and stabilizes PrP^C, blocked the conversion into PrP^{Sc}. The amino acid identity and similarity between the bovine and mouse PrPs are high, 89 and 93%, respectively. Moreover, the two sites of bPrP that are responsible for binding of R12 are located in the N-terminal unfolded region, as described later in the text, and the amino acid sequences of these two sites are almost identical for bovine and mouse PrPs. Therefore, it is likely that R12 selected as to bPrP also functioned for mouse PrP. Moreover, the amino acid sequences of these two sites for human PrP are also almost identical to those for bovine and mouse PrPs. There is the possibility that R12 may function for human PrP.

Recently, it has been reported that the segment 100–104 of mouse PrP^C plays a key role in the conversion after binding to mouse PrP^{Sc} (28). This segment corresponds to the 112–116 segment of bPrP, which is actually the K5 to P9 segment of a P16 peptide used in this work. Therefore, our observation that R12, which tightly binds to P16 (and P1) of bPrP, reduces the PrP^{Sc} level is consistent with this report.

It has been revealed this time that the unique R12 structure observed in a free form is preserved also in a complex form with bPrP. The structure determination of the complex with the binding peptide of bPrP has provided the detailed information on the interaction. It has been identified for the first time that the phosphate groups of G5, A9 and G11 of R12 are involved in the electrostatic interaction. Pairing for the interaction with one of three Lys residues of P16 has also been found. The structure has also revealed that the guanine base that makes stacking interaction with a Trp residue of P16 is that of G8 of R12. These kinds of information cannot be obtained without the structure determination of the complex. Then, the high affinity of R12 has been attributed to the simultaneous dual binding. For therapeutic application of R12 in future, chemical modification of either sugars or backbone may be required to increase the RNase resistance. In this case, however, the chemical modifications on R12 should not interfere with the R12–PrP interactions. The information on the interactions obtained from both the detailed complex structure and the overall architecture must be helpful to determine the position and the kind of modification.

We previously reported the structure of D12 (29). D12 forms a similar quadruplex structure to R12. In addition to the common G:G:G:G tetrad, the G(:A):G(:A):G(:A):G heptad is formed for the D12 quadruplex, instead of the G(:A):G:G(:A):G hexad for the R12 quadruplex. D12 also forms a similar dimer structure to R12, although the direction of the dyad axis differs between the D12 and R12 dimers. Therefore, it is likely that D12 interacts with bPrP in a similar way to R12. Some structural differences between the R12 and D12 quadruplexes may be responsible for the difference in affinity to bPrP. The lower anti-prion activity of D12 is supposed to be because of the lower affinity to and less stabilization of $bPrP^C$. Nonetheless, the finding of anti-prion activity for D12 is still interesting.

Compounds that inhibit the accumulation of PrP^{Sc} in prion-infected cells have been developed. Among them, GN8 binds to the hot spots in the C-terminal region of PrP (14), and a lysosomotropic agent, quinacrine, to the third helix in the C-terminal region of PrP (13). Pentosan polysulphate is reported to bind to the octapeptide repeats in the N-terminal region of PrP (15). Recently, Congo red was reported to bind to the lysine residue in the groove formed in Heterokaryon incompatibility protein s (HET-s) amyloid (17). The site and mode of interaction of R12 with PrP are different from those of these compounds. Therefore, the modulation of PrP conformational conversion by R12, and probably by D12 also, would be distinct from that by pre-existing anti-prion agents.

Recently, PrP^C has also been suggested to be a mediator of the amyloid-β-oligomer-induced synaptic dysfunction linked to Alzheimer's disease (30). It was revealed that anti-PrP antibodies prevent amyloid-β-oligomer binding to PrP^C, and that the antibodies rescue synaptic plasticity in hippocampal slices from oligomeric amyloid-β. An RNA aptamer also tightly binds to PrP^C like antibodies; hence, the RNA aptamer may also rescue synaptic plasticity by preventing amyloid-β-oligomer binding to PrP^C. Therefore, an RNA aptamer for PrP^C may have therapeutic potential not only for prion diseases but also for Alzheimer's disease.

In conclusion, we have demonstrated that a short RNA aptamer for PrP^C that comprises only 12 residues, R12, exhibits anti-prion activity in mouse neuronal cells persistently infected with the TSE agent. There is the possibility that R12 is also effective for bovine spongiform encephalopathy. The structural basis of the tight binding of R12 to PrP^C has been elucidated. These findings may be used to develop RNA aptamer-based drugs against the prion and Alzheimer's diseases.

ACCESSION NUMBERS

Protein Data Bank: Coordinates and restraints for the final ensemble of 10 structures of the R12:P16 complex have been deposited under the accession code 2rsk.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

The authors thank M. Imamura and T. Yokoyama for giving them the plasmid containing the bPrP gene, and T. Saeki for the technical help in evaluation of the anti-prion activity.

FUNDING

Ministry of Education, Culture, Sports, Science and Technology in Japan [23657072, 24121714 to M.K.; 23570146, 24113710 to T.N.]; JST [SENTAN and CREST]; Sumitomo-Denko Foundation; Iwatani Foundation. JSPS Research Fellow Program (to M.T.).

Funding for open access charge: Ministry of Education, Culture, Sports, Science and Technology in Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Prusiner,S.B. (1998) Prions. *Proc. Natl Acad. Sci. USA*, **95**, 13363–13383.
2. Pan,K.M., Baldwin,M., Nguyen,J., Gasset,M., Serban,A., Groth,D., Mehlhorn,I., Huang,Z., Fletterick,R.J., Cohen,F.E. *et al.* (1993) Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl Acad. Sci. USA*, **90**, 10962–10966.
3. Prusiner,S.B., Scott,M.R., DeArmond,S.J. and Cohen,F.E. (1998) Prion protein biology. *Cell*, **93**, 337–348.
4. Prusiner,S.B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science*, **216**, 136–144.
5. Prusiner,S.B. (1991) Molecular biology and transgenetics of prion diseases. *Crit. Rev. Biochem. Mol. Biol.*, **26**, 397–438.
6. Büeler,H., Aguzzi,A., Sailer,A., Greiner,R.A., Autenried,P., Aguet,M. and Weissmann,C. (1993) Mice devoid of PrP are resistant to scrapie. *Cell*, **73**, 1339–1347.
7. Sailer,A., Büeler,H., Fischer,M., Aguzzi,A. and Weissmann,C. (1994) No propagation of prions in mice devoid of PrP. *Cell*, **77**, 967–968.
8. Cashman,N.R. and Caughey,B. (2004) Prion diseases—close to effective therapy? *Nat. Rev. Drug. Disc.*, **3**, 874–884.
9. Mallucci,G. and Collinge,J. (2005) Rational targeting for prion therapeutics. *Nat. Rev. Neurosci.*, **6**, 23–34.
10. Sim,V.L. and Caughey,B. (2009) Recent advances in prion chemotherapeutics. *Infect. Disord. Drug Targets*, **9**, 81–91.
11. Proske,D., Gilch,S., Wopfner,F., Schätzl,H.M., Winnacker,E.L. and Famulok,M. (2002) Prion-protein-specific aptamer reduces PrP^{Sc} formation. *Chembiochem*, **3**, 717–725.
12. Rhie,A., Kirby,L., Sayer,N., Wellesley,R., Disterer,P., Sylvester,I., Gill,A., Hope,J., James,W. and Tahiri-Alaoui,A. (2003) Characterization of 2′-fluoro-RNA aptamers that bind preferentially to disease-associated conformations of prion protein and inhibit conversion. *J. Biol. Chem.*, **278**, 39697–39705.
13. Vogtherr,M., Grimme,S., Elshorst,B., Jacobs,D.M., Fiebig,K., Griesinger,C. and Zahn,R. (2003) Antimalarial drug quinacrine binds to C-terminal helix of cellular prion protein. *J. Med. Chem.*, **46**, 3563–3564.
14. Kuwata,K., Nishida,N., Matsumoto,T., Kamatari,Y.O., Hosokawa-Muto,J., Kodama,K., Nakamura,H.K., Kimura,K., Kawasaki,M., Takakura,Y. *et al.* (2007) Hot spots in prion protein for pathogenic conversion. *Proc. Natl Acad. Sci. USA*, **104**, 11921–11926.
15. Taubner,L.M., Bienkiewicz,E.A., Copie,V. and Caughey,B. (2010) Structure of the flexible amino-terminal domain of prion protein bound to a sulfated glycan. *J. Mol. Biol.*, **395**, 475–490.
16. Kimura,T., Hosokawa-Muto,J., Kamatari,Y.O. and Kuwata,K. (2011) Synthesis of GN8 derivatives and evaluation of their antiprion activity in TSE-infected cells. *Bioorg. Med. Chem. Lett.*, **21**, 1502–1507.
17. Schütz,A.K., Soragni,A., Hornemann,S., Aguzzi,A., Ernst,M., Böckmann,A. and Meier,B.H. (2011) The amyloid-Congo red interface at atomic resolution. *Angew. Chem. Int. Ed. Engl.*, **50**, 5956–5960.
18. Murakami,K., Nishikawa,F., Noda,K., Yokoyama,T. and Nishikawa,S. (2008) Anti-bovine prion protein RNA aptamer containing tandem GGA repeat interacts both with recombinant prion protein and its β isoform with high affinity. *Prion*, **2**, 73–78.
19. Nishikawa,F., Murakami,K., Matsugami,A., Katahira,M. and Nishikawa,S. (2009) Structural studies of an RNA aptamer containing GGA repeats under ionic conditions using microchip electrophoresis, circular dichroism, and 1D-NMR. *Oligonucleotides*, **19**, 179–190.
20. Mashima,T., Matsugami,A., Nishikawa,F., Nishikawa,S. and Katahira,M. (2009) Unique quadruplex structure and interaction of an RNA aptamer against bovine prion protein. *Nucleic Acids Res.*, **37**, 6249–6258.
21. García,F.L., Zahn,R., Riek,R. and Wüthrich,K. (2000) NMR structure of the bovine prion protein. *Proc. Natl Acad. Sci. USA*, **97**, 8334–8339.
22. Nishida,N., Harris,D.A., Vilette,D., Laude,H., Frobert,Y., Grassi,J., Casanova,D., Milhavet,O. and Lehmann,S. (2000) Successful transmission of three mouse-adapted scrapie strains to murine neuroblastoma cell lines overexpressing wild-type mouse prion protein. *J. Virol.*, **74**, 320–325.
23. Brünger,A.T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR.* New Haven, Yale University Press.
24. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
25. Furukawa,A., Nagata,T., Matsugami,A., Habu,Y., Sugiyama,R., Hayashi,F., Kobayashi,N., Yokoyama,S., Takaku,H. and Katahira,M. (2009) Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.*, **28**, 440–451.
26. Wüthrich,K. (1986) *NMR of Protein and Nucleic Acids.* John Wiley & Sons, New York, NY.
27. Campisi,D.M., Calabro,V. and Frankel,A.D. (2001) Structure-based design of a dimeric RNA-peptide complex. *EMBO J.*, **20**, 178–186.
28. Hara,H., Okemoto-Nakamura,Y., Shinkai-Ouchi,F., Hanada,K., Yamakawa,Y. and Hagiwara,K. (2012) Mouse prion protein (PrP) segment 100 to 104 regulates conversion of PrP^{C} to PrP^{Sc} in prion-infected neuroblastoma cells. *J. Virol.*, **86**, 5626–5636.
29. Matsugami,A., Ouhashi,K., Kanagawa,M., Liu,H., Kanagawa,S., Uesugi,S. and Katahira,M. (2001) An intramolecular quadruplex of $(GGA)_4$ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
30. Laurén,J., Gimbel,D.A., Nygaard,H.B., Gilbert,J.W. and Strittmatter,S.M. (2009) Cellular prion protein mediates impairment of synaptic plasticity by amyloid-β oligomers. *Nature*, **457**, 1128–1132.

Selected Papers

A Minimal Implementation of the AMBER-PAICS Interface for Ab Initio FMO-QM/MM-MD Simulation

Takuya Okamoto,[1] Takeshi Ishikawa,[2,†] Yoshiyuki Koyano,[1] Norifumi Yamamoto,[1,††] Kazuo Kuwata,[2] and Masataka Nagaoka*[1]

[1]Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601

[2]Division of Prion Research, Center for Emerging Infectious Disease, Gifu University, 1-1 Yanagido, Gifu 501-1194

Received August 9, 2012; E-mail: mnagaoka@is.nagoya-u.ac.jp

For the purpose of providing a realistic description of the reaction mechanisms in large molecular systems, we propose a quantum mechanical/molecular mechanical (QM/MM) method combined with the ab initio fragment molecular orbital (FMO) method, i.e., the ab initio FMO-QM/MM method. By connecting a molecular dynamics (MD) program AMBER with an FMO program PAICS, we have implemented an AMBER-PAICS interface (AP-IF). Using the AP-IF, we demonstrate three example applications: (a) a hydrogen fluoride and water molecular clusters, (b) an alanine dipeptide in aqueous solution, and (c) a prion protein–GN8 complex. From these results, it is confirmed that the FMO-QM/MM method offers a good compromise between chemical accuracy and computational cost and enables us to obtain in ab initio quality the inter- and intramolecular interaction energies between molecules or residues in large molecular systems such as solution and biomolecule, by using the dynamics-based interfragment interaction energy (IFIE) analysis.

In conjunction with the incremental technological advancements of computer power, molecular dynamics (MD) simulations of phenomena occurring in complex and heterogeneous environments in condensed phase are in very high demand in the field of computational chemistry. For the purpose of providing a realistic description of the reaction mechanisms, however, it is still challenging to execute ab initio quantum-mechanics (QM)-based MD simulation of chemical or biological systems in the condensed phase, where making and breaking chemical bonds results in significant changes to the electronic structure of solute molecules or active sites.[1] This is because the applications of the conventional ab initio molecular orbital (MO) methods to large molecular systems suffer from the rapid growth of computational time and resources scale with the system size N. The scaling value m of $O(N^m)$ for large molecular systems are typically 3 to 4 for Hartree–Fock (HF) or density functional theory (DFT), 5 and 6 for Møller–Plesset second-order perturbation (MP2) theory and coupled cluster singles and doubles (CCSD), respectively. CCSD with non-iterative triples (CCSD(T)) provides excellent results, but it is very time consuming even for medium-size molecules due to its unfavorable cost scaling of $m = 7$ without parallelization. Furthermore, in the MD simulation that is performed by integrating numerically Newton's equation of motion, a vast number of MO single-point calculations are required to obtain the energies and gradients of the whole system. Even for an isolated molecular cluster, therefore, the direct MD simulation retaining the QM description of the whole system has been until now limited to time scales of several tens or hundreds of picoseconds.

