

High performance computing for drug development on K computer

Hideaki Fujitani, Keiko Shinoda, Takefumi Yamashita and Tatsuhiko Kodama

Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba Meguro-ku Tokyo 153-8904 Japan

E-mail: fjtani@lsbm.org

Abstract. Massively parallel computations (MP-CAFEE) were developed to calculate absolute binding free energies of small molecules bound to a protein by all-atom molecular dynamics. It uses the nonequilibrium work measurement and Bennett acceptance ratio methods to calculate the free energy difference between the bound and unbound states. The FUJI force field was developed in order to assign force field parameters to arbitrary organic molecules in a unified manner including proteins and nucleic acids. Its dihedral parameters agree with the torsion energy profiles calculated by high-level ab initio molecular orbital theory for the model systems of protein backbone. Comparing with various force fields it agrees well with recent observations by vibrational spectroscopy on Ramachandran angle's population of alanine dipeptide in water. MP-CAFEE with FUJI force field has an efficient parallel algorithm and enough accuracy for computer aided drug design.

1. Introduction

One of the biggest challenges in drug development is to design small molecules of high affinity to a pharmaceutical target protein based on physical principles for all-atom molecular dynamics in water. In biology and pharmacology, a ligand means a substance (usually, a small molecule) bound to a protein which has a biological function. The binding occurs by intermolecular forces, caused by Coulomb and van der Waals interactions between all atoms. The binding of the protein and ligand is usually reversible without forming covalent bond. The Gibbs free energy difference between the bound and unbound states characterizes the binding affinity of the protein and ligand. The stronger the ligand binds, the more effectively the biological function works. Therefore, the binding affinity is an important criterion to design a new ligand to the target protein.

In order to explore dynamical behaviors of protein and ligand in water, it is popular to use molecular dynamics simulations with classical force field like AMBER, CHARMM, OPLS-AA, and so on. Based on the molecular force field many free energy calculation methods have been developed [1], but it was hard to predict the binding affinity with enough accuracy for drug design. In 2005 we reported the massively parallel computation of absolute binding free energy (MP-CAFEE) for FKBP and ligand systems using AMBER99 force field for the FKBP protein and general AMBER force field (GAFF) and Austin Model 1 bond charge correction (AM1BCC) charges for the ligands [2]. However, our calculated values were shifted by about 3.2 kcal/mol from experimental absolute binding free energies [3].



2. FUJI force field

The binding affinity of the protein and ligand is affected by thermal fluctuations. In order to rank molecules in computer aided drug design, we must predict the binding affinity within an error of 1 kcal/mol, while the thermal energy of the room temperature is 0.6 kcal/mol. From our early stage experiences in MP-CAFE calculations, we realized that we should refine the force field itself, instead of sticking to the sampling issue on the free energy calculation. Using a unified force field for proteins and organic molecules, we got better absolute binding free energies for the FKBP and ligands [3].

