except for 1LE3 have large RMSD values and form a completely different structure from the initial helix conformation. Hence, the energies were compared only for the 1LE3 protein. The Native structure is more stable than the All-helix structure in 1LE3. For the other proteins that could not retain the initial conformation, a sudden structure change caused by the electrostatic repulsion of charged amino acid residues is observed in the heating process. In the IGB = 1 (GBSA = 1) solvent model, the Native structure of 1J4M and the All-helix structure of 1B03, 1LE1, and 1NIZ form a completely different structure from the initial secondary conformation. Therefore, energy comparison between the helix and the β-sheet structures was performed only for 1LE0 and 1LE3 proteins. The Native structures of 1LE0 and 1LE3 are more stable than the All-helix structure. In the IGB = 1 (GBSA = 0) solvent model, energy comparison was performed only for 1LE1 protein, in which the Native structure was preferred.

Discussion

Modified Force Field

Reliability Evaluated from Amino Acid Conformations and Secondary Structure of Proteins

As for the difference in stability between αR and C5 conformations, the results of MD simulation are compatible with those of QC calculation for all amino acids except CYS, MET, THR, and VAL. The problem of excessive tendency to form a helix structure of ff03 force field has been modified. As long as THR residue, the β conformation is the most stable in both MD simulation and QC calculation. Therefore, the force field parameter of THR will also be acceptable.

The Native structure is more stable than the All-helix structure for all proteins except 1NIZ. Accordingly, the modified force field will be acceptable in terms of the secondary structure. 1NIZ is a part of the V3 loop region of gp120 of HIV-1. The full sequence of this region is "CTRPNYNKRKRIHIGPGRAFYTTKNIIGTIRQAHC" and it corresponds to the 301st to 335th residues of gp120 protein. A disulfide bond is formed between CYS301 and CYS335. Whereas 1NIZ is registered as a single molecule in PDB, this structure was determined by NMR spectroscopy in a complex with an HIV-1 neutralizing antibody.[33] Therefore, the partial model of ACE-KRIHIGPGRAFYTT-NME will not reflect the NMR experimental condition of 1NIZ.

The protein 1CE4, which has an amino acid sequence very similar to that of 1NIZ, was synthesized by a peptide synthesizer and was determined by NMR spectroscopy as a single molecule.[40] 1CE4 forms a disulfide bond like the V3 loop region and consists of bend, turn, and helix structures instead of the β-sheet structure seen in 1NIZ. Thus, there is no β-sheet structure in 1CE4. Furthermore, the C-terminal half side of 1CE4 has α helix structure (Fig. S2 in Supplementary Materials). Some amino acids on the N-terminal side also show a helix structure. The disulfide bond of 1CE4 strongly assists the folding of the protein and makes a twist at the center region of the protein. From this point of view, if there is no disulfide bond, most parts of 1CE4 will be stabilized in the helix structure. Accordingly, it is plausible that the All-helix structure is more stable than the Native structure in 1NIZ. In contrast, the proteins of 1B03, 1J4M, 1LE0, 1LE1, and 1LE3 were synthesized by a peptide synthesizer. Their structures were determined by NMR spectroscopy as a single molecule, and contains many β-sheet and β-turn conformations. As a result, the modified force field is confirmed to accurately describe the β-sheet structure, and the helix structure will be also reproducible as shown in the results for 1NIZ and 1CE4.

Reliability Evaluated in Terms of Tertiary Structure of Proteins and Application to Protein Structure Prediction

Protein structure prediction was performed for 1J4M and 1LE3 molecules containing the β-sheet structure and also for 1L2Y and 1VII molecules containing the α-helix structure. The accuracy of the structural prediction was satisfactory for all molecules. In the case of 1LE3, the side chains of four TRP residues gather in a row due to their π-π interactions in the PDB structure. However, the π-π interactions are not observed in the prediction structure (Fig. 4a). Since the kinetic energy is large because of the MD simulation at 375 K, the weak π-π interaction disappears and will not be observed in the predicted structure. The PDB structure of 1VII protein also shows two π-π interactions of PHE7-PHE18 and PHE11-PHE18 between the first helix and the second helix regions. These π-π interactions are not observed in the predicted structure of 1VII (Fig. 4b). For more precise protein prediction, the proteins containing weak π-π interaction should be simulated again at a lower temperature starting from the structures obtained in the present prediction, or using the simulation technique explicitly implementing the effect of π-π interaction.[41] For 1J4M, the whole structure was accurately predicted, but the β-sheet structure is still loose. 1J4M also needs to be simulated again at a lower temperature or by the π-π interaction implemented method.

Figure 5. Residue-averaged B-factors derived from the calculation with the modified force field (blue), with ff03 force field (green), and the experimental measurement by X-ray crystallography (red). The B-factors for β-lactamase computed in the explicit and implicit water conditions are shown in (a) and (a'), respectively. Those for HIV-1 protease in the explicit and implicit water conditions are in (b) and (b').

Figure 6. RMSD plot for main chain atoms during MD simulations with the modified force field (blue thick line) and the ff03 force field (green broken line). Simulations were executed at 310 K under the explicitly water-generated condition both for mini-proteins (a–d) and enzymatic proteins (e–f).

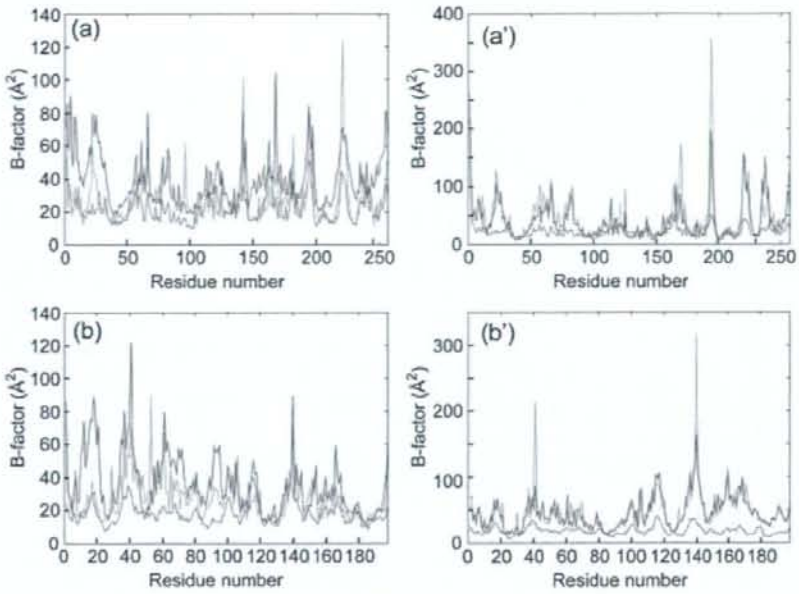

Figure 5.

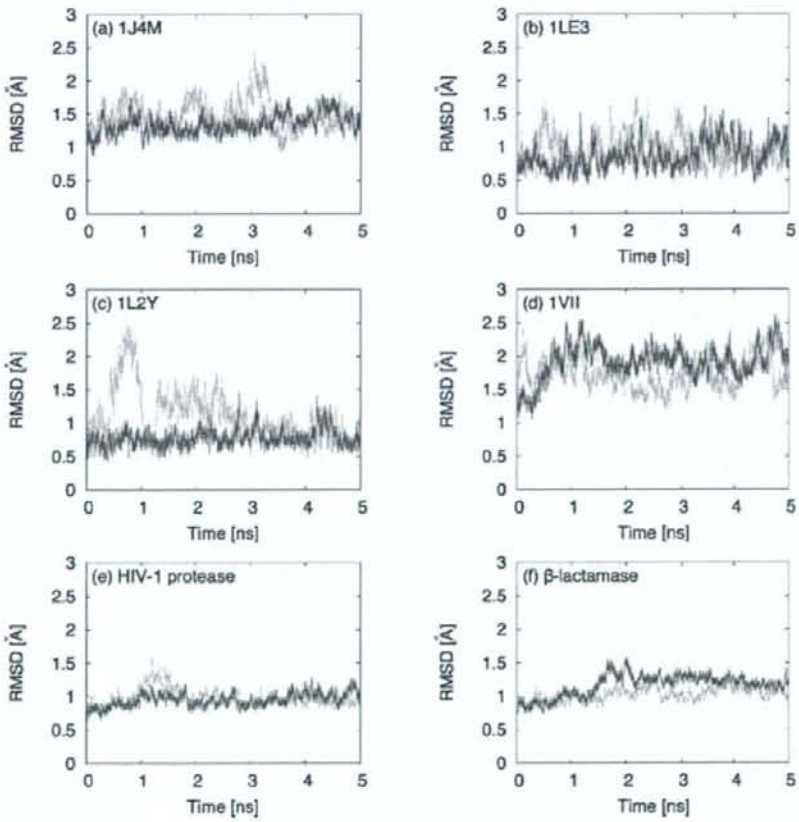

Figure 6.

The predicted structure of 1L2Y was accurate enough. Many previous studies have shown the success of structure prediction for 1L2Y protein with the conventional force field.[42–45] In the PDB structure of 1L2Y, the first helix region consists of eight residues of LEU2, TYR3, ILE4, GLU5, TRP6, LEU7, LYS8, and ASP9, the turn region consists of two residues of GLY10 and GLY11, the second helix region consists of three residues of PRO12, SER13, and SER14, and the strand region consists of five residues of GLY15, ARG16, PRO17, PRO18, and PRO19. Eight residues of the first helix region form an αR conformation and are stabilized as the helix structure. The main chain of GLY residue comprising the turn region is flexible and its conformational degree of freedom is the largest among the normal 20 amino acids. Three residues of the second helix region form the αR conformation and are stabilized as the 3-10 helix structure. In the strand region, GLY and ARG residues form the C5 conformation and three PRO residues form the β conformation. Considering the conformational stability of all amino acids with ff03 force field [Table S3(a–f)], five amino acids, ASP, ILE, LEU, TRP, and TYR, included in the first helix region are stabilized in the αR conformation in almost all solvent models. That is, the residues that stabilize the helix structure are fortunately assigned at the helix region of this protein. PRO and SER in the second helix region are stabilized in the β and αR conformations in ff03 force field, respectively. Hence, this second helix region has an excessive tendency to form the helix structure by SER. GLY, ARG, and PRO residues at the strand region are stabilized in the β conformation. Three PRO residues prefer the β conformation in ff03 force field. GLY and ARG are stabilized in the C5 conformation in almost all cases in Table S3. The residues whose parameters stabilize the strand structure are also fortunately located in the strand region of the protein. PRO is always stabilized in the β conformation because of its structural peculiarity of main chain. Accordingly, the structure of 1L2Y can be predicted easily although the force field that has the excessive helical tendency is used, which results in many reports on the success of 1L2Y prediction.[42–45] The modified force field gives an accurately predicted structure for 1L2Y without any excessive helical tendency in parameter. This suggests that the modified force field can stabilize the helix structure from the hydrogen bond formations.

Flexibility of Proteins

To evaluate the flexibility of proteins those consist of 200–300 amino acid residues, B-factors of β-lactamase from *Staphylococcus aureus* PC-1 (PDB code : 3BLM),[46,47] and human immunodeficiency virus type 1 (HIV-1) protease (PDB code : 1OHR)[48,49] have been compared among three methods: the calculation with the modified force field, the calculation with ff03 force field, and the measurement in X-ray crystallography (see Fig. 5). The B-factors from the calculations with the modified and ff03 force fields were obtained by using the Ptraj program after the 10 ns MD simulations both with the implicit and the explicit water conditions. The B-factors were estimated in the average for the last 1 ns MD simulation. The keywords of IGB = 5 and GBSA = 1 were employed in the simulation with the implicit water condition. The simulation with the explicit water condition was performed according to the Ross Walkers method.[50] The B-factors derived from the modified force field were similar to that from ff03 force field. As for the MD simulation with the implicit water condition (Figs. 5a' and b'), a large B-factor value in ff03 force field, that was over 300 Å^2 at the 194th residue of β-lactamase, was improved in the modified force field. The B-factor values of HIV-1 protease were also improved at the 41st and 140th residues by using the modified force field. The simulated B-factors tend to be large compared with the B-factors observed in X-ray crystallography as seen in the previous studies.[47,49,51] The results on the B-factor suggest that the modified force field can simulate and analyze protein function as effective as ff03 force field from the viewpoint of flexibility of proteins.

