## Conclusions

We proposed a mechanism of resistance against NFV due to the nonactive site mutation N88S in CRF01_AE PR through computational simulations. CRF01_AE PR has polymorphisms at nonactive sites, unlike subtype B PR. Nevertheless, the polymorphisms affect the binding affinities between NFV and PR variants that have the D30N or N88S mutation. The simulations suggest that N88S in CRF01_AE PR confers NFV resistance by reducing interaction energy between D30 and NFV. N88S creates hydrogen bonds between the D30 and S88 side chains and causes conformational changes at D30. These changes reduce the interactions between D30 and NFV. Furthermore, we proposed an explanation of why the emergence rates of D30N and N88S differ between subtypes B and AE HIV-1. The M36I mutation seen in the natural polymorphisms of CRF01_AE PR is particularly involved in the difference in the emergence rates of D30N and N88S.

## Experimental Section

**Force Field Parameters.** Before carrying out molecular dynamics (MD) simulations, we improved the torsional force field parameters for benzamide and determined the restrained electrostatic potential (RESP)[51] charges for NFV. The improved torsional parameters were generated in the same manner as that for the development of the AMBER ff03 force field.[43] First, RESP charges of benzamide were determined on the basis of data from quantum chemical calculations. Geometric optimization was performed at the HF/6-31G(d,p) level, and the electrostatic potential was calculated at the B3LYP/cc-pVTZ level under solvation conditions with ether ($\epsilon = 4$) by the IEFPCM method using the Gaussian 03 program.[52] The partial atom charges were determined using the RESP method so that the atom charges could reproduce the values of the calculated electrostatic potential at the surrounding points of the benzamide. Charges were equalized between two atoms if they were the same element and had the same bond coordination. Second, a potential energy curve was obtained by repeating the energy calculations with 5° stepwise changes in the torsional angle around the torsional axis. Energy calculations were executed at the MP2/cc-pVTZ level in the ether phase after geometric optimizations at the HF/6-31G(d,p) level. Third, the torsional parameters were obtained by fitting them to the potential energies from quantum chemical calculations. The torsional parameter for CA−CA−C−N was assumed to be equal to that for CA−CA−C−O. RESP charges for NFV were also determined in the same manner as that described above.

**Molecular Dynamics (MD) Simulations.** Minimizations and MD simulations were carried out using the Sander module of AMBER 8.[53] The AMBER ff03 force field[43] was used as the parameters for proteins, ions, and water molecules. The general AMBER force field[44] and our developed force field were used as the parameters for NFV.

We examined the structure of each of the six PRs in complex with NFV: wild-type (WT) PR, D30N PR, and N88S PR of subtype B HIV-1 (labeled B(WT), B(D30N), and B(N88S), respectively); the reference (Ref) PR, D30N PR, N88S PR of subtype AE HIV-1 in complex with NFV (AE(Ref), AE(D30N), and AE(N88S)). Additionally, we investigated the structures of M36I PR, M36I/N88S PR, and L10F/M36I/N88S PR of subtype B HIV-1 in complex with NFV (labeled B(M36I), B(M36I/N88S), and B(L10F/M36I/N88S), respectively). We used HXB2 as the WT sequence of subtype B HIV-1 and used NH1 as the reference sequence of subtype AE HIV-1.[54] Each initial structure for the PR in complex with NFV was modeled from the atom coordinates of an X-ray crystal structure (PDB code 1OHR)[42] using the LEaP module. Each model was placed in a rectangular box filled with about 8000 TIP3P water molecules,[55] with all of the crystal water molecules remaining. The cutoff distance for the long-range electrostatic and the van der Waals energy terms was set to 12.0 Å. The expansion and shrinkage of all covalent bonds connecting to hydrogen atoms were constrained using the SHAKE algorithm.[56] Periodic boundary conditions were applied to avoid the edge effect in all calculations. Energy minimization was achieved in three steps. First, movement was allowed only for water molecules and ions. Next, the ligand and the mutated residues were allowed to move in addition to the water molecules and ions. Finally, all atoms were permitted to move freely. In each step, energy minimization was executed by the steepest descent method for the first 10 000 steps and the conjugated gradient method for the subsequent 10 000 steps. After a 0.1 ns heating calculation until 310 K using the NVT ensemble, a 3.0 ns equilibrating calculation was executed at 1 atm and at 310 K using the NPT ensemble, with an integration time step of 2.0 fs. In the present calculations, the MD simulations showed no large fluctuations after about 2.0 ns of equilibrating calculations (Supporting Information Figures S8 and S9). Hence, atom coordinates were collected at an interval of 1.0 ps for the last 1.0 ns to analyze the structure in detail.

The protonation states of catalytic aspartates D25 and D25′ vary depending on the binding ligands or PRs.[57] Hence, the appropriate protonation states of catalytic aspartates should be determined for each model. Because NFV mimics a transition state of catalytic reaction by HIV-1 PR, we considered two kinds of protonation states.[58−60] One represented a combination of protonated D25 and unprotonated D25′ states, and the other represented the opposite combination. In order to determine the protonation states when NFV binds to each PR, the free energies of these two kinds of protonation states were compared using the calculation data obtained during 2.0−3.0 ns of MD simulations. The free energies were calculated on the basis of the MM/PBSA method.[61,62] We used the same parameter set for electrostatic and van der Waals energy terms as that used in the MD simulations, and no cutoff was applied for the calculation. Since the dielectric constants for the interior of proteins are considered to be in the range of 2−4, the interior dielectric constant was set to 2.0.[63] The outer dielectric constant was set to 80.0. The PBSA program was used to solve the Poisson−Boltzmann (PB) equation. B(D30N), B(M36I), B(L10F/M36I/N88S), and AE(D30N) were found to prefer the combination of protonated D25 and unprotonated D25′. The other five PRs (B(WT), B(N88S), B(M36I/N88S), AE(Ref), and AE(N88S)) preferred the combination of unprotonated D25 and protonated D25′ (Supporting Information Table S4).

**Hydrogen Bond Criteria.** The formation of a hydrogen bond was defined in terms of distance and orientation. The combination of donor D, hydrogen H, and acceptor A atoms with a D−H···A configuration was regarded as containing a hydrogen bond when the distance between donor D and acceptor A was shorter than 3.5 Å and the angle H−D−A was smaller than 60.0°.

**Binding Free Energy Calculation.** The binding free energy[64] was calculated by the following equation:

$$\Delta G_b = \Delta G_{int}^{ele} + \Delta G_{int}^{vdw} + \Delta G_{sol} - T\Delta S$$

where $\Delta G_b$ is the binding free energy in solution, $\Delta G_{int}^{ele}$ and $\Delta G_{int}^{vdw}$ are electrostatic and van der Waals interaction energies between a ligand and a protein, $\Delta G_{sol}$ is the solvation energy, and $-T\Delta S$ is the contribution of conformational entropy to the binding. In this study, assuming that the contribution of conformational entropy to the change in $\Delta G_b$ is negligible among mutants,[65] we neglected the entropy term in the energy estimation. $\Delta G_{int}^{ele}$ and $\Delta G_{int}^{vdw}$ were computed using the same parameter set as that used in the MD simulation, and no cutoff was applied to the calculation. Solvation energy $\Delta G_{sol}$ was calculated using the PBSA program. The interior dielectric constant was set to 2.0, and the outer dielectric constant was set to 80.0.[63] Furthermore, the contribution of each residue to the binding free energy was calculated. The total binding free energy was decomposed into the contribution from each individual residue by the MM/GBSA method. The modified GB model developed by Onufriev, Bashford, and Case[66] was used to calculate the solvation energy term. To ascertain whether or not

the MM/GBSA results were consistent with the MM/PBSA results, we compared the total binding free energy obtained by the MM/PBSA method with that obtained by the MM/GBSA method for all coordinates acquired through the MD simulation. The MM/GBSA results were confirmed to be highly correlated with the MM/PBSA results (correlation coefficient $r \geq 0.998$) (Supporting Information Figure S10).

**Acknowledgment.** This work was supported by a Health and Labor Science Research Grant for Research on HIV/AIDS from the Ministry of Health and Labor of Japan, by JSPS Research Fellowships for Young Scientists, and by a Grant-in-Aid for JSPS Fellows. A part of this work was also supported by a grant from the Japan Science and Technology Agency.

**Supporting Information Available:** A list of hydrogen bond networks in each model, determination of protonation states of catalytic aspartates, parameter development for the torsional parameters CA−CA−C−N and CA−CA−C−O, 3D plot of rmsd of the average structure from that of B(WT), contributions of each individual residue to the binding energy, rmsd plots during MD simulations, and results of principal component analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Krausslich, H. G.; Wimmer, E. Viral proteinases. *Annu. Rev. Biochem.* **1988**, *57*, 701−754.

(2) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolinick, E. M.; Sigal, I. S. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686−4690.

(3) Craig, J. C.; Duncan, I. B.; Hockley, D.; Grief, C.; Roberts, N. A.; Mills, J. S. Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral Res.* **1991**, *16*, 295−305.

(4) Vacca, J. P.; Dorsey, B. D.; Schleif, W. A.; Leven, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahy, O. M.; Roth, E.; Sardana, V. V.; Schlabach, A. J.; Graham, P. I.; Condra, J. H.; Gotlib, L.; Holloway, M. K.; Lin, J.; Chen, L.-W.; Vastag, K.; Ostvic, D.; Anderson, P. S.; Emini, E. A.; Huff, J. R. L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4096−4100.

(5) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentga, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X.; Wideburg, N. E.; Saldivar, A.; Ruiz, L.; Kati, W. M.; Sham, H. L.; Robins, T.; Stewart, K. D.; Hsu, A.; Plattner, J. J.; Leonard, J. M.; Norbeck, D. W. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2484−2488.

(6) Livington, D. J.; Pazhanisamy, S.; Porter, D. J.; Partaledis, J. A.; Tung, R. D.; Painter, G. R. Weak binding of VX-478 to human plasma proteins and implications for anti-human immunodeficiency virus therapy. *J. Infect. Dis.* **1995**, *172*, 1238−1245.

(7) Patick, A. K.; Mo, H.; Markowitz, M.; Appelt, K.; Wu, B.; Musick, L.; Kalish, V.; Kaldor, S.; Reich, S.; Ho, D.; Webber, S. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **1996**, *40*, 292−297 (Erratum, p 1575).

(8) Carrillo, A.; Stewart, K. D.; Sham, H. L.; Norbeck, D. W.; Kohlbrenner, W. E.; Leonard, J. M.; Kempf, D. J.; Molla, A. J. In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *J. Virol.* **1998**, *72*, 7532−7541.

(9) Robinson, B. S.; Riccardi, K. A.; Gong, Y. F.; Guo, Q.; Stock, D. A.; Blair, W. S.; Terry, B. J.; Deminie, C. A.; Djang, F.; Colonno, R. J.; Lin, P. F. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.* **2000**, *44*, 2093−2099.

(10) Larder, B. A.; Hertogs, K.; Bloor, S.; van den Eynde, C.; DeCian, W.; Wang, Y.; Freimuth, W. W.; Tarpley, G. Tipranavir inhibits broadly protease inhibitor-resistant HIV-1 clinical samples. *AIDS* **2000**, *14*, 1943−1948.

(11) Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J. F.; Boross, P.; Wang, Y. F.; Tie, Y.; Volarath, P.; Gaddis, L.; Harrison, R. W.; Weber, I. T.,; Ghosh, A. K.; Mitsuya, H. Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **2003**, *47*, 3123−3129.

(12) Cornelissen, M.; van den Burg, R.; Zorgdrager, F.; Lukashov, V.; Goudsmit, J. pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J. Virol.* **1997**, *71*, 6348−6358.

(13) Pieniazek, D.; Rayfield, M.; Hu, D. J.; Nkengasong, J.; Wiktor, S. Z.; Downing, R.; Biryahwaho, B.; Mastro, T.; Tanuri, A.; Soriano, V.; Lal, R.; Dondero, T. Protease sequences from HIV-1 group M subtypes A−H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group. *AIDS* **2000**, *14*, 1489−1495.

(14) Vergne, L.; Peeters, M.; Mpoudi-Ngole, E.; Bourgeois, A.; Liegeois, F.; Toure-Kane, C.; Mboup, S.; Mulanga-Kabeya, C.; Saman, E.; Jourdan, J.; Reynes, J.; Delaporte, E. Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients. *J. Clin. Microbiol.* **2000**, *38*, 3919−3925.

(15) Grossman, Z.; Vardinon, N.; Chemtob, D.; Alkan, M. L.; Bentwich, Z.; Burke, M.; Gottesman, G.; Istomin, V.; Levi, I.; Maayan, S.; Shahar, E.; Schapiro, J. M. Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. *AIDS* **2001**, *15*, 1453−1460 (Erratum, p 2209).

