Binding Free Energy of Protein/Ligand Complexes Calculated using Dissociation Parallel Cascade Selection Molecular Dynamics and Markov State Model

Hiroaki Hata[1†], Duy Phuoc Tran[1†], Mohamed Marzouk[1,2], Akio Kitao[1*]

[1] School of Life Sciences and Technology, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, Japan

[2] Physics Department, Faculty of Science, Ain Shams University, 11566, Cairo, Egypt

†These authors contributed equally.

*Corresponding author: Akio Kitao, School of Life Science and Technology, Tokyo Institute of Technology, M6-13, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, Japan; Tel: +81-3-5734-3373, Fax: +81-3-5734-3372, E-mail: akitao@bio.titech.ac.jp

Abstract

We recently proposed a computational procedure to simulate the dissociation of protein/ligand complexes using the dissociation Parallel Cascade Selection Molecular Dynamics simulation (dPaCS-MD) method and to analyze the generated trajectories using the Markov state model (MSM). This procedure, called dPaCS-MD/MSM, enables calculation of the dissociation free energy profile and the standard binding free energy. To examine whether this method can reproduce experimentally determined binding free energies for a variety of systems, we used it to investigate the dissociation of three protein/ligand complexes: trypsin/benzamine, FKBP/FK506, and adenosine A_{2A} receptor/T4E. First, dPaCS-MD generated multiple dissociation pathways within a reasonable computational time for all the complexes, although the complexes differed significantly in the size of the molecules and in intermolecular interactions. Subsequent MSM analyses produced free energy profiles for the dissociations, which provided insights into how each ligand dissociates from the protein. The standard binding free energies obtained by dPaCS-MD/MSM are in good agreement with experimental values for all the complexes. We conclude that dPaCS-MD/MSM can accurately calculate the binding free energies of these complexes.

Keywords:

PaCS-MD, binding free energy calculation, enhanced sampling method, Markov state model, molecular dynamics simulation

Significance

We benchmarked the combination of dissociation Parallel Cascade Selection Molecular Dynamics simulations and the Markov State Model, called dPaCS-MD/MSM, using three protein/ligand complexes: trypsin/benzamine, FKBP/FK506, and adenosine A2a receptor/T4E. The obtained standard binding free energies indicate that dPaCS-MD/MSM can efficiently calculate the binding affinities of the given complexes.

1. Introduction

Molecular binding/unbinding is involved in most molecular-level biological processes. Thus, understanding how biomolecules recognize each other and function upon binding is essential for furthering our understanding of biology and for drug development. Binding free energy is often used to determine the affinity of biomolecular interactions and the efficacy of drugs. Complete characterization of binding-competent protein conformations, ligand binding poses, and binding/unbinding kinetics is therefore needed for a thorough understanding of protein/ligand binding. Obtaining such information on drugs would significantly enhance the computational drug design process and maximize drug efficacy. Numerous methods for free energy calculation have been developed and applied to the binding of small compounds (ligands) with proteins, such as thermodynamic integration, free energy perturbation, and the adaptive biasing force technique [1,2].

Molecular dynamics (MD) is a computer simulation technique now routinely used in many fields related to biomolecules for examining dynamic properties and processes. Many features can be investigated by all-atom MD, such as the stability of biological macromolecules [3], conformational properties, the impact of dynamics on enzyme activity [4], molecular recognition and properties of complexes [5], protein association [6], protein folding [7], and other aspects. MD and related methods are widely used for binding free energy calculations [8]. Moreover, MD simulations can provide atomic details of interaction dynamics between proteins, ligands, and solvent molecules, and thus structural flexibility and entropic effects are explicitly considered in free energy calculations. However, MD simulations cannot sample across multiple energy minima within an affordable computational time if high energy barriers separate the minima. Accordingly, conventional MD simulations of protein/ligand binding and unbinding processes or other time-consuming events still require substantial computational resources, although the use of specialized MD engines for graphical processing units (GPUs) greatly reduce the computational cost [9,10]. Considerable recent research thus focuses on developing novel methods to improve the sampling of time-consuming events to overcome the limitations of conventional MD simulations [11].

The processes of binding/unbinding must be observed to calculate the free energy profile along the processes. To this end, various enhanced sampling methods have been developed, e.g.,

umbrella sampling [12], replica exchange MD [13], accelerated MD [14], steered MD [15], targeted MD [16], and metadynamics [17]. Some of the methods effectively enhance the movements of molecules by applying a bias force, which requires careful parameter tuning to avoid artifacts. Alternatively, parallel cascade selection molecular dynamics (PaCS-MD) is an enhanced simulation method that does not apply a bias force [18,19], similar to other simulation methods conducted as combinations of multiple unbiased MD methods, such as forward flux sampling [20,21], weighted ensemble [22,23], and milestoning [24,25]. These methods improve sampling efficiency and enable ligand binding/unbinding simulations of proteins. PaCS-MD comprises cycles of multiple parallel short (typically 0.1 ns) MD simulations combined with initial structure selection. The repetition of parallel MD simulations from selected promising structures with regenerated initial atom velocities drastically enhances the probability of observing the dissociation of protein/ligand complexes by selecting snapshots with longer protein–ligand distances [26–28]. Cycles of PaCS-MD generate dissociation pathways as a series of multiple MD trajectories that mutually overlap in conformational space.