Stability of Protein Structure

In this study, atom charges were determined from QC calculations in water phase. Atom charges in the standard force field had been derived from the vacuum- or ether-phase condition.[3] Hence, our modified parameters will be more suitable for describing the structure of such regions that are exposed to solvent. This advantage, however, may cause a drawback for overemphasizing hydrophilic effect and relatively underestimating hydrophobic one. The energy balance between hydrophilic and hydrophobic interactions is critically important to keep an appropriate protein structure in computer simulation. To assess the stability of protein structure in MD simulation with our charge-modified force field, MD simulations were performed and RMSD during the simulation was evaluated for all mini-proteins exemplified in the section "Evaluation of a Standard Force Field" and two enzymatic proteins instanced in the previous section.

MD simulations of 5 ns at 310 K were executed for all mini-proteins and enzymatic ones. All calculations were performed under the calculation condition described in the section "Execution of MD Simulation with the Explicitly Water-Generated Models." The change of RMSD values during the simulation are shown in Figure 6. The fluctuation in RMSD becomes small after 3.5 ns for 1J4M, 1LE3, and 1L2Y. It is confirmed from principal component analysis of Figure S3 in Supplementary Materials that the structure is satisfactorily equilibrated after 4.0 ns. For 1LE3 and 1L2Y, the averaged RMSD values are no longer than 1 Å measured from the PDB structure. For 1J4M, the averaged RMSD value is slightly large, 1.5 Å. In contrast, 1VII shows a certain degree of difficulty in equilibration as seen in both of RMSD plot and PCA map. This is due to the flexibility of N- and C- terminus of 1VII. In every case of mini-proteins, the fluctuation in structure during MD simulation is almost in the similar level between the charge-modified force field and the standard ff03 force field. As long as mini-proteins, no noticeable problem is found in the presently proposed force field from the viewpoints of structural stability.

To examine the structural stability using nonartificial proteins larger than mini-proteins, we executed MD simulation for two enzymatic proteins: HIV-1 protease and β-lactamase. HIV-1

Table 4. Comparison of Number Surface Atoms with That of Buried Ones for Several Proteins.

PDB code	Number of residues	Surface atoms/buried atoms	
2HBO	157	632/405	Thioesterase superfamily protein
1AJ6	219	834/625	DNA gyrase B
7HVP	99 * 2	878/669	HIV-1 protease
1A2F	291	1260/1091	Cytochrome C peroxidase
1OG5	208 * 2	1713/1567	Human cytochrome p450
1A7T	232 * 2	1805/1683	Metallo β-lactamase
117E	282 * 2	2270/2196	Pyrophosphatases
1A88	275 * 3	3017/3298	Chloroperoxidase L
12E8	214 * 4	3591/3071	Antibody Fab fragment
1ARZ	273 * 4	4201/3677	Dihydrodipicolinate reductase

Asterisk indicates that the protein structure is obtained in dimer, trimer, or tetramer.

protease consists of 198 amino acid residues and β-lactamase, class A β-lactamase in this calculation, contains 257 residues. In both cases, the change in RMSD value during MD simulation is fairly small compared with that for mini-proteins. A comparison of RMSD values between the charge-modified force field and the standard ff03 shows no significant difference in both cases of HIV-1 protease and β-lactamase. The RMSD values were also confirmed to be on a similar level to the previous computational studies on these proteins.[52–55] As far as the simulations for enzymatic proteins are concerned, the charge-modified force field leads to no serious error from the viewpoints of structural stability. Judging from the fluctuation in structure during MD simulations for mini-proteins and enzymatic proteins, the presently proposed force field satisfactorily provides the stably equilibrated structure and hardly induces serious error owing to the overestimation of solvent effect.

Importance of Solvent Effects of Waters

The dipeptide model is very useful for investigating the stability of the main chain torsion angles and the reliability of the force field parameters. For the ACE-ALA-NME model, detailed investigation has been performed on the conformational stability with gas-, ether- and water-phase QC calculations by Wang and Duan[56] According to their study, the dipeptide model was stabilized in the C7eq conformation in both gas- and ether phase and the αR, C5, and β conformations were less stable. In contrast, the αR, C5, and β conformations are more stable than the C7eq conformation in water phase. The C5 conformation is the most stable among these three conformations. The ϕ and ψ values of the main chain torsion angles are $(\phi, \psi) = (-70.5, -32.1)$ and $(-156.4, 143.8)$ in the optimized structures for αR and C5 conformations in water-phase calculation at the MP2/6-31G** level by Duan et al. In our QC calculations, the αR conformation is located at $(-77.9, -26.2)$ and the C5 conformation is located at $(-156.4, 149.5)$ at the HF/6-31G** level. Our results reproduced the results of Duan et al. in spite of our calculation level of HF/6-31G** being lower than theirs.

The gas- and ether-phase QC calculations by Wang and Duan [56] suggest that the protein folding structure cannot be correctly generated in a low-permittivity condition because electrostatic interaction is emphasized in a low-permittivity condition. In their results, the dipeptide model is strongly stabilized in the C7eq conformation, which forms a hydrogen bond between the O atom of ACE and the H atom bonding to N atom of NME. Therefore, it is reasonable to assume that the helix structure is easily stabilized in the condition of low permittivity because the intramolecular interaction is effective compared to the intermolecular interaction like hydrogen bonds between protein and water. Indeed, there are some proteins that have a strong tendency to form a helix structure in the low-permittivity condition.[57–59] In contrast, the C7eq conformation is less stable than the αR, C5, and β conformations in water phase. In particular, the C5 and β conformations are important in the condition of high permittivity since these conformations correspond to the β-sheet structure and make an intermolecular interaction with water. According to the results of the previous studies,[9,10] water molecules are indispensable for correct protein folding. Therefore, the protein folding structure will be accurately predictable by using the force field parameters reflecting the water-phase QC calculation, especially for the proteins such as mini-protein and the proteins containing β-sheet region. Accordingly, charge parameter modification is one of the promising approaches to precisely describe the solvent effect.

Water molecules play an important role in protein folding. In particular, the β-sheet structure is known to be significantly stabilized by the presence of water molecules.[9–10] In the present work, the reliability of protein structure prediction has increased by adopting the modified atom charges derived from the water-phase QM calculations. This implies that a fairly large number of residues in proteins are exposed to solvent water and located in the high-permittivity environment. For the purpose of examining the access of waters to the residues in proteins, we have calculated the ratio of the number of surface atoms to that of buried ones as shown in Table 4. Several globular proteins, whose number of residues is more than 150, are selected and the number of surface atoms are calculated by using a software program computing the solvent accessible surface area, GETAREA 1.1.[60] The calculated result indicated that the number of surface

atoms is almost comparable with that of buried ones or larger. That is, many atoms in proteins can be directly influenced by solvent waters. The permittivity will be fairly high due to the influence of waters although it may be true that the permittivity in the core of proteins is about 4.0 like in ether. Accordingly it is reasonable that the charge modification has improved the reliability in computation on the folding structure of soluble proteins.

Effective Solvent Model in GB Method

Six kinds of methods describing solvent effect in GB calculation were examined from the secondary conformations of proteins. In the three solvent models, IGB = 1 (GBSA = 0), IGB = 1 (GBSA = 1), and IGB = 2 (GBSA = 0), serious changes in protein structures were often observed. The initial structures of five proteins (1B03, 1J4M, 1LE0, 1LE1, 1NIZ) for IGB = 2 (GBSA = 0), three proteins (1B03, 1LE1, 1NIZ) for IGB = 1 (GBSA = 1), and four proteins (1B03, 1LE0, 1LE3, 1NIZ) for IGB = 1 (GBSA = 0) were drastically changed only for 5 ps MD simulation, and the RMSD values became large. 1B03 protein includes five positively charged amino acids: ARG1, LYS2, ARG5, ARG8, and ARG12. Strong repulsion forces among these five amino acids caused a sudden structural change. Five proteins, 1J4M, 1LE0, 1LE1, 1LE3 and 1NIZ, include several positively and negatively charged amino acids (Table S6 in Supplementary Materials). Sudden structural change caused by the electrostatic interaction has been frequently observed in these proteins. This suggests that the electrostatic interaction is excessively estimated in these three solvent models and that these solvent models are inadequate when used in combination with ff03 force field. In contrast, there was no sudden structural change in three solvent models: IGB = 2 (GBSA = 1), IGB = 5 (GBSA = 0), and IGB = 5 (GBSA = 1). In the IGB = 2 (GBSA = 1) solvent model, three proteins show a small energy difference less than 5 kcal/mol between $\Delta E_{\mathrm{GBTOT}}$ and $\Delta E_{\mathrm{PBTOT}}$. The energy difference is also small for two proteins in the IGB = 5 (GBSA = 0) model and for four proteins in the IGB = 5 (GBSA = 1) model. These results suggest that the solvent model IGB = 5 (GBSA = 1) is the most favorable among the six methods describing solvent effect from the viewpoint of energy and structure.

Protein Structure Calculated with ff03 Force Field

The charge parameters of ff03 force field were reported to be obtained by QC calculations in ether phase (ε = 4.335). The charge parameters provided before ff03 force field had been obtained by QC calculation in gas phase. In MD simulation, the solvent effects had been incorporated only by using dielectric cavity methods[12–17] or explicit water molecules.[35,61–64] Therefore, the advantage of ff03 force field is to include the solvent effect in itself. This advantage will appear especially in the improvement of the main chain torsion angles of proteins. According to the $\Delta E_{\mathrm{GBTOT}}$ calculation with the IGB = 5 (GBSA = 1) solvent model, the All-helix structure is more stable than the Native structure except for 1LE0 protein in the calculation with ff03 force field. The solvent models with IGB = 2 (GBSA = 1) and IGB = 5 (GBSA = 0) also show energetic preference for the All-helix structure in some proteins. These results clearly show that MD simulation using ff03 force field still cannot simulate the proteins accurately enough, especially when water molecules strongly influence the protein structure.

Both αR and β conformations are often observed as the most stable conformation in protein peptides because the ϕ angle of main chain is greatly stabilized at -60°. The C5 conformation, ϕ angle of which is around 180°, is rather unstable. The helix structure is normally stabilized by hydrogen bonds of main chain atoms between each residue and its fourth neighboring. On the other hand, the dipeptide model consists of only one amino acid residue, NME, and ACE. Hence, this model cannot form a hydrogen bond like the helix structure. However, the dipeptide model is stabilized in the αR conformation in MD simulations for many amino acids with every solvent method as shown in Table S3(a–f). MD simulation was also performed under the condition of explicit generation of water molecules for the ACE-ALA-NME dipeptide model. This dipeptide model resulted in stabilization of the αR conformation, and ϕ angle of main chain was also greatly stabilized at -60°. That is, no improvement was seen in the results of MD simulations in spite of the explicit generation of water molecules. These results suggest that ff03 force field seems intrinsically to have a tendency to lead the helix structure excessively in spite of the choice of solvent models or the incorporation of explicit waters.

For comparison, we have examined the conformational stability of secondary structures in MD simulation with ff99 force field using the ACE-ALA-NME dipeptide model. The ff99 force field was one of the most broadly used parameters before the release of ff03. The energies of each conformation estimated through MD simulation with ff99 force field are shown in Table S7 of Supplementary Materials. As long as ALA, the αR conformation is always the most energetically stable irrespective of the choice of GB calculation method. This indicates that ff99 force field also has a tendency to lead the α-helix structure excessively. A comparison of conformational stability between ff99 (Table S7) and ff03 (Table S3) force fields suggests that the tendency for helix is more serious in ff99. Hence, it is true that ff03 force field improved the overestimation of helix stability to a certain extent. In the conversion from ff99 to ff03 force field, a new atom type, H0, which is the hydrogen atom bonding to Cα atom of GLY, was introduced. As far as bonded terms are concerned, only the dihedral parameters for ϕ and ψ angles for main chain and the dihedral parameters concerning Cβ and three main chain atoms were modified. In our present study, the parameter values for the bonded terms are unchanged from ff03 force field, but only the atom charges are modified. This charge modification was demonstrated to diminish the tendency for helix conformation and to be effective for the prediction of protein folding structure.

Conclusions

Currently one of the most widely accepted force fields, ff03, still tends to overestimate the stabilization energy of the helix structure. Hence, a mini-protein structure or β-sheet structure, whose

stability is strongly affected by the presence of water molecules, is difficult to be predicted in a computer only from its amino sequence. We have presented a modified force field that is applicable even for these proteins. The advantage of our new force field was demonstrated in terms of conformational stability of amino acids, stability of the secondary structure, and predictability of the tertiary structure. The accuracy in predicting protein folding has been improved by this force field. A distinction between the modified force field and ff03 force field is the solvent effect. The effective charges have been calculated by QC calculations in water phase to deduce a good valance between the α-helix and β-sheet structures.