(16) Ariyoshi, K.; Matsuda, M.; Miura, H.; Tateishi, S.; Yamada, K.; Sugiura, W. Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. *JAIDS, J. Acquired Immune Defic. Syndr.* **2003**, *33*, 336−342.

(17) Clemente, J. C.; Coman, R. M.; Thiaville, M. M.; Janka, L. K.; Jeung, J. A.; Nukoolkarn, S.; Govindasamy, L.; Agbandje-McKenna, M.; McKenna, R.; Leelamanit, W.; Goodenow, M. M.; Dunn, B. M. Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry* **2006**, *45*, 5468−5477.

(18) Patick, A. K.; Duran, M.; Cao, Y.; Shugarts, D.; Keller, M. R.; Mazabel, E.; Knowles, M.; Chapman, S.; Kuritzkes, D. R.; Markowitz, M. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob. Agents Chemother.* **1998**, *42*, 2637−2644.

(19) Ziermann, E.; Limoli, K.; Das, K.; Arnold, E.; Petropoulos, C. J.; Parkin, N. T. A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J. Virol.* **2000**, *74*, 4414−4419.

(20) Sugiura, W.; Matsuda, Z.; Yokomaku, Y.; Hertogs, K.; Larder, B.; Oishi, T.; Okano, A.; Shiino, T.; Tatsumi, M.; Matsuda, M.; Abumi, H.; Takata, N.; Shirahata, S.; Yamada, K.; Yoshikura, H.; Nagai, Y. Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **2002**, *46*, 708−715.

(21) Resch, W.; Ziermann, R.; Parkin, N.; Gamarnik, A.; Swanstrom, R. Nelfinavir-resistant, amprenavir-hypersusceptible strains of human immunodeficiency virus type 1 carrying an N88S mutation in protease have reduced infectivity, reduced replication capacity, and reduced fitness and process the Gag polyprotein precursor aberrantly. *J. Virol.* **2002**, *76*, 8659−8666.

(22) Ode, H.; Ota, M.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Resistant mechanism against nelfinavir of human immunodeficiency virus type-1 proteases. *J. Phys. Chem. B* **2005**, *109*, 565−574.

(23) Mahalingam, B.; Louis, J. M.; Reed, C. C.; Adomat, J. M.; Krouse, J.; Wang, Y.-F.; Harrison, R. W.; Weber, I. T. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur. J. Biochem.* **1999**, *263*, 238−245.

(24) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci.* **2000**, *9*, 1898−1904.

(25) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T. Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 455−464.

(26) Mahalingam, B.; Boross, P.; Wang, Y.-F.; Louis, J. M.; Fischer, C. C.; Tozser, J.; Harrison, R. W.; Weber, I. T. Combining mutations in HIV-1 protease to understand mechanisms of resistance. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 107−116.

(27) Weber, J.; Mesters, J. R.; Lepsik, M.; Prejdova, J.; Svec, M.; Sponarova, J.; Mlcochova, P.; Skalika, K.; Strisovsky, K.; Uhlikova, T.; Soucek, M.; Machala, L.; Stankova, M.; Vondrasek, J.; Klimkait, T.; Kraeusslich, H.-G.; Hilgenfeld, R.; Konvalinka, J. Unusual binding mode of an HIV-1 protease inhibitor explains its potency against multi-drug-resistant virus strains. *J. Mol. Biol.* **2002**, *324*, 739−754.
(28) King, N. M.; Melnick, L.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Yang, S. S.; Gao, Y.; Nie, X.; Zepp, C.; Heefner, D. L.; Schiffer, C. A. Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci.* **2002**, *11*, 418−429.
(29) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *J. Virol.* **2003**, *77*, 1306−1315.
(30) Mahalingam, B.; Wang, Y.-F.; Boross, P. L.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*, 1516−1524.
(31) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; Béthune, M. P.; Schiffer, C. A. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J. Virol.* **2004**, *78*, 12012−12021.
(32) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J. Virol.* **2004**, *78*, 12446−12454.
(33) Skalova, T.; Dohnalek, J.; Duskova, J.; Petrokova, H.; Hradilek, M.; Soucek, M.; Konvalinka, J.; Hasek, J. HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site. *J. Med. Chem.* **2006**, *49*, 5777−5784.
(34) Rick, S. W.; Topol, I. A.; Erickson, J. W.; Burt, S. K. Molecular mechanisms of resistance: free energy calculations of mutation effects on inhibitor binding to HIV-1 protease. *Protein Sci.* **1998**, *8*, 1750−1756.
(35) Piana, S.; Carloni, P.; Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* **2002**, *11*, 2393−2402.
(36) Clemente, J. C.; Hermrajani, R.; Blum, L. E.; Goodenow, M. M.; Dunn, B. M. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry* **2003**, *42*, 15029−15035.
(37) Perryman, A. L.; Lin, J.-H.; McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **2004**, *13*, 1108−1123 (Erratum, p 1434).
(38) Wittayanarakul, K.; Aruksakunwong, O.; Saen-oon, S.; Chantratita, W.; Parasuk, V.; Sompornpisut, P.; Hannongbua, S. Insights into saquinavir resistance in the G48V HIV-1 protease: quantum calculations and molecular dynamic simulations. *Biophys. J.* **2005**, *88*, 867−879.
(39) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational simulations of HIV-1 proteases−Multi-drug resistance due to non-active site mutation L90M. *J. Am. Chem. Soc.* **2006**, *128*, 7887−7895.
(40) Meiselbach, H.; Horn, A. H. C.; Harrer, T.; Sticht, H. Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. *J. Mol. Model.* **2006**, *13*, 297−304.
(41) Batista, P. R.; Wilter, A.; Durham, E. H.; Pascutti, P. G. Molecular dynamics simulations applied to the study of subtypes of HIV-1 protease common to Brazil, Africa, and Asia. *Cell Biochem. Biophys.* **2006**, *44*, 395−404.
(42) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40*, 3979−3985.
(43) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(44) Wang, J.; Wolf, R. M.; Cladwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(45) http://structbio.vanderbilt.edu/archives/amber-archive/2003/0305.php.
(46) Johnson, V. A.; Brun-Vézinet, F.; Clotet, B.; Kuritzkes, D. R.; Pillay, D.; Schapiro, J. M.; Richman, D. D. Update of the drug resistance mutations in HIV-1: fall 2006. *Top. HIV Med.* **2006**, *14*, 125−130.
(47) Velazques-Campoy, A.; Todd, M. T.; Vega, S.; Freire, E. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6062−6067.
(48) Velazques-Campoy, A.; Vega, S.; Fleming, E.; Bacha, U.; Sayed, Y.; Dirr, H. W.; Freire, E. Protease inhibition in African subtypes of HIV-1. *AIDS Rev.* **2003**, *5*, 165−171.
(49) Ohtaka, H.; Schön, A.; Freire, E. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* **2003**, *42*, 13659−13666.
(50) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R. S.; Govindasamy, L.; Rutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. *Biochemistry* **2004**, *43*, 12141−12151.
(51) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004
(53) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.
(54) Sato, H.; Shiino, T.; Kodaka, N.; Taniguchi, K.; Tomita, Y.; Kato, K.; Miyakuni, T.; Takebe, Y. Evolution and biological characterization of human immunodeficiency virus type 1 subtype E gp120 V3 sequences following horizontal and vertical virus transmission in a single family. *J. Virol.* **1999**, *73*, 3551−3559.
(55) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(56) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(57) Zoete, V.; Michielin, O.; Karplus, M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **2002**, *315*, 21−52.
(58) Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; Lambert, R. W.; Merrett, J. H.; Millis, J. S.; Parkes, K. E. B.; Redshaw, S.; Ritchie, A. J.; Taylor, D. L.; Thomas, G. J.; Machin, P. J. Rational design of peptide-based HIV proteinase inhibitors. *Science* **1990**, *248*, 358−361.
(59) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (*R*)-hydroxyethylamine isostere. *J. Med. Chem.* **1991**, *34*, 3340−3342.
(60) Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. Hydrolysis mechanism of the phenylalanine−proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method. *J. Am. Chem. Soc.* **1999**, *121*, 7349−7354.
(61) Srinivasan, J.; Cheatham, T. E., III; Kollman, P.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(62) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(63) Wang, W.; Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **2000**, *303*, 567−582.
(64) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(65) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.
(66) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.

JM061158I

doi:10.1016/j.jmb.2007.04.081 J. Mol. Biol. (2007) 370, 598–607

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

# Computational Characterization of Structural Role of the Non-active Site Mutation M36I of Human Immunodeficiency Virus Type 1 Protease

**Hirotaka Ode[1]*, Shou Matsuyama[1], Masayuki Hata[1], Saburo Neya[1] Junko Kakizawa[2], Wataru Sugiura[2] and Tyuji Hoshino[1,3]**

[1]*Graduate School of Pharmaceutical Sciences Chiba University 1-33 Yayoi-cho, Inage-ku Chiba 263-8522, Japan*

[2]*AIDS Research Center National Institute of Infectious Diseases, 4-7-1 Gakuen Musashimurayama Tokyo 208-1011, Japan*

[3]*PRESTO, JST, 4-1-8 Honcho Kawaguchi, Saitama 332-0012 Japan*

**Corresponding author*

A prominent characteristic of human immunodeficiency virus type 1 (HIV-1) is its high genetic variability, which generates diversity of the virus and often causes a serious problem of the emergence of drug-resistant mutants. Subtype B HIV-1 is dominant in advanced countries, and the mortality rate due to subtype B HIV-1 has been decreased during the past decade. In contrast, the number of patients with non-subtype B viruses is still increasing in developing countries. One of the reasons for the prevalence of non-subtype B viruses is lack of information about the biological and therapeutic differences between subtype B and non-subtype B viruses. M36I is the most frequently observed polymorphism in non-subtype B HIV-1 proteases. However, since the 36th residue is located at a non-active site of the protease and has no direct interaction with any ligands, the structural role of M36I remains unclear. Here, we performed molecular dynamics (MD) simulations of M36I protease in complex with nelfinavir and revealed the influence of the M36I mutation. The results show that M36I regulates the size of the binding cavity of the protease. The reason for the rare emergence of D30N variants in non-subtype B HIV-1 proteases was also clarified from our computational analysis.



*Keywords:* HIV-1; protease; resistance; M36I; non-active site mutation

## Introduction

Human immunodeficiency virus type 1 (HIV-1) is one of the most hazardous viruses for humans. HIV-1 has a high genetic variability and has been classified into three groups: M, N, and O groups. Viruses in group M are further subdivided into subtypes, and circulating recombinant forms (CRFs). The subtype B virus is commonly found in HIV-1-infected patients in North America, Europe, and Japan. Some anti-HIV-1 drugs that target viral proteins (reverse transcriptase, aspartic protease, and fusion proteins) have been clinically used and have lowered the death rate due to acquired immune deficiency syndrome (AIDS) in advanced countries during the past decade. In contrast, developing countries are suffering from a growing epidemic of non-subtype B viruses. The number of patients infected with HIV-1 is still increasing in developing countries.[1]

HIV-1 proliferates under the assistance of its own aspartic protease, so-called HIV-1 protease (HIV-1 PR), in its life cycle.[2] HIV-1 PR is an enzyme composed of two identical polypeptides, each of which consists of 99 amino acid residues (Figure 1(a)), and has a function to process the viral Gag and Gag-Pol polyprotein precursors. Because this processing is essential for viral maturation, inhibition of the PR function leads to incomplete viral replication and prevents the transfer to other human cells.[3] Therefore, HIV-1 PR is an attractive target for anti-HIV-1 drugs. Nine PR inhibitors (PIs) have been approved by the Food and Drug Administration (FDA).[4–12] Nevertheless, the currently available PIs were developed and tested only against subtype B PRs. Some studies have shown that non-subtype B viruses show different responses than the responses of subtype B virus

Abbreviations used: HIV-1, human immunodeficiency virus type 1; PR, protease; PI, protease inhibitor; NFV, nelfinavir; FDA, Food and Drug Administration; WT, wild-type; MD, molecular dynamics.

E-mail address of the corresponding author: odehir@graduate.chiba-u.jp

0022-2836/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.