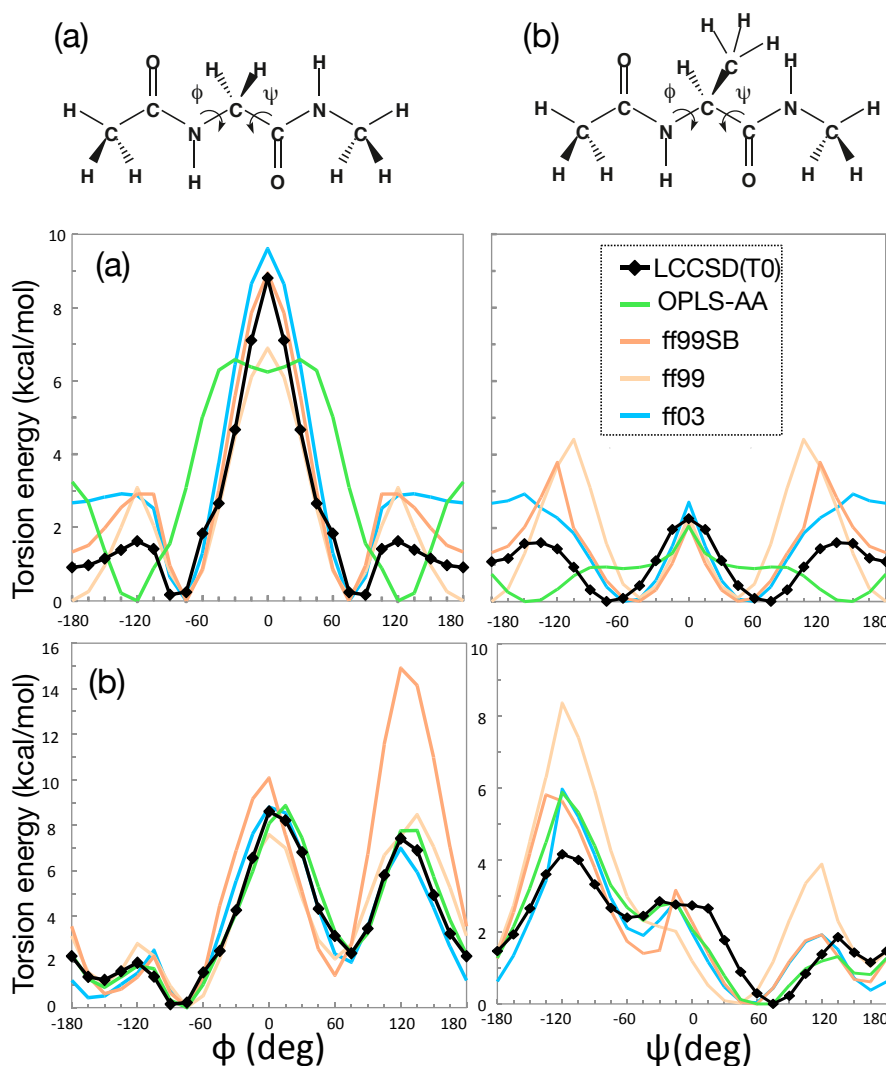


Figure 1. The torsional energy profiles of ϕ and ψ calculated with OPLS-AA and three AMBER force field variants (ff99SB, ff99, ff03) for (a) glycine dipeptide and (b) alanine dipeptide [4, 5, 6]. The filled squares with a black line show the torsional energy profile of ϕ and ψ calculated by MOLPRO quantum chemistry package with DF-LCCSD(T0)/Aug-cc-pVTZ//DF-LMP2/Aug-cc-pVTZ level.

In the 1990s Kollman's group developed the AMBER force field. The dihedral torsion parameters of protein backbone were determined to agree with quantum mechanical calculations for the model systems of glycine and alanine dipeptides [7]. But it was difficult to determine accurate torsion parameters by quantum mechanics, because of the shallow minima and very flat regions of potential energy surfaces with respect to the Ramachandran angles ϕ and ψ . We intensively investigated accuracy of various quantum mechanical theory levels to determine the appropriate theory level for the torsional energy profiles. The filled squares with black lines in figure 1 show the torsional energy profiles of glycine and alanine dipeptides calculated by the high level theory of DF-LCCSD(T0)/Aug-cc-pVTZ//DF-LMP2/Aug-cc-pVTZ. The colored lines show torsional energy profiles obtained by various molecular mechanical force fields, which all deviate from the ab initio torsional energy profile. We refined the relevant molecular mechanical dihedral parameters to fit the ab initio torsional energy profiles [8].