References

1. Tama, F.; Miyashita, O.; Kitao, A.; Go, N. Eur Biophys J 2000, 29, 472.
2. Samatey, F. A.; Matsunami, H.; Imada, K.; Nagashima, S.; Shaikh, T. R.; Thomas, D. R.; Chen, J. Z.; Derosier, D. J.; Kitao, A.; Namba, K. Nature 2004, 431, 1062.
3. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. J Comput Chem 2003, 24, 1999.
4. Lee, M. C.; Duan, Y. Proteins 2004, 55, 620.
5. Pedersen, A. K.; Peters, G. G.; Moller, K. B.; Iversen, L. F.; Kastrup, J. S. Acta Crystallogr D Biol Crystallogr 2004, 60, 1527.
6. Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B. Biochemistry 1998, 37, 17735.
7. Zhao, B.; Guengerich, F. P.; Voehler, M.; Waterman, M. R. J Biol Chem 2005, 280, 42188.
8. Cherbavaz, D. B.; Lee, M. E.; Stroud, R. M.; Koshland, D. E., Jr. J Mol Biol 2000, 295, 377.
9. Katagiri, D.; Tsuchiya, T.; Tsuda, M.; Hata, M.; Hoshino, T. J Phys Chem B 2002, 106, 9151.
10. Tsuchiya, T.; Katagiri, D.; Hata, M.; Hoshino, T.; Tsuda, M. J Mol Struct (Theochem) 2002, 589–590, 413.
11. Wang, Z. X.; Duan, Y. J Comput Chem 2004, 25, 1699.
12. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Chem Phys Lett 1995, 246, 122.
13. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 19824.
14. Tsui, V.; Case, D. A. Biopolymers (Nucl Acid Sci) 2001, 56, 275.
15. Tsui., V.; Case., D. A. J Am Chem Soc, 2000, 122, 2489.
16. Onufriev, A.; Bashford, D.; Case, D. A. Proteins 2004, 55, 383.
17. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., 3rd. J Comput Chem 2004, 25, 265.
18. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
19. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Jr., T. V.; 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, Inc, Wallingford, CT; 2004.
20. Tomasi, J.; Mennucci, B.; Cancès, E. J Mol Struct (Theochem) 1999, 464, 211.
21. Pomelli, C. S.; Tomasi, J.; Barone, V. Theor Chem Acc 2001, 105, 446.
22. Baronea, V.; Cossi, M.; Tomasi, J. J Chem Phys 1997, 107, 3210.
23. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. J Comput Chem 1995, 16, 1357.
24. Pastor, M. T.; Lopez de la Paz, M.; Lacroix, E.; Serrano, L.; Perez-Paya, E. Proc Natl Acad Sci USA 2002, 99, 614.
25. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Proc Natl Acad Sci USA 2001, 98, 5578.
26. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Nat Struct Biol 2002, 9, 425.
27. McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. Nat Struct Biol 1997, 4, 180.
28. Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; 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, AMBER8, P. A. University of California, San Francisco; 2004.
29. Pastor, R. W.; Brooks, B. R.; Szabo, A. Mol Phys 1988, 65, 1409.
30. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Biopolymers (Nucl Acid Sci) 1992, 32, 523.
31. Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. J Chem Phys 2001, 114, 2090.
32. Tugarinov, V.; Zvi, A.; Levy, R.; Anglister, J. Nat Struct Biol 1999, 6, 331.
33. Sharon, M.; Kessler, N.; Levy, R.; Zolla-Pazner, S.; Gorlach, M.; Anglister, J. Structure 2003, 11, 225.
34. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. V.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
35. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. J Chem Phys 1983, 79, 926.
36. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
37. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. J Chem Phys 1995, 103, 8577.
38. Crowley, M. F.; Darden, T. A.; Cheatham, T. E. III.; Deerfield, D. W. II. J Supercomput 1997, 11, 255.
39. Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. J Chem Phys 2000, 113, 10913.
40. Vranken, W. F.; Budesinsky, M.; Fant, F.; Boulez, K.; Borremans, F. A. FEBS Lett 1995, 374, 117.
41. Yuki, H.; Tanaka, Y.; Hata, M.; Ishikawa, H.; Neya, S.; Hoshino, T. J Comp Chem 2007, 28, 1091.
42. Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. J Mol Biol 2003, 327, 711.
43. Zhou, R. J Mol Graph Model 2004, 22, 451.
44. Pitera, J. W.; Swope, W. Proc Natl Acad Sci USA 2003, 100, 7587.
45. Zhou, R. Proc Natl Acad Sci USA 2003, 100, 13280.
46. Herzberg, O. J Mol Biol 1991, 217, 701.
47. Fujii, Y.; Okimoto, Y.; Hata, M.; Narumi, T.; Yasuoke, K.; Susukita, R.; Suenaga, R.; Futatsugi, N.; Koishi, T.; Furusawa, H.; Kawai, A.; Ebisuzaki, T.; Neya, S.; Hoshino, T. J Phys Chem B 2003, 107, 10274.
48. Kaldor, S.W.; Kalish, V. J.; Davies, J. F., II; 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. J Med Chem 1997, 40, 3979.
49. Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. J Am Chem Soc 2006, 128, 7887.
50. Walker, R. *Amber Tutorials*. Available at: http://amber.scripps.edu/tutorials/. Last accessed 3 March 2008.
51. Hata, M.; Fujii, Y.; Tanaka, Y.; Ishikawa, H.; Ishii, M.; Neya, S.; Tsuda, M.; Hoshino, T. Biol Pharm Bull 2006, 29, 2151.
52. Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. J Am Chem Soc 2006, 128, 7887.
53. Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. J Med Chem 2007, 50, 1768.
54. Ode, H.; Matsuyama, S.; Hata, M.; Neya, S.; Kakizawa, J.; Sugiura, W.; Hoshino, T. J Mol Biol 2007, 370, 598.
55. Hata, M.; Tanaka, Y.; Fujii, Y.; Neya, S.; Hoshino, T. J Phys Chem B 2005, 109, 16153.
56. Wang, Z.-X.; Duan, Y. J Comput Chem 2004, 25, 1699.
57. Hirota, N.; Mizuno, K.; Goto, Y. J Mol Biol 1998, 275, 365.
58. Luo, P.; Baldwin, R. L. Biochemistry 1997, 36, 8413.
59. Kurpin, S.; Graslund, A.; Ehrenberg, A.; Koch, M. H. J. Biochem Biophys Res Commun 1995, 217, 1151.
60. Fraczkiewicz, R.; Braun, W. J. Comp. Chem. 1998, 19, 319.
61. Jorgensen, W. L.; Madura, J. D. Mol Phys 1985, 56, 1381.
62. Mahoney, M. W.; Jorgensen, W. L. J Chem Phys 2000, 112, 8910.
63. Caldwell, J. W.; Kollman, P. A. J Phys Chem 1995, 99, 6208.
64. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J Phys Chem 1987, 91, 6269.

Japanese Journal of Applied Physics
Vol. 47, No. 5, 2008, pp. 3719–3725

Accurate Evaluation Method of Molecular Binding Affinity from Fluctuation Frequency

Tyuji HOSHINO[1,2*], Koji IWAMOTO[1], Hirotaka ODE[1], and Iwao OHDOMARI[3]

[1]*Graduate School of Pharmaceutical Sciences, Chiba University, Chiba 263-8522, Japan*
[2]*PRESTO, JST, Kawaguchi, Saitama 332-0012, Japan*
[3]*School of Science and Engineering, Waseda University, Tokyo 169-8555, Japan*

(Received November 22, 2007; accepted January 9, 2008; published online May 16, 2008)

Exact estimation of the molecular binding affinity is significantly important for drug discovery. The energy calculation is a direct method to compute the strength of the interaction between two molecules. This energetic approach is, however, not accurate enough to evaluate a slight difference in binding affinity when distinguishing a prospective substance from dozens of candidates for medicine. Hence more accurate estimation of drug efficacy in a computer is currently demanded. Previously we proposed a concept of estimating molecular binding affinity, focusing on the fluctuation at an interface between two molecules. The aim of this paper is to demonstrate the compatibility between the proposed computational technique and experimental measurements, through several examples for computer simulations of an association of human immunodeficiency virus type-1 (HIV-1) protease and its inhibitor (an example for a drug–enzyme binding), a complexation of an antigen and its antibody (an example for a protein–protein binding), and a combination of estrogen receptor and its ligand chemicals (an example for a ligand–receptor binding). The proposed affinity estimation has proven to be a promising technique in the advanced stage of the discovery and the design of drugs. [DOI: 10.1143/JJAP.47.3719]

KEYWORDS: binding energy, drug design, computation, affinity, interfacial fluctuation

1. Introduction

Computers play a significant role in the drug discovery and development of today. One of the major applications of computers in drug discovery is the virtual *in silico* screening to find hit chemicals interacting with a target protein, which reduces experimental efforts in the real *in vitro* screening. In the *in silico* screening, hit chemicals are searched from a database in terms of the similarity to several bioactive molecules which were already known to be bound to the target protein. Unless the information on these bioactive molecules is available in advance, we cannot perform the *in silico* screening. Further the screening cannot find hit chemicals, if the database has no entry on the chemicals. In order to overcome the difficulty in the *in silico* screening, the structure-based drug design has attracted much attention recently. Since the structure-based drug design takes both the chemical compound and the target protein into account, it demands heavy computational costs. Nevertheless, the structure-based drug design currently becomes a practical technology due to the rapid progress in the performance of computers.

The structure-based drug design requires an accurate evaluation of the binding affinity of chemicals to the target protein. One of the approaches to estimate the binding affinity in a computer is the energy calculations based on molecular dynamics (MD) simulations, in which molecular mechanics generalized Born surface area (MMGBSA)[1] or molecular mechanics Poisson Boltzmann surface area (MMPBSA) method[2] is a standard technique and frequently used in the field of computational biology. The energy computed by MMGBSA or MMPBSA method, however, usually show differences of at least more than several kcal/mol among snapshot structures during MD simulation. For the purpose of estimating more accurate binding energy, some advanced techniques like the free energy perturbation[3,4] or the thermodynamic integration[5] have been developed so far. The free energy perturbation method was shown to be effective[3] when an accurate binding free energy for one chemical to the target protein was already available and simulation was performed to evaluate the affinity of the designed compounds having a similar but another chemical structure. The other computational approaches for affinity estimation are surface complementarity,[6] B-factor analysis,[7] and molecular orbital approach.[8] While these techniques provide good estimation in some biomolecules, they have proved unsatisfactory in our trials for the estimation of drug efficacy of inhibitors against human immunodeficiency virus type-1 (HIV-1) protease. Hence, a more precise estimation method of molecular binding affinity is required for the success of the structure-based drug design and for the advance of drug discovery. In this study, HIV-1 protease and its inhibitor are firstly instanced to explain the relationship between the binding affinity and the fluctuation property at their interface. Next a combination of an hen egg-white lysozyme and its antibody HyHEL-10 is examined to demonstrate a compatibility between the estimation from the interfacial fluctuation and the experimentally measured binding energy. Furthermore, an association of human estrogen receptor α-subtype ligand binding domain (hERαLBD) and its ligand chemicals is picked as an example for the ligand–receptor complex.

2. Methods

2.1 *Construction of the model structures for the drug–enzyme, antigen–antibody, and ligand–receptor complexes*

All initial structures have been constructed from the X-ray crystal structures (PDB code: 1OHR for NFV-bound HIV-1 protease, 1C08, 1IC4, 1IC5, 1IC7, 1J1O, 1J1P, 1J1X for the wild type and HD32A, HD96A, HD32AD96A, LY50F, LS91A, LS93A mutants of HyHEL-10 antibody, 1ERE, 3ERD, 1ERR, 3ERT, 1X7R for EST, DES, RAL, OHT, or GEN-associated hERαLBD). Mutated residues in 1OHR crystal data have been converted to be identical with the wild-type sequence: HXB2cv.[9] Missing residues in the hERαLBD crystal data have been added manually to complete the protein structure. Model structures for inhib-

*E-mail address: hoshino@faculty.chiba-u.jp

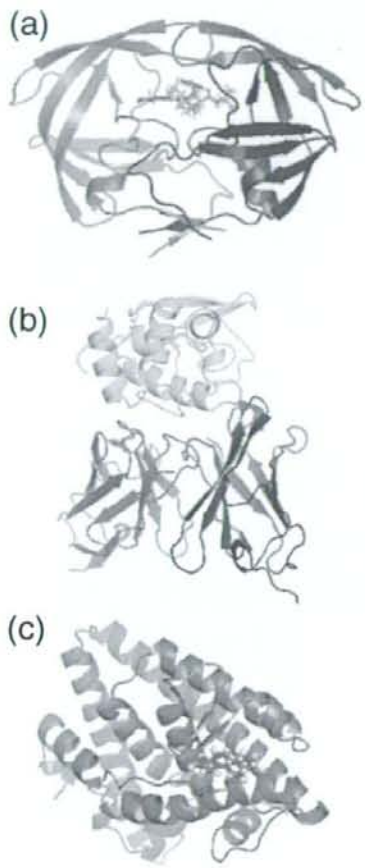

Fig. 1. (Color online) Structures of (a) NFV-bound wild type HIV-1 protease, (b) the complex of an antigen lysozyme and its antibody HyHEL-10, and (c) ligand-associated hERαLBD. Drug is represented by sticks in (a). The antigen is represented in yellow, and the heavy and light chains of the antibody are blue and green in (b). Ligands are represented by sticks in (c).

itor-bound HIV-1 protease, antigen-associated HyHEL-10 antibody, and ligand-bound hERαLBD are shown in Fig. 1.