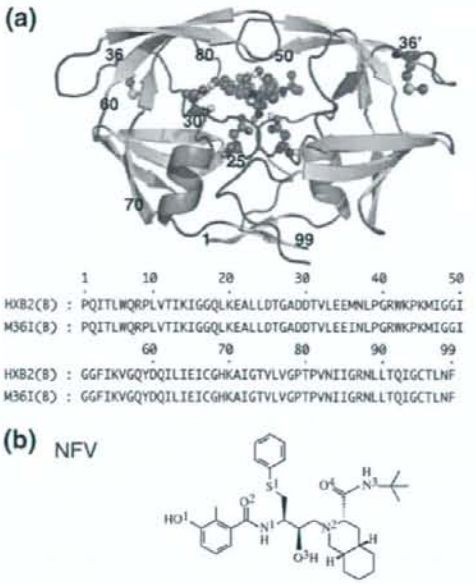


**Figure 1.** (a) Structure of HIV-1 PR in complex with NFV. Locations of two catalytic aspartate residues, the 30th and the 36th residues, and NFV are shown in ball and stick representations. The amino acid sequence of subtype B wild-type (WT) HIV-1 PR (HXB2) and that of M36I mutant are shown under the structure. (b) Chemical structure of NFV.

to those PIs.[13–18] Ariyoshi *et al.* reported that the patterns of emergence of resistant mutations against nelfinavir (NFV, Figure 1(b)) differed between subtype B and CRF01_AE HIV-1.[17] D30N predominantly appears in patients who are infected with subtype B HIV-1, and treatment with NFV has failed. D30N specifically confers resistance against NFV.[19,20] In contrast, N88S predominantly appears in patients with CRF01_AE HIV-1. D30N is not found in patients with CRF01_AE HIV-1. D30N has also rarely been found in patients with other non-subtype B HIV-1.[21–24] However, there have been few studies on the susceptibility of non-subtype B viruses to the presently available PIs, and a standard protocol of chemotherapy for non-subtype B viruses has not yet been established.[25]

Non-subtype B HIV-1 PRs have natural polymorphisms. Among those polymorphisms, M36I is the most frequently observed in patients with non-subtype B HIV-1.[24,26,27] M36I is known to be related to resistance against some FDA-approved PIs: NFV, indinavir, ritonavir, and atazanavir.[20] Therefore, M36I is a key mutation to understand the difference between subtype B and non-subtype B HIV-1 PRs. Nevertheless, the structural role of the non-active site mutation M36I remains unclear. This is because the 36th residue of PR is not located at the active site of PR and has no direct interaction with any substrates or any PIs.

Here, we investigated the structural role of M36I of PR in the binding of PR with NFV through computational simulations. We also examined the structural role of M36V in order to compare the effect of the side-chain at the 36th residue. Some previous computational simulations have revealed structural roles of non-active site mutations of HIV-1 PR.[28–34] Our simulations indicated that the mutations at the 36th residues influence the volume of the binding cavity of the active site of PR. The influence on the volume alters the interaction between D30 of PR and NFV. We also investigated the relationship between D30N and the mutation at the 36th residue of PR. Our results revealed that a combination of D30N and the mutation at the 36th residue induces the formation of hydrogen bonds between N30 of PR and NFV, which are not observed in PR with a single D30N mutation in complex with NFV. Our simulations suggest that M36I, despite the fact that it is a mutation at a non-active site, regulates the binding of ligands with PR. These findings will be helpful for developing more effective PIs against non-subtype B HIV-1.

## Results

### Hydrogen bond networks

We performed MD simulations of six subtype B HIV-1 PRs in complex with NFV: wild-type (WT) PR, M36I PR, M36V PR, D30N PR, D30N/M36I PR, and D30N/M36V PR (labeled as B(WT), B(M36I), B(M36V), B(D30N), B(D30N/M36I), and B(D30N/M36V), respectively). First, in order to clarify the structural role of the mutation at the 36th residue, we examined direct or one-water-molecule-mediated hydrogen bonds between each PR and NFV (Table 1, Supplementary Data Table S1). According to the X-ray structure of HIV-1 PR bound to NFV,[35] NFV has direct and one-water-molecule-mediated hydrogen bonds with D25/D25′, I50/I50′, D29′, and D30 (Supplementary Data Figure S1). In all of the six models, the side-chains of D25 and D25′ directly interacted with the central hydroxyl group of NFV (O3 in Figure 1(b)). All of models also have one-water-molecule-mediated hydrogen bonds of NFV with I50/I50′ and with D29′. Different hydrogen bonds are observed between the 30th residue and the phenol group of NFV. In B(WT) and B(M36V), either the main chain or the side-chain of D30 interacts with NFV through hydrogen bonds (Figure 2). In B(M36I), only the side-chain of D30 forms a hydrogen bond with NFV. In contrast, B(D30N) has no hydrogen bond with NFV. B(D30N/M36I) and B(D30N/M36V) show a different feature from that of B(D30N). They have direct hydrogen bonds between the main chain of N30 and NFV.

### Change in conformation of PR

Next, we investigated the difference in the average structure of each PR from that of B(WT). The average structure of each PR was generated from 1000 snapshot structures during the last 1.0 ns of molecular dynamics (MD) simulations. In order to

**Table 1.** Hydrogen bond networks of NFV with D30 or N30 in PR

| Donor | | Acceptor | | %[a] | Donor | | Acceptor | | % |
|---|---|---|---|---|---|---|---|---|---|
| WT | | | | | D30N | | | | |
| N | D30 | O1[b] | NFV | 30.7 | | | | | |
| O1 | NFV | OD1/OD2 | D30 | 31.6 | | | | | |
| O1 | NFV | O | D30 | 42.9 | | | | | |
| M36I | | | | | D30N/M36I | | | | |
| O1 | NFV | OD2 | D30 | 95.3 | O1 | NFV | O | N30 | 57.1 |
| M36V | | | | | D30N/M36V | | | | |
| N | D30 | O1 | NFV | 18.0 | N | N30 | O1 | NFV | 24.9 |
| O1 | NFV | OD1/OD2 | D30 | 46.1 | | | | | |
| O1 | NFV | O | D30 | 35.9 | O1 | NFV | O | N30 | 34.7 |

[a] Occupancy of hydrogen bonds during the 2.0–3.0 ns MD simulation.
[b] The atom names of NFV are shown in Figure 1.

make a comparison between each PR and B(WT), we superimposed the average structure of each PR onto that of B(WT) using the coordinates of N, $C^{\alpha}$, and C atoms. Since non-active site residues are more flexible than active site residues and the structure of non-active site residues is largely influenced by random atom motions in MD simulation (Supplementary Data Figure S2), we focused on conformational change of the active site residues. Figure 3 shows that the displacements of most of the active site residues are small. Detailed values are shown in Supplementary Data Figure S3. Exceptionally, a large conformational change (RMSD=1.5 Å) is observed at D29 in B(D30N/M36I). In the other PRs, D29 is shifted about 0.7 Å. D30 is noticeably displaced by 0.8 Å in only B(D30N), while displacement of the 30th residues is less than 0.5 Å in the

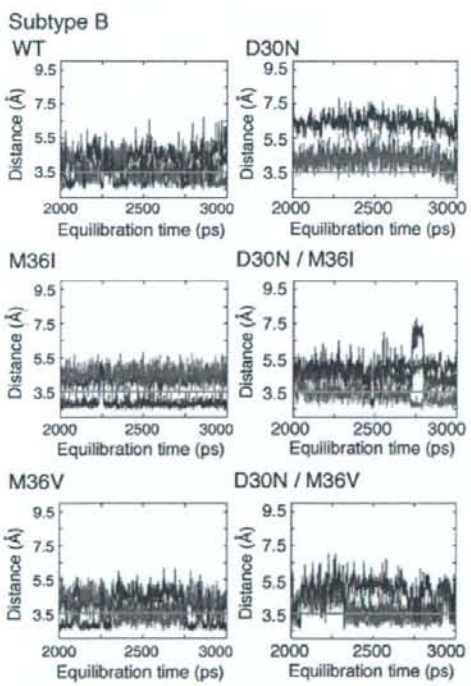


**Figure 2.** Distance between the 30th residue of PR and NFV. Red and green continuous lines correspond to the distance between N of the 30th residue and O1 atom of NFV and that between O of the 30th residue and O1 atom of NFV, respectively. Blue continuous lines of B(WT), B(M36I), and B(M36V) show the distance between OD1/OD2 of D30 and O1 atom of NFV, while those of B(D30N), B(D30N/M36I), and B(D30N/M36V) are the distance between OD1/ND2 of N30 and O1 atom of NFV.

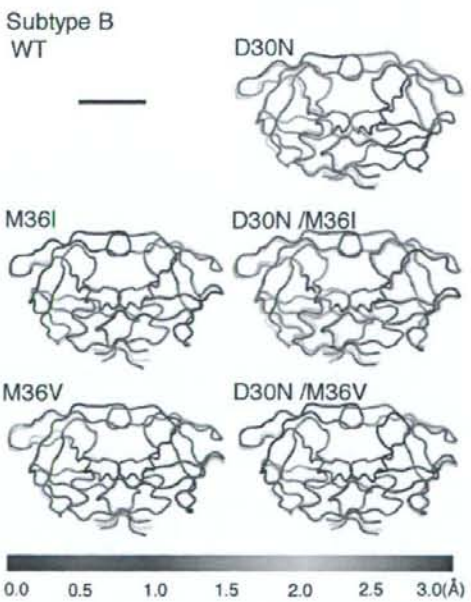


**Figure 3.** 3D plot of RMSD of the average structure of each model from that of B(WT). PR is shown in colored tube representations. Color indicates the magnitude of RMSD shown in the bottom bar. Each model was fitted to B(WT) using the coordinates of main chain atoms N, $C^{\alpha}$, and C of PR. The superimposed gray tubes represent the average structure of B(WT).

**Table 2.** Volume and surface area of the binding cavity in each model

| | WT | M36I | M36V | D30N | D30N/M36I | D30N/M36V |
|---|---|---|---|---|---|---|
| Volume (Å$^3$) | 375(±43)[a] | 345(±36) | 365(±43) | 403(±53) | 419(±53) | 375(±47) |
| Surface area (Å$^2$) | 490(±43) | 471(±35) | 469(±47) | 508(±40) | 506(±54) | 484(±48) |

[a] The standard deviations were written in the parentheses.

other PRs. At a flap region and around 80s loops, slight conformational changes are also observed. The flap and 80s loop consist of the 47th to 50th residues and of the 79th to 81st residues, respectively. Although HIV-1 PR is a homodimer, the asymmetric displacements are observed. The displacement in NFV was also examined (Supplementary Data Figure S4). Conformational changes in the benzamide group and the S-phenyl group (C6H6-S-) are observed. B(D30N) and B(D30N/M36I) show larger displacement than that of the other models. In contrast, the *tert*-butylcarboxamide moiety and the dodecahydroisoquinoline ring hardly change their locations in each model. The displacements of NFV are also asymmetric and correlate with the displacements of PR. The residues near the benzamide group of NFV, such as the 29th and 30th residues, show large displacement. On the other hand, the residues near the *tert*-butylcarboxamide moiety or the dodecahydroisoquinoline ring of NFV show small displacements. The asymmetric displacements observed in HIV-1 PR will be due to the structural collision with NFV.

Furthermore, we examined the effect of the mutation at the 36th residue on volume of the active site (Table 2). M36I reduces the volume of the active site, whereas both D30N and D30N/M36I increase it. M36V and D30N/M36V have almost no effect on the volume.

### Binding energy calculations

Binding energy between each PR and NFV is presented in Table 3. A single M36I or M36V mutation has almost no effect on the binding affinity with NFV, although the binding energy calculations have large standard deviations. In contrast, D30N and D30N/M36I mutations reduce the binding energies with NFV. Oppositely, D30N/M36V increases the affinity. Clemente *et al.* reported the IC50 values of B(WT), B(D30N), B(M36I), and B(D30N/M36I),[36] in which M36I was suggested to improve the binding affinity with NFV, and showed that D30N/M36I conferred resistance against NFV. Our results are compatible with those experimental results.

We additionally analyzed the contribution of each residue to the binding of NFV (Supplementary Data Figure S5). In B(M36I), B(M36V), and B(D30N/M36V), the respective residues except D25/D25′ have contributions similar to those in B(WT). The contribution of D25/D25′ depends on their protonation states. In contrast, B(D30N) reduces the binding energy between the 30th residue and NFV. B(D30N/M36I) lowers the interaction between D29 and NFV.

## Discussion

Here, we performed MD simulations of HIV-1 PRs in complex with NFV for the purpose of clarifying (1) the structural role of the non-active site mutation M36I and (2) the relationship between D30N and M36I mutations. M36V was also examined in order to analyze the effect of the side-chain of the 36th residue.