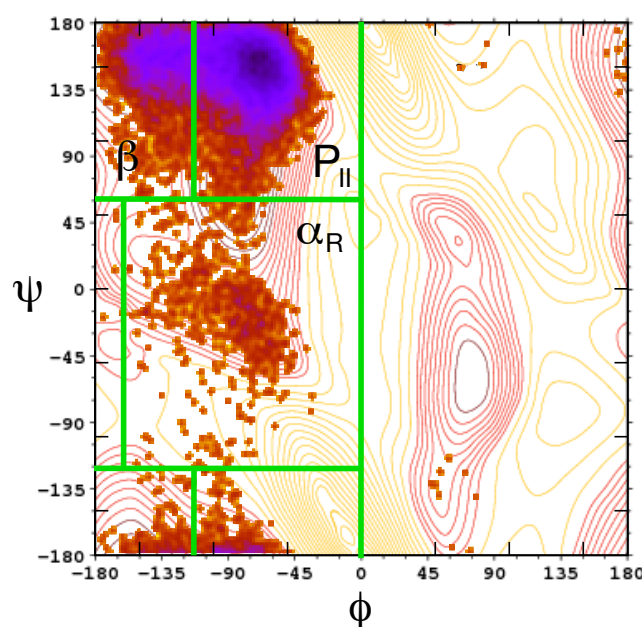


Figure 2. Ramachandran angle's population of alanine dipeptide in TIP3P water. The buried contour shows an energy landscape of alanine dipeptide in vacuum.

Grdadolnik *et al.* measured a short-lived population of Ramachandran angles of alanine dipeptide in aqueous solution using vibrational spectroscopy (infrared and Raman spectra) [9]. In order to compare with their experimental population we performed molecular dynamics simulation of alanine dipeptide in TIP3P water using the FUJI force field. Figure 2 shows an example of the Ramachandran angle's population in a 100 ns simulation at room temperature. The purple area in P_{II} region refers to dense population. The buried contour shows an energy landscape of alanine dipeptide in vacuum [8]. It has a metastable state around $(\phi = 60, \psi = -60)$, but the alanine dipeptide in water rarely takes such conformation.

Seabra *et al.* calculated the Ramachandran angle's population of alanine dipeptide in water by the replica exchange method using various semiempirical quantum mechanics and classical force field molecular mechanics. Their results were in contrast with the computational chemist's intuitive idea that the more expensive a method the better its accuracy. Actually, they found

Table 1. Ramachandran angle's population of alanine dipeptide in water. The top two rows (infrared and Raman) are experiments and others are theoretical results.

method	α_R	β	P_{II}	other	reference
infrared	0.11	0.29	0.60	–	[10]
Raman	0.09	0.29	0.62	–	[10]
ff94	0.84	0.04	0.11	0.01	[11]
ff99SB	0.32	0.24	0.40	0.04	[11]
ff03	0.45	0.19	0.35	0.01	[11]
PM3	0.14	0.51	0.33	0.02	[11]
CHARM27	0.53	0.07	0.40	0.00	[12]
FUJI	0.05	0.17	0.76	–	[13]
FUJI	0.06	0.23	0.71	0.00	this work

the ff99SB results were more accurate than most of the semiempirical quantum mechanical methods. Table 1 shows the Ramachandran angle's population in the three regions (α_R , β , P_{II}). Our population by the FUJI force field was derived from three 100 ns trajectories and agrees with the metadynamics analysis by Vymětal and Vondrášek [13].

Even taking into account some ambiguities in the experimental measurement and in the theoretical boundaries of the three regions, we can conclude that only the FUJI force field agrees with the experimental Ramachandran angle's population of alanine dipeptide in water. This conclusion is extremely significant for molecular dynamics simulations with classical force field, because the same dihedral parameters are used for the protein backbone of all amino acids.

3. MP-CAFE

There are important requirements for accurate absolute binding free energy calculations. The first requirement is a well-equilibrated bound structure including the conformational change of the protein induced by the binding of the small molecule. The second one is the convergence of the work distribution with a sufficient number of trajectories and dense spacing of the coupling constant λ between the small molecule and the rest of the system. Finally, the most important requirement is an accurate force field for the protein and small molecule. Our calculations were performed by GROMACS [14].

Using the FUJI force field we performed absolute binding free energy calculations for the major urinary protein (MUP) and two ligands: 2-methoxy-3-isopropylpyrazine (IPMP) and 2-methoxy-3-isobutylpyrazine (IBMP). MUP has a unique binding structure such that the whole ligand is wrapped in the protein. The binding site cannot be seen from outside (figure 3). Homans's group performed excellent works on this system including the binding free energy measurements by isothermal titration calorimetry [15]. It is a good test system for the absolute binding free energy calculations by all atom molecular dynamics.