For dealing with large molecular systems, some studies on fragment-based methods have been well-established and opened up an active field of research.[2] One is fragment molecular orbital (FMO) method,[3,4] which has been developed by Kitaura and co-workers since 1999. It is a closely related method to the explicit polarization (X-Pol) method,[5–7] which was initially proposed by Gao in 1997, with a similar idea, an identical fragmentation, and an identical self-consistent field (SCF) optimization procedure. By including correction terms of the fragment pairs and trimers, the FMO method can reproduce ab initio properties while showing an advantage of scaling to $O(N^2)$ or less. It can be used with various theoretical levels: restricted HF (RHF), DFT, MP2, multiconfigurational self-consistent field (MCSCF), configuration interaction with single excitations (CIS), and coupled cluster (CC) theory. ABINIT-MP,[8,9] GAMESS/FMO,[10–12] NWChem/FMO,[13,14] and PAICS,[15,16] are available as FMO calculation programs. As an application to MD simulation, the FMO-MD method developed by Komeiji et al.[17,18] was implemented by combining an MD program PEACH[19] with an FMO program ABINIT-MP or the latest version termed ABINIT-MPX,[20] and applied to several solvated molecular systems.[21–28] To our knowledge, however, little progress has been made on the FMO-MD simulation of a biomolecular system, since direct ab initio QM-MD

† Present address: Graduate School of Engineering, Tohoku University, 6-6 Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-8579

†† Present address: Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016

Published on the web February 6, 2013; doi:10.1246/bcsj.20120216

calculation of such a system demands great computational cost even with the FMO method.

On the other hand, multilevel treatments such as the quantum mechanical/molecular mechanical (QM/MM) method[29–32] and our own *n*-layered integrated molecular orbital and molecular mechanics (ONIOM) method[33,34] have been applied successfully to large molecular systems due to computational efficiency to afford expensive configuration sampling. In the QM/MM method, the chemically active region of primary interest is treated quantum mechanically, whereas the surroundings are described by a simple empirical molecular mechanics model. The QM/MM method requires some choices such as the theoretical level and the basis set of QM region, the force field of MM region, and the representation of QM/MM interaction terms (e.g., calibration of Lennard–Jones (LJ) parameter set).[35,36] In the case of covalently bonded systems, the selection of QM region necessarily results in the breaking of covalent bonds, which ameliorate by incorporating link atoms on the boundaries between QM and MM regions.[30]

There are some program packages that can execute ab initio QM/MM-MD simulation using build-in modules, i.e., CPMD/GROMACS,[37,38] CP2K,[39–41] and GAMESS/Tinker.[10,42,43] Depending on the intended use, any combination of the characteristic modules of already existing MO and MD programs can be selected by using a general-purpose interface program, i.e., ChemShell[44,45] and PUPIL.[46–48] On the other hand, an efficient approach to perform ab initio QM/MM-MD simulation is to connect semiempirical QM/MM-MD programs, such as AMBER[49] and CHARMM,[50] with ab initio MO programs. For applying to a number of theoretical methodologies based on the ab initio QM/MM-MD method, we provided an implementation of an AMBER-Gaussian interface (AG-IF)[51] connecting an MD program AMBER9[52,53] with an MO program Gaussian 03,[54] such that changes to the original code of AMBER program are minimal by adding a few subroutines to execute the MO program and transfer information on energies and forces.

Recently, ab initio QM/MM simulations of some condensed-phase systems have demonstrated the significant role of many-body effects within QM region, proving inadequacies of effective pairwise potential functions in the MM force field. From ab initio QM/MM-MD simulation of Li^+ ion in 18.4% aqueous ammonia solution, Tongraar and Rode showed that the ligand composition in the first solvation shell depends strongly on the size of QM region.[55] The composition of $Li^+[(H_2O)_3NH_3][(H_2O)_4(NH_3)_2]$ is obtained by using the QM region corresponding to the size of the first solvation shell, while the use of larger QM region leads to a clear water preference with that of $Li^+[(H_2O)_4][(H_2O)_4]$. Sumowski and Ochsenfeld investigated systematically how large the QM region needs to be chosen for a reliable description of the energetics for some peptidic systems by using the QM/MM calculations combined with linear-scaling QM methods both at the HF and atomic orbital-based MP2 (AO-MP2) levels and it was confirmed that enlarging the QM region can improve significantly the chemical accuracy of the QM/MM results.[56]

Under the circumstances, in this study, we present a QM/MM treatment based on the ab initio FMO method, i.e., the ab initio FMO-QM/MM method. As an application of the ab initio FMO-QM/MM method, Nagata et al. have already proposed the FMO/effective fragment potential (EFP) method,[57,58] which is still expensive. For the construction of one-electron terms in the effective QM/MM Hamiltonian, it would be practical to employ the MM atom-centered partial point charges, since it can obtain an improved trade-off between chemical accuracy and computational cost. To perform direct FMO-QM/MM-MD simulations of large molecular systems, we implement an AMBER-PAICS interface (AP-IF) combined with an FMO program PAICS.[15,16] As it is for the AG-IF,[51] it would enable us to develop and apply easily new theoretical approaches based on the ab initio QM/MM method to large molecular systems. Actually, we have applied the AG-IF to the isomerization of glycine in aqueous solution,[59] studied by the ab initio QM/MM free energy gradient (FEG) method, and the hydration dynamics of methanol in aqueous solution via the charges from interaction energy and forces (CHIEF) approach.[60]

This article is organized as follows. First, we explain briefly the original FMO method and the FMO-QM/MM method. Next, we present the implementation of the interfacing program AP-IF in the section "Implementation." In the section "Applications," we demonstrate three example applications: (a) a hydrogen fluoride and water molecular clusters, (b) an alanine dipeptide in aqueous solution, and (c) a prion protein (PrP)–GN8 complex in aqueous solution. Finally, concluding remarks are given in the last section.

Methods

FMO Method. We now describe briefly an outline of the FMO method.[3,4] In the method, the whole system is divided into N_f fragments and the electrons are assigned to each fragment. With the two-body FMO approximation (FMO2), only calculations of the fragments (refer to as monomers) and pair of the fragments (refer to as dimers) are required for evaluating those properties of the whole system.

The total energy E_{tot} of the whole system is represented as

$$E_{\mathrm{tot}} = \sum_{I}^{N_f} E'_I + \sum_{I<J}^{N_f} (E'_{IJ} - E'_I - E'_J) + \sum_{I<J}^{N_f} \mathrm{Tr}(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \quad (1)$$

where E'_X for $X = I$ (or $X = IJ$) represents the internal energy of the monomer (or dimer) within the total system, polarized by the environment, but without the energy of the environment. In eq 1, $\Delta \mathbf{D}^{IJ}$ and $\mathbf{V}^{IJ}$ are the dimer density matrix difference and the matrix form of the environmental electrostatic potential for dimer IJ, respectively. The matrix element of the environmental electrostatic potential $\mathbf{V}^{IJ}$ is

$$V^{IJ}_{\mu\nu} = \sum_{K \neq IJ} (u^K_{\mu\nu} + v^K_{\mu\nu}) \quad (2)$$

In eq 2, the one-electron and two-electron integrals of monomer $K \neq IJ$ are given, respectively, as

$$u^K_{\mu\nu} = \sum_{A \in K} \langle \mu | \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} | \nu \rangle \quad (3)$$

$$v^K_{\mu\nu} = \sum_{\lambda\sigma \in K} D^K_{\lambda\sigma} (\mu\nu|\lambda\sigma) \quad (4)$$

where $D^K_{\lambda\sigma}$ is the density matrix element of monomer K and $(\mu\nu|\lambda\sigma)$ is the two-electron integral in terms of basis functions. The dimer density matrix difference $\Delta \mathbf{D}^{IJ}$ is defined by

$$\Delta\mathbf{D}^{IJ} = \mathbf{D}^{IJ} - (\mathbf{D}^{I} \oplus \mathbf{D}^{J}) \tag{5}$$

where $\mathbf{D}^{I(/J)}$ is the monomer density matrix of fragment $I(/J)$, $\mathbf{D}^{IJ}$ being the dimer density matrix of fragment pair IJ. In eq 1, the internal energy E'_X has the following form,

$$E'_X = \sum_{\mu\nu\in X} D^X_{\mu\nu} h^X_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma\in X}\left[D^X_{\mu\nu}D^X_{\lambda\sigma} - \frac{1}{2}D^X_{\mu\lambda}D^X_{\nu\sigma}\right](\mu\nu|\lambda\sigma) + \sum_{\mu\nu\in X} D^X_{\mu\nu}P^X_{\mu\nu} + E^{\mathrm{nuc}}_X \tag{6}$$

where $h^X_{\mu\nu}$ is the one-electron Hamiltonian matrix element and the repulsion energy between nuclei E^{nuc}_X is represented by the following equation,

$$E^{\mathrm{nuc}}_X = \sum_{\substack{A<B\\ AB\in X}} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{7}$$

where $Z_{A(/B)}$ is the electric charge on the $A(/B)$th atomic nucleus located at a position $\mathbf{R}_A(/\mathbf{R}_B)$. The third term of eq 6 is the hybrid orbital projection (HOP) contribution,

$$\sum_{\mu\nu\in X} D^X_{\mu\nu}P^X_{\mu\nu} = \sum_{i\in X} 2\langle i|\hat{P}^X|i\rangle \tag{8}$$

The HOP operator $\hat{P}^X$ [61] is defined by

$$\hat{P}^X = \sum_{k\in X} B_k |\theta^X_k\rangle\langle\theta^X_k| \tag{9}$$

where $|\theta^X_k\rangle$ is the kth hybrid orbital and B_k is the universal constant. For example, the HOP operator for the fragmentation across a covalent C–C bond is based on an sp^3 orbital of a methane with C–H bond length of 1.09 Å.

The total energy in eq 1 can be rewritten as the following form,[9]

$$E_{\mathrm{tot}} = \sum_{I}^{N_f} E'_I + \sum_{I<J}^{N_f} \Delta E_{IJ} \tag{10}$$

$$\Delta E_{IJ} \equiv (E'_{IJ} - E'_I - E'_J) + \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) \tag{11}$$

where ΔE_{IJ} can be interpreted as the two-body interaction energy between monomers I and J, and is called as the inter-fragment interaction energy (IFIE)[62,63] or the pair interaction energy (PIE).[64] In the FMO method, this value is often utilized in the interaction analysis.[65]

The two-electron integral calculation for the environmental electrostatic potential is one of the most time-consuming steps in the FMO method. To save computational time, we adopted Mulliken and fractional point charge (esp-aoc and esp-ptc) approximations,[9] which were originally proposed by Gao,[5] and an electrostatic interaction approximation for the dimer calculation (dimer-es approximation).[66] These approximations are based on the interfragment distance R_{IJ} defined by the shortest contact atoms between monomers I and J, and the threshold values L are given in unit of the van der Waals (vdW) radius of the shortest contact atoms between two fragments. According to the previous study of Nakano et al.,[9] we used the threshold values of $L_{\text{esp-aoc}} = 0.0$, $L_{\text{esp-ptc}} = 2.0$, and $L_{\text{dimer-es}} = 2.0$ for the esp-aoc, the esp-ptc, and the dimer-es approximations, respectively.

The analytical gradient of the total energy with respect to the Ath atom located at a position $\mathbf{R}_A$ $(= (x_A, y_A, z_A))$ is calculated as follows,[61]

$$\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_A} = \sum_{I}^{N_f}\frac{\partial E'_I}{\partial \mathbf{R}_A} + \sum_{I<J}^{N_f}\left(\frac{\partial E'_{IJ}}{\partial \mathbf{R}_A} - \frac{\partial E'_I}{\partial \mathbf{R}_A} - \frac{\partial E'_J}{\partial \mathbf{R}_A}\right) + \sum_{I<J}^{N_f}\mathrm{Tr}\left(\Delta\mathbf{D}^{IJ}\frac{\partial}{\partial \mathbf{R}_A}\mathbf{V}^{IJ}\right) \tag{12}$$

Recently, Nagata et al. have presented a detailed formulation of the FMO energy gradient[67] and have also reported the fully analytical gradient including the response terms due to the electrostatic potentials.[68]

As shown in the next section, it is necessary to calculate the electric field $\mathbf{E}_{\mathrm{ele}}$ exerted by the electron distribution. In the case of the FMO method, the electric field at a position $\mathbf{R}_M$ can be obtained from the following equation,

$$\mathbf{E}_{\mathrm{ele}}(\mathbf{R}_M) = \sum_{I}\mathrm{Tr}\left(\mathbf{D}^I\frac{\partial}{\partial \mathbf{R}_M}\mathbf{V}^I(\mathbf{R}_M)\right) + \sum_{I<J}\mathrm{Tr}\left(\Delta\mathbf{D}^{IJ}\frac{\partial}{\partial \mathbf{R}_M}\mathbf{V}^{IJ}(\mathbf{R}_M)\right) \tag{13}$$

where the matrix element of the electrostatic potential $\mathbf{V}^X$ generated by the electrons is defined by

$$V^X_{\mu\nu}(\mathbf{R}_M) = \langle\mu|\frac{1}{\mathbf{R}_M - \mathbf{r}}|\nu\rangle \tag{14}$$

FMO-QM/MM Method. In the QM/MM method,[29–32] the whole system is partitioned into the QM and MM regions. Depending on the treatment of the electrostatic interaction between these two regions, the QM/MM schemes can be divided into two groups, i.e., mechanical embedding (ME) and electric embedding (EE) schemes.[69] In the present study, all QM/MM calculations were performed by using the latter scheme.

