Direct calculation of the binding free energies of FKBP ligands

Hideaki Fujitani,[a)] Yoshiaki Tanida, and Masakatsu Ito
Fujitsu Laboratories Ltd., 10-1 Morinosato-Wakamiya, Atsugi 243-0197, Japan

Guha Jayachandran
Department of Computer Science, Stanford University, Stanford, California 94305

Christopher D. Snow
Department of Biophysics, Stanford University, Stanford, California 94305

Michael R. Shirts,[b)] Eric J. Sorin, and Vijay S. Pande
Department of Chemistry, Stanford University, Stanford, California 94305

Direct calculations of the absolute free energies of binding for eight ligands to FKBP protein were performed using the Fujitsu BioServer massively parallel computer. Using the latest version of the general assisted model building with energy refinement (AMBER) force field for ligand model parameters and the Bennett acceptance ratio for computing free-energy differences, we obtained an excellent linear fit between the calculated and experimental binding free energies. The rms error from a linear fit is 0.4 kcal/mol for eight ligand complexes. In comparison with a previous study of the binding energies of these same eight ligand complexes, these results suggest that the use of improved model parameters can lead to more predictive binding estimates, and that these estimates can be obtained with significantly less computer time than previously thought. These findings make such direct methods more attractive for use in rational drug design. © *2005 American Institute of Physics.* [DOI: 10.1063/1.1999637]

I. INTRODUCTION

One of the biggest challenges in computational drug design is the accurate calculation of the free energy of binding of small ligands. If the ligand strongly binds to a target protein, it is a promising drug candidate, as it is likely to either interfere or enhance the activity of the target protein. Protein crystallography has been utilized since the early 1980's to discover the binding structure of protein-ligand complexes and has been instrumental in the development of structure-based drug design. Structure-based drug design has contributed significantly to the introduction of many compounds into clinical trials as well as drug approvals. Increasingly, computation has taken a significant role in drug design, from structure refinement using simulated annealing, to the development of the underlying molecular mechanics force fields, to structure visualization, and to the design molecular mechanics simulation of drug analogs.[1]

Many methods have been presented to calculate the binding free energy of biological ligands. Free-energy perturbation (FEP) and thermodynamic integration (TI) have been used frequently to obtain the free-energy difference between molecules by the "mutations" of small fragments,[2] but it is very inefficient for obtaining absolute binding energies. Recent improvements in methodology and computer power have made it possible to calculate with high precision the solvation free energy of small molecules using explicit solvent molecules.[3–6] However, the solute molecules in these studies consisted of less than 20 atoms. In general, it is extremely difficult to perform the binding free-energy calculations for the size of biomolecule and ligands commonly used as drugs with these methods.

Jarzynski proved that the distribution of nonequilibrium work values could yield an equilibrium free energy by taking the exponential average of the set of nonequilibrium work values.[7] This was shown as a particular example of a more general nonequilibrium identity[8] and FEP can be seen as a special case of "fast growth" nonequilibrium exponential work averaging.[9,10] But the results of exponential averaging strongly depend on the behavior at the tails of the distribution of work values. It is difficult to get the accurate free energy by the exponential averaging. Shirts *et al.* demonstrated that the Bennett acceptance ratio (BAR) method can be interpreted in terms of the maximum likelihood estimate of the free-energy difference given a set of nonequilibrium work values in the forward and reverse directions.[11,12] BAR yields a lower variance for equal simulation time than FEP or TI (Ref. 13) for many molecular simulations. Shirts *et al.* then applied the BAR to calculate the absolute binding free energies for eight FKBP ligand complexes[14] (Fig. 1) using the Folding@Home distributed computing system.[15] FKBP is best known as the target of the widely used immunosuppressive drug FK506 and is 107 residues long. Experimentally, the binding affinities of these ligands were determined directly by the inhibition constant (K_i) of rotamase activity,[16] all in the same manner, making it a good test system for binding affinities.