2.2 *Minimization and equilibration*

MD simulations are carried out using the AMBER8 program package with the Amber ff03 force field.[10] Complexes are solvated in a box and the periodic boundary condition is applied. An integration time step of the simulation is 1 fs. The particle mesh Ewald method and a cut-off distance of 12 Å for non-bonded interactions are employed, and the pressure is kept constant at 1 atm. First, potential energy minimizations are performed starting from the respective crystal structure. The earlier 3000 steps in minimization are executed by the steepest descent method and the later 7000 steps are by the conjugated gradient method. Next, MD simulation is performed from the energy minimized structure. The temperature of the complex is gradually increased by heating up to 300 K for 60 ps and then kept at 300 K for over next 2 ns. Judging from the changes in root mean square deviation and total potential energy and from the principal component analysis, the complex is confirmed to be in equilibration. Equilibration is very important for correct affinity evaluation. The trajectory at 300 K for the subsequent 128 ps simulation is considered to be the most probable structure under the physiological conditions and the energies are collected for affinity evaluation. The principal component analysis is again executed to confirm that the complex is in a single conformation during the acquired trajectory. MMGBSA and MMPBSA calculations are carried out by the mm-pbsa module[11] of AMBER8 using the acquired trajectory for the 128 ps MD simulation.

2.3 *Data acquisition, Fourier transformation, and affinity evaluation*

During the 128 ps MD simulation for data acquisition, Coulomb and van der Waals potential energies between the drug and enzyme, the antigen and antibody, or the ligand and receptor are calculated every integration time step (1 fs). No cut-off method is applied for this energy calculation. In the present work, only van der Waals energy is employed for the evaluation. Not only interfacial energy but also some specific inter-atomic distances can be a barometer to reflect the binding state at the interface.[12,13] Nevertheless, the interfacial van der Waals potential energy has been selected for data acquisition, because van der Waals energy collects the contributions from not only specific local bonds but all area of the interface. The energies are averaged every 1000 steps (1 ps), which generates a sequence of 128 energies along the simulation time. For normalization, 128 energies are scaled by the average. The deviation from the average, this average is equivalent to 1.0 at this stage, is calculated for the respective time point and 128 deviations are again scaled by the maximum one. These 128 sequential values are converted into the frequency components through discrete Fourier transformation. A sequence of 128 energies generates 2 non-periodic and 63 cosine and 63 sine wave components, each of which is characterized by the integer corresponding to the wave number; in other word, frequency. The cosine and sine waves with the same frequency can be combined into a single form. Two wave functions are combined by $A_i \cos(\omega_i t) + B_i \sin(\omega_i t) = q_i \cos(\omega_i t + \phi)$, where $q_i = \sqrt{A_i^2 + B_i^2}$ and $\tan\phi = -B_i/A_i$. A_i and B_i are Fourier coefficients and q_i is called Fourier amplitude. Each of the Fourier components is labelled by a wave number i, which is one of the integers from 0 to 63.

2.4 *Theory*

MD simulations were performed in our recent studies to investigate the action of HIV-1 protease inhibitors,[14,15] the effectiveness of antibiotics,[16] and the interaction of membrane proteins.[17–19] In the present work, these simulation results have been closely examined by converting the obtained dynamic atom motions into animations. The examination has suggested that the atoms locating at the interface between a drug and its target enzyme or between a protein and its conjugating protein rapidly fluctuate with the decrease in binding affinity. This is an important clue to finding a characteristic index of binding affinity. Figures 2(a) and 2(b) depict the situation that a drug is bound to its target enzyme, where several structures are super-

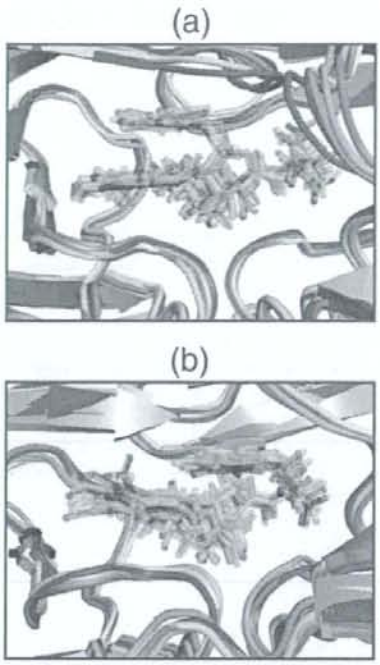

Fig. 2. (Color online) Structure of the binding pocket of HIV-1 protease associated with an inhibitor; NFV. Each image is obtained by the superimposition of six snapshot structures acquired every 10 ps for 50 ps MD simulation. The protease is depicted by a tube, gradated from dark blue to light green with the time-evolution. The inhibitor is represented by sticks with a gradation from light pink to bright red. (a) The firmly fit situation with the wild type protease. (b) The loosely bound situation with the D30N mutant. In the D30N mutant (b), NFV loses the hydrogen bond with the 30th residue of the protease and its side part starts to flicker.

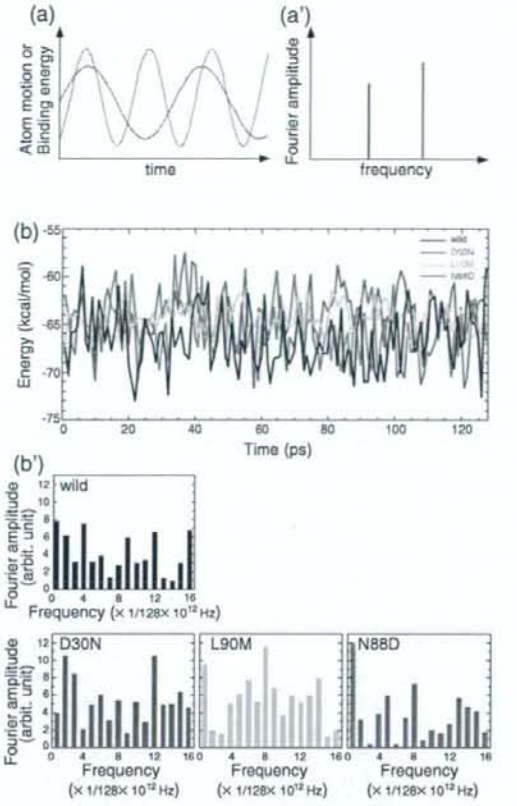

Fig. 3. (Color online) (a) Schematic representation of periodic fluctuations in atom motion and/or binding energy along the simulation time. Blue and red curves indicate the firmly fit and loosely bound cases. (a') Fourier components derived from Fourier transformation of (a). This is a core concept of binding affinity estimation from interfacial fluctuation. (b) Change of van der Waals potential energy at the interface between HIV-1 protease and its inhibitor; NFV, acquired through 128 ps MD simulation at 300 K after the equilibration of the drug–enzyme complex. (b') Fourier spectrum derived from the Fourier transformation of the 128 sampling energy data of (b). D30N, L90M, and N88D mutants are contrasted with the wild type protease.

imposed to represent the time-evolution of atom geometry. In the firmly fit case, the drug is steadily held inside the binding pocket and chemical groups of the drug seldom alter their orientation [Fig. 2(a)]. In contrast, the drug rapidly fluctuates and frequently changes its conformation in the loosely bound case [Fig. 2(b)].

Fourier transformation is one of the most promising approaches to analyze the dynamic motion of atoms. A concept of the presently proposing method of the binding affinity evaluation is explained in Fig. 3. Figure 3(a) shows schematic atom motion and/or binding energy seen in MD simulations, where the blue and red curves represent the firmly fit and the loosely bound cases. If the time-evolution of atom motion and/or binding energy shows the periodic change like Fig. 3(a), Fourier transformation gives a prominent difference in the Fourier components with respect to frequency. Namely, Fourier transformation will suggest the tendency that Fourier component appears at the higher frequency region in the loosely bound case, while at the lower frequency region in the firmly fit case [Fig. 3(a')]. The fluctuation in atom motion or binding energy is not so simple as Fig. 3(a) in the actual drug–enzyme complex. That is, the fluctuation consists of more motions with different vibration frequencies. Accordingly contributions from every motion with different frequencies should be taken into account to extract an index for the molecular binding affinity. The summation of the Fourier components weighted by a function of frequency will be one of the promising means.

Since the fluctuation frequency at the interface is closely correlated with the stability of the complex, an index for binding affinity will be derived from the time-evolved data on the interfacial energy.[20] Fluctuation can be explained from a collective set of vibrations, each of which is quantized by an oscillation frequency of ω_i and has an energy of $\hbar\omega_i$. The total vibration energy at a temperature T is given as the sum of the energies over all vibrational frequencies; $\sum_i \hbar\omega_i\langle n_i\rangle$. Here $\langle n_i\rangle$ is the thermal equilibrium occupancy of vibrations that is expressed by the distribution function $1/[\exp(\hbar\omega_i/kT) - 1]$ called the Planck formula. Since the interfacial fluctuation contains many vibrational modes even for one frequency ω_i, the number of possible modes of vibrations; in other words, the density of vibrational modes, should be taken into account in the estimation.

The number of possible vibrational modes is reflected in the respective Fourier amplitude. Hence, Fourier amplitude (q_i) contributes as a weighting factor for a frequency (ω_i) in the summation over i;

$$\sum_i q_i \frac{\hbar\omega_i}{\exp(\hbar\omega_i/kT) - 1} \quad (1)$$

It should be emphasized that the estimated value from eq. (1) becomes small with the increase of binding affinity, because this quantitative function reflects the instability of the interface of a complex. In eq. (1), $\omega_i = 2\pi/(128 \times 10^{-12}) \times i(s^{-1})$ and $T = 300$ K. k is Boltzmann's constant and $\hbar$ is Planck's constant divided by 2π. Since high frequency components may contain local libration of residues, they should be excluded in the evaluation to avoid an unfavourable error. Hence, the upper bound of the summation of eq. (1) is set to 16, which corresponds to the periodicity of 8 ps.

3. Results

HIV-1 protease is an enzyme essential for the reproduction of the virus. Nelfinavir (NFV) is one of the protease inhibitors currently used in the treatment of HIV infection. HIV-1 protease frequently causes point mutations and, for example, the D30N and L90M mutations are known to decrease the inhibitory ability of NFV; in contrast, the N88D mutant becomes more susceptible to NFV. These changes in drug efficacy of NFV were already evaluated in experiments. The effective concentration 90% (EC_{90}), which is one of the experimental measurements of drug efficacy, was reported to increase 6 and 5 times for the D30N and L90M mutants compared to the wild type and to decrease to two-thirds for the N88D mutant.[21] MD simulations were performed for the NFV-bound HIV-1 protease in the wild type and the D30N, L90M, and N88D mutants. After the complex structure of the NFV-bound HIV-1 protease became equilibrated in simulation, van der Waals energy at the interface between NFV and the protease was acquired as a function of time [Fig. 3(b)]. The number of the acquired sequential energies is 128 for the respective complex. The sequential data on the interfacial energy have been converted into the frequency components through the Fourier transformation. Figure 3(b′) shows the Fourier spectrum which represents the magnitude of the respective periodic components up to the wave number 16. It is obvious that larger Fourier amplitudes appear in the higher frequency regions in the D30N and L90M mutants compared to the wild type, while the large amplitudes in the higher frequency regions seem less prominent in the N88D mutant.