In the EE scheme, the total energy of the whole system E_{tot} is described as follows,

$$E_{\mathrm{tot}} = E_{\mathrm{QM}} + E_{\mathrm{QM/MM}} + E_{\mathrm{MM}} \tag{15}$$

where E_{QM} and E_{MM} are the energies for QM and MM regions, respectively. For the case where there are no covalent bonds between the atoms of the QM and MM regions, the QM/MM interaction term $E_{\mathrm{QM/MM}}$ is defined as a sum of electrostatic and nonelectrostatic interaction term,

$$E_{\mathrm{QM/MM}} = E^{\mathrm{est}}_{\mathrm{QM/MM}} + E^{\mathrm{vdW}}_{\mathrm{QM/MM}} \tag{16}$$

The QM/MM nonelectrostatic interaction term in eq 16 is described as a sum of LJ type potential functions,

$$E^{\mathrm{vdW}}_{\mathrm{QM/MM}} = \sum_{A}\sum_{M}\varepsilon_{AM}\left\{\left(\frac{R^{\mathrm{eq}}_{AM}}{R_{AM}}\right)^{12} - 2\left(\frac{R^{\mathrm{eq}}_{AM}}{R_{AM}}\right)^{6}\right\} \tag{17}$$

where R_{AM} $(= |\mathbf{R}_A - \mathbf{R}_M|)$ is the distance between the Ath QM and the Mth MM atoms, and ε_{AM} and R^{eq}_{AM} are a couple of LJ parameters for the Ath QM atom interacting with the Mth MM atom. For the system that has covalent bonds between the QM and MM regions, the QM/MM boundaries can be treated by the link atom approach.[30,53]

In the FMO-QM/MM method, QM region is subdivided into N_f QM fragments. In performing the individual MO calcu-

lation for monomer (or dimer), each QM fragment is interacting not only with the remaining QM fragments but also with all of the MM point charges located within a certain QM cutoff radius. The total energy of the whole system (eq 15) is rewritten by using the SCF energy E_{SCF} as follows,

$$E_{\mathrm{tot}} = E_{\mathrm{SCF}} + E^{\mathrm{vdW}}_{\mathrm{QM/MM}} + E_{\mathrm{MM}} \tag{18}$$

In eq 18, the SCF energy E_{SCF} is described by

$$\begin{aligned} E_{\mathrm{SCF}} &= E_{\mathrm{QM}} + E^{\mathrm{est}}_{\mathrm{QM/MM}} \\ &= \sum_{I}^{N_f} E''_I \sum_{I<J}^{N_f} (E''_{IJ} - E''_I - E''_J) \sum_{I<J}^{N_f} \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) \\ &\quad + \sum_{I}^{N_f} \{\mathrm{Tr}(\mathbf{D}^I\mathbf{V}^I_{\mathrm{ext}}) + E_I^{\mathrm{nuc\text{-}ext}}\} + \sum_{I<J}^{N_f} \mathrm{Tr}(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}_{\mathrm{ext}}) \end{aligned} \tag{19}$$

where E''_X for $X = I$ (or $X = IJ$) is the monomer (or dimer) internal energy under the QM environmental electrostatic potential and the MM external electrostatic potential. In eq 19, $\mathbf{V}^X_{\mathrm{ext}}$ for $X = I$ (or $X = IJ$) is the matrix form of the MM external electrostatic potential for monomer (or dimer). The matrix element of the MM external electrostatic potential $\mathbf{V}^X_{\mathrm{ext}}$ is defined as

$$V^X_{\mathrm{ext},\mu\nu} = \sum_M \langle\mu| \frac{q_M}{|\mathbf{r} - \mathbf{R}_M|} |\nu\rangle \tag{20}$$

In eq 19, the interaction energy $E_I^{\mathrm{nuc\text{-}ext}}$ between QM nucleus of monomer I and MM external point charge is described as follows,

$$E_I^{\mathrm{nuc\text{-}ext}} = \sum_{A\in I}\sum_M \frac{Z_A q_M}{|\mathbf{R}_A - \mathbf{R}_M|} \tag{21}$$

where Z_A is the Ath QM nuclear charge located at a position $\mathbf{R}_A$, q_M being the Mth MM point charge located at a position $\mathbf{R}_M$.

In the FMO-QM/MM method, the analytical gradient of the total energy (eq 18) with respect to the Ath QM atom located at a position $\mathbf{R}_A$ is calculated as follows,

$$\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_A} = \frac{\partial E_{\mathrm{SCF}}}{\partial \mathbf{R}_A} + \frac{\partial E^{\mathrm{vdW}}_{\mathrm{QM/MM}}}{\partial \mathbf{R}_A} + \frac{\partial E_{\mathrm{MM}}}{\partial \mathbf{R}_A} \tag{22}$$

where the gradient of the SCF energy is defined as

$$\begin{aligned} \frac{\partial E_{\mathrm{SCF}}}{\partial \mathbf{R}_A} &= \sum_I^{N_f} \frac{\partial E''_I}{\partial \mathbf{R}_A} + \sum_{I<J}^{N_f} \left(\frac{\partial E''_{IJ}}{\partial \mathbf{R}_A} - \frac{\partial E''_I}{\partial \mathbf{R}_A} - \frac{\partial E''_J}{\partial \mathbf{R}_A} \right) \\ &\quad + \sum_{I<J}^{N_f} \mathrm{Tr}\left(\Delta\mathbf{D}^{IJ} \frac{\partial}{\partial \mathbf{R}_A} \mathbf{V}^{IJ} \right) \\ &\quad + \sum_I^{N_f} \left\{ \mathrm{Tr}\left(\mathbf{D}^I \frac{\partial}{\partial \mathbf{R}_A} \mathbf{V}^I_{\mathrm{ext}} \right) + \frac{\partial E_I^{\mathrm{nuc\text{-}ext}}}{\partial \mathbf{R}_A} \right\} \\ &\quad + \sum_{I<J}^{N_f} \mathrm{Tr}\left(\Delta\mathbf{D}^{IJ} \frac{\partial}{\partial \mathbf{R}_A} \mathbf{V}^{IJ}_{\mathrm{ext}} \right) \end{aligned} \tag{23}$$

On the other hand, the analytical gradient of the total energy and the SCF energy with respect to the Mth MM point charge q_M located at a position $\mathbf{R}_M$ are given as

$$\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{R}_M} = \frac{\partial E_{\mathrm{SCF}}}{\partial \mathbf{R}_M} + \frac{\partial E^{\mathrm{vdW}}_{\mathrm{QM/MM}}}{\partial \mathbf{R}_M} + \frac{\partial E_{\mathrm{MM}}}{\partial \mathbf{R}_M} \tag{24}$$

$$\frac{\partial E_{\mathrm{SCF}}}{\partial \mathbf{R}_M} = -q_M(\mathbf{E}_{\mathrm{ele}} + \mathbf{E}_{\mathrm{nuc}}) \tag{25}$$

where $\mathbf{E}_{\mathrm{ele}}$ is the electrostatic field generated by the QM electrons (eq 13). In eq 25, the electrostatic field $\mathbf{E}_{\mathrm{nuc}}$ generated by the QM nuclei is obtained from the following equation,

$$\mathbf{E}_{\mathrm{nuc}}(\mathbf{R}_M) = -\frac{\partial}{\partial \mathbf{R}_M} \left(\sum_I \sum_{A\in I} \frac{Z_A}{|\mathbf{R}_M - \mathbf{R}_A|} \right) \tag{26}$$

The negative gradients of the total energy, i.e., the forces, are used to solve the simultaneous Newton's equations of motion in the MD simulation.

In FMO-QM/MM simulation of a large molecular system such as protein, it would be inevitable to partition the whole system into QM and MM regions. As is the case in the typical QM/MM method, we can apply the link atom approach[30,53] to the FMO-QM/MM method.

Implementation

We present here an AMBER-PAICS interface (AP-IF) to connect an MD program AMBER9 with an FMO program PAICS[15,16] parallelized with the message-passing interface (MPI).[70] The main MD engine of AMBER, which is called as SANDER module, consists of Program main "`multisander`," MD main "`sander`," Integrator subroutine "`runmd`," Forces subroutine "`force`," QM/MM forces subroutine "`qm_mm`," and other subroutines (Figure 1). As shown in the central flowchart of Figure 1, SCF energy and QM-QM and QM-MM forces are presented in several independent routines called by the subroutine "`qm_mm`." In the original code of AMBER program, there exist routines for the QM/MM calculation where the energies and forces for the QM part can be derived from a semiempirical method or density functional tight-binding (DFTB) method. Thus, as pointed out in our article of the AG-IF,[51] we can make it easy to implement a QM/MM interface by adding those routines to execute an external MO program and transfer information on energies and forces. The latest version of AMBER (AMBER12) could be also used as an MD engine in the AP-IF, since the fundamental architecture of AMBER12 remains unchanged as that of AMBER9.

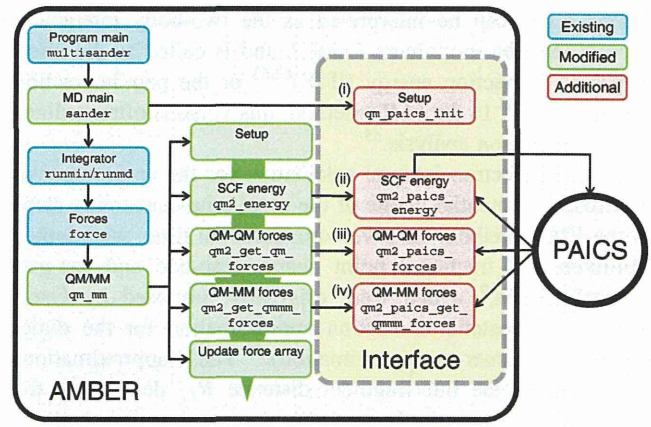


Figure 1. Schematic view of AMBER, AP-IF, and PAICS, including the flowcharts of the QM/MM subroutine "`qm_mm`" of AMBER (left center) and our developed AP-IF (right center). In the AP-IF, four subroutines (i)–(iv) were added to execute PAICS program and transfer information on energies and forces.

In the present study, the following 4 subroutines are provided for implementing of the AP-IF (Figure 1).

(i) Subroutine `qm_paics_init`: specify PAICS keywords and variables.

(ii) Subroutine `qm2_paics_energy`: execute PAICS program and update energy.

(iii) Subroutine `qm2_paics_forces`: update force exerted on QM atoms.

(iv) Subroutine `qm2_paics_get_qmmm_forces`: update force exerted on MM atoms.

Prior to an FMO-QM/MM-MD simulation, we should prepare a PAICS reference file "`paics.ref`" containing the FMO input keywords and variables, since there is no information concerning the FMO calculation in AMBER input and topology files. By these keywords and variables in "`paics.ref`," we can define detailed information concerning the fragmentation, e.g., the net charges on each fragment, the atom numbers assigned to each fragment, and the atom numbers of the detached covalent bonds. Also, we can request the theoretical level, the basis set on each atom, the approximations related to the environmental electrostatic potentials, and the number of CPU cores used in the MPI parallel calculation for each fragment monomer or dimer. Subroutine "`qm_paics_init`" reads a PAICS reference file "`paics.ref`," and corrects, if necessary, the information concerning the fragmentation.

As shown in Figure 1, the subroutines (ii)–(iv) are called by Subroutine "`qm2_energy`," "`qm2_forces`," and "`qm2_get_qmmm_forces`," respectively. In Subroutine "`qm2_paics_energy`," a PAICS input file is created for the instantaneous configuration generated at each MD time step. Then, in addition to the above-mentioned PAICS keywords and variables, the Cartesian coordinates of QM atoms and the Cartesian coordinates and point charges of MM atoms around the QM region are written in the PAICS input file. After executing the MPI run of PAICS through the system call, the SCF energy $E_{\mathrm{SCF}}^{\mathrm{PAICS}}$ (eq 19), the QM gradients $\nabla E_{\mathrm{SCF}}^{\mathrm{PAICS}}$ (eq 23), and the electric field $\mathbf{E}_{\mathrm{ele}}^{\mathrm{PAICS}}$ from the QM electrons (eq 13) can be obtained from the PAICS output file. If necessary, the QM Mulliken charges can be required. The QM forces $\mathbf{F}_{\mathrm{QM}}^{\mathrm{AMBER}}$ and the MM electrostatic forces $\mathbf{F}_{\mathrm{MM}}^{\mathrm{AMBER}}$ on an MM point charge exerted by the QM atoms are updated in Subroutine "`qm2_paics_forces`" and "`qm2_paics_get_qmmm_forces`," respectively. The relations between AMBER internal variables and PAICS output ones are as follows,

$$E_{\mathrm{SCF}}^{\mathrm{AMBER}} = E_{\mathrm{SCF}}^{\mathrm{PAICS}} \tag{27}$$

$$\mathbf{F}_{\mathrm{QM}}^{\mathrm{AMBER}} = -\frac{\partial E_{\mathrm{SCF}}^{\mathrm{PAICS}}}{\partial \mathbf{R}_A} \tag{28}$$

and

$$\mathbf{F}_{\mathrm{MM}}^{\mathrm{AMBER}} = q_M^{\mathrm{AMBER}}(\mathbf{E}_{\mathrm{ele}}^{\mathrm{PAICS}} + \mathbf{E}_{\mathrm{nuc}}) \tag{29}$$

where q_M^{AMBER} is the Mth MM point charge located at a position $\mathbf{R}_M$. The electric field $\mathbf{E}_{\mathrm{nuc}}$ exerted by QM nuclei in the second term of eq 29 is obtained in the AP-IF.

Here, the definition of fragments in QM region is shown briefly. Each QM fragment should form a closed shell so as not to break the bond electron pairs.[3] From a previous FMO study,[66] it is found that detaching a double C–C, triple C–C, or peptide C–N bond results in low accuracy. In general, the fragmentation of a polypeptide is suitable for detaching single C–C bonds. As seen in Figures 2a and 2b, it is noticeable that the segment positions of amino acid residues are different from those of minimal fragments. For a detached covalent C–C bond, the two electrons forming the bond are assigned to one fragment, and none to the other (Figure 2b). Thus, the bond fragmentation is performed at a C atom, which is called the bond detached atom (BDA). The other atom forming the C–C bond is assigned to the other fragment, and is referred to as the bond attached atom (BAA). The BDA keeps all other electrons except the one which is assigned to the BAA. The original HOP scheme[61,67] is applied in the present study. The details of fragmentation schemes used in the FMO method are described in Ref. 71. Since the fragmentation for a large molecule such as protein is complicated, we created a PAICS reference file "`paics.ref`" on a graphical user interface (GUI) software "PaicsView," which was designed to visualize the structure of a biomolecular system and to edit information on the fragmentation.