The direct calculation of the absolute binding free en-

a)Electronic mail: fjtani@labs.fujitsu.com
b)Present address: Havemeyer Hall Mail Code 3131, Department of Chemistry Columbia University, New York, NY 10027.
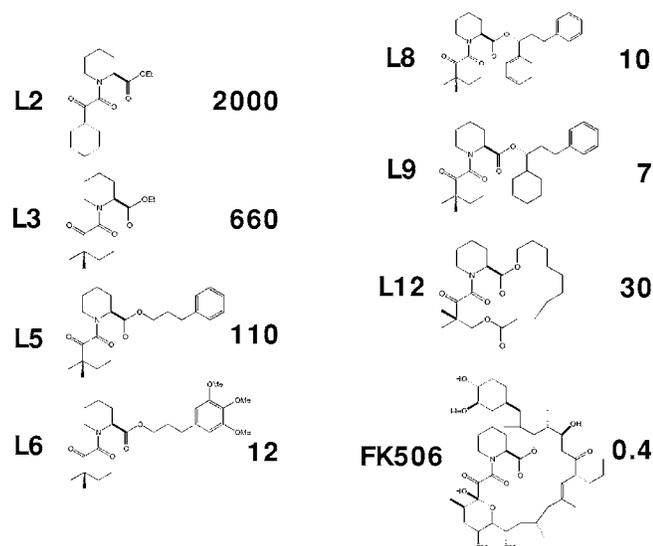
123, 084108-1

FIG. 1. Structures of eight ligands investigated. Experimental inhibition constants in nM are on the right side of the structure.

ergy for pharmaceutical target proteins like FKBP is not only a significant scientific challenge, but also a potentially extremely valuable technology for the pharmaceutical industry. Fujitsu developed a new massively parallel machine (known as the BioServer) to perform these types of calculations.

The huge computational power of the Folding@Home system makes it possible to perform very comprehensive sampling using molecular dynamics of the complex biochemical systems. Shirts *et al.* used about 400 000 CPU days for the eight binding-energy calculations and obtained binding free energies with a rms error from a linear fit of about 1 kcal/mol and a maximum deviation from the fit of 2 kcal/mol. We performed similar calculations on the Fujitsu BioServer and obtained more predictive binding free energies with a rms error from a linear fit of only 0.4 kcal/mol and a maximum deviation from the fit of 0.6 kcal/mol, using very similar methodologies as the previous study, but with significantly less computational power (about 50 000 CPU days).

II. METHOD

A sufficiently accurate force field is vital in order to get an accurate binding free energy of a protein and ligand. Wang *et al.* recently developed the general AMBER force field (GAFF) for organic molecules.[17] GAFF is designed to be compatible with the existing AMBER force fields for proteins and nucleic acids[18] and has parameters for most organic and pharmaceutical molecules that are composed of H, C, N, O, S, P, and halogens. We incorporated the Amber99 force field and the new GAFF parameters with the GROMACS package.[19,20] The ligand charges were calculated by the AM1-BCC method[21] using MOPAC2002. The previous Folding@Home calculation used older GAFF parameters which was distributed with AMBER version 7. This difference in ligand parameters appears to have improved the the predictability of the binding energies relative to the previous calculation.[14]

Simulation methodology and changes in GROMACS were taken directly from the previous study,[14] except as described here. The single precision version of GROMACS-3.1.4 was used instead of the double precision version for greater speed, and the double-annihilation binding-energy method[22] was performed instead of double decoupling.[23–25] All molecular-dynamics simulations were performed at 298 K with the Nose-Hoover temperature control[26,27] and at 1.0 atm using the Berendsen pressure control.[28] We used TIP3P water model, LINCS with order 8 to constrain all bonds, and 2-fs time step.[29] A long-range correction for the finite cutoff of the Lennard-Jones potential was taken into account for energy and pressure corrections, but not for the binding free energy afterward. The previous study used modified terms for the backbone torsions and excluded the protein 1-4 H, H terms from the calculation, whereas this study used AMBER(ff99) including the 1-4 H, H terms.[18]

All pre-equilibration binding configurations were the same as the previous work.[14] To prepare the initial configuration after the insertion of water around the solutes, we performed energy minimization using the conjugate gradient method and then a molecular-dynamics simulation for 30 ps with the solute positions restrained. The double-annihilation method requires two free-energy calculations: one for the solvated ligand system and the other for the solvated FKBP-ligand system. The former had about 500 TIP3P water molecules and the latter had about 5000, the exact number depending on the ligand. Before the free-energy calculation, we performed a long equilibration at 298 K with the full interactions between the ligands and their surroundings, 5 ns for the solvated ligand system and from 10 to 20 ns for the solvated FKBP-ligand system. This initial equilibration is an important difference between this study and the previous study,[14] where only 2-ns equilibration (though for each intermediate state) was performed from the initial structure before collecting data.