The use of an appropriate method to analyze these trajectories is a critical step in determining the binding free energy of a complex. The Markov state model (MSM) is a powerful analysis method in computational biology for identifying stationary states and kinetic details of protein dynamics from MD simulation data [29]. MSM provides information on the physical process defined as a set of transitions between discretized metastable states. A dynamic description of simulated unbinding processes can be obtained by constructing an MSM from many short PaCS-MD trajectories. By selecting snapshots with longer intermolecular distances, our group previously used dissociation PaCS-MD (dPaCS-MD) to generate dissociation pathways of tri-N-acetyl-d-glucosamine from hen egg white lysozyme [26], of the transactivation domain of p53 protein (p53-TAD) from murine double-minute clone 2 protein (MDM2) protein [30], and of an N-terminal fragment of the bacterial flagellar rotor protein FliM from the signaling protein CheY [28]. We built an MSM model using dPaCS-MD trajectories to investigate the dissociation mechanisms by describing states based on ligand/protein geometry [29,31]. Using this combination, we obtained binding free energies of the protein/ligand complexes in accordance with the experimentally determined values. These successes motivated us to extend dPaCS-MD/MSM to different types of

protein/ligand complexes to demonstrate accuracy and computational efficiency in binding free energy calculations.

In this work, we investigated the dissociation of three protein/ligand complexes: trypsin/benzamine, FKBP/FK506, and adenosine A_{2A} receptor/T4E. The first complex comprises the inhibitor benzamidine bound to bovine trypsin, an enzyme that degrades dietary proteins [32]. We investigated the dissociation process of a relatively small ligand using this complex. The second complex investigated was FK506 (tacrolimus), which inhibits T-cell activation and is effective in organ transplantation because it binds to FK506 binding protein (FKBP). FKBP is an immunophilin protein involved in the regulation of T-cell activation and inhibition of the enzymatic activity, thus affecting different signal transduction pathways [33]. FK506 is larger and more flexible than benzamine. Both of these systems are commonly used as benchmarks for calculating binding energy and kinetic rates and have been measured both experimentally and computationally. The third complex involves adenosine A_{2A} receptor (A_{2A}), a member of the G protein-coupled receptor (GPCR) superfamily. A_{2A} drugs have been developed to address wound healing, vascular diseases such as atherosclerosis, restenosis, and platelet activation, and inflammation and cancer [34]. 4-(3-Amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (T4E) was developed as an antagonist of A_{2A} [35]. The A_{2A} receptor/T4E complex is a more challenging target compared to the other two because T4E binds deeply inside the binding cavity of A_{2A}. Here, we applied the dPaCS-MD/MSM scheme to each of the three complexes by conducting multiple dPaCS-MD and free energy calculations with MSM. The unbinding pathways obtained by dPaCS-MD showed differences between the unbinding mechanisms of the complexes. Subsequent MSM analyses provided free energy profiles along the dissociation pathway. Finally, we compared the standard finding free energies calculated by dPaCS-MD/MSM with the experimental values, thereby demonstrating that dPaCS-MD/MSM can accurately calculate the binding free energies of these complexes.

2. Methods

2.1. Simulation systems

The simulated protein/ligand complexes are listed in Table 1. For the trypsin/benzamine complex, the crystal structure (PDB ID: 3ATL [36]) was immersed in a cubic water box with an edge length of 111 Å and containing KCl at a concentration of 150 mM (approximately 140,000 atoms, Fig. 1a). Protonation states at pH 8.5 were chosen using PDB2PQR [37] according to the crystallization condition. For the FKBP/FK506 complex, the crystal structure (PDB ID: 1FKF [33]) was placed in a cubic water box with an edge length of 117 Å and containing NaCl at a concentration of 150 mM (~ 120,000 atoms, Fig. 2a). Protonation states were chosen using AmberTools [38]. For the A_{2A}/T4E complex, the crystal structure (PDB id: 3UZC [35]) was embedded in a membrane using CHARMM-GUI [39]), which contains 210 DMPC lipid molecules. The initial box size was 82 × 82 × 138 $Å^3$. The Amber ff14SB force field [40] and the SPC/E_b water model [41] were used for all simulations. The ligand parameters were generated using the Antechamber module in the AMBER18 package [42,43] with GAFF and AM1-BCC [44].

2.2. Simulation procedure

The soluble trypsin/benzamine and FKBP/FK506 complexes were simulated using the GPU implementation [45] of the PMEMD module in the Amber package, and the A_{2A}/T4E complex embedded in a membrane was simulated by using GROMACS [46]. For the first two cases, the system was energy minimized and equilibrated in an NPT ensemble (300 K, 1 bar) for 10 ns by conventional MD simulation. For the A_{2A}/T4E complex, a 100 ns relaxation MD was conducted with positional restrains imposed on the protein and ligand, followed by free equilibration by 100 ns MD without restraints. A Langevin thermostat [47] with a friction parameter of 2 ps^{-1} and a Berendsen barostat [48] were used for temperature and pressure control, respectively. Equations of motion were integrated with a time step of 2 fs. Covalent bonds involving hydrogens were constrained using the SHAKE algorithm [49], and the water molecules were kept rigid using the SETTLE algorithm [50]. The long-range Coulomb energy was evaluated

using the particle mesh Ewald method [51], while real space non-bonded interactions were evaluated at a cutoff distance of 12 Å. For the A2A/T4E complex, the same settings were used except that the covalent bonds involving hydrogens were constrained using the LINCS algorithm [52] and the temperature and pressure were controlled by a Hoover-Nose thermostat [53,54] and a MTTK barostat [55], respectively. Three trials of PaCS-MD were conducted for the first two cases and five trials were performed for the A2A/T4E complex.