The binding affinity between NFV and HIV-1 protease has been estimated from eq. (1), based on the Fourier components shown in Fig. 3(b′). The estimations for the D30N, L90M, and N88D mutants are presented relative to the wild type protease. Figure 4(a) shows a correlation between the estimated binding affinity and the experimental virus resistivity. The resistivity is given in logarithmic transform of the ratio of drug concentrations in the same inhibitory effect relative to the wild type; $\log(EC_{90\_mutant}/EC_{90\_wild})$, which has the dimension of energy and is usually defined as the drug potency. This drug potency shows a clear linearity with the computational binding affinity in Fig. 4(a). The correlation is excellent between the computational affinity and the experimental measurement ($R = 0.99$).

A combination of an antigen (lysozyme) and its antibody (HyHEL_10) has been well studied experimentally on the wild type antibody and several mutants.[22] This is a good example for testing the validity of our affinity estimation and a typical instance for protein–protein interaction. In order to execute the computational affinity evaluation, MD simulations have been performed for the wild type HyHEL_10 and its mutants (HD32A, HD96A, and HD32AD96A). Through these MD simulation, van der Waals potential energies have been acquired as sequential data. The binding affinities between the antigen and the wild type or mutated antibodies have been estimated using eq. (1) and compared with the experimentally measured binding free energy [Fig. 4(b)]. Our affinity evaluation again shows excellent consistency with the experimental measurement ($R = 0.99$). Another experimental measurement was reported on the mutations in the light chain of the HyHEL_10 antibody; LS91A, LS93A, LY50F.[23] The binding affinities between the antigen and the wild type and mutated antibodies have been evaluated by eq. (1) [Fig. 4(c)]. Our affinity evaluation also shows a linear correlation with experimentally measured binding energy ($R = 0.91$).

Due to the therapeutic interest and the concern as an environmental hormone, estrogens have been extensively studied from the viewpoint of the affinity with their receptor. Because an estrogen receptor is capable of binding a variety of ligands and because experimental measurement of their binding affinity is available, the association of an estrogen receptor with its ligand chemicals has been a suitable subject for the trial of the computational binding assay.[8,24] The computational evaluation has been carried out, acquiring the sequential data on the interfacial energy through the MD simulations on the human estrogen receptor α-subtype ligand binding domain (hERαLBD) complexed with several kinds of ligand chemicals; 17β-estradiol (EST), diethylstilbestrol (DES), raloxifene (RAL), 4-hydroxytamoxifen (OHT), genistein (GEN). Our affinity evaluation for the above ligand chemicals provides a sound compatibility with the experimental ligand activity[25] ($R = 0.91$) [Fig. 4(d)]. In all present trials with HIV-1 protease, HyHEL_10 antibody, and estrogen receptor, the approach proposed in this study shows better compatibility with the experimental measurements compared to the conventional MMGBSA and MMPBSA methods (Fig. 5).

4. Discussion

The proposed affinity evaluation from interfacial fluctuation shows a keen analogy to the thermal energy of solids. A solid contains a lot of lattice vibrations, each of which is quantized by a frequency ν associating with an energy $h\nu$. The thermal energy of the solid at a temperature T is given as the total sum of the energies over all vibrational modes;

$$\int_0^{\nu_D} Z(\nu) \frac{h\nu}{\exp(h\nu/kT) - 1} \, d\nu, \quad (2)$$

where $1/[\exp(h\nu/kT) - 1]$ is the Planck distribution function, $Z(\nu)\,d\nu$ is the number of possible modes of vibrations in the frequency interval between ν and $\nu + d\nu$, and ν_D is the upper bound for the vibration frequency. Debye derived his

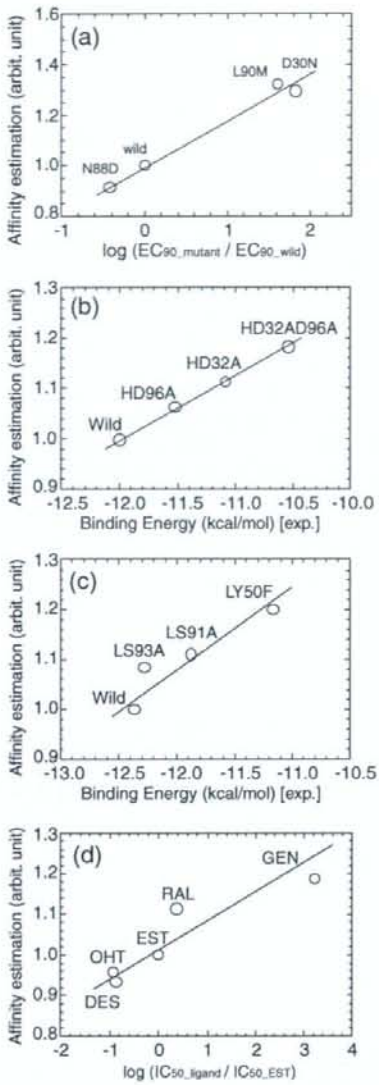

Fig. 4. Comparison of the presently proposed affinity evaluation with the experimentally measured effective concentration 90% (EC_{90}), binding energy, or inhibitory concentration 50% (IC_{50}). (a) Affinity of NFV with the wild type HIV-1 protease and its mutants. (b) Affinity of antigen–antibody complexes. HyHEL_10, which is a typical antibody for lysozyme, is exemplified with its variants mutated in the heavy chain. The affinity estimation is presented relative to that for the wild type. (c) Affinity of the lysozyme-bound HyHEL_10 antibody and its variants mutated in the light chain. (d) Affinity of ligand–receptor complex, exemplified by human estrogen receptor α ligand binding domain and its ligand chemicals. The binding affinities are presented relative to the estimation for 17β-estradiol (EST).

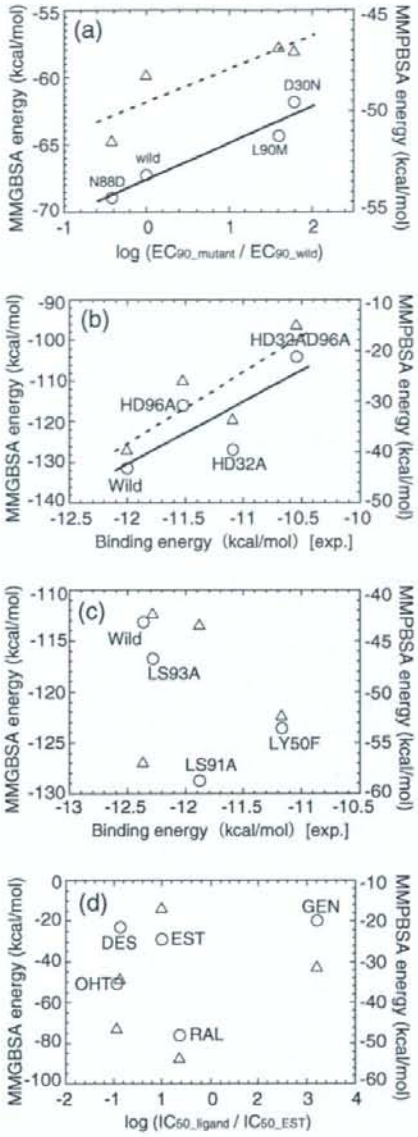

Fig. 5. Comparison of the computational energy with the experimental binding affinity. The binding energies computed by MMGBSA method are presented by triangles and the energies by MMPBSA method are by circles. (a) NFV-bound HIV-1 protease. Both MMGBSA and MMPBSA methods show good correlation with experimentally measured virus resistivity ($R = 0.87$ and 0.96). (b) Complex of an antigen lysozyme and its antibody HyHEL_10, with the variants mutated in the heavy chain. A compatibility with the experimental binding energy is good both in MMGBSA ($R = 0.83$) and MMPBSA ($R = 0.78$) methods. (c) Antigen-complexed HyHEL_10 antibody, with the variants mutated in the light chain. Computations are inconsistent with the experimental binding energy in both methods. (d) Ligand-associated hERαLBD. Neither method gives a clear compatibility with the experimental measurement.

T^3 law on the heat capacity from eq. (2) and achieved a perfect explanation on the experimental thermal property of a solid crystal.[26] Equation (1) proposed in this study is equivalent to eq. (2) when the summation is displaced by integration. The value estimated from eq. (2) is in a dimension of energy, which is stored in a solid as lattice vibration. Likewise, eq. (1) is interpreted to represent a value in a dimension of energy, which is accumulated at an interface as atom fluctuation.

The high frequency components, whose wave number is over 16; i.e., whose periodicity is below 8 ps, have been excluded in the affinity estimation of our present trials, which gives a better compatibility with the experimental binding energy compared to the inclusion of all frequency components. This exclusion is related to the atom motions detected by the nuclear magnetic resonance (NMR). NMR is an experimental technique essential for analyzing the internal motions of proteins and is sensitive not only to the magnitude but also to the time scale of the motions. A typical NMR time-correlation function for internal motion in proteins shows two major time regions.[27] One is a large peak at the region below several picoseconds and the other is a moderate plateau at the region from tens to a few hundred picoseconds. The former corresponds to the librational motion of residues and the latter to the dynamic motion of domains. The latter dynamics is responsible for the hold of molecules as a whole, while the former libration scarcely characterizes the binding state at the interface. This means that the upper bound of the frequency in the summation of eq. (1) is effective to focus only on the fluctuations relevant to the molecular binding.

Our approach using the Fourier transformation provides two important pieces of information. One is the summation of Fourier components weighted by a function of frequency like eq. (1). The other is the composition of frequency spectrum. For example, the character of the inter-molecular interaction appears in the frequency spectra for the D30N L90M, and N88D mutants of Fig. 2(b′). The 30th residue of HIV-1 protease makes a direct hydrogen bond with NFV. This hydrogen bond disappears due to the D30N mutation and the binding ability of NFV is seriously reduced. In contrast, the L90M mutation causes serious deformation of the binding pocket.[14] Due to this deformation, the orientation of NFV alters inside the binding pocket to avoid the fatal collision with the protease. The N88D mutant also causes noticeable deformation of the binding pocket. Interestingly the fitness of NFV for the deformed pocket is improved in the N88D mutant. We notice that the frequency spectrum of the D30N mutant resembles that of the wild type. On the other hand, the frequency spectra of the L90M and N88D mutants are quite different from that of the wild type. This means that the L90M and N88D mutants change their binding constraints on NFV compared to the wild type protease.

As exemplified in the estimation of binding affinity between HIV-1 protease and its inhibitor, the D30N or L90M mutation causes the decrease of drug efficacy. In case of an approved anti-HIV-1 drug; NFV, EC_{90} was reported to increase 6 and 5 times for the D30N and L90M mutants compared to the wild type.[21] Hence more than 5-fold dosage is required to suppress the enzymatic activity of the mutants at the same level as the suppression of the wild type virus. This means that NFV cannot be prescribed for the patient who was infected with the D30N or L90M HIV-1 variants. Since the virus easily causes amino acid mutation and the diversity of variants is high, it is preferable to assay the efficacy of drugs for individual patient before prescription. The computational assay will be one of the helpful techniques for analyzing the efficacy of drugs for individuals and will contribute the progress of the tailor-made medicine because the process can be automated to treat a large number of samples and the cost for assay will be reduced due to the increase of the performance of computer. For the purpose of drug screening, rapidness should be weighted at the cost of accuracy, since subsequent experimental assay will cover more precise screening of potent candidates. In contrast, accuracy is critically important for the choice of drugs for patient. Therefore, an accurate computational technique is indispensable for the therapy in future and the approach proposed in this study will be one of the promising methodologies.

In our present approach to estimate the binding affinity, Fourier transformation is executed to convert the data on the fluctuation of van der Waals energy into frequency components. Discrete Fourier transformation is usually employed when applied to the sampling data acquired by monitoring some signal. Since Fourier transformation can be fundamentally applied only to the periodic function, discontinuity between the values at the first and last sampling points sometimes causes an unfavourable error. Similar difficulty will also arises when a complex of some ligand and its receptor has two stable configurations and the binding configuration frequently alters. While confirming the structural equilibration is one of the fundamental notes to avoid such accidental errors, some technique to solve this difficulty should be developed in our future study.

5. Conclusions

In order to solve the problem that conventional energy calculations to estimate the ligand–receptor binding affinity are not accurate enough to distinguish a slight difference in drug efficacy, we proposed a concept of estimating molecular binding affinity, focusing on the fluctuation at a ligand–receptor interface. The proposed affinity estimation coincides well with the available experimental results such as an association of a virus protease and its inhibitor, a complexation of an antigen and its antibody, and a combination of estrogen receptor and its ligand chemicals, and has proven to be one of the promising techniques in drug design and discovery.

Acknowledgment

We thank Professor Y. Shinoda of Waseda University for his critical reading of the manuscript. This work was supported by a Grant-in-Aid for COE research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Health and Labour Science Research Grant for Research on HIV/AIDS from the Ministry of Health and Labour of Japan. A part of this work is also supported by the grants from Japan Science and Technology Agency.