The nonbonding interaction between the ligand and other molecules is parametrized by λ. We used 12 λ points to reduce Coulomb interaction and 21 λ points to reduce van der Waals interaction with a soft core potential.[14] In both the solvated ligand system and the solvated FKBP-ligand system, we used λ points of 0.0, 0.1, 0.25, 0.45, 0.55, 0.65, 0.7, 0.75, 0.8, 0.9, 0.95, and 1.0 for the Coulomb interaction and λ points of 0.0, 0.1, 0.2, 0.275, 0.375, 0.45, 0.55, 0.65, 0.675, 0.725, 0.75, 0.775, 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, and 1.0 for the van der Waals interaction. These 33 (12+21) independent molecular-dynamics simulations had the same initial equilibrated coordinates and the same randomly generated initial momentum distribution at 298 K, but different λ values.

In order to calculate the free energy, we used 12 independent molecular-dynamics simulations with different initial momentum distributions and the same equilibrated structures described above. Thus, 396 (33×12) independent molecular-dynamics simulations were performed for each system. Potential-energy differences between independent samples neighboring in λ were used as work values to find the free energy by BAR between each λ state,[11,14] and the results from the 12 simulations were averaged for each intermediate state.
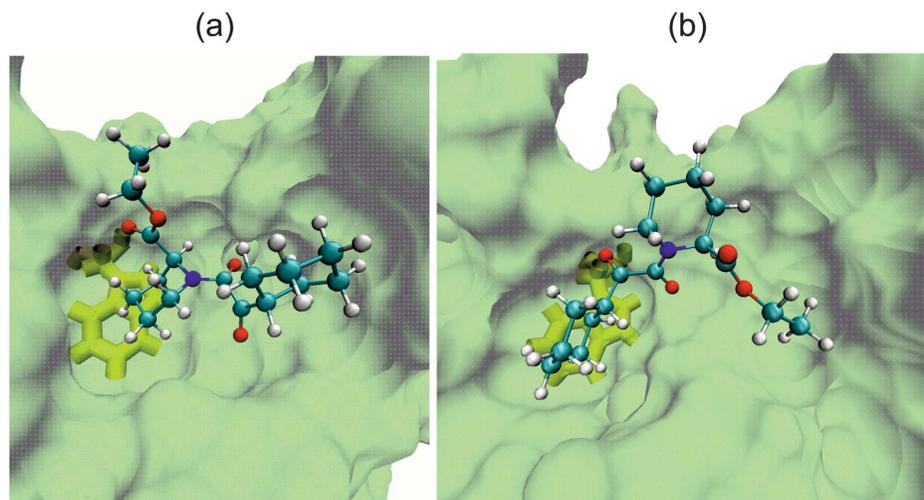
(a)   (b)

FIG. 2. (Color) Bound structures of the FKBP-L2 complex. (a) Initial modeled structure and (b) final structure after the 20-ns equilibration simulation. The indole of Trp-59 at the bottom of the hydrophobic pocket is transparently shown in yellow.

The BioServer has 1920 FR-V processors (high-performance and low-power processor with 8-way very-long instruction word (VLIW) architecture made by Fujitsu[30]) in a rack $(90 \times 60 \times 200 \text{ cm}^3)$. Each processor has 256-MB memory running the LINUX operating system, allowing us to run independent GROMACS molecular-dynamics simulations in the rack, thereby making it very suitable to perform the binding free-energy calculations described here.

III. RESULT AND DISCUSSION

The atomic structures for the L8, L9, and FK506 complexes with FKBP have been determined by x-ray crystallography,[16] but no such structures existed for other ligands. As all eight ligands have common binding elements, a pipecolate and an α-keto amide region (Fig. 1), we assumed that the binding atomic structures of the other ligands are similar to the known three complexes. Modeled initial structures are described in the previous study,[14] after which the several nanosecond-long initial equilibration of the complexes and ligands described in Sec. II were performed.

At the beginning of the simulations, the pipecolate in all the ligands sat atop of the indole of 59th Tryptophan (Trp-59) at the bottom of the hydrophobic pocket of FKBP. During the long equilibration at 298 K the pipecolate of L2 and L3 moved from the bottom to the top within the hydrophobic pocket, while the pipecolate of the other ligands stayed at the bottom of the hydrophobic pocket. L2 rotated around until 7 ns and then stabilized at about 13 ns (Fig. 2) during the 20-ns equilibration.