Non-subtype B HIV-1 is still pandemic in the world.[1] Nevertheless, a standard protocol of chemotherapy for non-subtype B viruses has not yet been established,[25] and little is known about the difference between susceptibilities of non-subtype B viruses and subtype B virus to clinically available drugs.[13–18] M36I is the most frequently observed polymorphism among non-subtype B HIV-1 PRs.[24,26,27] Therefore, M36I is a key mutation to clarify the reason why the efficacy of PIs varies among subtypes. Some studies have indicated that M36I is related to the resistance against NFV and other FDA-approved PIs by complementing the effects of other resistant mutations.[17,20,34,36] For example, M36I has been shown to contribute to an increase in the emergence rate of the NFV-resistant mutation N88S.[17,34] It has also been reported that a

**Table 3.** Binding free energy of each model

| | $\Delta G_{int}^{ele}$ (kcal/mol) | $\Delta G_{int}^{vdw}$ (kcal/mol) | $\Delta G_{sol}$ (kcal/mol) | $\Delta G_b$[a] (kcal/mol) | $\Delta\Delta G_b$ (kcal/mol) | IC50[b] (nM) |
|---|---|---|---|---|---|---|
| WT | −12.5±1.4 | −71.8±3.8 | 15.1±1.4 | −69.2±3.7 | – | 1.2±0.2 |
| M36I | −11.8±1.2 | −73.3±3.7 | 15.4±1.1 | −69.7±3.5 | −0.5 | 0.9±0.1 |
| M36V | −13.2±1.5 | −72.1±3.8 | 16.5±1.3 | −68.8±3.6 | 0.4 | ND[c] |
| D30N | −6.9±1.2 | −70.5±4.1 | 10.9±0.9 | −66.5±3.9 | 2.7 | 6.8±0.9 |
| D30N/M36I | −8.1±1.2 | −66.8±3.6 | 9.8±1.1 | −65.0±3.5 | 4.2 | 6.0±1.0 |
| D30N/M36V | −10.8±1.2 | −72.8±3.9 | 13.3±1.0 | −70.3±4.0 | −1.1 | ND |

[a] $T\Delta S$ is not included.
[b] Reference from the report by Clemente *et al.*[35]
[c] ND denotes no data.

single M36I mutation does not confer resistance against FDA-approved PIs.[36] The structural role of M36I is, however, not clear because the 36th residue of PR is located at a non-active site of HIV-1 PR. According to the results of the present simulations, a single M36I mutation reduces the volume of the binding cavity of subtype B HIV-1 PR. Energetically, M36I PR slightly increases the binding affinity with NFV, compared with WT PR. This result is compatible with an experimental finding.[36] A single M36V mutation also reduces the volume of the cavity, although the effect of M36V on the cavity volume is not as great as that of M36I. When we carried out an additional 2.0 ns MD simulation of each model, the reduction of the volume of the active site of M36I PR and that of M36V PR were also observed (Supplementary Data Table S2). Therefore, the side-chain of the 36th residue will indeed influence the shape of the active site of PR, despite its location.

In order to reveal the mechanism by which the non-active site mutation M36I regulates the volume of the active site of PR, we additionally calculated the interaction energy between each residue of PR and the 36th residue. The energy calculations suggest that the 36th residue mainly interacts with I15/I15′, Q18/Q18′, K20/K20′, L33/L33′, and V77/V77′ (Supplementary Datal Figure S6). Among these residues, L33/L33′ and V77/V77′ are located near active site residues (Figure 4). V77/V77′ is close to the 80s loop, which consists of the residues from P79/P79′ to P81/P81′. L33/L33′ is in the vicinity of T31/T31′. Interestingly, mutations at L33 and V77 (L33F and V77I) are also related to the resistance against some PIs.[20] These mutations would play a role similar to that of M36I. We speculate that the slight inward shifts of L33/L33′ and V77/V77′ due to the mutation M36I trigger the reduction in volume of the cavity of PR. In order to confirm this speculation, we further examined the shift of each residue and extracted the shift only toward the center of the active site (Figure 5, Supplementary Data Figures S7–S9). The results show that M36I causes a shrinkage of the active site around P79 and around T31′. The 31st residue T31/T31′ creates stable hydrogen bond networks with D29, T74, and N88 or with D29′, T74′, and N88′ (Supplementary Data Table S3).[28,34] also indirectly influences the conformations around D29/D29′. A28′, D29′, D30′, and T31′ in one monomer are all displaced inward. In contrast, A28 and D29 show outward positional shifts, while the residues around L33 in the other monomer move inward. It is notable that these residues rotate on D30. M36I hardly shrinks or expands the cavity at D30. However, D30 shifts toward the α-helix region (R87–I93) in B(M36I) (Supplementary Data Figure S10). Although this displacement is very slight (0.4 Å), the shift enlarges the distance from the main chain of D30 to the *m*-phenol group of NFV, and it also shortens the distance from the side-chain of D30 to NFV (Figure 2). Therefore, M36I changes the interaction between D30 and NFV. D30 is an important residue for binding with NFV.[28,34,35] B(M36I) has a hydrogen bond of the *m*-phenol group of NFV only with the side-chain of D30, unlike B(WT). B(WT) forms hydrogen bonds of the *m*-phenol group of NFV with either the main chain or the side-chain of D30. These results show that the non-active site mutation M36I influences the shape of the active site of PR by the following mechanism. M36I mutation shifts L33/L33′ and V77/V77′ inward. Subsequently, these shifts cause changes in conformation at the active site, especially around T31/T31′ and P79/P79′.

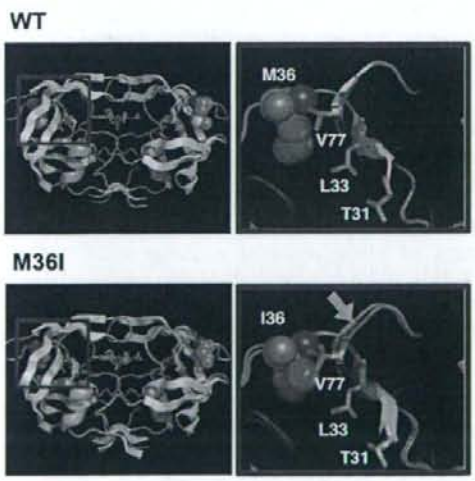


**Figure 4.** Interactions of the 36th residues with L33 and with V77. The average structure of B(WT) was superimposed onto that of B(M36I) using the coordinates of N, $C^{\alpha}$, and C atoms. The superimposed structure of B(WT) is shown in gray sticks and cartoons representation. The orange arrow indicates the locations of the 80s loop, where a prominent change of conformation occurred by M36I.

It is also informative to investigate the relationship between D30N and M36I. As stated in Introduction, most of the non-subtype B HIV-1 PRs carry M36I as a polymorphism.[24,26,27] In contrast, D30N rarely appears in non-subtype B HIV-1 PRs.[21–24] This rare emergence of D30N in non-subtype B PRs has been assumed to be due to the low viral replication ability of D30N mutants in non-subtype B viruses,[22,37] because L89M, which is a polymorphism of some non-subtype B PRs, and D30N decrease the replication ability.[22,37] On the other hand, our simulations provide a novel insight into the relationship between D30N and M36I. It should be noted that B(D30N/M36I) forms a hydrogen bond between the main chain of N30 and NFV (Figure 2), which is not seen in B(D30N). M36I enforces the interaction between N30 and NFV in D30N mutant PR. A clear difference is observed at D29 among B(WT), B(M36I), B(D30N), and B(D30N/M36I) (Figure 5). D29 in B(D30N/M36I) shows a large outward positional shift from that of B(WT). The shift of B(D30N/M36I) is two-times larger than that of B(M36I) and B(D30N). Binding energy calculations show that D30N/M36I lowers the binding affinity of NFV with subtype B PR (Table 3) and greatly

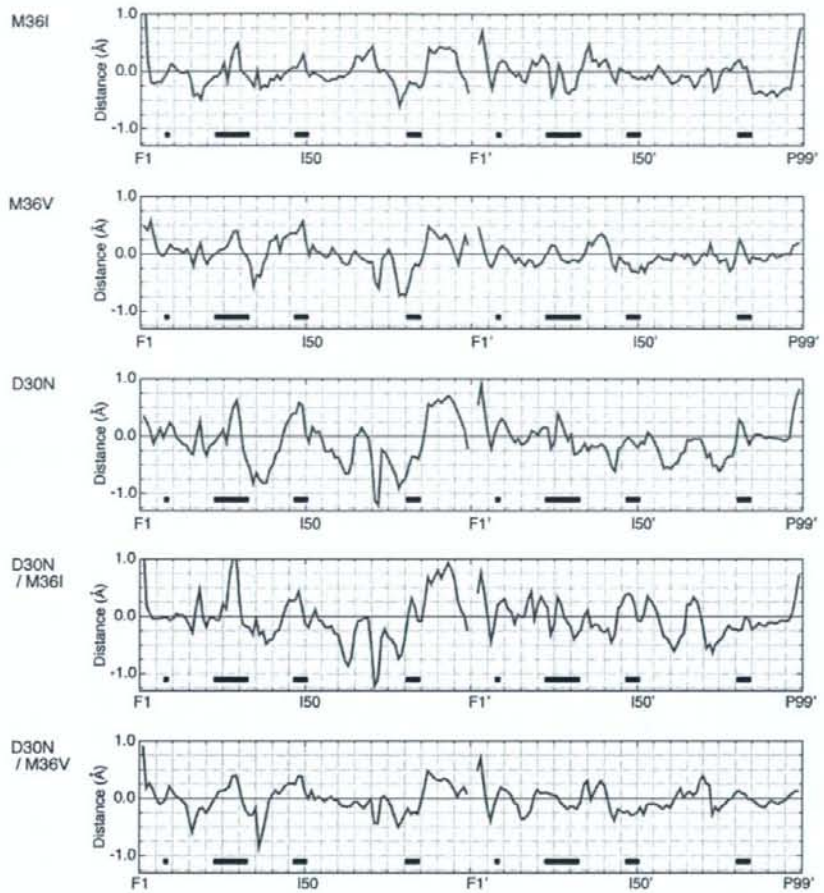


**Figure 5.** Positional shift of each residue measured from the center of the binding cavity in the average coordinate of WT PR. Negative values indicate contraction of distances measured from the center of the cavity, and positive values indicate elongation of the distances. Bottom black lines represent the locations of the active site residues.

reduces the interaction between D29 and NFV. D30N/M36I mutant of subtype B PR was reported to confer resistance against NFV.[36] This energetical result is compatible with experimental findings.[36] These results suggest that the resistance mechanism due to D30N/M36I is different from that of D30N. D30N causes resistance against NFV because of the loss of hydrogen bonds between N30 and NFV.[28,34] On the other hand, D30N/M36I confers resistance against NFV due to the large change in conformation at D29. NFV still interacts with N30 in B(D30N/M36I). It is known that D30N is not observed in CRF01_AE PR.[17] As we previously reported,[34] this observation is explained by the finding that the affinity of CRF01_AE D30N PR with NFV is similar to that of CRF01_AE reference PR. CRF01_AE D30N PR has stronger interaction with NFV than does subtype B D30N PR. CRF01_AE D30N PR has one-water-molecule-mediated hydrogen bonds between NFV and N30, whereas subtype B D30N PR forms no hydrogen bond between them. M36I enforces the interaction between N30 and NFV, and polymorphisms other than M36I will release the distortion at D29. NFV will be able to maintain its efficacy against D30N mutant PR in non-subtype B viruses. Therefore, D30N mutation is rarely observed in non-subtype B HIV-1 PRs.

Here, we performed MD simulations of six HIV-1 PRs in complex with NFV and clarified the structural role of the non-active site mutation M36I. M36I influences the shape of the active site of PR in spite of its location. The changes in conformation at the active site are caused by the alteration of interaction of the 36th residue with L33 and V77. We also examined the relationship between D30N and M36I. A combination of D30N and M36I enforces the interaction between N30 and NFV, causing distortion on the conformation around D29. D30N/M36I shows a different mechanism of resistance against NFV from that of D30N. M36I is the most frequently observed mutation in non-subtype B PRs. Therefore, M36I is a key mutation to understand the differences

between subtype B and non-subtype B PRs. The findings of this study provide valuable information for developing the drugs that are more effective for non-subtype B PRs. Furthermore, accumulation of information on structural roles of key residues of PR will enable us to predict the effectiveness of PIs against non-subtype B viruses as well as against subtype B ones.