1) D. Bashford and D. A. Case: Annu. Rev. Phys. Chem. **51** (2000) 129.
2) P. Kollman: Chem. Rev. **93** (1993) 2395.
3) K. M. Merz and P. A. Kollman: J. Am. Chem. Soc. **111** (1989) 5649.
4) S. Hirono and P. A. Kollman: J. Mol. Biol. **212** (1990) 197.
5) W. L. Jorgensen: Acc. Chem. Res. **22** (1989) 184.
6) M. C. Lawrence and P. M. Colman: J. Mol. Biol. **234** (1993) 946.
7) C. B. Post, C. M. Dobson, and M. Karplus: Proteins: Struct. Funct. Genet. **5** (1989) 337.
8) K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma, and T. Nakano: J. Comput. Chem. **26** (2005) 1.

9) W. Sugiura, Z. Matsuda, Y. Yokomaku, K. Hertogs, B. Larder, T. Oishi, A. Okano, T. Shiino, M. Tatsumi, M. Matsuda, H. Abumi, N. Takata, S. Shirahata, K. Yamada, H. Yoshikura, and Y. Nagai: Antimicrob. Agents Chemother. **46** (2002) 708.
10) D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A. Kollman: AMBER8 (University of California, San Francisco, CA, 2004).
11) A. Onufriev, D. A. Case, and D. Bashford: J. Comput. Chem. **23** (2002) 1297.
12) S. Y. Sheu, D. Y. Yang, H. L. Selzle, and E. W. Schlag: Proc. Natl. Acad. Sci. U.S.A. **100** (2003) 12683.
13) H. Nojima, M. Takeda-Shikata, Y. Kurihara, M. Adachi, S. Yoneda, K. Kamiya, and H. Umeyama: Chem. Pharm. Bull. **50** (2002) 1209.
14) H. Ode, M. Ota, S. Neya, M. Hata, W. Sugiura, and T. Hoshino: J. Phys. Chem. B **109** (2005) 565.
15) H. Ode, M. Hata, S. Neya, W. Sugiura, and T. Hoshino: J. Am. Chem. Soc. **128** (2006) 7887.
16) Y. Fujii, N. Okimoto, M. Hata, T. Narumi, K. Yasuoka, R. Susukita, A. Suenaga, N. Futatsugi, T. Koishi, H. Furusawa, A. Kawai, T. Ebisuzaki, S. Neya, and T. Hoshino: J. Phys. Chem. B **107** (2003) 10274.
17) Y. Sato, M. Hata, S. Neya, and T. Hoshino: Protein Sci. **14** (2005) 183.
18) K. Mori, M. Hata, S. Neya, and T. Hoshino: Chem-Bio Inf. J. **4** (2004) 15.
19) K. Mori, M. Hata, S. Neya, and T. Hoshino: J. Am. Chem. Soc. **127** (2005) 15127.
20) K. Iwamoto, H. Ode, M. Ohta, T. Misu, M. Hata, S. Neya, and T. Hoshino: Jpn. J. Appl. Phys. **44** (2005) L1370.
21) A. K. Patick, M. Duran, Y. Cao, D. Shugarts, M. R. Keller, E. Mazabel, M. Knowles, S. Chapman, D. R. Kuritzkes, and M. Markowitz: Antimicrob. Agents Chemother. **42** (1998) 2637.
22) M. Shiroishi, A. Yokota, K. Tsumoto, H. Kondo, Y. Nishimiya, K. Horii, M. Matsushima, K. Ogasawara, K. Yutani, and I. Kumagai: J. Biol. Chem. **276** (2001) 23042.
23) A. Yokota, K. Tsumoto, M. Shiroishi, H. Kondo, and I. Kumagai: J. Biol. Chem. **278** (2003) 5410.
24) P. D. Kirchhoff, R. Brown, S. Kahn, M. Waldman, and C. M. Venkatachalam: J. Comput. Chem. **22** (2001) 993.
25) G. G. J. M. Kuiper, J. G. Lemmen, B. Carlsson, J. C. Corton, S. H. Safe, P. T. van der Saag, B. van der Burg, and J. Gustafsson: Endocrinology **139** (1998) 4252.
26) A. J. Dekker: *Solid State Physics* (Prentice-Hall, New York, 1957).
27) C. L. Brooks III, M. Karplus, and B. M. Pettitt: in *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (John Wiley & Sons, New York, 1988) p. 201.

Mechanism of Drug Resistance Due to N88S in CRF01_AE HIV-1 Protease, Analyzed by Molecular Dynamics Simulations

Hirotaka Ode,*,† Shou Matsuyama,† Masayuki Hata,† Tyuji Hoshino,†,‡ Junko Kakizawa,§ and Wataru Sugiura§

Graduate School of Pharmaceutical Sciences, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan, PRESTO, JST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan, and AIDS Research Center, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama, Tokyo 208-1011, Japan

Received October 6, 2006

Nelfinavir (NFV) is a currently available HIV-1 protease (PR) inhibitor. Patients in whom NFV treatment has failed predominantly carry D30N mutants of HIV-1 PRs if they have been infected with the subtype B virus. In contrast, N88S mutants of HIV-1 PRs predominantly emerge in patients in whom NFV treatment has failed and who carry the CRF01_AE virus. Both D30N and N88S confer resistance against NFV. However, it remains unclear why the nonactive site mutation N88S confers resistance against NFV. In this study, we examined the resistance mechanism through computational simulations. The simulations suggested that despite the nonactive site mutation, N88S causes NFV resistance by reducing interactions between PR and NFV. We also investigated why the emergence rates of D30N and N88S differ between subtype B and CRF01_AE HIV-1. The simulations suggested that polymorphisms of CRF01_AE PR are involved in the emergence rate of the drug-resistant mutants.

Introduction

Human immunodeficiency virus type 1 (HIV-1) is one of the most hazardous viruses for humans, and there is still a risk of a worldwide HIV-1 pandemic. HIV-1 has high genetic variability and has been classified into three groups labeled M, N, and O. Viruses in group M are further divided into subtypes, subsubtypes, and circulating recombinant forms (CRFs). The subtype B virus is commonly found in HIV-1-infected patients in the Americas, Europe, and Japan. In contrast, developing countries suffer from a growing epidemic of nonsubtype B viruses.

HIV-1 proliferates with the assistance of its own aspartic protease, so-called HIV-1 protease (HIV-1 PR), in its life cycle.[1] HIV-1 PR is an enzyme composed of two identical polypeptides each consisting of 99 amino acid residues, and its function is to process the viral Gag and Gag-Pol polyprotein precursors (Figure 1A). Because this processing is essential for viral maturation, inhibition of PR function leads to incomplete viral replication and prevents the transfer to other cells.[2] Therefore, HIV-1 PR is an attractive target for anti-HIV-1 drugs. Nine PR inhibitors (PIs)[3−11] have been approved by the FDA and have successfully lowered the death rate due to acquired immune deficiency syndrome (AIDS) in advanced countries during the past decade. However, the currently available PIs were developed and tested only against subtype B PRs. Few studies have examined the susceptibility of nonsubtype B viruses to those PIs, and no standard protocol of chemotherapy for nonsubtype B viruses has been established.[12−17]

Recently, Ariyoshi et al. reported that the pattern of drug-resistant mutations differed between subtype B and CRF01_AE (subtype AE) HIV-1.[16] Mutations of L10F, K20I, L33I, and N88S in PR were more frequently seen in patients infected with subtype AE HIV-1 than in patients infected with subtype B HIV-1. Mutations of D30N, A71V, and N88D in PR were found in patients with subtype B HIV-1. Most of the charac-

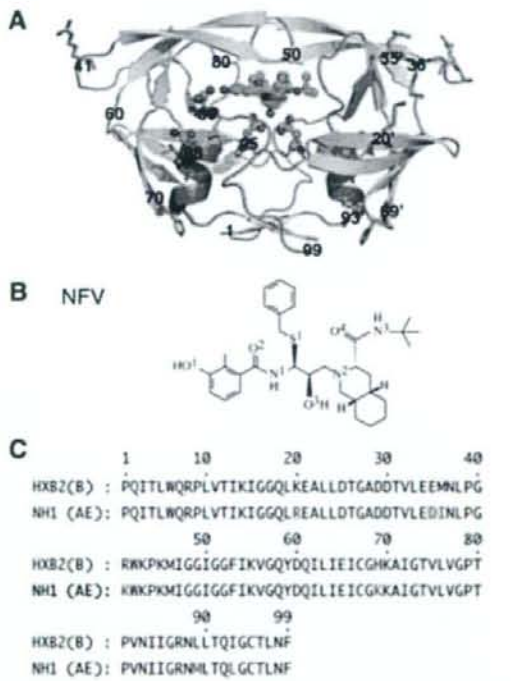

Figure 1. (A) Structure of HIV-1 PR. Locations of two catalytic aspartates, the 30th and the 88th residues, are shown in ball and stick representation. Locations of polymorphisms in subtype AE PR (K20R, E35D, M36I, R41K, H69K, L89M, and I93L) are shown in stick representation. (B) Chemical structure of NFV. (C) Amino acid sequences of a wild-type (WT) subtype B HIV-1 PR (HXB2) and a reference sequence of subtype AE HIV-1 PR (NH1). The polymorphisms in subtype AE PR are highlighted in red letters.

* To whom correspondence should be addressed. Phone: +81-43-290-2926. Fax: +81-43-290-2925. E-mail: odehir@graduate.chiba-u.jp.
† Chiba University.
‡ PRESTO.
§ National Institute of Infectious Diseases.

10.1021/jm061158i CCC: $37.00
Published on Web 03/17/2007

teristic mutation patterns in that study were associated with a history of treatment with nelfinavir (NFV, Figure 1B), an FDA-approved PI. D30N and N88S are particularly related to resistance against NFV.[18−21] Interestingly, N88S is also known to cause hypersensitivity to another PI, amprenavir. We have previously suggested by computational simulations that D30N in subtype B PR confers resistance against NFV by canceling hydrogen bonds between NFV and N30.[22] In addition, we and other groups have proposed an explanation of why some mutations confer resistance against PIs by not only X-ray crystallography[23−33] but also computational studies.[34−41] However, it has not been clarified why N88S in subtype AE PR confers resistance against NFV. Since N88S occurs at a nonactive site of PR, it is difficult to speculate on the mechanism of resistance. Furthermore, it has not been understood why N88S emerges more predominantly than D30N in patients with subtype AE HIV-1 or why D30N emerges more predominantly than N88S in patients with subtype B HIV-1. Subtype AE HIV-1 has natural polymorphisms, K20R, E35D, M36I, R41K, H69K, L89M, and I93L, in PR unlike subtype B HIV-1 (Figure 1C). These polymorphisms are also located at the nonactive site of PR. It is still uncertain whether or not the polymorphisms affect the emergence rates of those mutations.

In this study, we investigated the mechanism of resistance against NFV due to N88S in subtype AE HIV-1 PR through computational simulations. Our simulations indicated that the N88S mutation creates hydrogen bonds between the D30 and S88 side chains. Therefore, N88S mutation reduces interactions between D30 and NFV. We also investigated the reason for the difference between the two subtypes in the emergence rate of D30N as well as that of N88S. The results indicated that, in subtype B HIV-1, D30N PR has a lower affinity for NFV than does N88S PR. In subtype AE HIV-1, on the other hand, D30N PR has a higher affinity than does N88S. Our findings suggest that despite the nonactive site mutations, the polymorphisms regulate the emergence rates of these drug-resistant mutants.

Results

Reconsideration of Torsional Force Field Parameters for Benzamide. Before carrying out molecular dynamics (MD) simulations, we reconsidered torsional force field parameters for benzamide: CA−CA−C−N and CA−CA−C−O (Supporting Information Figure S1). The benzamide group comprises a part of NFV. The benzamide moiety in NFV has an important interaction with D30 of PR.[22,42] Nevertheless, the AMBER ff03[43] and general AMBER force fields[44] cause a much higher energy barrier around the rotatable bond between the benzene and amide groups in benzamide than that based on quantum chemical calculations. This is a serious problem for our simulations. The force field parameters for benzamide need to be carefully examined and preferably changed from the original AMBER force fields, as described in the AMBER Archive in 2003.[45] Since these force field parameters have not been changed in the AMBER force fields yet, we improved the torsional force field parameters for the benzamide moiety in NFV. The torsional parameters were generated in the same manner as that for the development of the AMBER ff03 force field. The obtained parameters are listed in Table 1. We executed MD simulations using these newly developed force field parameters.