Using these equilibrated binding structures, we performed massively parallel binding free-energy calculations on the BioServer described in the method. Figure 3 shows the convergence of the free energy of FK506 over time. The upper figure shows the solvated ligand system and the lower one shows the solvated FKBP-ligand system. Each circle represents the free energy calculated over 100-ps windows averaged over all 12 samples. The error bar shows the minimum and maximum values from the 12 samples. Since the coupling of the ligands' nonbonding interaction (λ) changes from one to the 33 different values at the beginning of the free-energy calculation, the 100-ps free energy in Fig. 3

changes mainly at the beginning and fluctuates a little afterward. Because of this need for equilibration to the intermediate λ states, we use work values after the equilibration in order to calculate the time-averaged free energy by BAR. The equilibration time depends on the ligand. For FK506 we used work values from 200 ps to 1 ns for the solvated ligand
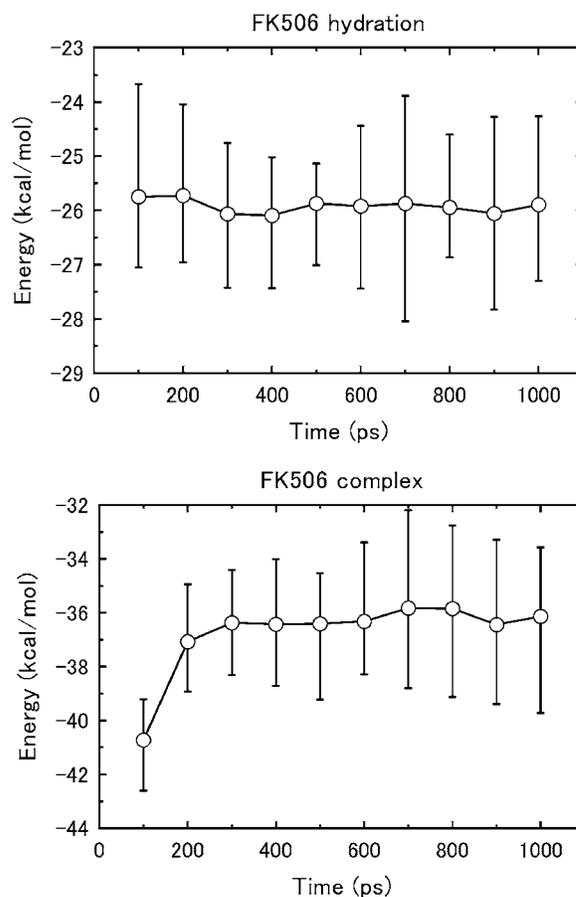

FIG. 3. Free energy vs simulation time for the FKBP-FK506 complex. Each circle represents the free energy calculated using BAR over 100-ps simulation, then averaged among the 12 independent simulations. The error bar indicates minimum and maximum values of the 12 samples. The upper figure shows the free energy of solvation of the FK506 alone, and the lower figure shows the free energy of decoupling FK506 from the solvated FKBP-FK506 complex.
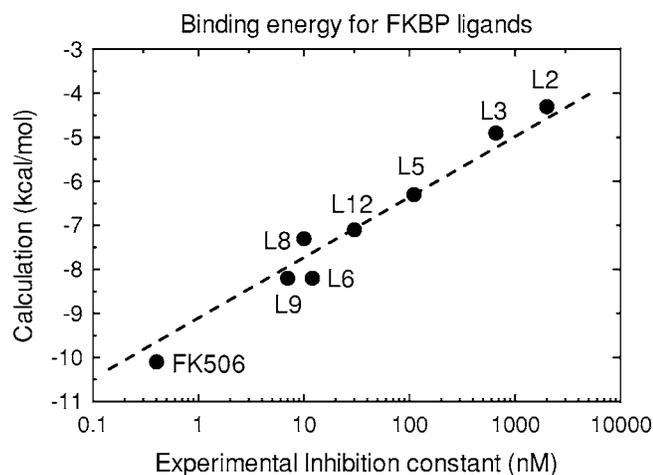
FIG. 4. Calculated absolute binding free energies vs experimental inhibition constants for eight ligands. The dashed line indicates $\Delta G = RT \ln(K_i) + 3.2$.

system and from 600 ps to 1 ns for the solvated FKBP-ligand system. The estimated free-energy value of the annihilation of the ligand from complex water system minus the free energy of annihilation of the ligand from the solvated ligand system gives our estimation of the final absolute binding free energy of the models of FKBP and FK506.