The dissociation of protein/ligand complexes simulated using PaCS-MD [56] according to the procedure was described earlier [28]. Ten parallel MDs (replicas) were used for the trypsin/benzamine and FKBP/FK506 complexes and 30 replicas were employed for A_{2A}/T4E. As the initial structures for dPaCS-MD, ten structures were selected every 1 ns from the 10 ns conventional MD trajectories for the first two cases. For the A_{2A}/T4E complex, the equilibration MD was extended for 1 ns, the obtained trajectories were clustered into 30 clusters, and these representative 30 structures were used as the initial structures. For each cycle of dPaCS-MD, 0.1 ns MD simulations were conducted in parallel. Molecular structures were saved every 40 fs and used for MSM analysis. The structures obtained every 1 ps from each cycle were rank-ordered according to the distance d between the center-of-mass positions of the protein and ligand, and the top ten structures were selected as the initial structures for the next PaCS-MD cycle. dPaCS-MD was stopped when two molecules were sufficiently separated (d = 35, 40 and 60 Å for trypsin/benzamine, FKBP/FK506, and A_{2A}/T4E, respectively). MD trajectories were analyzed using the cpptraj module in the AmberTools14 package. The number of intermolecular contacts was calculated from heavy-atom contacts between the protein and ligand with a distance threshold of 4.5 Å.

2.3. Free energy calculation

The Gibbs free energy change ΔG with the dissociation of a protein-ligand complex was calculated by MSM analysis of the MD trajectories generated by each PaCS-MD trial according to the procedure described previously [28]. To achieve sufficient statistics to construct a reasonable

MSM, molecular structures were additionally sampled by multiple 0.1 ns MD simulations started from the PaCS-MD initial structures of each cycle with different random velocities until the implied time scale sufficiently converged to a plateau value [57]. For the trypsin/benzamidine complex, the total number of MD structures was tripled by additional MD simulations for two of the three observed pathways, whereas the third pathway did not require additional MD simulation. For the FKBP/FK506 complex, the total number of MD structures was tripled for all three pathways by additional MD simulations. For the A2A/T4E complex, 30 replicas were used in dPaCS-MD, and no additional MD simulation was required.

MSM analysis was performed using MSMBuilder 3.5.0 [58]. Microstates were determined by k-means clustering ($k = 25$) for d values. Lag times of 48 ps (trypsin/benzamidine), 32 ps (FKBP/FK506) and 50 ps (A_{2A}/T4E) were selected based on the implied time scale test [57]. The ΔG value for the i-th microstate ΔG_i was calculated from the stationary probability of the microstate π_i:

$$\Delta G_i = -k_B T \ln \frac{\pi_i}{max_j \pi_j},$$

where k_B is the Boltzmann constant and T is the absolute temperature [59]. The standard binding free energy $\Delta G°$ was obtained from the free energy profile and the correction term ΔG_v originating from the difference between the sampled bound volume V_b and the standard volume $V°$ ($= 1661$ Å) [60]:

$$\Delta G° = -\Delta G + \Delta G_v,$$

where ΔG is the average value of ΔG_i included in the unbound state. ΔG_v was calculated by:

$$\Delta G_v = -k_B T \ln \frac{V_b}{V°}.$$

The bound state was defined as a region before the free energy curve becomes flat, i.e., a *d* of 22 Å for the trypsin/benzamine and FKBP/FK506 complexes, and a *d* of 27 Å for A2a/T4E. The range in which the energy curve is flat was defined as the unbound state, which was 30 Å

(trypsin/benzamidine), 35 Å (FKBP/FK506), and 45 Å (A_{2A}/T4E). The value of V_b was obtained as the volume of the convex hull defined by the center-of-mass coordinates of the ligand in the bound state relative to the center-of-mass of the protein at the origin. The convex hull calculation was performed for each dissociation simulation using Qhull [61].

3. Results and discussion

3.1. Trypsin/benzamidine complex

Trypsin is a protease found in many vertebrate species and benzamidine is a competitive inhibitor of trypsin [62]. The binding between trypsin and benzamidne has been thoroughly investigated as an exemplar of biomolecular binding [63], and the binding free energy of the complex has been measured both experimentally [64] and computationally [60,65].

Dissociation of the trypsin/benzamidine complex was simulated three times by dPaCS-MD. The distance between the center-of-mass positions of trypsin and benzamidine, d, gradually increased with increasing number of PaCS-MD cycles, with back-and-forth motions of benzamidine around trypsin (Fig. 1b). At $d = 27.1 \pm 0.7$ Å (average $\pm$ SD of the three simulations), the complex dissociated completely, i.e., the number of inter-molecular contacts became zero (Fig. 1b, broken line). The number of PaCS-MD cycles required for the complete dissociation was 27 ± 10 Å. The corresponding computational time was 2.7 ns (0.1 ns × 27 cycles) on average, and the total computational cost was 107 ns (0.1 ns × 27 cycles × (10 + an additional 30 replicas)). The experimental dissociation rate constant k_{off} was 600 s^{-1}, and the corresponding dissociation time was 1.67 ms [64], which means that PaCS-MD observed the dissociation within a computational time shorter by six orders of magnitude than the actual dissociation time.