The AP-IF is designed on the assumption that the definition of the QM fragments remains unchanged during the MD simulation, which is called *static fragmentation*.[26] In some FMO-MD calculations of micro-hydrated cluster models, however, it was reported that a proton transfer between two H_2O molecules could sometimes take place during heating-up process.[21,23] By executing the "auto-fragmentation" program in PAICS, it is possible to divide solvent water molecules into appropriate fragments of some species, e.g., H_2O, H_3O^+, and OH^-, or their complexes.

If a large molecule is divided between QM and MM regions, an explicit link atom (hydrogen atom) is placed along the covalent bond that crosses the boundary between the QM and MM regions, at distance of 1.09 Å from the QM atom (Figure 2c).[53] The force exerted on the link atom is redistributed between the connecting QM and MM atoms. To avoid significant energy drift caused by the increase or decrease in the number of MM atoms within QM cutoff radius, we implemented not the atom-based QM cutoff but the residue-based one, which can keep the whole MM residue attribute even if only one of its MM atoms might exist within the QM cutoff radius.

Applications

In this section, we show three example applications; (a) a hydrogen fluoride and water molecular clusters, (b) an alanine dipeptide in aqueous solution, and (c) a prion protein (PrP)–GN8 complex. For applications (b) and (c), those intermolecular interaction energies between molecules or residues were investigated by combining the FMO-QM/MM-MD calculation and the IFIE analysis, which is one of the advantages of the FMO method.

Comparison of Computational Speed and Accuracy. The computational speed and accuracy were investigated by executing FMO and conventional MO calculations for molecular clusters of a hydrogen fluoride and $n-1$ water molecules, i.e., $HF(H_2O)_{n-1}$, for $n = 1$, 2, 5, 8, 17, 31, and 57. The theoretical level was adopted at the FMO2-HF/6-31G or HF/6-31G level. In the present FMO calculation, each molecule is

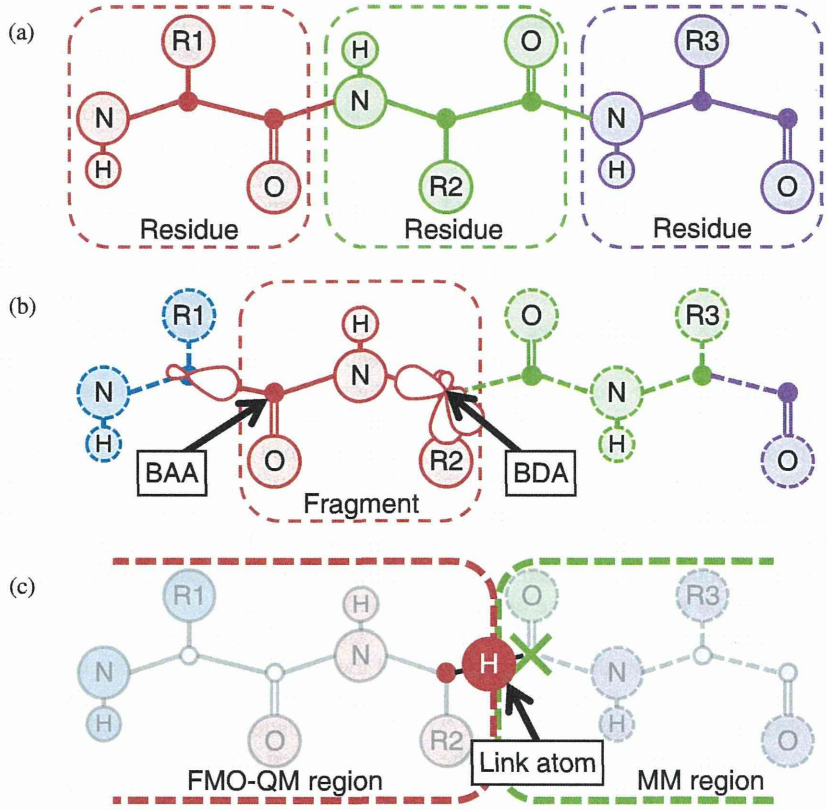


Figure 2. Schematic view of (a) amino acid residues and (b) minimal fragments in a model polypeptide and (c) a link atom between the QM and MM regions. Each residue or fragment is distinguished by different colors.

Table 1. Convergence of Total Energies of $HF(H_2O)_{n-1}$ Clusters ($n = 1$, 2, 5, 8, 17, 31, and 57) Obtained by FMO and Conventional MO Calculations

Number of molecules n	Total energy E/hartree		$\Delta E \times 10^3$ /hartree
	FMO calculation	MO calculation	
1	−99.9833946	−99.9833946	0.0000
2	−175.9853010	−175.9853010	0.0000
5	−403.9754709	−403.9767145	1.2436
8	−631.9350835	−631.9357563	0.6727
17	−1315.8822379	−1315.8812407	−0.9973
31	−2379.8446589	−2379.8257818	−18.8771
57	−4355.7860210	−4355.7399728	−46.0482

regarded as respective fragment monomer. To reduce calculation time, we can select parallel or concurrent MO calculation depending on the situation. For example, a parallel MO calculation using multiple CPU cores is suitable for a system or a large-size fragment that consists of a large number of atoms, while in an FMO calculation, MO calculations for the fragment monomer and the dimer can be performed concurrently as each independent task on respective CPU cores. All of the calculations were performed using the Intel Xeon processor (E3410 2.33 GHz) on 8 CPU cores.

Using PAICS, we computed FMO total energy and conventional MO one for each molecular cluster model. Table 1 shows the energies and their differences between these two calculations for cluster models. The total energies of $HF(H_2O)_{n-1}$ clusters for $n = 1$ and 2 correspond to the monomer energy E_I and the dimer one E_{IJ}, respectively. The total energies for $n = 5$, 8, and 17 are almost the same for the FMO and MO methods, in the range of 10^{-3} hartree. However, the energy differences are not negligible for $n = 31$ and 57. These poor results are due to well-known limitations of the FMO method with the two-body approximation. As previously discussed in some studies,[26,72,73] the FMO method can be, however, improved systematically by including higher-body terms and increasing the fragment size. Especially in the case of aqueous solution, energy difference from the conventional MO calculation can be reduced by about 1/10 with the three-body FMO approximation (FMO3). Figure 3 shows computational time required to perform the FMO and MO methods for these cluster models. In all of the clusters, the computational time is reduced drastically by applying the FMO method. In particular, the FMO calculations of the large-sized clusters for $n = 31$ and 57 can be more than 10 times faster than the corresponding MO calculations.

In addition, we investigate conservation of the total energy by using direct FMO-MD and MO-MD calculations of $HF(H_2O)_{n-1}$ cluster for $n = 5$. The MD simulations with two different time steps of 0.5 and 2.0 fs are executed in the *NVE* ensemble. With a time step of 2.0 fs, the SHAKE algorithm[74] is adopted to constrain the bond length. Figure 4 presents time variation of the total energy starting from the same initial configuration. It is found that the direct FMO-MD calculation in the *NVE* ensemble can conserve the total energy of the system, and the root mean square deviation (RMSD) of the

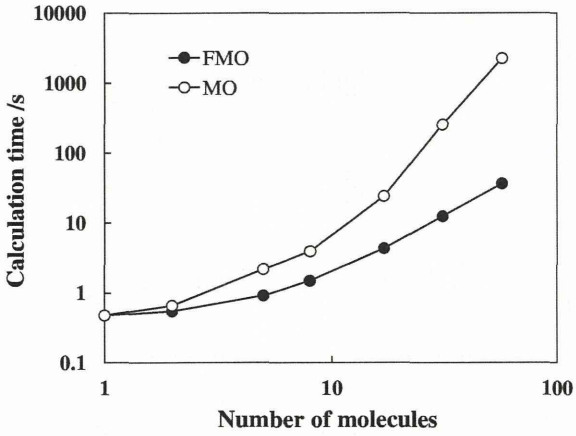


Figure 3. Comparison of computational time between the FMO calculation (closed circles: —●—) and the conventional MO one (open circles: —○—) of $HF(H_2O)_{n-1}$ cluster ($n = 1$, 2, 5, 8, 17, 31, and 57). The FMO2-HF/6-31G or HF/6-31G level of theory was applied to all molecular clusters.

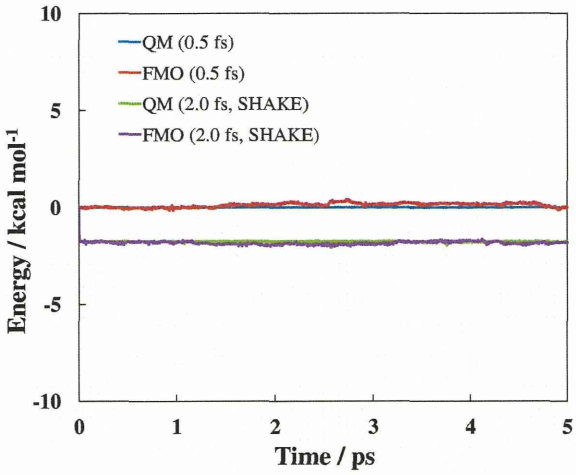


Figure 4. Time variation of total energies of $HF(H_2O)_{n-1}$ cluster for $n = 5$ obtained from the FMO-MD and MO-MD trajectories. The SHAKE algorithm was applied to geometric constraints to chemical bonds in the MD simulations with a time step of 2.0 fs.

total energy during the 5 ps FMO-MD calculation is within 0.1 kcal mol^{-1}, which is an acceptable value. With a relatively large time step of 0.5 or 2.0 fs, the total energy of the FMO-MD calculation shows a similar tendency with that obtained from the corresponding MO-MD trajectry. This is because the errors caused by the approximate energy gradient[67] are sufficiently small compared with those associated with the numerical integration scheme in the MD simulation.

As pointed out above, the inclusion of the higher-body terms is significant to treat for a large molecular system with many fragments. In an application of a chemical reaction process that takes place spontaneously on a short time scale, the fully analytical gradient including the response terms[68] ensures that the FMO-MD simulation can be carried out without the accumulation of errors as the time evolves. To expand the applications of our FMO-QM/MM interface, the FMO3

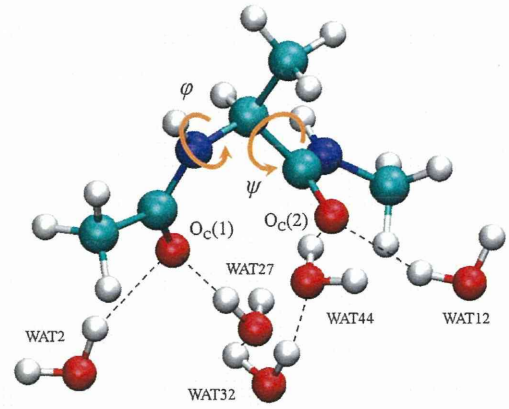


Figure 5. Typical snapshot structure of the alanine dipeptide in aqueous solution. Two peptide C=O groups of the alanine dipeptide are forming directly hydrogen bonds with four H_2O molecules (WAT2, WAT12, WAT27, and WAT44). The local backbone conformation of the alanine dipeptide was defined by two dihedral angles, φ and ψ.

approximation in the energy calculation[26,72,73] and the response term in the energy gradient[68] are going to be implemented in the near future, although they are not available in the current version of PAICS.

Based on the above results of $HF(H_2O)_{n-1}$ clusters, we consider here applications of the direct FMO-QM/MM-MD simulation to the other systems. In the QM/MM-MD simulation, the total computational time can be estimated roughly from the computational time required for the QM single-point calculation at each time step. Also, the forces exerted on the MM point charges are calculated by using the electric field from the QM nuclei and electrons and the time of calculating the electric field is negligible in comparison to that of a typical QM single-point calculation.[51] In modeling the aqueous solution system, the above FMO results of $HF(H_2O)_{n-1}$ clusters will help us to determine the appropriate size of the QM region. For example, to describe accurately solute–solvent many-body interactions, the size of the QM region can be chosen in such a way that the first and second hydration shells around a small solute molecule or ion might be included. For a biological system, it is to be noted that the size of each fragment must be relatively large, and MO calculations required for the monomers (or dimers) take longer computational time, compared to those of the fragment containing such a small molecule as HF or H_2O. Thus, it is important to check the sensitivity of the QM/MM results with respect to enlarging the QM region and dividing chemical bonds between QM fragments. In the following subsections, we will examine the FMO-QM/MM simulations of an alanine dipeptide and a PrP–GN8 complex as simple examples of solution system and biological one.