Figure 4 shows experimental inhibition constants[16] (K_i) versus our calculated binding free energies (ΔG) for eight FKBP ligand complexes. We here make the approximation that the K_i values are equal to the binding constant (K_d), although this is not always strictly true for the rotamase assay used in these binding experiments, especially for weakly binding ligands. The calculated values show excellent agreement to a linear fit to the log values of the inhibition constant, with a rms error from the linear fit of only 0.4 kcal/mol and a maximum deviation from the fit of 0.6 kcal/mol. This is significantly better than the previous study, with a rms error of near 1 kcal/mol and a maximum deviation near 2 kcal/mol, despite the fact that this study used significantly less computer time.

There are two main explanations for this improved performance. First is the improved parametrizations for the ligand. Wang *et al.* reported the performance tests of the new GAFF parametrization. They compared the relative energies of 71 conformation pairs with experimental values which were used in the development of AMBER99 force field. The new GAFF gave the rms error of 0.5 kcal/mol to experimental values.[17] This improved parametrization suggests that we may be able to get within experimental predictability. The second reason may be the improved equilibration of the initial structures. As for L2 and L3 complexes, we used the binding structures obtained after the long molecular-dynamics equilibration. Calculating the free energy from the correct binding mode may be necessary to get accurate free energies.

The center line in Fig. 4 indicates $\Delta G = RT \ln(K_i) + 3.2$. this means that while the slope of the correlation is 1, our binding energies are uniformly 3.2 kcal/mol smaller than the experimental value. Solvation energy calculations of the amino acid side chain analogs give unfavorable hydration energies.[3–5] As the ligands are built of the same small mol-

ecules as the side chains, we would expect ligands to have the same behavior. Unfortunately, unfavorable predicted hydration behavior would result in binding that was too tight compared to the experiment. This can be seen as follows: if we assume that solute-solute interactions are approximately correct (as we expect them to be, as most molecular mechanics solutes were parametrized for solute-solute interactions such as bulk liquid properties) and the hydration of computed ligands is disfavored with respect to hydration, then the computed ligands would be driven to stronger interactions with the protein. The fact that the binding energies are less strong than experiment is a matter that requires further study.

IV. SUMMARY

We calculated the absolute binding free energies for eight FKBP ligand complexes using the Fujitsu BioServer massively parallel computer. Using the latest GAFF parameters and with significantly less computation power than was used in a similar Folding@Home study, we obtained pharmaceutically useful predictability between the calculated and experimental absolute binding energies. The ability to perform these calculations demonstrates the utility of massively parallel computation resources such as the BioServer in rational drug design.

[1] W. L. Jorgensen, Science **303**, 1813 (2004).
[2] P. F. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[3] A. Villa and A. E. Mark, J. Comput. Chem. **23**, 548 (2002).
[4] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).
[5] J. L. Maccallum and D. P. Tieleman, J. Comput. Chem. **24**, 1930 (2003).
[6] J. T. Wescott, L. R. Fisher, and S. Hanna, J. Chem. Phys. **116**, 2361 (2002).
[7] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[8] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[9] D. A. Hendrix and C. Jarzynxki, J. Chem. Phys. **114**, 5974 (2001).
[10] G. Hummer, J. Chem. Phys. **114**, 7330 (2002).
[11] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[12] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[13] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 144107 (2005).
[14] M. R. Shirts, G. Jayachandran, C. D. Snow, H. Fujitani, E. J. Sorin, and V. S. Pande, J. Comput. Chem. (unpublished).
[15] B. Zagrovic, E. J. Sorin, and V. S. Pande, J. Mol. Biol. **313**, 151 (2001).
[16] D. A. Holt, J. I. Luengo, D. S. Yamashita *et al.*, J. Am. Chem. Soc. **115**, 9925 (1993).
[17] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[18] J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).
[19] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).

[20] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).

[21] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132 (2000); A. Jakalian, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **23**, 1623 (2002).

[22] W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, J. Chem. Phys. **89**, 3742 (1988).

[23] B. Roux, M. Nina, R. Pomes, and J. C. Smith, Biophys. J. **71**, 670 (1996).

[24] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. **72**, 1047 (1997).

[25] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, J. Phys. Chem. B **107**, 9535 (2003).

[26] S. Nose, Mol. Phys. **52**, 255 (1984).

[27] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[28] H. J. Berendsen, J. P. Postma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[29] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[30] H. Okano, A. Suga, T. Shiota *et al.*, Dig. Tech. Pap.-IEEE Int. Solid-State Circuits Conf., 374 (2002).