## Materials and Methods

### Molecular dynamics simulation

We performed minimizations and MD simulations in a manner similar to that described elsewhere.[34] Minimizations and MD simulations were carried out using the Sander module of the AMBER8 package.[38] The AMBER ff03 force field[39] was used as the parameters for proteins, ions, and water molecules. Our originally developed torsion parameters for the benzamide moiety in NFV, CA-CA-C –N and CA-CA-C –O, were applied.[34] The general AMBER force field[40] was used as other parameters for NFV. RESP charges for NFV were determined on the basis of data obtained from quantum chemical calculations.[34]

We performed simulations of six proteases in complex with NFV: wild-type (WT) PR, M36I PR, M36V PR, D30N PR, D30N/M36I PR, and D30N/M36V PR of subtype B HIV-1 (labeled as B(WT), B(M36I), B(M36V), B(D30N), B(D30N/M36I), and B(D30N/M36V), respectively). B(M36V) and B(D30N/M36V) were examined in order to analyze the effect of the side-chain of the 36th residue clearly. We used HXB2 as the WT sequence of subtype B HIV-1. Each initial structure for the PR in complex with NFV was modeled from the atom coordinates of an X-ray crystal structure (PDB code: 1OHR[35]) and the respective mutations were introduced using the LEaP module. First, we obtained the PDB file of the PR bound to NFV from Protein Data Bank†. Second, we edited the PDB file to change the residue names of the mutated residues and to delete the information on the coordinates of the side-chain atoms of the mutated residues. Third, the coordinates of the side-chain atoms of the mutated residues were automatically generated by LEaP module. Fourth, each model was placed in a rectangular box filled with about 8000 TIP3P water molecules,[41] with all of the crystal water molecules remaining. The cutoff distance for the long-range electrostatic and the van der Waals energy terms was set to 12.0 Å. The expansion and shrinkage of all covalent bonds connecting to a hydrogen atom were constrained using the SHAKE algorithm.[42] Periodic boundary conditions were applied to avoid the edge effect in all calculations. Energy minimization was achieved in three steps. First, movement was allowed only for water molecules and ions. Next, the ligand and the mutated residues were allowed to move in addition to the water molecules and ions. In this step, steric collisions of the automatically generated residues were minimized and favorable configurations of the side-chains of the mutated residues were obtained. Finally, all atoms were permitted to move freely. In each step, energy minimization was executed by the steepest descent method for the first 10,000 cycles and the conjugated gradient method for the subsequent 10,000 cycles. After a 0.1 ns heating calculation until 310 K using the NVT ensemble, a 3.0 ns equilibrating calculation was executed at 1 atm and at 310 K using the NPT ensemble, with an integration time step of 2.0 fs. In the present calculations, the MD simulations showed no large fluctuations after about 2.0 ns equilibrating calculation (Supplementary Data Figures S11 and S12). Hence, atom coordinates were collected at intervals of 1.0 ps for the last 1.0 ns to analyze the structure in detail. Furthermore, we performed an additional 2.0 ns simulation for each model and confirmed equilibrium of each simulation (Supplementary Data Figure S13).

Protonation states of the catalytic aspartic acids D25 and D25′ vary depending on the binding ligands or PRs.[43] Hence, appropriate protonation states of the catalytic aspartic acids should be determined for each model. Since NFV mimics a transition state of catalytic reaction by HIV-1 PR, we considered two kinds of protonation states.[44–46] One complex represented a combination of protonated D25/unprotonated D25′ states, and the other represented the opposite combination. In order to determine the protonation states when NFV is bound to each PR, the free energies of two kinds of complexes were compared using calculation data obtained during the 2.0–3.0 ns MD simulations (Supplementary Data Table S4). The free energies were calculated on the basis of the MM/PBSA method.[47,48] We used the same parameter set for electrostatic and van der Waals energy terms as that used in the MD simulations, and no cutoff was applied for the calculation. Since the dielectric constants for the interior of proteins is considered to be in the range of 2 to 4, the interior dielectric constant was set to 2.0.[49] The outer dielectric constant was set to 80.0. The pbsa program was used to solve the Poisson-Boltzmman (PB) equation. B(D30N), B(M36I), and B(D30N/M36I) have been found to favor the combination of protonated D25 and unprotonated D25′. The other three PRs, B(WT), B(M36V) and B(D30N/M36V), prefer the combination of unprotonated D25 and protonated D25′.

### Hydrogen bond criteria

The formation of a hydrogen bond was defined in terms of distance and orientation. The combination of donor D, hydrogen H, and acceptor A atoms with a D - H ... A configuration was regarded as a hydrogen bond when the distance between donor D and acceptor A was shorter than 3.5 Å and the angle H-D-A was smaller than 60.0 degrees.

### Calculations of volume and surface area of the binding cavity

We employed the Pocket program[50] to estimate the volume and the surface area of the active site of PR. The program is based on the Alpha Shape theory,[50] which provides an analytical method for detecting pockets in proteins and measuring their volume and surface area. The ligand-binding cavity of HIV-1 PR is not completely separated from solvent. However, it should be noted that, in the Pocket program, a pocket is defined as a cavity that is inaccessible to the solvent outside of a protein. If a water molecule is trapped within the pocket, the water molecule cannot escape to the outside of the protein. Hence, we can define the ligand-binding cavity as a pocket. Supplementary Data Figure S14 shows the binding pocket of HIV-1 PR visualized by MAGE program.[51]

† http://www.pdb.org/

### Binding free energy calculation

The binding free energy[52] was calculated by the following equation:

$$\Delta G_b = \Delta G_{int}^{ele} + \Delta G_{int}^{vdw} + \Delta G_{sol} - T\Delta S,$$

where $\Delta G_b$ is the binding free energy in solution, $\Delta G_{int}^{ele}$ and $\Delta G_{int}^{vdw}$ are the electrostatic and van der Waals interaction energies between a ligand and a protein, $\Delta G_{sol}$ is the solvation energy, and $-T\Delta S$ is the contribution of conformational entropy to the binding. Here, assuming that the contribution of conformational entropy to the change in $\Delta G_b$ is negligible among mutants,[53] we disregarded the entropy term in the energy estimation. $\Delta G_{int}^{ele}$ and $\Delta G_{int}^{vdw}$ were computed using the same parameter set as that used in the MD simulation, and no cutoff was applied for the calculation. Solvation energy $\Delta G_{sol}$ was calculated using the pbsa program. The interior dielectric constant was set to 2.0[49] and the outer dielectric constant was set to 80.0. Furthermore, the contribution of each residue to the binding free energy was calculated. The total binding free energy was decomposed into the contribution from each individual residue by the MM/GBSA method. The modified GB model developed by Onufriev *et al.*[54] was used to calculate the solvation energy term. The MM/GBSA results were highly correlated with the MM/PBSA results, as we described previously.[34]

## Acknowledgements

This work was supported by a Health and Labor Science Research Grant for Research on HIV/AIDS from the Ministry of Health and Labor of Japan and by JSPS Research Fellowships for Young Scientists and a grant-in-aid for JSPS Fellows. A part of this work was also supported by a grant-in-aid from Japan Society for the Promotion of Science, No. 19590467, and by a grant from the Japan Science and Technology Agency.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.04.081

## References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS). (2006). 2006 Report on the Global AIDS Epidemic. *UNAIDS*, Geneva, Switzerland.
2. Krausslich, H. G. & Wimmer, E. (1988). Viral proteinases. *Annu. Rev. Biochem.* **57**, 701–754.
3. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. *et al.* (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl Acad. Sci. USA*, **85**, 4686–4690.
4. Craig, J. C., Duncan, I. B., Hockley, D., Grief, C., Roberts, N. A. & Mills, J. S. (1991). Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral Res.* **16**, 295–305.
5. Vacca, J. P., Dorsey, B. D., Schleif, W. A., Leven, R. B., McDaniel, S. L., Darke, P. L. *et al.* (1994). L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc. Natl Acad. Sci. USA*, **91**, 4096–4100.
6. Kempf, D. J., Marsh, K. C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentga, C. A. *et al.* (1995). ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl Acad. Sci. USA*, **92**, 2484–2488.
7. Livington, D. J., Pazhanisamy, S., Porter, D. J., Partaledis, J. A., Tung, R. D. & Painter, G. R. (1995). Weak binding of VX-478 to human plasma proteins and implications for anti-human immunodeficiency virus therapy. *J. Infect. Dis.* **172**, 1238–1245.
8. Patick, A. K., Mo, H., Markowitz, M., Appelt, K., Wu, B., Musick, L. *et al.* (1996). Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **40**, 292–297 (Erratum, **40**, 1575).
9. Carrillo, A., Stewart, K. D., Sham, H. L., Norbeck, D. W., Kohlbrenner, W. E., Leonard, J. M. *et al.* (1998). In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *J. Virol.* **72**, 7532–7541.
10. Robinson, B. S., Riccardi, K. A., Gong, Y. F., Guo, Q., Stock, D. A., Blair, W. S. *et al.* (2000). BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.* **44**, 2093–2099.
11. Larder, B. A., Hertogs, K., Bloor, S., van den Eynde, C., DeCian, W., Wang, Y. *et al.* (2000). Tipranavir inhibits broadly protease inhibitor-resistant HIV-1 clinical samples. *AIDS*, **14**, 1943–1948.
12. Koh, Y., Nakata, H., Maeda, K., Ogata, H., Bilcer, G., Devasamudram, T. *et al.* (2003). Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **47**, 3123–3129.
13. Cornelissen, M., van den Burg, R., Zorgdrager, F., Lukashov, V. & Goudsmit, J. (1997). pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J. Virol.* **71**, 6348–6358.
14. Pieniazek, D., Rayfield, M., Hu, D. J., Nkengasong, J., Wiktor, S. Z., Downing, R. *et al.* (2000). Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV variant working group. *AIDS*, **14**, 1489–1495.
15. Vergne, L., Peeters, M., Mpoudi-Ngole, E., Bourgeois, A., Liegeois, F., Toure-Kane, C. *et al.* (2000). Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients. *J. Clin. Microbiol.* **38**, 3919–3925.
16. Grossman, Z., Vardinon, N., Chemtob, D., Alkan, M. L., Bentwich, Z., Burke, M. *et al.* (2001). Genotypic variation of HIV-1 reverse transcriptase and protease:

comparative analysis of clade C and clade B. *AIDS*, **15**, 1453–1460, (Erratum, **15**, 2209).
17. Ariyoshi, K., Matsuda, M., Miura, H., Tateishi, S., Yamada, K. & Sugiura, W. (2003). Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. *J. AIDS*, **33**, 336–342.
18. Clemente, J. C., Coman, R. M., Thiaville, M. M., Janka, L. K., Jeung, J. A., Nukoolkarn, S. *et al.* (2006). Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry*, **45**, 5468–5477.
19. Sugiura, W., Matsuda, Z., Yokomaku, Y., Hertogs, K., Larder, B., Oishi, T. *et al.* (2002). Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **46**, 708–715.
20. Johnson, V. A., Brun-Vézinet, F., Clotet, B., Kuritzkes, D. R., Pillay, D., Schapiro, J. M. & Richman, D. D. (2006). Update of the drug resistance mutations in HIV-1: Fall 2006. *Top. HIV Med.* **14**, 125–130.
21. Cane, P. A., Ruiter, A., Rice, P., Wiselka, M., Fox, R. & Pillay, D. (2001). Resistance-associated mutations in the human immunodeficiency virus type 1 subtype C protease gene from treated and untreated patients in the United Kingdom. *J. Clin. Microbiol.* **39**, 2652–2654.
22. Grossman, Z., Paxinos, E. E., Averbuch, D., Maayan, S., Parkin, N. T., Engelhard, D. *et al.* (2004). Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrob. Agents Chemother.* **48**, 2159–2165.
23. Gonzalez, L. M. F., Brindeiro, R. M., Aguiar, R. S., Pereira, H. S., Abreu, C. M., Soares, M. A. & Tanuri, A. (2004). Impact of nelfinavir resistance mutations on in vitro phenotype, fitness and replication capacity of human immunodeficiency virus type 1 with subtype B and C proteases. *Antimicrob. Agents Chemother.* **48**, 3552–3555.
24. Kantor, R., Katzenstein, D. A., Efron, B., Carvalho, A. P., Whynhoven, B., Cane, P. *et al.* (2005). Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *Pros. Med.* **2**, e112.
25. Billingam, D. (2006). Does the world need another AIDS authority? *Lancet*, **368**, 1639–1640.
26. Kantor, R. & Katzenstein, D. (2003). Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS Rev.* **5**, 25–35.
27. Nkengasong, J. N., Adje-Toure, C. & Weidle, P. J. (2004). HIV antiretroviral drug resistance in Africa. *AIDS Rev.* **6**, 4–12.
28. Ode, H., Ota, M., Neya, S., Hata, M., Sugiura, W. & Hoshino, T. (2005). Resistant mechanism against nelfinavir of human immunodeficiency virus type-1 proteases. *J. Phys. Chem. B*, **109**, 565–574.
29. Skalova, T., Dohnalek, J., Duskova, J., Petrokova, H., Hradilek, M., Soucek, M. *et al.* (2006). HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site. *J. Med. Chem.* **49**, 5777–5784.
30. Piana, S., Carloni, P. & Rothlisberger, U. (2002). Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* **11**, 2393–2402.
31. Ode, H., Neya, S., Hata, M., Sugiura, W. & Hoshino, T. (2006). Computational simulations of HIV-1 proteases-multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.* **128**, 7887–7895.
32. Meiselbach, H., Horn, A. H. C., Harrer, T. & Sticht, H. (2006). Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. *J. Mol. Model.* **13**, 297–304.
33. Batista, P. R., Wilter, A., Durham, E. H. & Pascutti, P. G. (2006). Molecular dynamics simulations applied to the study of subtypes of HIV-1 protease common to Brazil, Africa, and Asia. *Cell Biochem. Biophys.* **44**, 395–404.
34. Ode, H., Matsuyama, S., Hata, M., Hoshino, T., Kakizawa, J. & Sugiura, W. (2007). Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *J. Med. Chem.* **50**, 1768–1777.
35. Clemente, J. C., Hermrajani, R., Blum, L. E., Goodenow, M. M. & Dunn, B. M. (2003). Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry*, **42**, 15029–15035.
36. Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K. *et al.* (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **40**, 3979–3985.
37. Abecasis, A. B., Deforche, K., Snoeck, J., Bacheler, L. T., McKenna, P., Carvalho, A. P. *et al.* (2005). Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. *AIDS*, **19**, 1799–1806.
38. Case, D. A., Darden, T. A., Cheatham, T. E. III, Simmerling, C. L., Wang, J., Duke, R. E. *et al.* (2004). *AMBER 8*, University of California, San Francisco.
39. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W. *et al.* (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012.
40. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). Development and testing of a general Amber force field. *J. Comput. Chem.* **25**, 1157–1174.
41. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
42. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.
43. Zoete, V., Michielin, O. & Karplus, M. (2002). Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **315**, 21–52.
44. Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B. *et al.* (1990). Rational design of peptide-based HIV proteinase inhibitors. *Science*, **248**, 358–361.
45. Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J. & Hatada, M. H. (1991). Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)hydroxyethylamine isostere. *J. Med. Chem.* **34**, 3340–3342.
46. Okimoto, N., Tsukui, T., Hata, M., Hoshino, T. & Tsuda, M. (1999). Hydrolysis mechanism of the