Alanine Dipeptide in Aqueous Solution. As an example of solution systems, we treated an alanine dipeptide (*N*-(*N*-acetyl-L-alanyl)methylamine) in aqueous solution. This system is known as one of the simplest models of protein backbones with two dihedral angles, φ and ψ (Figure 5). Past experimental results indicated that the polyproline II (PP_{II}) conformation is dominant in aqueous solution.[75–78] The relative populations of the different conformers obtained by using

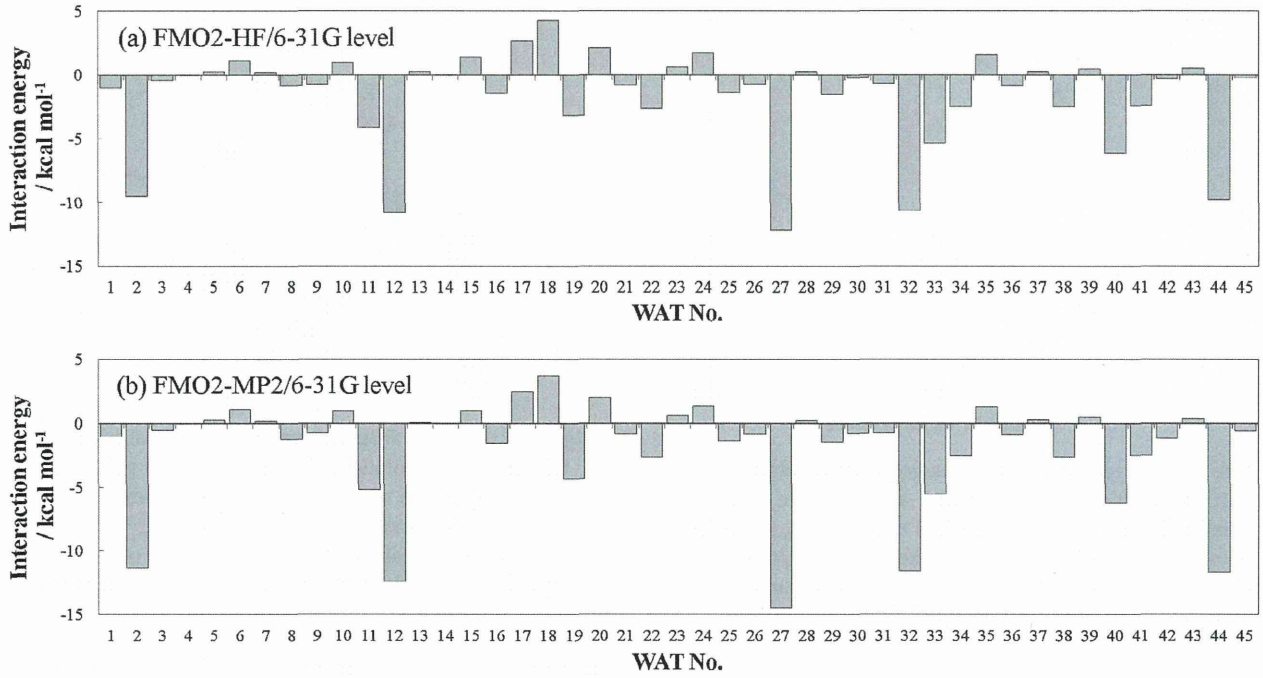


Figure 6. Averaged values of IFIEs of the alanine dipeptide with the solvent H_2O molecule calculated from the 0.5 ps MD simulation. All solvent H_2O molecules in the QM region are labeled from No. 1 to No. 45. The IFIE analyses were executed at (a) FMO2-HF/6-31G and (b) FMO2-MP2/6-31G levels of theory.

classical MD simulations depend strongly on the force fields employed, as summarized by Kwac et al.[79]

We investigated the stability of an alanine dipeptide in the PP_{II} conformation. In a cubic unit cell ($47.53 \times 47.53 \times 47.53$ Å^3) under the periodic boundary condition, 3510 solvent H_2O molecules were arranged around an alanine dipeptide. The initial configuration was prepared by energy minimization and classical MD simulation prior to FMO-QM/MM-MD. To execute direct FMO-QM/MM-MD simulations, we used two kinds of QM regions for this system. First, all molecules within 5 Å from any atoms of the alanine dipeptide were adopted as QM region and the number of the QM H_2O molecules is 45. The definition of the QM fragments remains unchanged during the QM/MM-MD simulation. Second, we considered a simple QM region made of the alanine dipeptide. In the QM fragmentation, the alanine dipeptide was regarded as a fragment monomer. The remaining MM H_2O molecules were described by the TIP3P model. The SHAKE algorithm[74] was used to constrain the covalent bonds involving hydrogen atoms. The MD time step was set to 2 fs. The QM cutoff radius of 9 Å was used for the electrostatic interaction between the QM and MM atoms. The direct FMO-QM/MM simulations were executed in the *NVT* ensemble at 300 K using the Berendsen thermostat.[80] A preliminary equilibration run of 3.5 ps was executed in order to conserve the total energy of the system and the total simulation time is set to 6 ps (= 3000 sampling steps). In the QM/MM-MD simulation, the energies and the gradients of the QM molecules were described at the FMO2-HF/6-31G level of theory. The IFIEs for the sampling structures from the MD simulation were calculated at the FMO2-HF/6-31G and FMO2-MP2/6-31G levels. All of the calculations were performed using the Intel Xeon processor (E3410 2.33 GHz) on 16 CPU cores.

For the large QM system, the IFIEs between the alanine dipeptide and the surrounding QM H_2O molecules are given as the averaged values during a 0.5 ps MD simulation (Figure 6). For 5 H_2O molecules, i.e., WAT2, WAT12, WAT27, WAT32, and WAT44, the averaged values of the IFIEs at the FMO2-HF/6-31G level (ca. -10 kcal mol^{-1}) are stronger than those of the other H_2O molecules. As can be seen from Figures 6a and 6b, the IFIEs observed at the FMO2-HF/6-31G level are quite similar to the ones at the FMO2-MP2/6-31G level. Thus, it should be noted that the interaction energies obtained at the HF level are meaningful in qualitative discussion.

A typical snapshot structure of the alanine dipeptide and these 5 H_2O molecules is illustrated in Figure 5. It is found that WAT2 and WAT27 interact with $O_C(1)$ atom while WAT12 and WAT44 interact with $O_C(2)$ atom, where O_C stands for the oxygen atom in two peptide C=O groups of the alanine dipeptide (Figure 5). Also, a *triple-water bridge* is observed between two peptide C=O groups, i.e., C=O···$(H_2O)_3$···O=C. We have calculated the radial distribution functions (RDFs) from a 2.5 ps trajectory data. In Figures 7a and 7b, shown are the RDFs $g(R)$s and the integrated coordination numbers (ICNs) $n(R)$s with respect to the lengths $R(O_C–H_W)$ and $R(O_C–O_W)$, where the oxygen and hydrogen atoms of solvent H_2O molecule are abbreviated to O_W and H_W, respectively. For the $O_C–H_W$ RDF, the first peak is observed at $R(O_C–H_W) = 1.8$ Å and the ICN up to the first minimum ($R(O_C–O_W) = 2.5$ Å) is 4.4 (Figure 7a). The first peak of the $O_C–H_W$ RDF reflects the formation of the direct hydrogen bonds with 4 H_2O molecules. The position and height of first peak in the $O_C–H_W$ RDF are similar to those of the PP_{II}-β conformation in the previous first principle DFT-based MD simulations.[81] For the $O_C–O_W$ RDF (Figure 7b), the first peak and minimum are located at $R(O_C–$

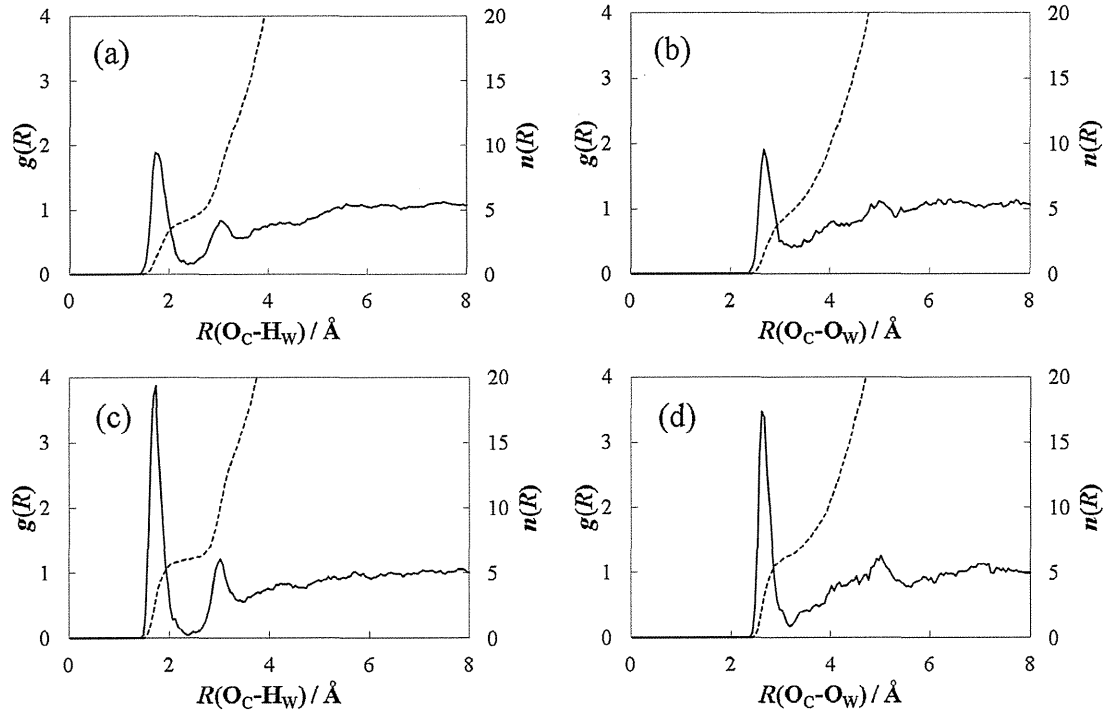


Figure 7. RDF (solid curve) and ICN (broken curve) between solute alanine dipeptide and solvent H_2O molecule with respect to O_C–H_W and O_C–O_W for (a, b) large QM system and (c, d) smaller QM system.

O_W) = 2.7 and 3.4 Å, respectively, and the integration of the first peak gives an ICN of 5.4 (Figure 7b). It should be noted that WAT32 contributes the stabilization of the alanine dipeptide being simultaneously within both first hydration shells of $O_C(1)$ and $O_C(2)$ atom, although no direct hydrogen bond exists between WAT32 and the peptide C=O groups of the alanine dipeptide. From these results, we can recognize that these 5 H_2O molecules contribute significantly to the stabilization of the alanine dipeptide in the PP_{II} conformation in aqueous solution.

On the other hand, for the smaller QM system, three kinds of *double-water bridges* were observed, i.e., C=O···$(H_2O)_2$···O=C, C=O···$(H_2O)_2$···H–N, and N–H···$(H_2O)_2$···O=C, as reported in a previous QM/MM-MD study,[82] where the alanine dipeptide was treated at the HF/3-21G level and all solvent H_2O molecules were described by the classical TIP3P model. For this system, the RDFs and the ICNs for O_C–H_W and O_C–O_W are depicted in Figures 7c and 7d, respectively. For the O_C–H_W RDF, the first peak and minimum are observed at $R(O_C$–$H_W) = 1.7$ and 2.4 Å, respectively, and integration of the first peak yields an ICN of 6.0 (Figure 7c). For the O_C–O_W RDF, the ICN integrating over the first peak ($R(O_C$–$H_W) = 2.6$ Å) to the minimum ($R(O_C$–$H_W) = 3.2$ Å) is 6.3 (Figure 7d). The well-defined first peak and minimum for O_C–H_W and O_C–O_W RDFs correspond to the formation of *double-water bridges* around the alanine dipeptide. For the smaller QM system, the O_C–H_W and O_C–O_W RDFs appear more structured than those for the large QM system, with a higher first peak and a deeper first minimum.

We here compare our results with the other MD ones. Some droplet models of ionic or neutral solute were investigated by using FMO-MD calculations employing various theoretical levels and FMO approximations.[27,28] However, they give poor results on structural properties because it is influenced by the edge effect of droplet model employed in the FMO-MD simulation. Thus, it is difficult to discuss the chemical accuracy of the FMO method on the structural properties of solution systems. On the other hand, García-Prieto et al. investigated the electronic structure of the QM alanine dipeptide at the DFT and MP2 levels of theory under the averaged electrostatic potential generated by the MM solvent H_2O molecules for understanding the solvent effect on the conformation equilibrium.[83] Unexpectedly, their method yields structured RDFs for the smaller QM system of the alanine dipeptide, although the QM alanine dipeptide was described at the DFT and MP2 levels of theory. Therefore, it is concluded that the influence of the QM region size is quite significant. In addition, the hydration structure of the alanine dipeptide could depend strongly on the formulation and calibration employed to the QM/MM interaction term.

PrP–GN8 Complex in Aqueous Solution. Prion protein (PrP) is an infectious agent of prion diseases, which results from the conformational conversion of a normal cellular form of the protein (PrP^C) into an alternatively folded scrapie isoform (PrP^{Sc}).[84] Recently, an antiprion compound termed GN8 (Figure 8) was found to suppress efficiently local fluctuations of PrP^C by forming intermolecular hydrogen bonds with the two residues, i.e., N159 and E196, and interfere with the pathological conversion of PrP^C to PrP^{Sc}.[85] In this study, we demonstrate the intermolecular interaction analysis between PrP and GN8 using the FMO-QM/MM-MD simulation. In particular, we focus our attention on the key interactions for two GN8-binding residues of N159 and E196 (Figure 9).

The initial structure of PrP–GN8 complex was generated as follows. The globular domain of PrP including residues 124–

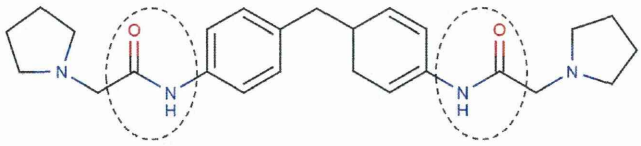


Figure 8. Chemical structure of GN8. Two amide groups of GN8 (dotted circles) can bind to PrP by crosslinking two distant residues, N159 and E196.

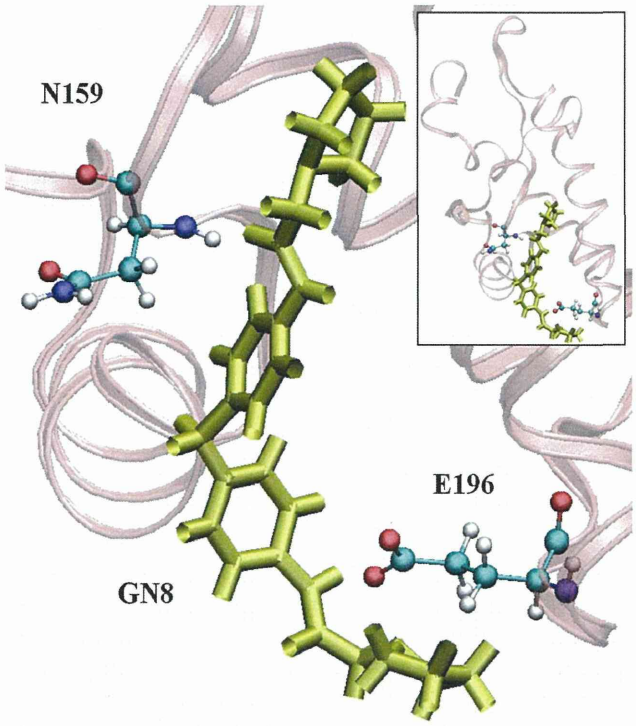


Figure 9. Graphical representation of two amino acid residues in PrP (N159 and E196) and GN8. The upper right panel shows the overall structure of the PrP–GN8 complex. GN8 is illustrated with a yellow stick model, and N159 and E196 are represented by CPK models.