Figure 1c depicts the dissociation pathways for benzamidine from trypsin shown by the representative PaCS-MD trajectories (from each cycle, the trajectory with maximum d was selected). Since PaCS-MD did not apply any artificial bias force to enhance dissociation, the

observed possible dissociation pathways indicate how benzamidine unbinds from trypsin. All the pathways behaved similarly at the beginning of dissociation, but clearly differed after detachment from trypsin. Since benzamidine is mostly buried in trypsin in the bound state, the dissociation direction is strictly restrained in the initial process of dissociation. This is in agreement with the binding process reported previously using massive MD simulations combined with MSM, showing several binding pathways from the surface of trypsin to its binding pocket [60].

The Gibbs free energy change ΔG along the dissociation pathway was obtained by MSM analysis (Fig. 1d). Statistically reliable MSMs were constructed using MD trajectories obtained from PaCS-MD and additional MD simulations (see Methods for details). In the range $d < 22$ Å, the ΔG value increased, then flattened where the intermolecular interactions were insubstantial. Thus, MD structures included in the flat region were defined as the unbound state, and the other region was considered as the bound state. We found a local free energy minimum at $d \sim 21$ Å in the position where benzamidine was located around the surface of trypsin (Fig. 1d, inset), in agreement with several metastable states in a simulation of the binding process [60].

The standard binding free energy $\Delta G°$ was calculated from the free energy difference between the bound and unbound states with a correction for ligand concentration (Table 1). The calculated $\Delta G°$ value was -6.1 ± 0.1 kcal/mol, in good agreement with the experimental values of -6.4 [66] and -7.3 kcal/mol [67]. The difference between the calculated and measured values is within the limit of force field accuracy, believed to be 1 to 2 kcal/mol [68].

3.2. FKBP/FK506 complex

FK506 binding protein (FKBP) has been identified in many eukaryotes, from yeast to human, and possesses prolylisomerase activity. FK506 is a flexible ligand and is much larger than benzamidine. The dissociation of FK506 from FKBP was simulated three times by PaCS-MD (Fig. 2b). The dissociation completed at $d = 26.0 \pm 3.6$ Å (Fig. 2b, dashed line) corresponds to a total computational time of 6 ± 2 ns, which is twice that required for trypsin/benzamidine dissociation.

The computational time spent observing dissociation should be related to the probability of escaping the free energy minima, i.e., the affinity of the protein/ligand complex. The experimental binding free energies are -12.9 kcal/mol [32] for the trypsin–benzamidine complex, respectively (Table 1). However, previous studies reported that the number of dPaCS-MD cycles strongly depends on the number of trials when a small number of MD replicas run in each PaCS-MD cycle due to low probabilities of fluctuation occurrence toward dissociation [26,28]. If the number of replicas is not sufficient for MSM, additional MD simulations should be conducted as in the cases of trypsin/benzamine and FKBP/FK506. Estimating the relative affinities of protein/ligand complexes directly from the PaCS-MD trajectories may require a sufficient number of MD replicas (≥ 30 at least).

The obtained three FKBP/FK506 pathways were different from the beginning of dissociation, unlike those for the trypsin-benzamidine complex (compare Fig. 2c and Fig. 1c). Since FK506 is bulky, it is not buried in the binding pocket of FKBP (Fig. 2a), and thus from the beginning, FK506 can move in multiple directions towards outside the pocket.

The free energy landscape of FKBP/FK506 dissociation obtained by PaCS-MD/MSM is shown in Fig. 2d. The ΔG value continuously increased along the dissociation pathways until $d \sim 22$ Å and then became flat, showing no metastable states during dissociation. The structures with $d < 22$ Å were defined as the bound state and the outside structures were considered as the unbound state. The calculated $\Delta G°$ value of -13.6 ± 1.6 kcal/mol agrees well with the experimental value of -12.9 kcal/mol (Table 1).

3.3. Adenosine A2a receptor/T4E complex

The third target in this study was the T4E ligand dissociating from its complex with the adenosine A_{2A} receptor (Fig. 3b). The dissociation processes was complete at $d = 41.0 \pm 2.2$ $\mathring{A}$ (dashed line in Fig. 3b). The computational time required to reach the dissociated state was 4.3 ± 1.1 ns. Since T4E is located deep inside the well-defined ligand binding pocket of the A_{2A} receptor in the bound state, the ligand requires significant time in the initial dissociation process to move from the native binding position to the entrance of the binding pocket (from a d value of 17 to 21

Å). Interestingly, all the trials required many cycles (132 ± 18 cycles) for d between 16.9 to 18.2 Å, during which the ligand tried to find an escape path to reach bulk water by exchanging positions with water molecules and attempting to break hydrogen bonds, specifically with ALA266 in extracellular loop 3 (EL3) and ILE81 in transmembrane helix 3 (TM3). In addition, water-mediated contacts with ASN254 in transmembrane helix 6 (TM6) via two water molecules contributed to this trap. In all trials, T4E first escaped the binding pocket by breaking interactions of its NH_2 group with the A2a receptor.