phenylalanine-proline peptide bond specific to HIV-1 protease: investigation by the ab initio molecular orbital method. *J. Am. Chem. Soc.* **121**, 7349–7354.
47. Srinivasan, J., Cheatham, T. E. III, Kollman, P. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
48. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accts. Chem. Res.* **33**, 889–897.
49. Wang, W. & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **303**, 567–582.
50. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.
51. Richardson, J. S. & Richardson, D. C. (2001). MAGE, PROBE, and Kinemages, Chapter 25.2.8. In *International Tables for Crystallography* (Rossmann, M. G. & Arnold, E., eds), vol. F, pp. 727–730, Kluwer Publishers, Dordrecht.
52. Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417.
53. Massova, I. & Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **18**, 113–135.
54. Onufriev, A., Bashford, D. & Case, D. A. (2004). Exploring protein native states and largescale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Genet.* **55**, 383–394.

***Edited by D. Case***

*(Received 12 February 2007; received in revised form 27 April 2007; accepted 29 April 2007)*
Available online 10 May 2007

Jpn. J. Infect. Dis., 59, 77-84, 2006

**Original Article**

# Mutations of Conserved Glycine Residues within the Membrane-Spanning Domain of Human Immunodeficiency Virus Type 1 gp41 Can Inhibit Membrane Fusion and Incorporation of Env onto Virions

Kosuke Miyauchi, Rachael Curran[1], Erin Matthews[1], Jun Komano, Tyuji Hoshino[2], Don M. Engelman[1] and Zene Matsuda*

*Laboratory of Virology and Pathogenesis, AIDS Research Center, National Institute of Infectious Diseases, Tokyo 162-8640;*
*[2]Graduate School of Pharmaceutical Sciences, Chiba University, Chiba 263-8522, Japan; and*
*[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA*

(Received December 28, 2005. Accepted January 19, 2006)

**SUMMARY**: The membrane-spanning domain (MSD) of HIV-1 envelope protein (Env) has an additional glycine residue within a well-conserved putative transmembrane helix-helix interaction motif, GXXXG, and forms a $G^{690}G^{691}XXG^{694}$ sequence (G, glycine; X, any residues; the numbering indicates the position within the Env of an infectious molecular clone, HXB2). Different from vesicular stomatitis virus G (VSV-G), the glycine residues of the GXXXG motif of HIV-1 showed higher tolerance against mutations, and a simultaneous substitution of G690 and G694 with leucine residues only modestly decreased fusion activity and replication capacity of HIV-1. When G691 was further substituted with alanine, phenylalanine or leucine residue while G690 and G694 were substituted with leucine residues, the efficiency of membrane fusion decreased, with the decrease greatest occurring with the leucine substitution, a less severe decrease with phenylalanine, and the least severe decrease with alanine. Substitution with leucine residue also decreased the incorporation of Env onto virions, and the mutant showed the most delayed replication profile. Thus the presence of the extra glycine residue, G691, may increase the tolerance of the other two glycine residues against mutations than VSV-G. The fact that a more severe defect was observed for the leucine residue than the phenylalanine residue suggested that the function of Env depended on the steric nature rather than on the simple volume of the side chain of the amino acid residue at position 691. Based on this result, we propose a hypothetical model of the association among MSDs of gp41, in which $G^{691}$ locates itself near the helix-helix interface.

## INTRODUCTION

The envelope protein (Env) of human immunodeficiency virus type 1 (HIV-1) is a trimer of non-covalently associated heterodimers of gp120 and gp41. As with other retroviruses, gp120 (SU) and gp41(TM) play key roles in the determination of host range and membrane fusion, respectively.

For the three subdomains of HIV-1 gp41, the structure-function relationship of the ectodomain during membrane fusion has been elucidated at the molecular level (1,2). After gp120 binds to the receptor/coreceptor, the ectodomain undergoes a conformational change to form a six-helix bundle, a common structure observed for the class I fusion protein (3,4). Information on the structure-function relationships of the membrane-spanning domain (MSD) is rather limited. Although the amino acid sequence of the MSD of HIV-1's gp41 is highly conserved among different clades, the significance of the specific amino acid sequence within MSD has been underestimated, because some heterologous MSDs can substitute functionally for the native MSD of gp41 (5,6). However, truncation of HIV-1 MSD by a glycosylphosphatidylinositol anchor abolished the fusion activity (7). Furthermore, in viruses such as simian immunodeficiency virus and the influenza virus, mutations introduced into the MSD have been shown to impede the late stage of membrane fusion (8-11). These data suggest the importance of the MSD for the function of Env.

A glycine-containing helix-helix interaction motif, a GXXXG motif, has been found in MSDs of many membrane proteins such as glycophorin A (GpA) (12,13) and the hepatitis C virus envelope glycoproteins (14). In the case of HIV-1, it occurs as a $G^{690}G^{691}XXG^{694}$ sequence (G, glycine residue; X, any amino acid residue; the number indicates the position of each glycine residue in the Env of a molecular clone HXB2 [15]) within the MSD of gp41. The glycine residues within the GXXXG motif are critical for the proper fusogenicity of vesicular stomatitis virus G (VSV-G) (16). In a previous study we have shown that the point mutations of the individual glycine residue of the GXXXG motif of gp41 MSD were well tolerated (17). The molecular basis for this high tolerance of gp41 MSD against mutations was not identified.

Here we hypothesize that the GXXXG motif of gp41, like in other transmembrane helices, forms the helix-helix interface of gp41 MSDs. We reevaluated the role of glycine residues within the gp41 MSD by introducing a simultaneous substitution of several glycine residues. We confirmed the high tolerance of gp41 MSD against mutations through the finding that any combinatorial mutations of two glycine residues with leucine residues were well tolerated. By using one of the mutants, the $LG^{691}XXL$ mutant, we evaluated the role

*Corresponding author: Mailing address: Laboratory of Virology and Pathogenesis, AIDS Research Center, National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku, Tokyo 162-8640, Japan. Tel: +81-3-5285-1111, Fax: +81-3-5285-5037, E-mail: zmatsuda@nih.go.jp

of the extra glycine residue by substituting it with several different amino acid residues, such as alanine, leucine, and phenylalanine residue. These substitutions negatively affected the function of Env, such as its fusogenicity or virion incorporation. We also found that there was a correlation between the steric characteristics of the side chain of the residue replacing $G^{691}$ and the alteration in the function of Env. Based on these findings, we propose a potential association model of gp41 MSD.

## MATERIALS AND METHODS

**Construction of plasmids**: The substitution mutants were generated by using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, Calif., USA) using the subclone containing the 1.2-kb *Nhe*I-*Bam*HI fragment covering the *env* portion of HXB2RU3ΔN as described previously (17). The complementary oligonucleotide pairs used were as follow: 691F, ATGATAGTAG GATTCTTGGT AGGTTTA/ TAAA CCTACC AAGAATCCTA CTATCAT; 690/691-2L, GTACT GCTCT TGGTAGGTTT AAGAATAGTT TTTG/ CAAA AACTAT TCTTAAACCT ACCAAGAGCA GTAC; 690/694-2L, ATTCATAATG ATAGTACTGG GCTTGGTACT TTTAAG/ CTTAAAAGT ACCAAGCCCA GTACTATCAT TATGAAT; 691/694-2L, TTCATAATGA TAGTAGGACT CTTGGTACTT TTAAG/ CTTAAAAGTA CCAAGAGTCC TACTATCATT ATGAA. PCR was performed using PfuTurbo (Stratagene). The three glycine (G691) substitution mutants were created by site-directed mutagenesis using one subclone of the 2L mutants, 690/694-2L, as a PCR template and the following complementary oligonucleotide pairs: 690/694-2L + 691A, ATTCATAATG ATAGTACTGG CCTTGGTACT TTTAAG/ CTTAAAAGTA CCAAGGCCAG TACTATC ATT ATGAAT; 690/694-2L + 691L, CATAATGATA GTA CTGCTCT TGGTACTTTT AAGAAT/ ATTCTTAAAA GTACCAAGAG CAGTACTATC ATTATG; 690/694-2L + 691F, ATTCATAATG ATAGTACTGT TCTTGGTACT TTTAAG/ CTTAAAAGTA CCAAGAACAG TACTATC ATT ATGAAT). Following mutagenesis, the 1.2-kb *Nhe*I-*Bam*HI fragments were sequenced and cloned back into the pSP65HXB2RU3ΔN or pElucEnv (17) plasmid. The entire *Nhe*I-*Bam*HI portion, together with the junction, was verified by sequencing after the fragments were back.

**Cells and antibodies**: COS7 cells, 293 cells and 293CD4 cells (17) were grown in Dulbecco's modified essential medium (DMEM; Sigma, St. Louis, Mo., USA) supplemented with 10% fetal bovine serum (FBS) (HyClone Laboratories, Logan, Utah, USA) and penicillin-streptomycin (Gibco-BRL, Rockville, Md., USA). Jurkat cells were grown in RPMI 1640 (Sigma) supplemented with 10% FBS and penicillin-streptomycin. Cells were kept under 5% $CO_2$ in a humidified incubator. Anti-gp120 polyclonal antibody was obtained from Fitzgerald Industries International, Inc. (Concord, Mass., USA). The hybridoma 902 and Chessie 8 were obtained from Bruce Chesebro and George Lewis, respectively, through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health (18-20). Serum from a patient infected with HIV-1 was kindly provided by T. H. Lee of Harvard School of Public Health, Boston, Mass., USA.

**Protein analysis**: The provirus DNA constructs were transfected into COS7 cells by electroporation or lipofection. In electroporation, COS7 cells (4 $\mu$g proviral DNA per $1 \times 10^7$ cells) were suspended in serum-free DMEM and electroporated at a 250-kV, 950-$\mu$F setting using Gene Pulsar II (Bio-Rad, Hercules, Calif., USA). In lipofection, COS7 cells ($3 \times 10^6$ cells) were transfected with 7 $\mu$g proviral DNA by FuGene6 (Roche Molecular Biochemicals, Mannheim, Germany). At 72 h after transfection, transfected COS7 cells were collected by scraping and were centrifuged at 2,000 $\times g$ for 10 min (Allegra 6KR system; Beckman Coulter, Fullerton, Calif., USA). The cell pellets were dissolved in radioimmunoprecipitation assay (RIPA) lysis buffer (0.05 M Tris-Cl [pH 7.2] including 0.15 M NaCl, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% sodium dodecyl sulfate [SDS]) and the clear lysates were centrifuged at 314,00 $\times g$ for 45 min at 4°C (Himac CS 120fx system; Hitachi, Tokyo, Japan). The virus was sedimented from pre-cleared supernatants (centrifuged at 2,000 $\times g$ for 20 min; Allegra 6KR system, then filtered through 0.45-$\mu$m-pore-size filters; Millipore, Bedford, Mass., USA) by centrifuging at 113,000 $\times g$ for 1.5 h at 4°C on 3 ml of a 20% sucrose cushion (SW28 rotor; Beckman Coulter). Virus pellets were dissolved in RIPA lysis buffer. Both cell and virus lysates were run on a 7.5-15% gradient in a SDS-polyacrylamide electrophoresis gel (PAGE) system (DRC, Tokyo, Japan), and proteins were blotted onto Immobilon-P (Millipore) by passive transfer, as described previously (21). The immunoblotting procedure was as described previously (17). Enhanced chemiluminescence (Roche Molecular Biochemicals) and a LAS-3000 (Fuji Photo Film, Kanagawa, Japan) were used to detect the bands.