226 was modeled from the NMR structure of mouse PrP^C (PDB code: 1AG2). The missing hydrogen atoms were added to the protein structure and a disulfide bond between C179 and C214 in PrP was connected by using LEaP module of AMBER9 program. The initial GN8-binding structure was generated by using the NMR structure as the reference. In the unit cell, 10309 water molecules and 2 sodium ions were arranged around the PrP–GN8 complex. The AMBER parm99SB force field[86] and the general force field (GAFF)[87] were adopted for PrP and GN8 as the MM force field parameters, respectively. For equilibrating the system, energy minimization and MD simulation were executed using the MM force fields. In the subsequent FMO-QM/MM-MD simulation, the ligand GN8 molecule and two GN8-binding residues of N159 and E196 in PrP were treated quantum mechanically based on the FMO scheme, and all of the remaining atoms were described molecular mechanically (Figure 9). Here the GN8 molecule was divided into four QM fragments in accord with a previous FMO study.[16] Two residues of N159 and E196 were regarded as individual QM fragments whose terminals were capped each

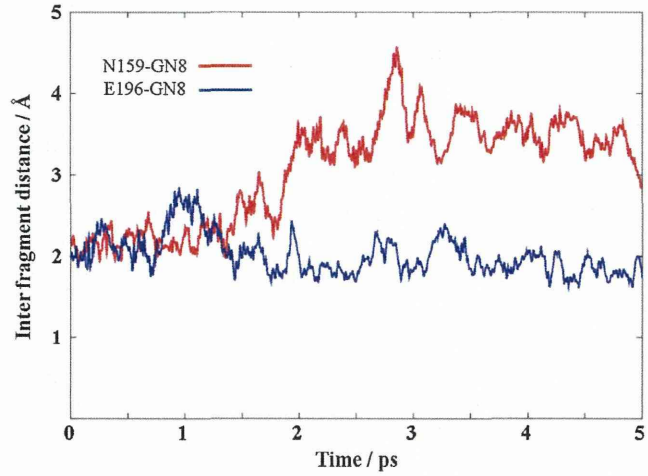


Figure 10. Time variation of interfragment distances for two hydrogen-bonding pairs of N159–GN8 (red curve) and E196–GN8 (blue curve) obtained from the direct FMO-QM/MM-MD trajectory.

by a link hydrogen atom. At each time step, the energies and the gradients in QM region were calculated at the FMO2-HF/6-31G level. After the MD simulation, the single-point calculations for the IFIE analysis were performed on the trajectory at the FMO2-MP2/6-31G level. The QM cutoff distance was set to 12 Å. The FMO-QM/MM-MD simulation was carried out in *NVT* ensemble, where the time step and total simulation time were set to 1 fs and 5 ps (=5000 sampling steps), respectively. The SHAKE algorithm[74] was applied to fix the length of the bonds involving hydrogen atoms. The temperature of the system was maintained at 300 K using the Berendsen thermostat.[80] The unit cell was a rectangular parallelepiped (69.47 × 64.96 × 72.79 $Å^3$) with periodic boundary condition and the mass density was adjusted to be about 1.00 g cm^{-3}. It took about 20 days to execute the direct FMO-QM/MM-MD simulation for 5000 steps using an Intel Xeon processor (E5429, 3.0 GHz) on 16 CPU cores.

Based on the resulting MD trajectory, we investigated the intermolecular interaction energies in the PrP–GN8 complex. Figure 10 shows time variation of the interfragment distances for two hydrogen-bonding pairs of N159–GN8 and E196–GN8. The interfragment distance between N159 and GN8 is kept within the range of 2.0–2.5 Å for the initial period, but increases abruptly at $t = 1.3$ ps. It indicates that the hydrogen bond of GN8 with N159 was disrupted within the 5 ps simulation. On the other hand, the interfragment distance for E196–GN8 is within the range of 1.5–3.0 Å, indicating that the hydrogen bond between E196 and GN8 persisted during this simulation.

Figure 11 presents time variation of the interaction energies for N159–GN8 and E196–GN8, where the interaction energies were obtained at the FMO2-MP2/6-31G level by summing the IFIEs between a fragment of each residue and 4 fragments of GN8. The interaction energy for E159–GN8 is −11.074 kcal mol^{-1} at the initial time but becomes weakened after $t =$ 1.3 ps and slowly reaches zero. On the other hand, the interaction energy for E196–GN8 retains within the range from −15 to −35 kcal mol^{-1} during the simulation. These results dem-

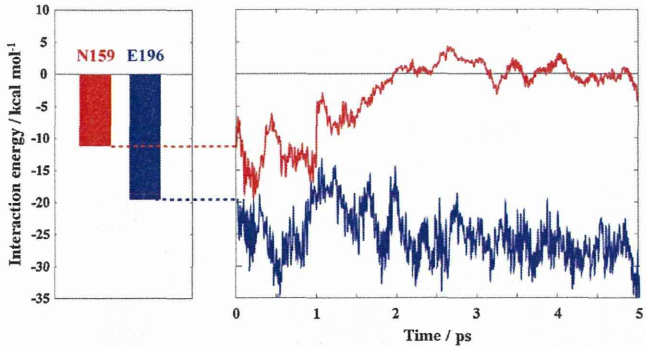


Figure 11. Time variation of IFIEs for two hydrogen-bonding pairs of N159–GN8 (red curve) and E196–GN8 (blue curve) obtained from the direct FMO-QM/MM-MD trajectory. These IFIEs were calculated at FMO2-MP2/6-31G level of theory.

onstrate that the two residues of N159 and E196 show different behaviors in binding GN8, i.e., the hydrogen bond between N159 and GN8 is relatively weak and disrupted easily, but the hydrogen-bonding interaction with E196 retains strongly and would play an important role in the antiprion ability of GN8. Such information will be helpful in further developing therapeutic compounds for prion diseases, aided by dynamics-based intermolecular interaction analysis.

For the present PrP–GN8 system, the FMO energies and gradients at the HF/6-31G level were adopted to integrate numerically Newton's equation of motion, but it should be inadequate for more quantitative discussion. In future, we would employ some better MO levels including electron correlation effects and larger basis sets, as the dispersion interaction has a pronounced role in many pharmacophores. Recently, Mochizuki et al. reported the FMO-MP2 gradient calculations and applied it to the FMO-MD simulation of water cluster.[28] In addition, a high-accuracy MP2 gradient[88] and the resolution of identity (RI) approximation (RI-MP2 gradient)[89] have been proposed within the FMO framework. By applying such a FMO-MP2 gradient, the dynamics-based interaction analysis could provide more reliable results.

Concluding Remarks

In the present study, we have developed an ab initio FMO-QM/MM interface "AP-IF" connecting AMBER with PAICS. Whereas PAICS was left unchanged, only AMBER was modified by adding some subroutines not only to generate automatically PAICS input file but also to calculate the energy and the gradients, as is the case in the our previously implemented AG-IF.[51] In addition to some subroutines for executing FMO-QM/MM calculations, we configured a subroutine "`qm_paics_init`" to prepare an FMO input file, since there is no information regarding the QM fragmentation in AMBER input and topology files.

We have also demonstrated three applications using the AP-IF. First, FMO and conventional MO calculations for molecular clusters of $HF(H_2O)_{n-1}$ were executed so as to investigate the computational speed and accuracy. The total energies for $n = 5$, 8, and 17 were almost the same for the FMO and MO methods. The computational time was reduced drastically by applying the FMO method. Even for larger molecular systems, it should be possible to retain the calculation accuracy by adopting the larger fragments or the FMO3 approximation.[26,72,73] In addition, the total energy of the cluster for $n = 5$ was conserved in an *NVE*-ensemble simulation with a relatively large time step. In the FMO-MD simulation with a short time step, the accumulation of error during the continuous evolution would be much smaller by using the fully analytical gradient including the response terms.[68] Second, we treated interaction energies between the alanine dipeptide and the surrounding solvent H_2O molecules in aqueous solution. The 5 H_2O molecules around the two peptide groups of the alanine dipeptide, i.e., WAT2, WAT12, WAT27, WAT32, and WAT44, are important for the stabilization of the alanine dipeptide in the PP_{II} conformation in aqueous solution. Third, the interaction energies between PrP and GN8 were analyzed by using instantaneous configurations obtained from the FMO-QM/MM-MD trajectory. While the interaction for N159–GN8 increases slowly to zero, the interaction for E196–GN8 retains within the range from -15 to $-35\,\mathrm{kcal\,mol^{-1}}$. The FMO-QM/MM method enables us to compute in ab initio quality those interaction energies between molecules or residues in large molecular systems, by using the dynamic-based IFIE analysis.

However, there still remain two major problems in the current implementation of the FMO-QM/MM method. The first one is how to update the fragmentation in the QM region during MD simulation. In the current version of AP-IF, the "auto-fragmentation" program in PAICS can perform the redefinition of fragments, which is called *dynamic fragmentation*,[21,23,26] only for some chemical species such as H_2O, H_3O^+, and OH^-. For simulating a chemical process associated with making or breaking bonds of the other chemical species, it is necessary to check whether the fragmentation might be appropriate at each MD time step and, if necessary, to update it by merging or separating the species in a chemically reasonable manner. The second one is how to obtain a proper boundary representation between QM and MM regions. In a QM/MM-MD simulation in solution containing both QM and MM solvent molecules, the nature of the solvent diffusion causes QM and MM molecules to move in and out of the QM region. Furthermore, those forces exerted on the atoms near the QM/MM boundary are inaccurate due to transferability problems between the QM and MM force models. To overcome the drawback of a partitioning of the system, some adaptive approaches have been proposed.[90–93] Recently, our group presented a new adaptive QM/MM approach to conserve satisfactorily the total energy of a liquid water system in the *NVE* ensemble and to reduce considerably computational time.[94,95] As just described, there is still room to improve the present FMO-QM/MM method, but it should become one of the most promising approaches in theoretical studies of large molecular systems.

This work was supported by a Grant-in-Aid for Science Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, and also by the Core Research for Evolutional Science and Technology (CREST) "High Performance Computing for Multi-scale and Multi-physics Phenomena" from the Japan Science and Technology Agency (JST).

References

1 D. Marx, J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge, **2009**. doi:10.1017/CBO9780511609633.

2 M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem. Rev.* **2012**, *112*, 632.

3 K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701.

4 *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, ed. by D. G. Fedorov, K. Kitaura, CRC Press, Boca Raton, **2009**. doi:10.1201/9781420078497.

5 J. Gao, *J. Phys. Chem. B* **1997**, *101*, 657.

6 J. Gao, *J. Chem. Phys.* **1998**, *109*, 2346.

7 S. J. Wierzchowski, D. A. Kofke, J. Gao, *J. Chem. Phys.* **2003**, *119*, 7365.

8 ABINIT-MP. http://moldb.nihs.go.jp/abinitmp/.

9 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Umebayasi, K. Kitaura, *Chem. Phys. Lett.* **2002**, *351*, 475.

10 Generalized Atomic and Molecular Electronic Structure System (GAMESS). http://www.msg.chem.iastate.edu/gamess/.

11 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, Jr., *J. Comput. Chem.* **1993**, *14*, 1347.

12 FMO in GAMESS. http://staff.aist.go.jp/d.g.fedorov/.

13 H. Sekino, Y. Sengoku, S. Sugiki, N. Kurita, *Chem. Phys. Lett.* **2003**, *378*, 589.

14 S.-i. Sugiki, N. Kurita, Y. Sengoku, H. Sekino, *Chem. Phys. Lett.* **2003**, *382*, 611.

15 Parallelized Ab Initio Calculation System based on FMO (PAICS). http://www.paics.net/.

16 T. Ishikawa, T. Ishikura, K. Kuwata, *J. Comput. Chem.* **2009**, *30*, 2594.

17 Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D. G. Fedorov, K. Kitaura, *Chem. Phys. Lett.* **2003**, *372*, 342.

18 Y. Komeiji, Y. Inadomi, T. Nakano, *Comput. Biol. Chem.* **2004**, *28*, 155.

19 Y. Komeiji, M. Uebayasi, R. Takata, A. Shimizu, K. Itsukashi, M. Taiji, *J. Comput. Chem.* **1997**, *18*, 1546.

20 Y. Mochizuki, K. Yamashita, T. Murase, T. Nakano, K. Fukuzawa, K. Takematsu, H. Watanabe, S. Tanaka, *Chem. Phys. Lett.* **2008**, *457*, 396.

21 Y. Mochizuki, Y. Komeiji, T. Ishikawa, T. Nakano, H. Yamataka, *Chem. Phys. Lett.* **2007**, *437*, 66.

22 M. Sato, H. Yamataka, Y. Komeiji, Y. Mochizuki, T. Ishikawa, T. Nakano, *J. Am. Chem. Soc.* **2008**, *130*, 2396.

23 Y. Komeiji, T. Ishikawa, Y. Mochizuki, H. Yamataka, T. Nakano, *J. Comput. Chem.* **2009**, *30*, 40.

24 Y. Komeiji, Y. Mochizuki, T. Nakano, D. G. Fedorov, *THEOCHEM* **2009**, *898*, 2.

25 M. Sato, H. Yamataka, Y. Komeiji, Y. Mochizuki, T. Nakano, *Chem.—Eur. J.* **2010**, *16*, 6430.

26 Y. Komeiji, Y. Mochizuki, T. Nakano, *Chem. Phys. Lett.* **2010**, *484*, 380.

27 T. Fujiwara, Y. Mochizuki, Y. Komeiji, Y. Okiyama, H. Mori, T. Nakano, E. Miyoshi, *Chem. Phys. Lett.* **2010**, *490*, 41.

28 Y. Mochizuki, T. Nakano, Y. Komeiji, K. Yamashita, Y. Okiyama, H. Yoshikawa, H. Yamataka, *Chem. Phys. Lett.* **2011**, *504*, 95.

29 A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.

30 U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718.

31 M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700.

32 *Combined Quamtum Mechanical and Molecular Mechanical Methods* in *Symposium Series*, ed. by J. Gao, M. A. Thompson, American Chemical Society, Washington, DC, **1998**, No. 712. doi:10.1021/bk-1998-0712.

33 M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* **1996**, *100*, 19357.

34 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, M. J. Frisch, *THEOCHEM* **1999**, *461–462*, 1.