The obtained free energy plot of the dissociation process of T4E out of the A_{2A} receptor is shown in Fig. 3d. The free energy increased from $d \sim 12.5$ Å and reached a constant value at 32 Å. The subtle local minimum at ~ 17 Å is consistent with the position of the trap observed during the PaCS-MD simulations. The calculated $\Delta G°$ value of −14.3±1.2 kcal/mol is in good agreement with the experimental value of −13.2 kcal/mol [35].

4. Conclusion

In this study, we applied dPaCS-MD/MSM to three distinct protein/ligand complexes and evaluated the binding free energies. dPaCS-MD generated multiple dissociation pathways within a reasonable computational time for all three complexes, despite the complexes comprising significantly different sized molecules and involving different intermolecular interactions. The MSM analyses produced free energy profiles for the dissociations, providing insights into how the ligand dissociates from the protein. The calculated standard binding free energies agreed well with previously published experimental values. We therefore conclude that dPaCS-MD/MSM can accurately calculate the binding free energies of these complexes.

Although dPaCS-MD/MSM can be applied to different protein/ligand complexes, some factors should be considered. First, dPaCS-MD is conducted in a cubic water box, which freely allows rotation and dissociation of the complex in any direction. Howenver, PaCS-MD has a higher computational cost for larger non-globular complexes. One way to avoid this is the use of positional

restraints to stop the rotation of the protein, and the use of a rectangular box, with the long axis positioned along the direction of dissociation [26], although the dissociation directions are limited in this case. Second, a larger number of PaCS-MD cycles might be needed to observe the dissociation of a ligand bound deep inside the protein. In this case, dissociation might be enhanced by introducing other corrective variables for structure selection in PaCS-MD [69]. However, T4E situated deeply inside the binding cavity of adenosine A_{2A} receptor successfully dissociated only by the use of *d* as shown in this work. This was also confirmed for other A_{2A} ligands which are in progress in our laboratory.

Many MD calculations are distributed as replicas and conducted in parallel in PaCS-MD, and synchronization among the replicas within each cycle is not required. This advantage enables the production of unbiased dissociation pathways and accurate free energy profiles of protein/ligand dissociation within a reasonable computational time. Although the total computational cost might be large for larger systems or strongly bound complexes, parallelizability is suited for calculations using parallel computing supercomputers such as Summit, Sierra, and Fugaku, as well as distributed computing. In addition, PaCS-MD/MSM can be used to investigate not only the thermodynamics but also the kinetics of biomolecular interactions [27]. Calculations of kinetic rate constants (k_{on}, k_{off}, residence time, etc.) using PaCS-MD/MSM are currently under investigation in our laboratory.

Conflicts of Interest

The authors declare no competing financial interest.

Author Contributions

H.H., D.P.T., M.M., and A.K. contributed to the design of the work, conducted the overall analyses, and wrote the manuscript; H.H. and D.P.T. performed the computations.

Acknowledgments

This research was supported by MEXT/JSPS KAKENHI Nos. JP19H03191 and JP20H05439 to A.K. and JP19K23721 to D.P.T., and by MEXT "Program for Promoting Researches on the Supercomputer Fugaku" (Application of Molecular Dynamics Simulation to Precision Medicine Using Big Data Integration System for Drug Discovery) to A.K. This work used computational resources of the supercomputer TSUBAME provided by Tokyo Institute of Technology, FUGAKU through the HPCI System Research Project (Project ID: hp210029, hp210172, hp210177), the RCCS, The National Institute of Natural Science, and ISSP, The University of Tokyo.

References

[1] Lelièvre, T., Rousset, M. & Stoltz, G. *Free Energy Computations: A Mathematical Perspective*. (2010). DOI: 10.1142/P579

[2] Plattner, N., Noé, F., Bleicher, K. H., Bohm, H., Muller, K., Alanine, A., *et al.* Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* 6, 7653 (2015). DOI: 10.1038/ncomms8653

[3] Tiana, G. Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Sci.* 13, 113–124 (2004). DOI: 10.1110/ps.03223804

[4] Wang, W., Donini, O., Reyes, C. M. & Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.* 30, 211–243 (2001). DOI: 10.1146/annurev.biophys.30.1.211

[5] Brooijmans, N. & Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 32, 335–373 (2003). DOI:

10.1146/annurev.biophys.32.110601.142532

[6] Elcock, A. H. Molecular Simulations of Diffusion and Association in Multimacromolecular Systems. *Methods Enzymol.* 383, 166–198 (2004). DOI: 10.1016/S0076-6879(04)83008-8

[7] Day, R. & Daggett, V. All-atom simulations of protein folding and unfolding. *Adv. Protein Chem.* 66, 373–403 (2003). DOI: 10.1016/S0065-3233(03)66009-2

[8] De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* 59, 4035–4061 (2016). DOI: 10.1021/acs.jmedchem.5b01684

[9] Harvey, M. J., Giupponi, G. & Fabritiis, G. De. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* 5, 1632–1639 (2009). DOI: 10.1021/ct9000685

[10] Buch, I., Harvey, M. J., Giorgino, T., Anderson, D. P. & De Fabritiis, G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. *J. Chem. Inf. Model.* 50, 397–403 (2010). DOI: 10.1021/ci900455r

[11] Armacost, K. A., Riniker, S. & Cournia, Z. Novel Directions in Free Energy Methods and Applications. *J. Chem. Inf. Model.* 60, 1–5 (2020). DOI: 10.1021/acs.jcim.9b01174

[12] Deng, Y. & Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* 113, 2234–2246 (2009). DOI: 10.1021/jp807701h

[13] Sugita, Y. & Okamoto, Y. Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol. Biol.* 350, 205–223 (2007).