**Infection study**: For the infection study, the virus seed was prepared by transfecting 1 $\mu$g of the proviral DNA into $10^6$ of the COS7 cells using FuGene6 (Roche Molecular Biochemicals). Jurkat cells were infected with each virus adjusted by the p24 amount (10 ng per $10^6$ cells). The infection was monitored by measuring the amount of p24 in the culture supernatant at specific time points after infection (0, 8, 15, 22 and 25 days). A p24 ELISA was performed using a p24 RETRO-TEK ELISA kit (ZeptoMetrix, Buffalo, N.Y., USA).

**Flow cytometry**: The level of Env expressed on the cell surface was monitored by fluorescence-activated cell sorting analysis as described previously (17). Briefly, 48 h after the COS7 cells were transfected with each Env expression vector by FuGene6, the cells were stained with the 902 monoclonal antibody for 1 h at 4°C (10 $\mu$g/ml in phosphate-buffered saline [PBS] with 2% FBS), incubated with biotin-XX goat anti-mouse immunoglobulin G (Molecular Probes, Eugene, Oreg., USA) for 30 min at 4°C, treated with streptavidin Alexa Fluor 555 (Molecular Probes) for 30 min at 4°C, and finally fixed with 1% paraformaldehyde in PBS. Cells were suspended in PBS with 2% FBS and analyzed with Becton Dickinson FACSCalibur and CellQuest software (BD Biosciences Immunocytometry Systems, San Jose, Calif., USA). A double gate was defined by forward versus side scatter and by the amount of GFP (FL-1). A total of 10,000 events within this gate were collected for analysis. An Env KO that fails to express Env was used as a negative control, as described previously (17).

**T7 RNA polymerase (RNApol) transfer assay**: The efficiency of fusion pore formation was examined using T7 RNApol transfer assay as described previously (17). Briefly, COS7 cells were transfected with each Env expression vector together with the reporter plasmid pTM3hRL using FuGene6. The reporter plasmid contains the T7 promoter-driven renilla luciferase. At 48 h after transfection, the transfected COS7 cells were cocultured with the 293CD4 cells

that had been transfected with the T7 RNApol expression vector, pCMMP T7RNApoliresGFP (the ratio of cells was 1:1). At 12 h after the coculture, the cells were lysed. Firefly luciferase activities, derived from the Env expression vector, and renilla luciferase activities, activated by the T7 RNApol transferred from 293CD4 cells through the generated fusion pores, were determined using the Dual-Glo luciferase reporter assay system (Promega, Madison, Wis., USA).

## RESULTS

**Mutagenesis of glycine residues within the GGXXG sequence**: In our previous study, mutations to an individual glycine residue within gp41 MSD were well tolerated, and high tolerance of gp41 MSD against mutations was expected (17). Therefore this time we simultaneously mutated several glycine residues within the MSD (Fig. 1). First, we substituted two glycine residues with leucine residues to create three forms of 2L mutants: 690/691-2L, 690/694-2L, and 691/694-2L. In the mutant 690/694-2L, the conserved glycine residues corresponding to the GXXXG motif were substituted. A similar mutation introduced into that of VSV-G resulted in functionally defective VSV-G (16). Next, to address the significance of the additional glycine residue at position 691, we substituted G691 with alanine, phenylalanine, or leucine residue while G690 and G694 were substituted with leucine residues. These substitutions formed the 2L + 691X mutants 690/694-2L + 691A, 690/694-2L + 691F, and 690/694-2L + 691L. Thus the sequence context of 2L + 691X mutants is LXxxL (the small x represents the original sequence of HXB2). As a control, the single substitution of the G691 with phenylalanine (691F) in which the other two glycine residues were left intact was generated (Fig. 1). The single substitution of the G691 to alanine (691A) or leucine residue (691L) was well tolerated and described in our previous study as mentioned above (17).

**The replication profile of the MSD mutants**: To examine replication capacity, Jurkat cells were infected with the mutant viruses after the p24 amount was adjusted for. Virus replication was monitored by measuring the amount of p24 released into the culture supernatants at 0, 8, 15, 22, and 25 days after infection. A representative result of the Jurkat cell experiment is shown in Figure 2. The single substitution mutant, 691F, was replicated with a slight delay compared with the wild type (WT). Other single substitution mutants, 691A or 691L, were replicated as efficiently as the WT (described in a previous study, data not shown) (17). The 2L mutants showed slightly delayed replication kinetics compared to those of the WT. The replication of the 2L + 691X mutants was less efficient than that of the WT. Substitution with leucine residue (690/694-2L + 691L) produced the slowest replication profile in repeated experiments. These data confirmed the high tolerance of the glycine residues within the GXXXG motif of gp41 MSD against mutation. Although the replication kinetics were generally slower in the H9 cells, a similar replication profile was observed (data not shown).

**Analysis of the protein profiles of MSD mutants**: The protein profiles of the mutants depicted in Figure 1 were examined by immunoblotting analysis using the serum from an individual infected with HIV-1 and anti-gp120 polyclonal antibody for both cell and virus lysates. Similar protein profiles were observed for all mutants and the WT in cell lysates (Fig. 3A). Almost equivalent amounts of gp160 and gp120 were observed for all constructs, and there was no obvious alteration in the processing. There was no difference found in the processing of Gag and Pol products. In the virus lysates, as in the cell lysates, no differences were found in the profiles of Gag and Pol products (Fig. 3B). However the amounts of gp120 found on 690/694-2L + 691L virions were 50 - 60% those of the WT (51.9% ± 14.4 %, $n = 3$, the amount of the incorporated Env was estimated by determining the ratio of the intensity of the Env to the p24 bands). This result is not due to the shedding of gp120, because the probing of the virus lysates with anti-gp41 monoclonal antibody detected a smaller amount of gp41 in 690/694-2L + 691L than in the WT (Fig. 3C). Therefore the alteration of the replication profiles observed for 690/694-2L + 691L (Fig. 2) may be accounted for by a defect in the incorporation of Env onto the virions. The reason other mutants manifested the slower replication is not evident from the protein profiles.

**Fusion activity of mutants evaluated by the efficiency of fusion pore formation**: Our previous study as well as others (11,17,22), have shown that mutations of the MSD sometimes negatively affect the fusogenicity of the Env. Therefore, to investigate the reason for the delayed replication observed in Figure 2, we evaluated the fusion efficiency of our mutants using the Env expression vector (Fig. 4A).

```
WT: yikLFIMIVGGLVGLRIVFAVLSIVnrv
691F: yikLFIMIVGFLVGLRIVFAVLSIVnrv
2L mutants
  690/691-2L: yikLFIMIVLLLVGLRIVFAVLSIVnrv
  690/694-2L: yikLFIMIVLGLVLLRIVFAVLSIVnrv
  691/694-2L: yikLFIMIVGLLVLLRIVFAVLSIVnrv
690/694-2L+691 mutants
  690/694-2L+691A: yikLFIMIVLALVLLRIVFAVLSIVnrv
  690/694-2L+691L: yikLFIMIVLLLVLLRIVFAVLSIVnrv
  690/694-2L+691F: yikLFIMIVLFLVLLRIVFAVLSIVnrv
```

Fig. 1. The mutants of gp41 MSD studied. The primary structures of the MSD mutants used in this study are shown using the one-letter abbreviation of amino acid residues. The position numbering is based on that used for HXB2 Env. The portion of the predicted MSD is shown in capital letters. WT corresponds to the wild type HXB2. Mutated residues are underlined.

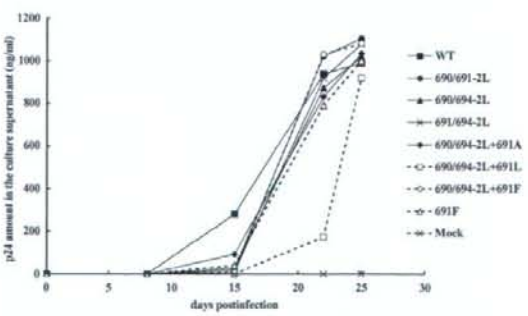


Fig. 2. The replication profile of gp41 MSD mutants in Jurkat cells. The replication of HIV-1 was monitored by measuring the p24 amount in the culture medium at 0, 8, 15, 22 and 25 days. The replication of 691A and 691L was similar to that of WT, as reported previously (17) and not shown here.

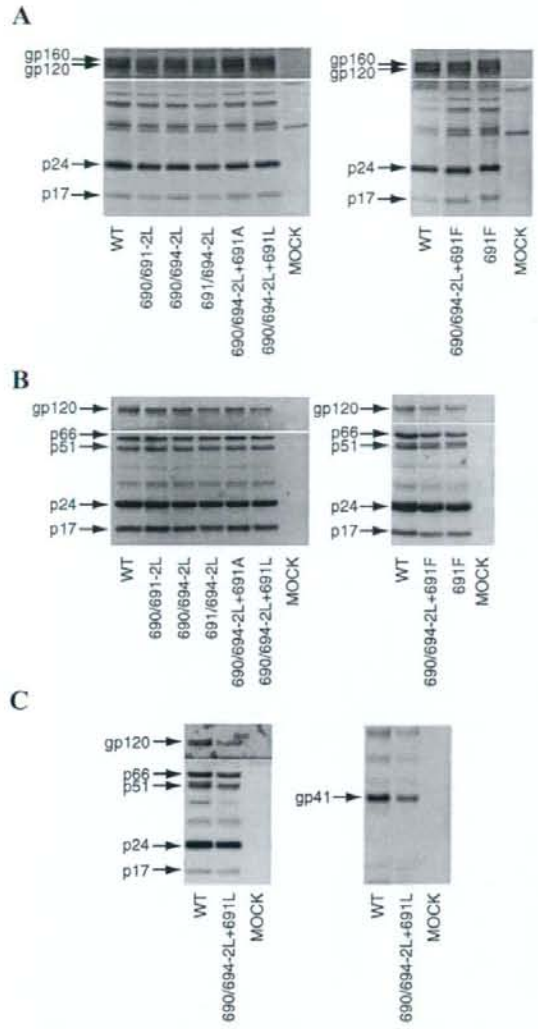


Fig. 3. Protein profiles of cell and virus lysates of the WT and mutants. The protein profiles of mutants were examined by immunoblotting. The cell (A) and virus (B, C) lysates prepared from transfected COS7 cells were used. The Env proteins (gp160, gp120 and gp41) were detected using anti-gp120 polyclonal antibody or anti-gp41 monoclonal antibody, respectively. Gag (p17 and p24) and Pol (p51 and p66) were detected using serum from an individual infected with HIV-1. The names of the mutants are shown.

First, we confirmed that the surface expression level of Env was similar by using the 902 monoclonal antibody in a flow cytometric analysis (Fig. 4B). This confirmation further supported the hypothesis that 690/694-2L + 691L has a defect in the incorporation of Env onto the virions rather than a defect in the intracellular transport of Env. We then analyzed the fusion activity of Env mutants using the T7 RNApol transfer assay. In this assay, the T7 RNApol that is transferred through the fusion pore between the Env- and receptor-expressing cells activates the T7 promoter-driven renilla luciferase. The renilla luciferase activity was normalized for transfection efficiency by the firefly luciferase activity derived from the Env expression vector. The representative data are shown in Figure 4C. Compared with WT, the 2L mutants showed a decrease in fusion efficiency of about 30%. There were no significant differences in fusogenicity among the different 2L mutants, a finding that was consistent with the replication profile shown in Figure 2. Therefore, the delayed replication observed for 2L mutants may be due to a defect in fusion. The delayed replication of 691F may also be due to a similar mechanism, although the observed decrease in the fusogenicity of 691F is not statistically significant (Fig. 4C).