35 M. E. Martín, M. A. Aguilar, S. Chalmet, M. F. Ruiz-López, *Chem. Phys.* **2002**, *284*, 607.

36 Y. Koyano, N. Takenaka, Y. Nakagawa, M. Nagaoka, *Bull. Chem. Soc. Jpn.* **2010**, *83*, 486.

37 CPMD. http://cpmd.org/.

38 P. K. Biswas, V. Gogonea, *J. Chem. Phys.* **2005**, *123*, 164114.

39 CP2K. http://www.cp2k.org/.

40 T. Laino, F. Mohamed, A. Laio, M. Parrinello, *J. Chem. Theory Comput.* **2005**, *1*, 1176.

41 T. Laino, F. Mohamed, A. Laio, M. Parrinello, *J. Chem. Theory Comput.* **2006**, *2*, 1370.

42 J. R. Shoemaker, L. W. Burggraf, M. S. Gordon, *J. Phys. Chem. A* **1999**, *103*, 3245.

43 J. Shoemaker, L. W. Burggraf, M. S. Gordon, *J. Chem. Phys.* **2000**, *112*, 2994.

44 Computational Chemistry Shell (ChemShell). http://www.chemshell.org/.

45 P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *THEOCHEM* **2003**, *632*, 1.

46 Program for User Package Interfacing and Linking (PUPIL). http://pupil.sourceforge.net/.

47 J. Torras, Y. He, C. Cao, K. Muralidharan, E. Deumens, H.-P. Cheng, S. B. Trickey, *Comput. Phys. Commun.* **2007**, *177*, 265.

48 J. Torras, G. de M. Seabra, E. Deumens, S. B. Trickey, A. E. Roitberg, *J. Comput. Chem.* **2008**, *29*, 1564.

49 Assisted Model Building with Energy Refinement (AMBER). http://ambermd.org/.

50 Chemistry at HARvard Macromolecular Mechanics (CHARMM). http://www.charmm.org/.

51 T. Okamoto, K. Yamada, Y. Koyano, T. Asada, N. Koga, M. Nagaoka, *J. Comput. Chem.* **2011**, *32*, 932.

52 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, Jr., D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. E. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. E. A. F. Schafmeister, W. S. Ross, P. A. Kollman, *AMBER 9*, University of California, San Francisco, **2006**.

53 R. C. Walker, M. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019.

54 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,

M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03 (Revision C.02)*, Gaussian Inc., Wallingford CT, **2004**.

55 A. Tongraar, B. M. Rode, *Chem. Phys. Lett.* **2008**, *466*, 61.

56 C. V. Sumowski, C. Ochsenfeld, *J. Phys. Chem. A* **2009**, *113*, 11734.

57 T. Nagata, D. G. Fedorov, K. Kitaura, M. S. Gordon, *J. Chem. Phys.* **2009**, *131*, 024101.

58 T. Nagata, D. G. Fedorov, T. Sawada, K. Kitaura, M. S. Gordon, *J. Chem. Phys.* **2011**, *134*, 034110.

59 N. Takenaka, Y. Kitamura, Y. Koyano, T. Asada, M. Nagaoka, *Theor. Chem. Acc.* **2011**, *130*, 215.

60 K. Yamada, Y. Koyano, T. Okamoto, T. Asada, N. Koga, M. Nagaoka, *J. Comput. Chem.* **2011**, *32*, 3092.

61 K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji, M. Uebayasi, *Chem. Phys. Lett.* **2001**, *336*, 163.

62 Y. Mochizuki, K. Fukuzawa, A. Kato, S. Tanaka, K. Kitaura, T. Nakano, *Chem. Phys. Lett.* **2005**, *410*, 247.

63 S. Amari, M. Aizawa, J. Zhang, K. Fukuzawa, Y. Mochizuki, Y. Iwasawa, K. Nakata, H. Chuman, T. Nakano, *J. Chem. Inf. Model.* **2006**, *46*, 221.

64 D. G. Fedorov, K. Kitaura, *J. Comput. Chem.* **2007**, *28*, 222.

65 I. Kurisaki, K. Fukuzawa, Y. Komeiji, Y. Mochizuki, T. Nakano, J. Imada, A. Chmielewski, S. M. Rothstein, H. Watanabe, S. Tanaka, *Biophys. Chem.* **2007**, *130*, 1.

66 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2000**, *318*, 614.

67 T. Nagata, D. G. Fedorov, K. Kitaura, *Chem. Phys. Lett.* **2009**, *475*, 124.

68 T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, M. S. Gordon, *J. Chem. Phys.* **2011**, *134*, 124115.

69 D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580.

70 Message Passing Interface (MPI). http://www.mpi-forum.org/.

71 D. G. Fedorov, J. H. Jensen, R. C. Deka, K. Kitaura, *J. Phys. Chem. A* **2008**, *112*, 11808.

72 D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.

73 D. G. Fedorov, K. Kitaura, *Chem. Phys. Lett.* **2006**, *433*, 182.

74 J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.

75 J. Grdadolnik, S. G. Grdadolnik, F. Avbelj, *J. Phys. Chem. B* **2008**, *112*, 2712.

76 V. Madison, K. D. Kopple, *J. Am. Chem. Soc.* **1980**, *102*, 4855.

77 M. A. Mehta, E. A. Fry, M. T. Eddy, M. T. Dedeo, A. E. Anagnost, J. R. Long, *J. Phys. Chem. B* **2004**, *108*, 2777.

78 C.-D. Poon, E. T. Samulski, C. F. Weise, J. C. Weisshaar, *J. Am. Chem. Soc.* **2000**, *122*, 5642.

79 K. Kwac, K.-K. Lee, J. B. Han, K.-I. Oh, M. Cho, *J. Phys. Chem.* **2008**, *128*, 105106.

80 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.

81 M.-P. Gaigeot, *J. Phys. Chem. B* **2009**, *113*, 10059.

82 R. Jono, Y. Watanabe, K. Shimizu, T. Terada, *J. Comput. Chem.* **2010**, *31*, 1168.

83 F. F. García-Prieto, I. F. Galván, M. A. Aguilar, M. E. Martín, *J. Chem. Phys.* **2011**, *135*, 194502.

84 S. B. Prusiner, *Science* **1982**, *216*, 136.

85 K. Kuwata, N. Nishida, T. Matsumoto, Y. O. Kamatari, J. Hosokawa-Muto, K. Kodama, H. K. Nakamura, K. Kimura, M. Kawasaki, Y. Takakura, S. Shirabe, J. Takata, Y. Kataoka, S. Katamine, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11921.

86 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.

87 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

88 T. Nagata, D. G. Fedorov, K. Ishimura, K. Kitaura, *J. Chem. Phys.* **2011**, *135*, 044110.

89 T. Ishikawa, K. Kuwata, *J. Phys. Chem. Lett.* **2012**, *3*, 375.

90 T. Kerdcharoen, K. R. Liedl, B. M. Rode, *Chem. Phys.* **1996**, *211*, 313.

91 T. Kerdcharoen, K. Morokuma, *Chem. Phys. Lett.* **2002**, *355*, 257.

92 A. Heyden, H. Lin, D. G. Truhlar, *J. Phys. Chem. B* **2007**, *111*, 2231.

93 R. E. Bulo, B. Ensing, J. Sikkema, L. Visscher, *J. Chem. Theory Comput.* **2009**, *5*, 2212.

94 N. Takenaka, Y. Kitamura, Y. Koyano, M. Nagaoka, *Chem. Phys. Lett.* **2012**, *524*, 56.

95 N. Takenaka, Y. Kitamura, Y. Koyano, M. Nagaoka, *J. Chem. Phys.* **2012**, *137*, 024501.

THE PROTEIN SOCIETY

Characterizing antiprion compounds based on their binding properties to prion proteins: Implications as medical chaperones

Yuji O. Kamatari,[1] Yosuke Hayano,[2] Kei-ichi Yamaguchi,[2,3] Junji Hosokawa-Muto,[2] and Kazuo Kuwata[2,3,4]*

[1]Life Science Research Center, Gifu University, Gifu 501-1194, Japan

[2]Center for Emerging Infectious Diseases, Gifu University, Gifu 501-1194, Japan

[3]United Graduate School of Drug Discovery and Medical Information Sciences, Gifu University, Gifu 501-1194, Japan

[4]CREST, Japan

Received 25 June 2012; Revised 12 October 2012; Accepted 15 October 2012
DOI: 10.1002/pro.2180
Published online 18 October 2012 proteinscience.org

Abstract: A variety of antiprion compounds have been reported that are effective in *ex vivo* and *in vivo* treatment experiments. However, the molecular mechanisms for most of these compounds remain unknown. Here we classified antiprion mechanisms into four categories: I, specific conformational stabilization; II, nonspecific stabilization; III, aggregation; and IV, interaction with molecules other than PrP^C. To characterize antiprion compounds based on this classification, we determined their binding affinities to PrP^C using surface plasmon resonance and their binding sites on PrP^C using NMR spectroscopy. GN8 and GJP49 bound specifically to the hot spot in PrP^C, and acted as "medical chaperones" to stabilize the native conformation. Thus, mechanisms I was predominant. In contrast, quinacrine and epigallocathechin bound to PrP^C rather nonspecifically; these may stabilize the PrP^C conformation nonspecifically including the interference with the intermolecular interaction following mechanism II. Congo red and pentosan polysulfate bound to PrP^C and caused aggregation and precipitation of PrP^C, thus reducing the effective concentration of prion protein. Thus, mechanism III was appropriate. Finally, CP-60, an edarabone derivative, did not bind to PrP^C. Thus these were classified into mechanism IV. However, their antiprion activities were not confirmed in the GT + FK system, whose details remain to be elucidated. This proposed antiprion mechanisms of diverse

Abbreviations: AEBSF, 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride; DMSO, dimethyl sulfoxide; DSS, 4,4-dimethyl-4-silapentane-1-sulfonic acid; EDTA, ethylenediaminetetraacetic acid; EGCG, epigallocatechin gallate; HSQC, Heteronuclear Single Quantum Coherence; IC_{50}, inhibitory concentration for 50% reduction of PrP^{Sc}; K_D, dissociation constant; PPS, pentosan polysulfate; PrP, prion protein; PrP(23–231), PrP residues 23–231; PrP(121–231), PrP residues 121–231; PrP*, the intermediate state between PrP^C and PrP^{Sc}; PrP^C, the cellular form of PrP; PrP^{Sc}, the scrapie form of PrP; SPR, surface plasmon resonance; TSEs, transmissible spongiform encephalopathies.

Additional Supporting Information may be found in the online version of this article.

Junji Hosokawa-Muto's current address is First Department of Forensic Science, National Research Institute of Police Science, 6-3-1 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan.

Grant sponsors: Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation, the Molecular Imaging Project of Japan Society and Technology Agency, and Ministry of Education, Culture, Sports, Science & Technology, and the Research Committee of Prion Disease and Slow Virus Infection, Ministry of Health, Labour and Welfare, Japan.

Correspondence to: Kazuo Kuwata, Center for Emerging Infectious Diseases, Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan. E-mail: kuwata@gifu-u.ac.jp

antiprion compounds could help to elucidate their antiprion activities and facilitate effective antiprion drug discovery.

Keywords: prion protein; anti-prion compounds; action mechanism; medical chaperones

Introduction

Transmissible spongiform encephalopathies (TSEs) are neurodegenerative diseases that include Creutzfeldt–Jakob disease, chronic wasting disease, scrapie, bovine spongiform encephalopathy, and others. Although the conversion of prion protein (PrP) from the normal cellular form (PrP^{C}) to the misfolded scrapie form (PrP^{Sc}), and the accumulation of PrP^{Sc} in the central nervous system are the characteristic features of these diseases,[1,2] the detailed structure of PrP^{Sc} and the details of the conversion reaction remain unknown. The occurrence of TSEs is associated with specific mutations in PrP, inoculation with infectious material, or spontaneous onset. Because there are currently no established therapies, it is important to identify compounds with therapeutic or prophylactic activity against TSEs. The purpose of this study was to understand the various mechanisms of action of previously reported antiprion compounds and obtain further insights into optimizing their antiprion activities as well as the conversion mechanism of prion proteins.

To date, no purification method for a unique PrP^{Sc} strain has been fully established;[3] thus, quantitative binding experiments for a single PrP^{Sc} conformer is not feasible. Ideal stoichiometric determinations of ligand binding akin to the oxygen–hemoglobin binding scheme[4] are rather restricted because of the limitation of quantifying PrP^{Sc}.

In contrast, populations of PrP^{Sc} are considered to be kinetically regulated.[5] If the reaction rate of the $PrP^{C} \rightarrow PrP^{Sc}$ conversion could be successfully reduced to a certain level, the population of available PrP^{Sc} would gradually be reduced.[6] Following this strategy, we can completely examine the binding properties of ligands with PrP^{C} and evaluate their stabilization effects on PrP^{C}, their interference effects on the interaction between PrP^{C} and PrP^{Sc}, and such other effects. These activities are well represented by the notion of chemical chaperones.[6]

We previously reported a variety of antiprion compounds that were discovered based on the structures and dynamics of prion proteins.[6,7] Among these antiprion compounds, GN8[6] and its derivatives[8] are unique because of their affinities for PrP^{C}, their binding sites at atomic resolution, and their antiprion activities that were clarified in both *ex vivo* and *in vivo* experiments.[6,8] Other compounds have also been reported to bind to PrP^{C} and inhibit its pathogenic conversion.[7,9] Several groups have also reported various antiprion compounds based on *ex vivo* and/or *in vivo* experiments.[10–13] However, the details of their antiprion activities remain unknown. Thus, we attempted to characterize their antiprion mechanisms based on their binding properties to PrP^{C}.

Although previous studies on antiprion compounds primarily emphasized the biological effects of these compounds (IC_{50}), this study examined the various mechanisms of antiprion effects based on their binding properties to PrP^{C}. Thus, this study may offer new insights into the mechanisms of prion diseases. To characterize these antiprion compounds, we determined their binding affinities to PrP^{C} using surface plasmon resonance (SPR) and their binding sites on PrP^{C} using NMR spectroscopy.