[14] Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929 (2004).

[15] Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* vol. 11 pp. 224–230 (2001). DOI: 10.1016/S0959-440X(00)00194-9

[16] Schlitter, J., Engels, M. & Krüger, P. Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *J. Mol. Graph.* 12, 84–89 (1994). DOI:

10.1016/0263-7855(94)80072-3

[17] Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 826–843 (2011). DOI: 10.1002/wcms.31

[18] Harada, R. & Kitao, A. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* 139, 035103 (2013). DOI: 10.1063/1.4813023

[19] Harada, R. & Kitao, A. Nontargeted parallel cascade selection molecular dynamics for enhancing the conformational sampling of proteins. *J. Chem. Theory Comput.* 11, 5493–5502 (2015). DOI: 10.1021/acs.jctc.5b00723

[20] Allen, R. J., Frenkel, D. & Ten Wolde, P. R. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *J. Chem. Phys.* 124, (2006). DOI: 10.1063/1.2198827

[21] Allen, R. J., Valeriani, C. & Rein Ten Wolde, P. Forward flux sampling for rare event simulations. *J. Phys. Condens. Matter* 21, (2009). DOI: 10.1088/0953-8984/21/46/463102

[22] Zuckerman, D. M. & Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys* 46, 43–57 (2017). DOI: 10.1146/annurev-biophys-070816-033834

[23] Huber, G. A. & Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* 70, 97–110 (1996). DOI: 10.1016/S0006-3495(96)79552-8

[24] Vanden-Eijnden, E. & Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* 130, (2009). DOI: 10.1063/1.3129843

[25] Vanden-Eijnden, E., Venturoli, M., Ciccotti, G. & Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* 129, 174102 (2008). DOI: 10.1063/1.2996509

[26] Tran, D. P., Takemura, K., Kuwata, K. & Kitao, A. Protein–Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics. *J. Chem. Theory Comput.* 14, 404–417 (2018). DOI: 10.1021/acs.jctc.7b00504

[27] Tran, D. P. & Kitao, A. Dissociation Process of a MDM2/p53 Complex Investigated by Parallel Cascade Selection Molecular Dynamics and the Markov State Model. *J. Phys.*

Chem. B 123, 2469–2478 (2019). DOI: 10.1021/acs.jpcb.8b10309

[28] Hata, H., Nishihara, Y., Nishiyama, M., Sowa, Y., Kawagishi, I. & Kitao, A. High pressure inhibits signaling protein binding to the flagellar motor and bacterial chemotaxis through enhanced hydration. *Sci. Rep.* 10, 2351 (2020). DOI: 10.1038/s41598-020-59172-3

[29] Nagel, D., Weber, A. & Stock, G. MSMPathfinder: Identification of Pathways in Markov State Models. *J. Chem. Theory Comput.* 16, 7882 (2020). DOI: 10.1021/acs.jctc.0c00774

[30] Tran, D. P. & Kitao, A. Dissociation Process of a MDM2/p53 Complex Investigated by Parallel Cascade Selection Molecular Dynamics and the Markov State Model. *J. Phys. Chem. B* 123, 2469–2478 (2019). DOI: 10.1021/acs.jpcb.8b10309

[31] Noé, F. Markov Models of Molecular Kinetics. *Encycl. Biophys.* 1385–1394 (2013). DOI: 10.1007/978-3-642-16712-6_726

[32] Bierer, B. E., Mattila, P. S., Standaert, R. F., Herzenberg, L. A., Burakoff, S. J., Crabtree, G., *et al.* Two distinct signal transmission pathways in T lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9231–9235 (1990). DOI: 10.1073/pnas.87.23.9231

[33] Van Duyne, G., Standaert, R., Karplus, P., Schreiber, S. & Clardy, J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science (80-.).* 252, 839–842 (1991). DOI: 10.1126/science.1709302

[34] Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci.* 108, 13118–13123 (2011). DOI: 10.1073/pnas.1104614108

[35] Congreve, M., Andrews, S. P., Doré, A. S., Hollenstein, K., Hurrell, E., Langmead, C. J., *et al.* Discovery of 1,2,4-Triazine Derivatives as Adenosine A2A Antagonists using Structure Based Drug Design. *J. Med. Chem.* 55, 1898–1903 (2012). DOI: 10.1021/JM201376W

[36] Yamane, J., Yao, M., Zhou, Y., Hiramatsu, Y., Fujiwara, K., Yamaguchi, T., *et al.* In-crystal affinity ranking of fragment hit compounds reveals a relationship with their inhibitory activities. *J. Appl. Crystallogr.* 44, 798–804 (2011). DOI:

10.1107/S0021889811017717

[37] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667 (2004). DOI: 10.1093/nar/gkh381

[38] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688 (2005). DOI: 10.1002/jcc.20290

[39] Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865 (2008). DOI: 10.1002/jcc.20945

[40] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713 (2015). DOI: 10.1021/acs.jctc.5b00255

[41] Takemura, K. & Kitao, A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J. Phys. Chem. B* 116, 6279–6287 (2012). DOI: 10.1021/jp301100g

[42] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174 (2004). DOI: 10.1002/JCC.20035

[43] Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260 (2006). DOI: 10.1016/J.JMGM.2005.12.005

[44] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174 (2004). DOI: 10.1002/jcc.20035

