and complete alkylation with the second alkylating reagent were observed. Minor N-terminal alkylation was also observed

Enzyme kinetics: Kinetics of each protease variant was assayed by its ability to hydrolyze the fluorogenic substrate HiLyte Fluor 488-Lys-Ala-Arg-Val-Leu_Ala-Glu-Ala-Met-Ser-Lys (QXL-520) (Anaspec, Inc., Fremont, CA) corresponding to the HIV-1 CA-p2 substrate. The assay was conducted in 96-well plates with 50 to 100nM protease in a total volume of 100 µl; the enzymatic reaction was initiated by adding 0-50µM substrate. Hydrolysis was monitored at 535 nm, while exciting at 485 nm using a Victor microplate reader (PerkinElmer, Waltham, MA). The reaction buffer comprised 0.1 M sodium acetate, 1 M NaCl, 1 mg/ml BSA, 2% DMSO with and without 5 mM TCEP at a pH of 5.5. Relative fluorescence units were measured every 7 seconds and converted to concentration using standard calibration curves generated for HiLyte Fluor 488 at each substrate concentration. The K_m values could not be obtained because of the limited solubility of fluorogenic peptide in reaction conditions.

Molecular dynamics simulations, setup and analysis: MD simulations were performed using the program Sander in the AMBER 8[12] (Assisted Model Building with Energy Refinement) package. Initial coordinates for the MD simulation of the "oo" protease were provided by the crystal structure $_{DRV}(G16C/L38C)_{or}$ with the ligand removed and an in-silico disulfide established between the cysteines in the second monomer. The initial coordinates for the MD simulation of the "rr" protease were provided by the crystal structure $_{DRV}(G16C/L38C)_{rr}$ with the ligand removed. Three independent simulations each were carried out on G16C/L38C in both the cross-linked (oo) and non-cross-linked (rr) states. Each structure was solvated with the TIP3P water cubic box to allow at least 8 Å of solvent on each face of the protease. The dimensions of the final periodic box are 63 by 55 by 78 Å. A three-step energy minimization process with the steepest descent method was used to allow the system to reach an energetically favorable conformation. In the first energy minimization step, all the heavy atoms of the protease were restrained with a harmonic force constant of 10 kcal $mol^{-1}Å^{-2}$. In the second step, only the backbone nitrogen, oxygen, and carbon atoms were restrained. In the third step, all atoms were allowed to move. Each of the three steps had 2000 cycles. The temperature of the energy-minimized system was then gradually raised from 50ºK to 300ºK in the NVT ensemble. Initial velocities were assigned according to the Maxwellian distribution with a random seed. In the thermalization process, heavy atoms were restrained with a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$. The whole process was 50 picoseconds (50,000 steps, each 1 fs). A 50-ps equilibration was then performed in the NPT ensemble without restraining heavy atoms. In the subsequent sampling MD simulations, each step was 2 fs, and the trajectory was recorded every 100 fs. The total simulation time was 20 ns. For the

thermalization, equilibration, and sampling simulations, the SHAKE algorithm was applied to constrain all hydrogen atoms. At every 1 ps, a snapshot was taken to be analyzed for the production phase.

Double difference plots[13] were generated for 0 ns and 20 ns simulated structures of G16C/L38C in (oo) and (rr) to graphically visualize structural differences between them. Double difference plot analysis compares backbone shifts in closely related structures without adding bias towards methods used for superposition. Briefly, distances between all C_α atoms within the dimer were calculated for each structure. A distance-difference matrix was then computed for each atom for a given pair of complexes. The distance-difference matrix was then plotted as a contour plot using MATLAB (MathWorks, Natick, Massachusetts). RMSF (root mean square fluctuation) values were calculated per C_α residue over the 20ns MD simulation trajectories for G16C/L38C in (oo) and (rr) states and mapped on to representative protease structure using PyMol.

	C_β- C_β distance, Å		
	0 ns	10 ns	14.5 ns
G16-L38	5.71	4.67	2.99
R14-E65	4.32	4.68	4.79

***Table SI1*.** C_β- C_β distance between the cysteines modeled at sites G16, L38 and R14, E65, respectively at 0 ns, 10 ns and 14.5 ns MD simulation snapshots[4]

Variable	DRV-WT	DRV-G16C/L38C (rr)	DRV-G16C/L38C (or)	DRV-G16A/L38A Ala (control)	Apo-G16C/L38C (oo)	DRV-R14C/E65C (rr)
Resolution, Å	1.5	1.95	1.3	1.9	2.8	1.8
Temperature, °K	100	100	100	100	100	100
Space group	$p2_12_12_1$	$p2_12_12_1$	$p2_12_12_1$	$p2_12_12_1$	$p2_12_12_1$	$p2_12_12_1$
Unit cell parameters						
a, Å	50.9	50.9	51.2	51.9	51.5	50.7
b, Å	57.8	57.9	58.5	57.4	58.2	57.9
c, Å	61.6	61.9	61.8	60.7	61.0	61.5
Z	4	4	4	4	4	4
R_{merge}, %	4.6	7.3	7.6	7.1	9.0	7.1
Completeness, %	92.2	96.3	92.4	99.8	97.9	98.6
Total reflections	195859	92311	331698	103492	30570	120045
Unique reflections	26088	12766	40660	14074	4535	16582
I/σ	15.1	10.4	18.4	11.1	9.0	10.8
RMSD, Å						
Bond length, Å	.008	.008	.009	.009	.008	.009
Bond angles	1.43	1.44	1.36	1.48	1.39	1.37
R_{factor}, %	17.0	18.2	20.4	18.4	19.8	17.03
R_{free}, %	19.6	22.9	21.9	23.6	26.2	21.9
Redundancy, %	7.1	6.9	7.7	7	6.4	6.9
PDB ID codes	4DQB	4DQC	4DQE	4DQF	4DQG	4DQH

***Table SI2*.** Crystallographic statistics for DRV complexes of WT, G16C/L38C (or, rr), R14C/E65C (rr), G16A/L38A and apo G16C/L38C (oo) HIV-1 proteases**.**

Corresponding Author

* University of Massachusetts Medical School, Worcester, MA.

Email: Celia.Schiffer@umassmed.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by grants from the National Institutes of Health P01 GM66524.

REFERENCES

(1) Scientific, T.; Ellman's Reagent: 22582.
(2) Ellman, G. L. *Arch Biochem Biophys* **1959**, *82*, 70-7.
(3) Seeger, M. A.; von Ballmoos, C.; Eicher, T.; Brandstatter, L.; Verrey, F.; Diederichs, K.; Pos, K. M. *Nat Struct Mol Biol* **2008**, *15*, 199-205.
(4) Ferrige, A. G.; M.J., S.; S., J.; Skilling, J.; Aplin, R. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 374-377.
(5) Prabu-Jeyabalan, M.; Nalivaika, E.; King, N. M.; Schiffer, C. A. *J Virol* **2004**, *78*, 12446-12454.
(6) Otwinowski, Z.; Minor, W. *Methods in Enzymology* **1997**, *276*, 307-326.
(7) *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 760-3.
(8) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658-674.
(9) Painter, J.; Merritt, E. A. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 439-50.
(10) Emsley, P.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126-32.
(11) Wong, J. W.; Hogg, P. J. *J Thromb Haemost*.
(12) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J Comput Chem* **2005**, *26*, 1668-88.
(13) Prabu-Jeyabalan, M.; Nalivaika, E. A.; Romano, K.; Schiffer, C. A. *Journal of Virology* **2006**, *80*, 3607-16.