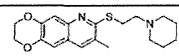
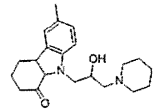
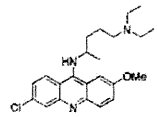
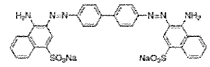
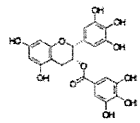
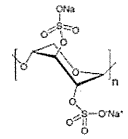
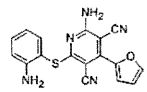
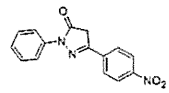
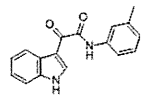
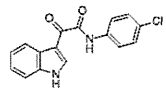
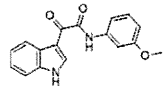
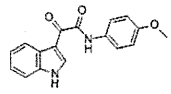
We selected 13 typical antiprion compounds, including GJP49, GJP14, quinacrine, Congo red, epigallocatechin gallate (EGCG), pentosan polysulfate (PPS), CP-60, Edaravone derivative 13, D-PEN, and Indole-3-glyoxylamide derivatives (see Table I). We classified these compounds into several classes based on their binding properties to PrP^{C}. GJP49 and GJP14 are antiprion compounds that were discovered through *in silico* screening.[7] Quinacrine, which has been used as an antimalarial drug for over 60 years, has also been shown to be an antiprion compound.[10] Congo red, a dye used for staining amyloid fibrils, has been reported to have antiprion activity.[14] EGCG is a polyphenolic compound in tea that has been reported to have antiprion activity,[15] and PPS, which is used as a therapeutic agent for chronic cystitis, has also been reported to be an antiprion compound.[16,17] We also examined CP-60,[11] edaravone derivative 13,[18] D-PEN,[19] and indole-3-glyoxylamide derivatives.[20] In addition, we discuss a novel concept, "medical chaperone," in the context of further optimization of these previously identified antiprion compounds.

Results

GJP49 and GJP14 specifically bind to PrP^{C}

We previously reported the binding affinity of GJP49 for PrP^{C} as determined by SPR.[7] The SPR data were fit using a 1:1 binding model of equilibrium analysis, and the dissociation constant (K_D) was estimated to be 50.8 μM. To identify the interaction sites on PrP^{C} for GJP49, we measured the ^{1}H-^{15}N HSQC spectra with or without GJP49 and superimposed them, as shown in Figure 1(A). We also calculated the chemical shift perturbations ($\Delta\delta$) of ^{1}H and ^{15}N nuclei in PrP^{C} upon binding with this compound [Fig. 1(B)], and significantly perturbed

Table I. *Antiprion Compounds Used in This Study*

Compound	Chemical structure	M.W.	Relative PrP^{Sc} level[a] (%)	Antiprion mechanism	References
GJP49		344.5	47.4	I	7
GJP14		342.5	51.3	I	7
Quinacrine		400	Toxic[b]	II	10
Congo red		696.7	39.6	II + III	14
Epigallo-catechin gallate (EGCG)		458.4	65.6	II + III	15
Pentosan polysulfate (PPS)		366.3n	ND[c]	II + III	16,17
CP-60		333.4	94.3	IV[d]	11
Edaravone derivative 13		281.3	93.0	IV[d]	18
D-PEN		149.2	95.5	IV[d]	19
Indole-3-glyoxylamide derivative 2		278.3	81.8	IV[d]	20
Indole-3-glyoxylamide derivative 10		298.7	110	IV[d]	20
Indole-3-glyoxylamide derivative 12		294.3	103	IV[d]	20
Indole-3-glyoxylamide derivative 13		294.3	Toxic[e]	IV[d]	20

[a] Mean PrP^{Sc} level (%) using 10 μM of each compound on Fukuoka-1 strain-infected GT + FK cells.
[b] Relative PrP^{Sc} level is 37.1 at 2 μM for quinacrine.
[c] Though PrP^{Sc} level at 10 μM cannot determined because the compound concentration could not be determined, relative PrP^{Sc} level at 10 μg/mL is 24.4 for PPS.
[d] Binding to PrP^{C} nor antiprion activities in GT + FK system were not confirmed.
[e] Relative PrP^{Sc} level is 144 at 5 μM for indole-3-glyoxylamide derivative 13.

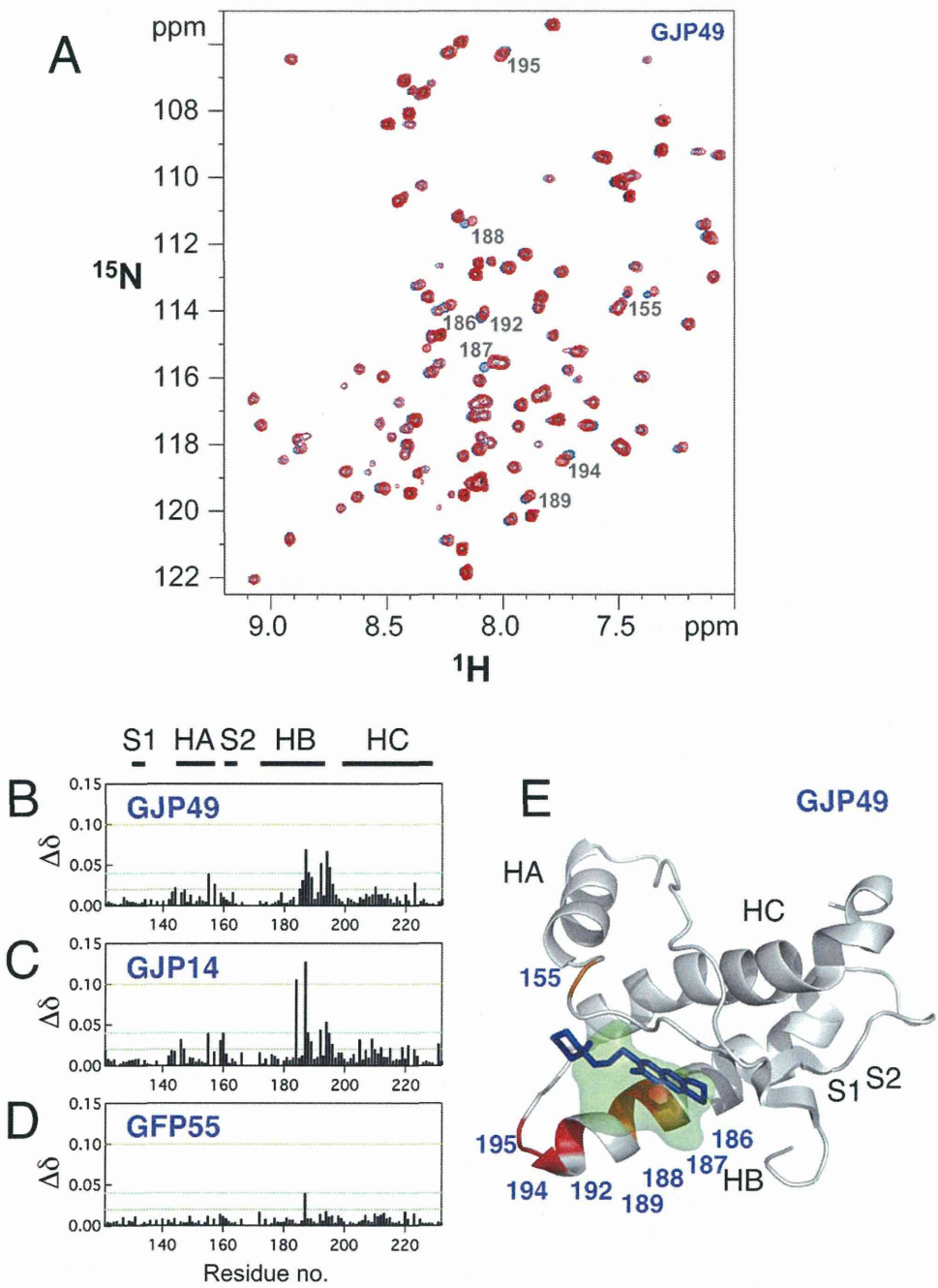


Figure 1. Characteristics of the binding sites for the antiprion compound GJP49. GJP49 was discovered by *in silico* screening using Autodock.[7] (A) ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled recombinant prion protein, mouse PrP(121–231) (33 μ*M*), with (red) or without (blue) GJP49 (500 μ*M*). (B–D) Plots of chemical shift perturbations ($\Delta\delta = [(\Delta\delta_{1H})^2 + 0.17(\Delta\delta_{15N})^2]^{1/2}$) as a function of the residue number. (E) Mapping of the significantly perturbed residues on the three-dimensional structure of the prion (1AG2). The perturbed residues with $\Delta\delta$ values of >0.04 ppm are shown in red, and those with $0.4 < \Delta\delta < 0.3$ ppm are shown in orange. The binding pocket is overlaid in green. S1, HA, S2, HB, and HC indicate S1 strand, helix A, S2 strand, helix B, and helix C, respectively. The image was created using PyMol.

PrP^C residues were mapped onto the three dimensional structure of mouse PrP (PDB ID: 1AG2). Figure 1(E) shows the specific binding of GJP49 to the C-terminal region of helix B and at part of the B-C loop of PrP^C.

We also examined the interaction sites for GJP14, another effective antiprion compound that was found by *in silico* screening.[7] This compound also bound to the same region of PrP^C [Fig. 1(C)]. The region of PrP^C that binds GJP49 and GJP14 corresponds to that of GN8, another antiprion compound that we examined in a previous article.[6] This region of PrP^C undergoes a global fluctuation on a time scale of micro- to milliseconds.[21]

As a negative control for these binding experiments, we also examined the interaction sites of

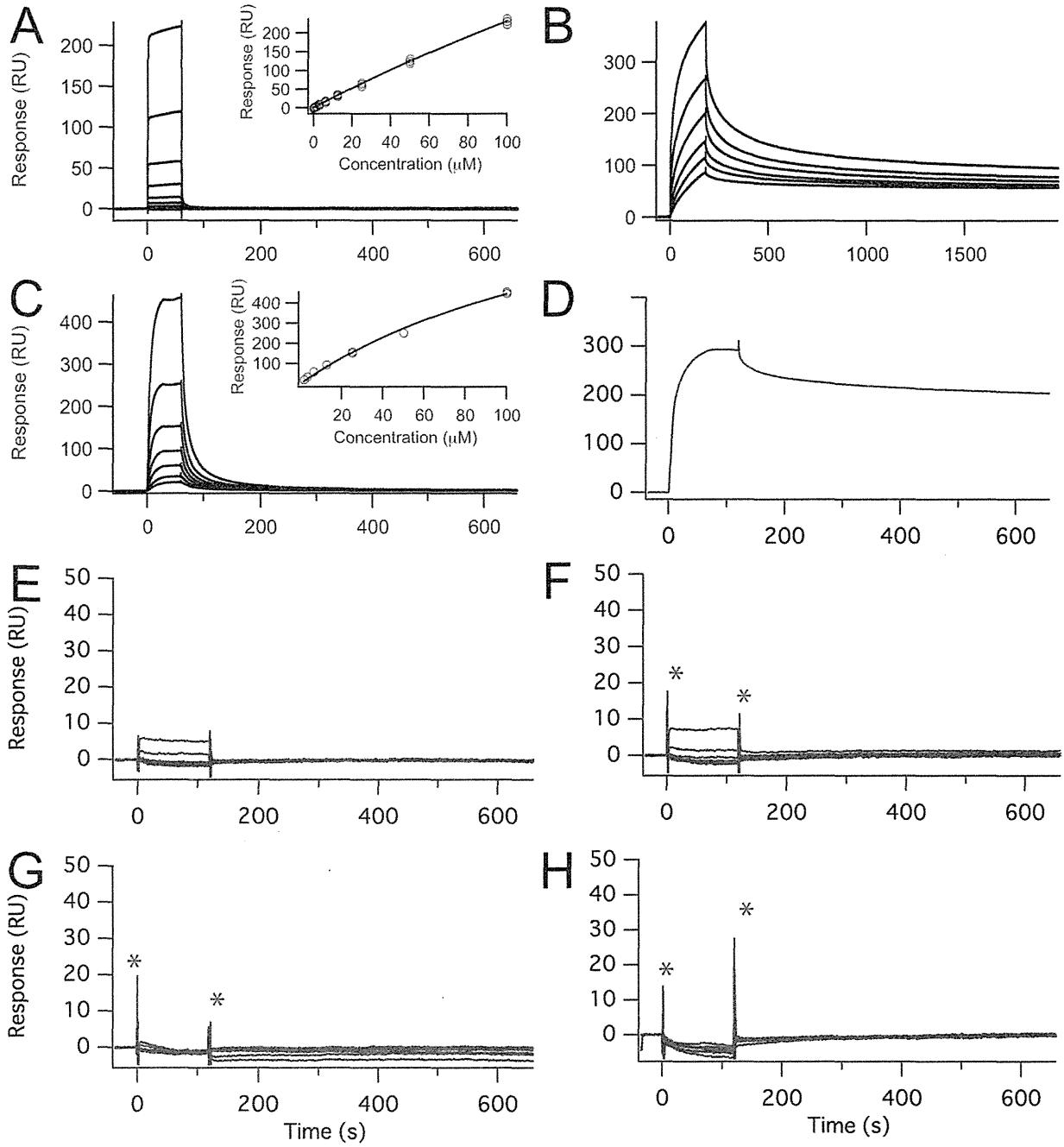


Figure 2. Surface plasmon resonance sensorgrams for various antiprion compounds. (A) Quinacrine; the concentrations from bottom to top are: 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, and 100 μ*M*. (B) Epigallocathechin gallate; the concentrations from bottom to top are: 3.91 7.81, 15.6, 31.3, 62.5, and 125 μ*M*. (C) Congo red; the concentrations from bottom to top are: 1.56, 3.13, 6.25, 12.5, 25, 50, and 100 μ*M*. (D) Pentosan polysulfate; the concentrations is: 1.95 μg/mL. (E–H) CP60, Edarabone derivative 13, D-PEN, and indole-3-glyoxylamide derivative 10; the concentrations from bottom to top are: 0.75, 1.25, 2.5, 5.0, 10, 20, and 40 μ*M*. The symbol * indicates spike noise. Insets show the compound concentration dependence of the Biacore response. For quinacrine, Congo red, and pentosan polysulfate binding experiments, recombinant mouse PrP(121–231) was fixed on the surface of the sensor chip. For epigallocatechin gallate, CP60, edarabone derivative 13, D-PEN, and indole-3-glyoxylamide derivative 10 binding experiments, recombinant mouse PrP(23–231) was fixed on the surface of the sensor chip.

GFP55, which belongs to a low-binding, ineffective group of compounds.[7] In contrast to GJP49 and GJP14, the addition of GFP55 did not result in a significant chemical shift change [Fig. 1(D)]. Therefore, we considered that GJP49 and GJP14 specifically bound to a flexible pocket of PrP^{C} and inhibited the pathogenic conversion of PrP^{C} to PrP^{Sc}.