[45] Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8, 1542–1555 (2012). DOI: 10.1021/ct200909j

[46] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., *et al.* GROMACS

4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854 (2013). DOI: 10.1093/bioinformatics/btt055

[47] Prusty, M. & Cheong, S. A. Stochastic boundary conditions for molecular dynamics simulations. *Chem. Phys. Lett.* 105, 495–500 (2009). DOI: 10.1016/j.cpc.2011.11.004

[48] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690 (1984). DOI: 10.1063/1.448118

[49] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341 (1977). DOI: 10.1016/0021-9991(77)90098-5

[50] Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962 (1992). DOI: 10.1002/jcc.540130805

[51] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593 (1995). DOI: 10.1063/1.470117

[52] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472 (1997). DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

[53] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81, 511–519 (1984). DOI: 10.1063/1.447334

[54] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697 (1985). DOI: 10.1103/PhysRevA.31.1695

[55] Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101, 4177 (1998). DOI: 10.1063/1.467468

[56] Kitao, A., Harada, R., Nishihara, Y. & Tran, D. P. Parallel cascade selection molecular dynamics for efficient conformational sampling and free energy calculation of proteins. in *AIP Conference Proceedings* vol. 1790 p. 020013 (AIP Publishing LLC AIP Publishing , 2016). DOI: 10.1063/1.4968639

[57] Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* 134, 174105 (2011). DOI: 10.1063/1.3565032

[58] Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S. & Pande, V. S. MSMBuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* 7, 3412–3419 (2011). DOI: 10.1021/ct200463m

[59] Nishihara, Y., Harada, R. & Kitao, A. Cascade-type Massive Parallel Simulation for Protein Conformational Transition Pathway Search. *Proc. Inst. Stat. Math.* 62, 273–284 (2014). DOI: 10.1017/CBO9781107415324.004

[60] Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci.* 108, 10184–10189 (2011). DOI: 10.1073/pnas.1103547108

[61] Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The Quickhull Algorithm for Convex Hulls. *ACM Trans. Math. Softw.* 22, 469–483 (1996). DOI: 10.1145/235815.235821

[62] Shaw, E., Mares-Guia, M. & Cohen, W. Evidence for an Active-Center Histidine in Trypsin through Use of a Specific Reagent, l-Chloro-3-tosylamido-7-amino-2-heptanone, the Chloromethyl Ketone Derived from Nα-Tosyl-L-lysine. *Biochemistry* 4, 2219–2224 (1965). DOI: 10.1021/bi00886a039

[63] Teo, I., Mayne, C. G., Schulten, K. & Lelièvre, T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *J. Chem. Theory Comput.* 12, 2983–2989 (2016). DOI: 10.1021/acs.jctc.6b00277

[64] Guillain, F. & Thusius, D. Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J. Am. Chem. Soc.* 92, 5534–5536 (1970). DOI: 10.1021/ja00721a051

[65] Doerr, S. & de Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* 10, 2064–2069 (2014). DOI: doi: 10.1021/ct400919u

[66] Talhout, R. & Engberts, J. B. F. N. Thermodynamic analysis of binding of p-substituted benzamidines to trypsin. *Eur. J. Biochem.* 268, 1554–1560 (2001). DOI: 10.1046/j.1432-

1327.2001.01991.x

[67] Katz, B. A., Elrod, K., Luong, C., Rice, M. J., Mackman, R. L., Sprengeler, P. A., *et al.* A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J. Mol. Biol.* 307, 1451–1486 (2001). DOI: 10.1006/jmbi.2001.4516

[68] Mobley, D. L. & Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics* vol. 46 pp. 531–558 (2017). DOI: 10.1146/annurev-biophys-070816-033654

[69] Harada, R. & Shigeta, Y. On-the-Fly Specifications of Reaction Coordinates in Parallel Cascade Selection Molecular Dynamics Accelerate Conformational Transitions of Proteins. *J. Chem. Theory Comput.* 14, 3332–3341 (2018). DOI: 10.1021/acs.jctc.8b00264

[70] Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38 (1996). DOI: 10.1016/0263-7855(96)00018-5

Tables

Table 1.

Free energies calculated by dPaCS-MD/MSM and comparison with experimental values. The values after '±' indicate standard errors.

Complex	$-\Delta G$ (kcal/mol)	ΔG_v (kcal/mol)	$\Delta G°$ (kcal/mol)	ΔG_{exp} (kcal/mol)
Trypsin/benzamidine	-6.6 ± 0.2	0.5 ± 0.2	-6.1 ± 0.1	-6.4 [66] -7.3 [67]
FKBP/FK506	-14.2 ± 1.5	0.6 ± 0.1	-13.6 ± 1.6	-12.9 [32]
Adenosine A_{2A}/T4E	-15.5 ± 1.2	1.2 ± 0.2	-14.3 ± 1.2	-13.2 [35]

Figure captions

Figure 1.