Table 2. The calculated binding free energy $\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{sol}}$ and experimental binding free energy ΔG_{exp} in cal/mol.

ligand	ΔG_{sol}	$\Delta G_{\text{complex}}$	ΔG_{bind}	ΔG_{exp}
IPMP	-1.6	-9.9	-8.3	-8.1
IBMP	-1.9	-11.3	-9.4	-9.2

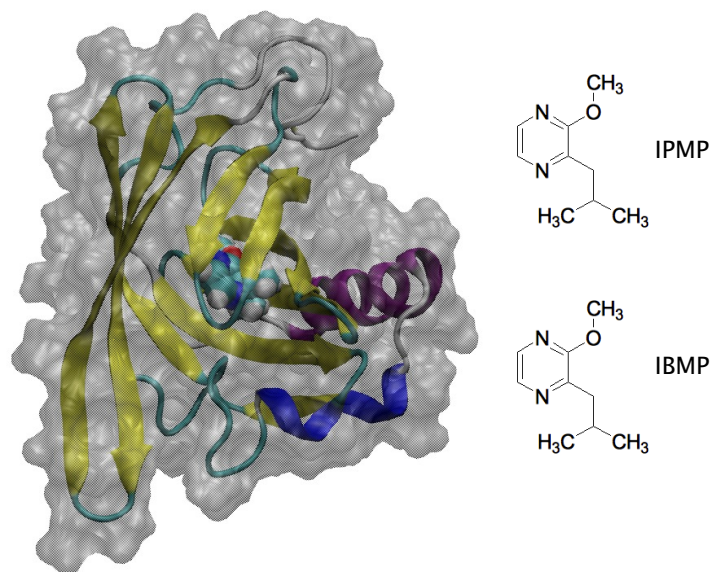


Figure 3. Binding structure of IPMP to MUP and two chemical formulas of IPMP and IBMP.

We used the FUJI force field for MUP and ligands. The point charges of the ligands were determined by the restrained electrostatic potential fit (RESP) using two conformations for each ligand [16]. After 20 ns equilibration from X-ray binding structures (PDB: 1QY1, 1QY2) we performed free energy calculations using 32 λ points and 12 trajectories for each λ point. The details of the MP-CAFEE procedure are explained in [3]. Calculated free energies and experimental binding free energies are listed in table 2. They show excellent agreement between theory and experiment.

In order to examine how much the force field difference affects the binding free energy, we performed the same binding free energy calculation using the ff99SB-ildn force field which was improved by ff99SB by D. E. Shaw's research [17]. The difference is only the force field for MUP. The ligand force field and other calculation details are the same as the FUJI force field calculation. The ff99SB-ildn gave 1.9 kcal/mol stronger binding free energy than the FUJI force field (figure 4). It seems reasonable because ff99SB has larger torsional energy barrier between the lowest energy position ($\phi = -80$) and the second lowest energy position ($\phi = -180$) in figure 1(a). But its deviation from the experimental value is too large for the computer aided drug design.

4. Drug Design on K computer

We believe that the most important issue for computational physics is to prepare correct input data. "Garbage in, garbage out" is always true on computers. The force field is vital to all atom molecular dynamics on biomolecules. As the FUJI force field seems enough accurate for computer aided drug design, we start to design small molecules for pharmaceutical target proteins with MP-CAFEE.

K computer is one of the most powerful computers; it has 88,128 CPU nodes. We calculated the binding free energies of 40 small molecules to a pharmaceutical target protein in one parallel job using 15,360 CPU nodes of K computer. Some of the 40 molecules were synthesized and their calculated binding free energies agreed with experimental measurements. Now is the time to utilize the MP-CAFEE with FUJI force field in order to accelerate the drug development by K computer.