Hydrogen Bonds between NFV and PRs. Hydrogen bonds play an important role in protein−ligand bindings. First, we examined the hydrogen bonds between NFV and PR in each complex: 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, and N88S PR of subtype AE HIV-1 in complex with NFV (AE(Ref), AE(D30N), and AE(N88S)). We examined 1000 snapshot structures for the last 1.0 ns and identified direct or one-water-molecule-mediated hydrogen bonds (Table 2 and Supporting Information Table S1 and Figure S2). All six PRs create similar hydrogen bond networks. The side chains of both D25 and D25′ interact with the central hydroxyl group of NFV (the atom corresponding to O3 in Figure 1B). One water molecule mediates the interaction between the main chains of I50/I50′ and NFV. Furthermore, another water molecule mediates the interaction between D29′ and NFV. However, the interaction between NFV and the 30th residue has variations among the six PRs. In B(WT) and AE(Ref), either the main chain or the side chain of D30 makes a direct hydrogen bond with NFV. D30N and N88S models show different interactions between subtype B and AE PRs. B(D30N) has no hydrogen bond between N30 and NFV, whereas AE(D30N) has direct or one-water-molecule-mediated hydrogen bonds. B(N88S) frequently creates a direct hydrogen bond between the main chain of D30 and NFV. On the other hand, AE(N88S) mainly creates one-water-molecule-mediated hydrogen bonds between the main chain of D30 and NFV. Interestingly, the side chain of N30 in AE(D30N) is clearly closer to the phenol group of NFV than that of B(D30N) (Figure 2). In contrast, the side chain of D30 in AE(N88S) is more distant from the phenol group of NFV than that of B(N88S).

Table 1. Force Field Parameters for the Torsional Parameters of the Benzamide Part of NFV[a]

CA−CA−C−N			CA−CA−C−O		
$V_n/2$	n	γ	$V_n/2$	n	γ
Developed Parameters					
0.90	2	180.0	0.90	2	180.0
0.05	4	0.0	0.05	4	0.0
AMBER ff03 Force Field					
3.63	2	180.0	3.63	2	180.0

[a] Torsional energy is given by $E = (V_n/2)[1 + \cos(n\phi - \gamma)]$.

Hydrogen Bonds of the Side Chain of the 30th Residue with PR Residues. In B(D30N) and in AE(N88S), the side chain of the 30th residue does not create any hydrogen bonds with NFV. To clarify the effects of the D30N and N88S mutations in detail, the interactions of the side chain of the 30th residue with other residues of PR were investigated as shown in Table 3. B(WT) and AE(Ref) have an interaction between the side chains of D30 and K45. B(D30N) has direct hydrogen bonds from the side chain of N30 to T31 and T74. On the other hand, AE(D30N) has one-water-molecule-mediated hydrogen bonds from N30 to T31, T74, and N88. T31, T74, and N88 also create hydrogen bond networks at the nonactive sites in B(WT) and AE(Ref), although D30 is not involved in the networks (Supporting Information Table S2). The side chains of D30 in both B(N88S) and AE(N88S) have either a direct hydrogen bond with the side chain of S88 or one-water-molecule-mediated hydrogen bonds with T31, T74, and S88. The mutations D30N and N88S affect those hydrogen bond networks.

Comparison of the Structures with B(WT). To clarify the effects of mutations at the 30th and the 88th residues on the active site conformations, the average structure of each model for the last 1.0 ns was compared with that of B(WT). Each model was fitted to B(WT) using the coordinates of main chain atoms N, Cα, and C, and the root mean squared deviation (rmsd) value was calculated (Figure 3). When the active site residues of each PR are compared with those of B(WT), conformational

Table 2. Hydrogen Bond Networks of NFV with D30 or N30 in PR

subtype B					subtype AE				
donor		acceptor		%[a]	donor		acceptor		%
		B(WT)					AE(ref)		
N	D30	O1[b]	NFV	30.7	N	D30	O1	NFV	5.7
O1	NFV	OD1/OD2	D30	31.6	O1	NFV	OD1/OD2	D30	69.2
O1	NFV	O	D30	42.9	O1	NFV	O	D30	9.0
		B(D30N)					AE(D30N)		
					O1	NFV	OD1	N30	25.2
					O1	NFV	O	WAT766	12.6
					O	WAT766	OD1	N30	12.1
					O1	NFV	O	WAT1770	8.4
					O	WAT1770	OD1	N30	6.8
					O1	NFV	O	WAT8063	13.7
					O	WAT8063	OD1	N30	9.1
		B(N88S)					AE(N88S)		
N	D30	O1	NFV	26.7	N	D30	O1	NFV	7.4
O1	NFV	OD2	D30	12.4	O1	NFV	O	D30	27.9
O1	NFV	O	D30	56.7	O1	NFV	O	WAT6715	28.8
					N	D30	O	WAT6715	28.3
					O	WAT6715	N	D30	12.6
					O	WAT6715	O	D30	31.8
					O1	NFV	O	WAT7886	13.6
					O	WAT6715	OD1	D30	5.2
					O	WAT6715	O	D30	25.5

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

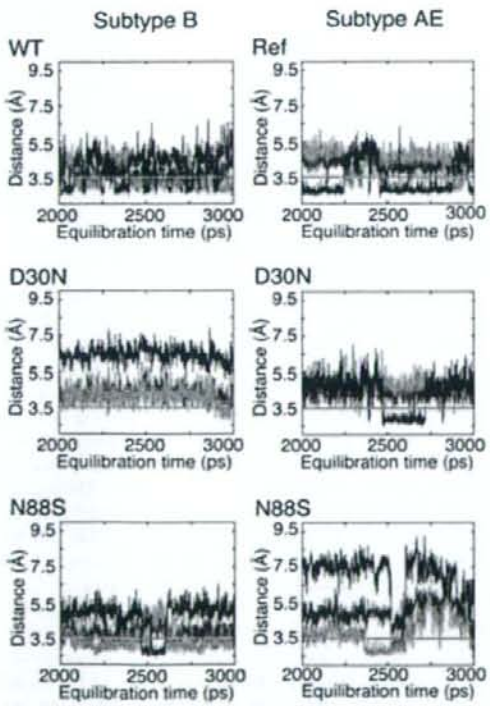

Figure 2. Distance between NFV and the 30th residue. Each red and green solid line corresponds to the distance between N of the 30th residue and the O1 atom of NFV and to the distance between O of the 30th residue and the O1 atom of NFV. Blue solid lines of B(WT), AE(Ref), B(N88S), and AE(N88S) show the distances between O1 of NFV and OD1/OD2 of D30, while those of B(D30N) and AE(D30N) show the distances between O1 of NFV and OD1/ND2 of N30.

changes are observed only on the active site residues around the 30th residue. AE(N88S) shows a large conformational change at D30 (rmsd = 1.5 ± 0.4 Å, Figure 4). The other four models (B(D30N), B(N88S), AE(Ref), and AE(D30N)) show a slight conformational change at D30. When each subtype B PR is compared with the corresponding subtype AE PR, B(D30N) is found to have larger conformational changes on N30 than AE(D30N). AE(N88S) shows larger conformational changes on D30 than B(N88S). Next, we compared the location of NFV in each model with that of B(WT). The benzamide group of NFV, which interacts with the 30th residue, shows a larger positional deviation than do other parts of NFV in every model (Supporting Information Figure S3).

Binding Free Energy Calculations. The influence of mutation or polymorphism on the binding free energy ΔG_b was examined for each model. Table 4 shows the results of MM/PBSA calculations for all of the PRs in complex with NFV. In subtype B HIV-1, B(D30N) reduces the binding energy with NFV from B(WT) more than B(N88S) does. On the other hand, in subtype AE HIV-1, AE(D30N) shows affinity with NFV, similar to AE(Ref), and has a higher affinity with NFV than AE(N88S). The results correspond to the emergence rates of subtypes B and AE variants in patients in whom NFV treatment has failed. D30N predominantly emerges in patients with subtype B HIV-1, whereas N88S predominantly emerges in patients with subtype AE HIV-1. We also investigated the contributions of the respective residues to binding free energy (Figure 5). In all six models, the active site residues stabilize the complex of each PR and NFV. When we focus on the binding energy due to the 30th residue, D30 or N30, B(D30N) reduces the contribution to the binding free energy in comparison with B(WT) (Figure 6). AE(N88S) also reduces the contribution to the binding energy compared with AE(Ref). B(N88S) shows a contribution similar to that of B(WT), and AE(D30N) shows a contribution similar to that of AE(Ref).

Discussion

In this study, we performed MD simulations of HIV-1 PRs in complex with NFV for the purpose of clarifying (1) the mechanism of resistance against NFV due to N88S in subtype

Table 3. Hydrogen Bond Networks of the Side Chain of D30 or N30 with PR Residues

subtype B					subtype AE				
donor		acceptor		%[a]	donor		acceptor		%
		B(WT)					AE(Ref)		
NZ	K45	OD1/OD2	D30	67.0	NZ	K45	OD1	D30	38.3
		B(D30N)					AE(D30N)		
ND2	N30	O	T74	89.7	ND2	N30	O	WAT224	76.9
ND2	N30	N	T31	59.5	N	T31	O	WAT224	72.6
ND2	N30	O	T31	98.1	O	WAT224	O	T31	70.7
N	T31	ND2	N30	67.9	O	WAT224	O	T74	76.8
					ND2	N88	O	WAT224	39.4
		B(N88S)					AE(N88S)		
OG	S88	OD1	D30	34.7	OG	S88	OD2	D30	77.8
O	WAT1142	OD1/OD2	D30	46.2	O	WAT226	OD2	D30	33.7
N	T31	O	WAT1142	38.3	OG1	T31	O	WAT226	18.1
O	WAT1142	O	T74	45.1	N	T31	O	WAT226	31.4
OG	S88	O	WAT1142	40.9	O	WAT226	O	T74	33.8

[a] Occupancy of hydrogen bonds during 2.0−3.0 ns of MD simulation.

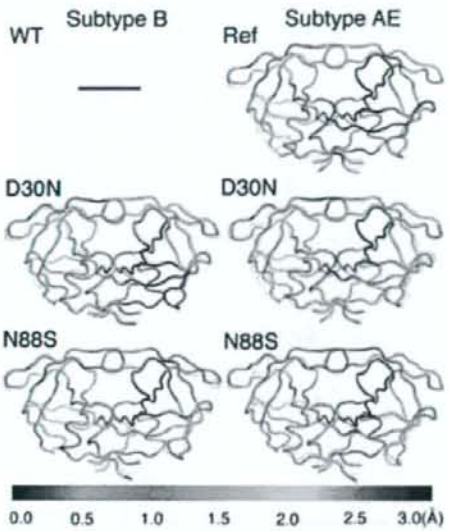

Figure 3. A 3D plot of rmsd of the average structure of each model from that of B(WT). The PR in each model is shown in colored tube representation. The color refers to the magnitude of rmsd shown in the bottom bar. Each model was fitted to B(WT) using the coordinates of main chain atoms N, Cα, and C of PR. The superimposed gray sticks and tubes represent the structure of B(WT).

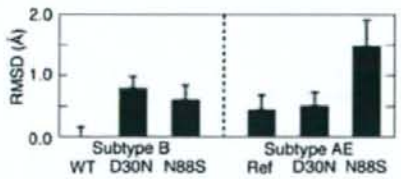

Figure 4. The rmsd value on the 30th residue of the average structure of each PR from that of B(WT). Error bars show root-mean-squared fluctuations (rmsf).

AE PR and (2) the reason that the emergence rates of D30N and N88S differ between subtypes B and AE HIV-1.

The 88th residue is located at a nonactive site of HIV-1PR. Thus, it is difficult to speculate on the mechanism of resistance due to N88S. Our simulations indicate that N88S mutant PR has a lower affinity with NFV than does Ref PR in subtype AE HIV-1, owing to the following mechanism. First, a hydrogen bond between the side chain of D30 and the side chain of S88 is created (Figure 7). Second, the location of D30 is changed. Finally, the interaction between D30 and NFV is reduced. N88S indirectly affects the binding between NFV and D30. Accordingly, both N88S and D30N are thought to confer specific resistance against NFV because the interaction with D30 is an essential factor for NFV binding. Indeed, it has been reported that the emergence of the N88S mutation is highly related to resistance against NFV.[18−21] N88D is another frequently observed mutation at the 88th residue of PR. It has also been reported that N88D changes the interactions of the 88th residue with D30, T31, and T74.[22,26] However, N88D hardly affects the ligand binding at the active site and does not cause resistance against NFV. N88S changes the interactions in a manner different from that of N88D.