The effect of the mutations of the conserved glycine residues within the GXXXG motif of gp41 MSD was less severe than in the case of VSV-G, in which a simultaneous substitution of two glycine residues corresponding to G690 and G694 resulted in almost complete elimination of fusion activity (16). The presence of an additional glycine residue within the GXXXG motif generates the GG$^{691}$XXG sequence in gp41 MSD, which led us to evaluate the effect of substituting other amino acid residues for the glycine residue at position 691 in the 690/694-2L context for membrane fusion. Substituting glycine residue with alanine residue did not reduce the fusion efficiency further (for example, compare 690/694-2L with 690/694-2L + 691A in Fig. 4C). However, changing the glycine residue to phenylalanine reduced the fusion efficiency significantly (690/694-2L versus 690/694-2L + 691F in Fig. 4C). The introduction of leucine residue in place of the glycine residue had the most severely negative effect. Thus the presence of a glycine residue at position 691 seems to produce an apparently higher fusion efficiency for gp41 even when the two other glycine residues constituting the GXXXG motif were replaced with leucine residue. Furthermore, because substitution with leucine resulted in less fusogenic gp41 than did substitution with phenylalanine residue, it seemed that the steric nature of the side chain at position 691, not simple bulkiness, affected the outcome.

## DISCUSSION

The importance of the GXXXG motif in helix-helix association has been well established through the studies of GpA. The glycine residues within the GXXXG motif play a critical role in helix-helix interaction (13,23). Here we showed that the mutations introduced in the conserved glycine residues within the GXXXG motif of gp41 MSD were well tolerated. This finding is quite different from that of the previous study of VSV-G MSD, where a similar mutation corresponding to the 690/694-2L mutant almost abolished the fusion activity (16). Furthermore, we also observed that the substitution of any two glycine residues within the GGXXG sequence of gp41 MSD only modestly decreased the fusion activity (Fig. 4C). The expression, processing, and transport of Env proteins to the cell surface were preserved (Fig. 3A and Fig. 4B). Consistent with the modest decrease in fusion activity, the replication profiles of these mutant viruses were only slightly retarded when compared to that of the WT virus (Fig. 2). This result confirmed the high tolerance of gp41 MSD that allows the substitution of any two glycine residues within the GGXXG sequence. The results for the 690/694-2L mutant may suggest that the GXXXG motif of gp41 MSD does not play a role as a helix-helix interaction motif or that there is another mechanism that cancels out the introduced mutations.

The presence of the additional glycine residue within the GXXXG motif is a well-conserved feature of gp41's MSD (15), and this feature is absent from VSV-G. The three

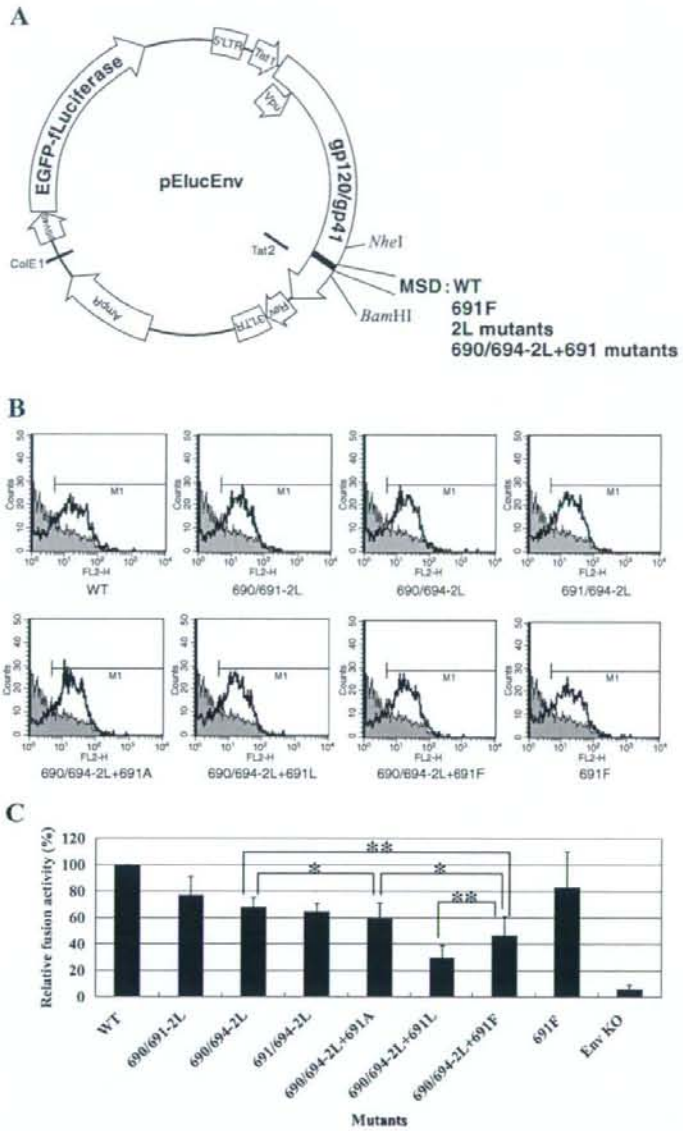


Fig. 4. Evaluation of the cell surface expression and measurement of the fusion activity of mutant Env. (A) The map of the Env expression vector, pElucEnv. pElucEnv supports the expression of HIV-1 *env* (gp120/gp41) and the gene of the EGFP-firefly luciferase (EGFP-fLuciferase) hybrid protein, respectively, from two separate promoters. The *Nhe*I and *Bam*HI sites used for cloning are indicated. The LTR, *tat*, *vpu*, and *rev* of HIV-1 are shown. AmpR, $\beta$-lactamase gene; SV40, SV40 late promoter; ColE1, ColE1 replication origin. (B) Cell-surface expression of Env. FACS analysis of Env expressed on the surface of COS7 cells transfected with each pElucEnv construct was accomplished as described in the Materials and Methods section. The signal for each Env is shown as a gray line. The filled area depicts the signal obtained for the control vector, Env KO. (C) Fusion activity of the mutants evaluated by T7 RNApol transfer assay. The cell-cell fusion assay between the Env expressing cells (containing T7 promoter-driven plasmid) and CD4+ cells (bearing the T7 RNApol expression plasmid) was used to evaluate the fusion efficiency of mutant Envs. A representative result of four independent experiments is shown (*statistically not significant, **statistically significant difference: $P < 0.05$ by Student's $t$ test). The results shown are means ± s.d. ($n = 4$).

glycine residues will cluster within the gp41 MSD helix (Fig. 5A, prepared by the program Insight II; Accelrys, San Diego, Calif., USA). Having a hydrogen atom as its side chain, the clustering of three glycine residues may give the region more flexibility to accommodate mutations. To test whether the high tolerance of gp41MSD against mutations is attributable to the presence of the extra glycine residue within the gp41 GXXXG motif, we further mutated the extra glycine residue (G691) while the other two glycine residues were mutated to leucine residues under the LG$^{691}$XXL context. This also allows us to obtain information on a potential helix-helix interface among gp41 MSDs. We would expect the mutation

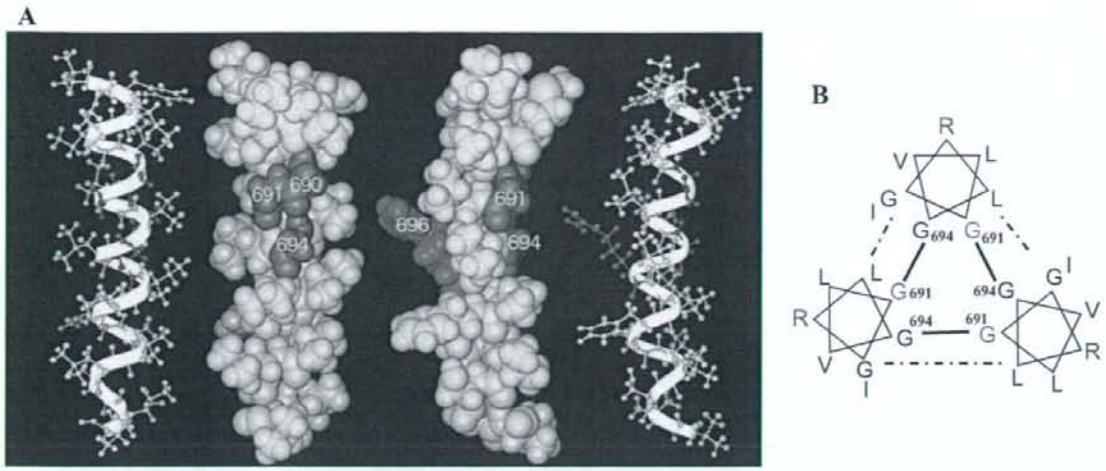


Fig. 5. A structural model of MSD and a hypothetical arrangement of the MSD helices. (A) The MSD region (position 684-705 of HXB2 Env, the amino acid sequence: LFIMIVGGLVGLRIVFAVLSIV) of the WT HXB2 was modeled in an a-helix conformation. Molecular dynamics simulations were carried out in the lipid-like environment, using the Generalized Born method with a relative permittivity of 4.0 (25, 26). The most representative structure seen in the simulation was extracted by a principal component analysis and is shown graphically. Three glycine residues (shown in red) colocalize each other and generate a hollow surface on one aspect of the helix. Two different views (ball-and-stick and CPK) of the same helix are shown. (B) A hypothetical arrangement of the MSD helices. The glycine residue at positions 691 and arginine residue at position 696 are shown in red and blue, respectively. The potential interactions between amino acid residues are shown by the bold and dotted lines, respectively.

to have no significant effect if G691 was facing outward to the lipid environment and was not involved in the helix-helix interaction. We took a genetic approach and we created the mutants 690/694-2L + 691A, 690/694-2L + 691F, and 690/694-2L + 691L. None of these mutations changed the processing or surface expression of Env (Figs. 3A and 4B). This suggests that these mutations did not induce drastic conformational changes that could be detected by a quality control mechanism of endoplasmic reticulum. Indeed, our molecular dynamics analyses of these mutants failed to detect severe deformation of the structure of the individual helix (data not shown).

When the fusion activity of the 690/694-2L + 691X mutants was evaluated, there was no statistically significant difference between glycine and alanine residue (Fig. 4C). This finding is consistent with the observation that alanine residue can functionally substitute glycine residue in transmembrane helices (23). The substitution of glycine residue with a bulky residue, namely phenylalanine residue (690/694-2L + 691F), resulted in a significant reduction in fusion activity. The fusion efficiency of 690/694-2L + 691F was about 50% that of WT. Substitution with leucine residue (690/694-2L + 691L) had an even greater negative effect on fusogenicity, and the fusion efficiency of 690/694-2L + 691L was about 30% that of WT (Fig. 4C). The replication efficiency somewhat reflected these changes in fusion activity, although we could not attribute the defect of 690/694-2L + 691L to the defect in fusion alone because 690/694-2L + 691L had an additional defect in Env incorporation (Figs. 3B and C). These results showed that the presence of glycine residue at position 691 seemed to reduce the negative effects of the leucine residue substitution of glycine residue at positions 690 and 694.

It is noteworthy that we have observed the phenotypic changes of gp41 according to the nature of the substituted amino acid residue at the position 691 (Figs. 2-4). Interestingly, the negative effect of mutations was not dependent on the mere volume of the side chain (F > L), but rather on the steric nature of the side chain. The introduction of another branched amino acid residue, isoleucine residue, at position 691 also resulted in decreased fusion activity (data not shown). This may suggest that the mutation may affect the interhelix association of MSDs, and the changes in association among MSDs may affect the function of gp41. Indeed our preliminary analysis of the measurement of helix-helix interaction by means of TOXCAT (12) analysis indicated that 690/694-2L + 691L had a slightly stronger association than WT (E. Matthews and D. M. Engelman, unpublished data). Based on these data, we present our hypothetical model of the association of the three MSDs in which the GXXXG motif is facing inward as shown in Figure 5B. Under this configuration, the G691 locates itself near the helical interface. Although this model is consistent with our observation, it places the highly conserved arginine residue downstream of the GXXXG motif toward the lipid environment. This may not be a thermodynamically favorable arrangement despite the recent finding that suggested that arginine residue can be accommodated into a lipid bilayer more easily than previously expected (24). Therefore we cannot rule out alternative arrangements of the gp41 MSDs at present. The observed high tolerance of glycines constituting this hypothetical interface may also suggest that the potential interaction among gp41 MSDs may be a rather weak one. To prove our model, the physical structure of the gp41 MSDs in lipid environments must be determined.

Here we have shown that the efficiency of the Env-mediated fusion pore formation and the incorporation of Env onto virions were affected by alterations within the MSD of gp41. This confirms that the specific primary structure of MSD is important for its proper function. Our findings also suggest that a subtle change in the structure within MSD can affect the conformations of other subdomains of gp41. Conversely, they may suggest that the conformational changes