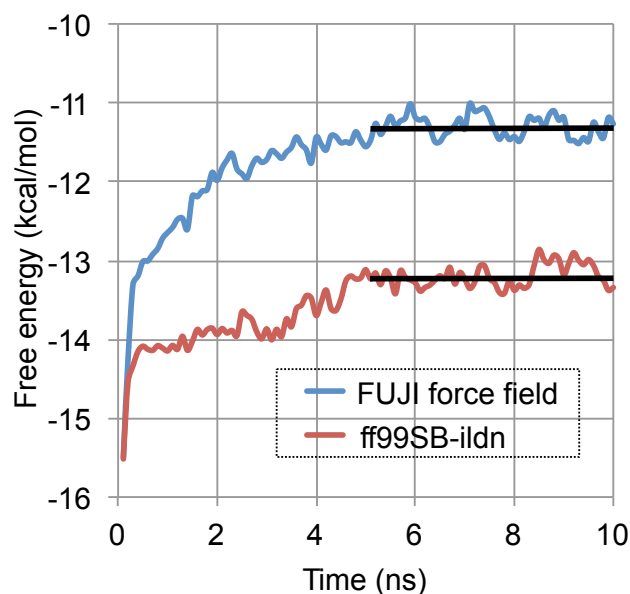


Figure 4. Binding free energy convergence of IBMP to MUP calculated with two different force fields: FUJI force field and ff99SB-ildn.

Acknowledgments

We like to thank all our collaborators on the technology development for the computer aided drug design. This research has been funded by MEXT SPIRE Supercomputational Life Science and by FIRST Kodama project in Japan.

References

- [1] Kollman P A 1993 *Chem. Rev.* **93** 2395–2417
- [2] Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow C D, Shirts M R, Sorin E J and Pande V S 2005 *J. Chem. Phys.* **123** 084108
- [3] Fujitani H, Tanida Y and Matsuura A 2009 *Phys. Rev. E* **79** 021914
- [4] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C 2006 *Proteins: Struct. Funct. Bioinf.* **65** 712–725
- [5] Wang J, Cieplak P and Kollman P A 2000 *J. Comput. Chem.* **21** 1049–1074
- [6] Duan Y, Wu C, Chowdhury S, Lee M C, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J and Kollman P 2003 **24** 1999–2012
- [7] Cornell W D, Cieplak P, Bayly C I, Gould I R, Merz, Jr K M, Ferguson D M, Spellmeyer D C, Fox T, Caldwell J W and Kollman P A 1995 *J. Am. Chem. Soc.* **117** 5179–5197
- [8] Fujitani H, Matsuura A, Sakai S, Sato H and Tanida Y 2009 *J. Chem. Theory Comput.* **5** 1155–1165
- [9] Grdadolnik J, Grdadolnik S G and Avbelj F 2008 *J. Phys. Chem. B* **112** 2712–2718
- [10] Grdadolnik J, Mohacek-Grosec V, Baldwin R L and Avbelj F 2011 *Proc. Natl. Acad. Sci.* 1794–1798
- [11] de M Seabra G, Walker R C and Roitberg A E 2009 *J. Phys. Chem. A* **113** 11938–11948
- [12] Vymětal J and Vondrášek J 2010 *J. Phys. Chem. B* **114** 5632–5642
- [13] Vymětal J and Vondrášek J 2011 *Chem. Phys. Lett.* **503** 301–304
- [14] Hess B, Kutzner C, van der Spoel D and Lindahl E 2008 *J. Chem. Theory Comput.* **4** 435–447
- [15] Bingham R J, Findlay J B C, Hsieh S Y, Kalverda A P, Kjellberg A, Perazzolo C, Phillips S E V, Seshadri K, Trinh C H, Turnbull W B, Bodenhausen G and Homans S W 2004 *J. Am. Chem. Soc.* **126** 1675–1681
- [16] Bayly C I, Cieplak P, Cornell W D and Kollman P A 1993 *J. Phys. Chem.* **97** 10269
- [17] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J L, Dror R O and Shaw D E 2010 *Proteins: Struct. Funct. Bioinf.* **78** 1950–1958