We then pose another question: Why does N88S emerge more frequently in patients with subtype AE HIV-1 in whom NFV treatment has failed than in patients with subtype B HIV-1? Ariyoshi et al. reported that D30N emerged predominantly in patients with subtype B HIV-1 whereas N88S appeared predominantly in patients with subtype AE HIV-1.[16] Subtype AE HIV-1 PR has some natural polymorphisms (K20R, E35D, M36I, R41K, H69K, L89M, and I93L) unlike subtype B PR. These amino acids are located at nonactive sites of PR. To reveal whether the polymorphisms affect NFV binding or what causes the difference in the emergence rates of D30N and N88S, we carried out simulations of NFV complexes of WT PR, D30N PR, and N88S PR of subtype B (B(WT), B(D30N), B(N88S)) and Ref PR, D30N PR, and N88S PR of subtype AE (AE(Ref), AE(D30N), AE(N88S)). AE(Ref) has an interaction with NFV similar to that of B(WT). On the other hand, D30N and N88S mutations show different effects between subtypes B and AE PRs. D30N in subtype B PR greatly reduces the binding affinity with NFV because the hydrogen bonds between N30 and NFV are canceled, as we previously reported.[22] In contrast, D30N in subtype AE PR hardly affects the affinity with NFV. AE(D30N) has direct or one-water-molecule-mediated hydrogen bonds between N30 and NFV. On the other hand, N88S in subtype AE PR significantly reduces the binding affinity with NFV, whereas N88S in subtype B PR hardly affects the affinity with NFV. In both B(N88S) and AE(N88S), a hydrogen bond is created between the side chain of D30 and the side chain of S88. However, the interactions of NFV with D30 differ between

Table 4. Binding Free Energy of Each Model[a]

		ΔG_{int}^{ele}	ΔG_{int}^{vdw}	ΔG_{sol}	ΔG_b [b]	$\Delta\Delta G_b$ [c]	$\Delta\Delta G_b$ [d]
WT	B	-12.5 ± 1.4	-71.8 ± 3.8	15.1 ± 1.4	-69.2 ± 3.7		
Ref	AE	-12.8 ± 1.7	-70.8 ± 4.0	15.1 ± 1.5	-68.6 ± 3.7	0.6	
D30N	B	-6.9 ± 1.2	-70.5 ± 4.1	10.9 ± 0.9	-66.5 ± 3.9	2.7	
	AE	-7.5 ± 1.3	-70.2 ± 3.9	9.5 ± 1.0	-68.2 ± 3.7	1.0	0.4
N88S	B	-12.0 ± 1.3	-71.7 ± 3.8	15.0 ± 1.2	-68.7 ± 3.7	0.5	
	AE	-10.6 ± 1.4	-67.8 ± 4.0	12.8 ± 1.9	-65.6 ± 3.9	3.6	3.0

[a] Energy is presented in units of kcal/mol. [b] $T\Delta S$ is not included. [c] Difference from B(WT). [d] Difference from AE(Ref).

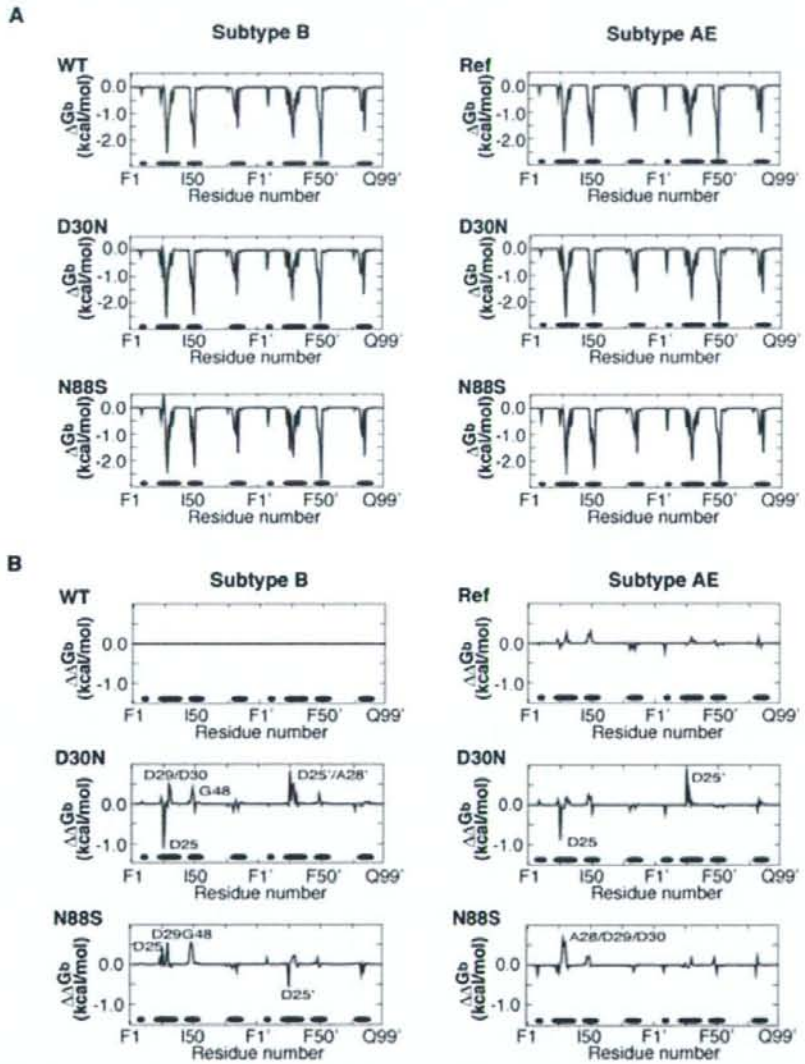

Figure 5. (A) Contribution of each individual residue to binding free energy. (B) Difference in contribution of each residue to the binding energy between the respective mutant and B(WT). The energies of contributions of the residues correspond to red solid lines, and those of B(WT) to green lines. The bottom black lines indicate the locations of the active site residues (R8, L23-V32, I47-I50, P81-I84, R8′, L23′-V32′, I47′-I50′, P81′-I84′).

subtype B and AE PRs. B(N88S) has a direct hydrogen bond between the main chain of D30 and NFV, whereas AE(N88S) mainly has one-water-molecule-mediated hydrogen bonds between the main chain of D30 and NFV. D30N PR has lower affinity with NFV than does N88S PR in subtype B HIV-1. In contrast, D30N PR has higher affinity than N88S PR in subtype AE HIV-1. These results are compatible with the results of a study by Ariyoshi et al.[16] Both D30N and N88S mutations in HIV-1 PRs exhibit significant losses of viral fitness.[20,21] Therefore, D30N and N88S mutants of HIV-1 have low growth kinetics relative to WT or Ref variants under the condition without any PIs. Nevertheless, it is frequently observed that the D30N mutant emerges in patients with subtype B HIV-1 in whom NFV treatment has failed and that the N88S mutant

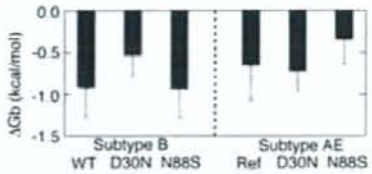

Figure 6. Contribution of the 30th residue to binding free energy in each model. Error bars stand for standard deviation.

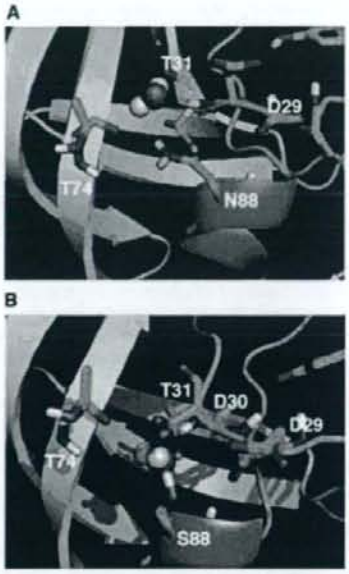

Figure 7. Hydrogen bond networks around the 88th residue of PR: (A) hydrogen bond networks in B(WT); (B) those in AE(N88S).

emerges in patients with subtype AE HIV-1.[16] These results indicate that the effectiveness of NFV is significantly reduced for these mutants. In contrast, N88S mutants of subtype B PR and D30N mutants of subtype AE PR have rarely been seen clinically. This is thought to be due not only to their low degree of fitness but also to their affinities with NFV comparable to those of subtype B WT or subtype AE Ref variants. Our simulations suggest that the natural polymorphisms of subtype AE PR, in spite of the nonactive site mutations, reduce the emergence rate of D30N and increase that of N88S.

The polymorphisms in subtype AE PR increase the emergence rate of N88S. However, there remains the question of which is the key mutation that affects the emergence rate of N88S among the polymorphisms K20R, E35D, M36I, R41K, H69K, L89M, and I93L. In this study, we focused on M36I for three reasons. First, M36I is related to the resistance against NFV.[46] Second, N88S has been observed in combination with mutations at various positions, including 20, 36, 46, 63, and 77.[19] Third, M36I is frequently observed as a polymorphism in other subtypes, namely, A and C.[13,14] We executed additional simulations 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). L10F is a mutation that is frequently seen in CRF01_AE HIV-1 accompanied by N88S.[16] Our simulations suggest that the single M36I mutation in subtype B PR does not affect NFV binding. B(M36I) has stable hydrogen bonds between NFV and D30 (Supporting Information Table S3 and Figure S4). In contrast, the combination of M36I and N88S mutations in subtype B PR reduces the binding affinity with NFV. B(M36I/N88S) has fewer hydrogen bonds with NFV than does B(M36I) or B(N88S). Furthermore, the conformational change at D30 is larger in B(M36I/N88S) than in B(M36I) or B(N88S) (Supporting Information Figures S5 and S6). B(L10F/M36I/N88S) also creates fewer hydrogen bonds between NFV and D30 and causes conformational alteration at D30. The polymorphism M36I reduces the contribution of D30 to the binding with NFV (Supporting Information Figure S7). Our simulations suggest that N88S in subtype B PR reduces the binding affinity with NFV when it appears together with M36I.

It is interesting that both D30N and N88S confer resistance against NFV by decreasing the interaction between the 30th residue and NFV. Both D30N and N88S affect the active site residues around the 30th residue. Other active site residues hardly change their interaction with NFV or their conformations. As can be seen in Figure 2 and Figure S4, the NFV-resistant PRs (B(D30N), AE(N88S), B(M36I/N88S), and B(L10F/M36I/N88S)) each show an increase in distance between the 30th residue and NFV. The NFV-resistant N88S mutants (AE(N88S), B(M36I/N88S), and B(L10F/M36I/N88S)) each have a stable direct hydrogen bond between the side chain of S88 and the side chain of D30. Therefore, N88S does not appear simultaneously with D30N clinically.

Prior to the MD simulations, we reconsidered torsional force field parameters for the benzamide moiety in NFV. This moiety has essential hydrogen bonds with D30 of HIV-1 PR.[22,42] Thus, those torsional parameters are expected to greatly affect the results of the simulations. Nevertheless, the AMBER ff03[43] and general AMBER force fields[44] cause a much higher energy barrier around the rotatable bond between the benzene and amide groups in benzamide than that based on quantum chemical calculations (Supporting Information Figure S1). This was a serious problem for our simulations. Therefore, we improved the torsional force field parameters for the benzamide moiety in NFV by fitting them to the energy curve obtained from quantum chemical calculations. Our newly developed parameters enabled us to carry out precise simulations of HIV-1 PR in complex with NFV.

In this study, we not only proposed the mechanism of resistance against NFV of N88S in subtype AE PR but also examined the influence of the polymorphisms in subtype AE PR on the emergence rates of D30N and N88S mutations. N88S and the polymorphisms in subtype AE PR are all classified as nonactive site mutations. Nevertheless, these mutations affect the binding of NFV. We and other groups have reported that the nonactive site mutations affect the binding affinity of some inhibitors, the emergence rate of mutants, and the catalytic activity of the protease.[17,22,33,36,39,41,47-50] For example, the polymorphisms in subtype C HIV-1 enhance the catalytic efficacy. However, there have been few studies on the influence of nonactive site mutations from structural viewpoints. There have also been few studies on the differences between HIV-1 subtypes. Clarification of the roles of nonactive site mutations and polymorphisms will enable us to design potent drugs, since the currently available PIs were developed and tested only against subtype B PRs. Accumulation of data on the susceptibilities of nonsubtype B viruses to the currently available PIs is also needed in order to establish an effective HIV-1 therapy strategy. Clarification of these susceptibilities will also be useful for selecting more appropriate drugs for patients.