Binding free energy calculation for the trypsin/benzamidine complex. (a) Initial configuration of the complex, ions, and water oxygen in the simulation box. A close-up view of benzamidine with trypsin is shown in the inset. (b) The distance between the center-of-mass (COM) positions of trypsin and benzamidine, d, plotted as a function of the PaCS-MD cycles. Three independent simulations are shown in different colors. The dashed line indicates the average value of d where the complex totally dissociated. (c) Representative dissociation pathways of benzamidine from trypsin obtained by PaCS-MD simulations. Colored lines show the trace of the benzamidine COM position along representative concatenated trajectories. Colors are the same as in (b). The initial and final structures of benzamidine in the first PaCS-MD trial are shown in the stick model, and in transparent color for the other trials. (d) Free energy profile against d calculated by PaCS-MD/MSM. The error bars indicate the standard deviation of three PaCS-MD/MSM trials. Green lines indicate the experimentally determined values of the binding free energy. The inset shows a close-up view of a benzamidine structure with $d = 21.1\ \mathring{A}$. In this paper, the molecular structure was visualized using VMD [70].

Figure 2.

Binding free energy calculation for the FKBP/FK506 complex, shown in a similar manner to that for the trypsin/benzamidine complex in Fig. 1. (a) Overall view of the system simulated. (b) Evolution of d against the PaCS-MD cycles. (c) Dissociation pathways obtained by three PaCS-MD simulations. (d) The average free energy profile from three independent PaCS-MD/MSM calculations. The experimentally determined binding free energy is indicated by a green line.

Figure 3.

Binding free energy calculation for the adenosine A_{2A}/T4E complex shown in a similar manner to that for the trypsin/benzamidine complex in Fig. 1. (a) Configuration of the simulated system after careful relaxation. The protein is shown in a cartoon representation while the T4E ligand (atom-type coloring) and lipid membrane (yellow) are shown as licorice models. The pink and cyan spheres represent chloride and sodium ions, respectively, and water molecules are shown as line representations with atom-type coloring. (b) Evolution of d against the PaCS-MD cycles. (c) Unbinding pathways obtained by five PaCS-MD simulations. (d) The inter-COM free energy profile from five independent PaCS-MD/MSM calculations. The experimentally determined binding free energy is indicated by a green line. The inset of panel (d) displays the binding to the entrance of binding pocket of T4E.

Figures

Figure 1.

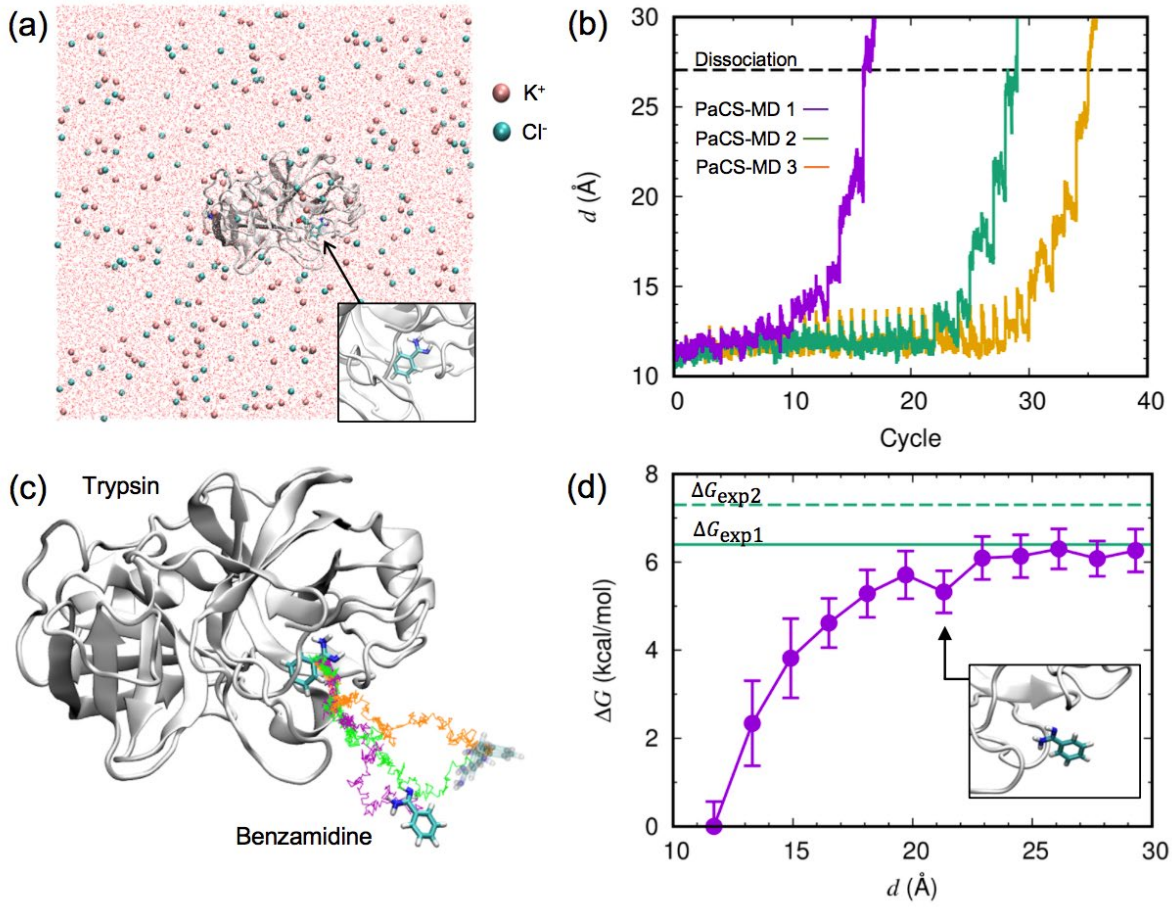
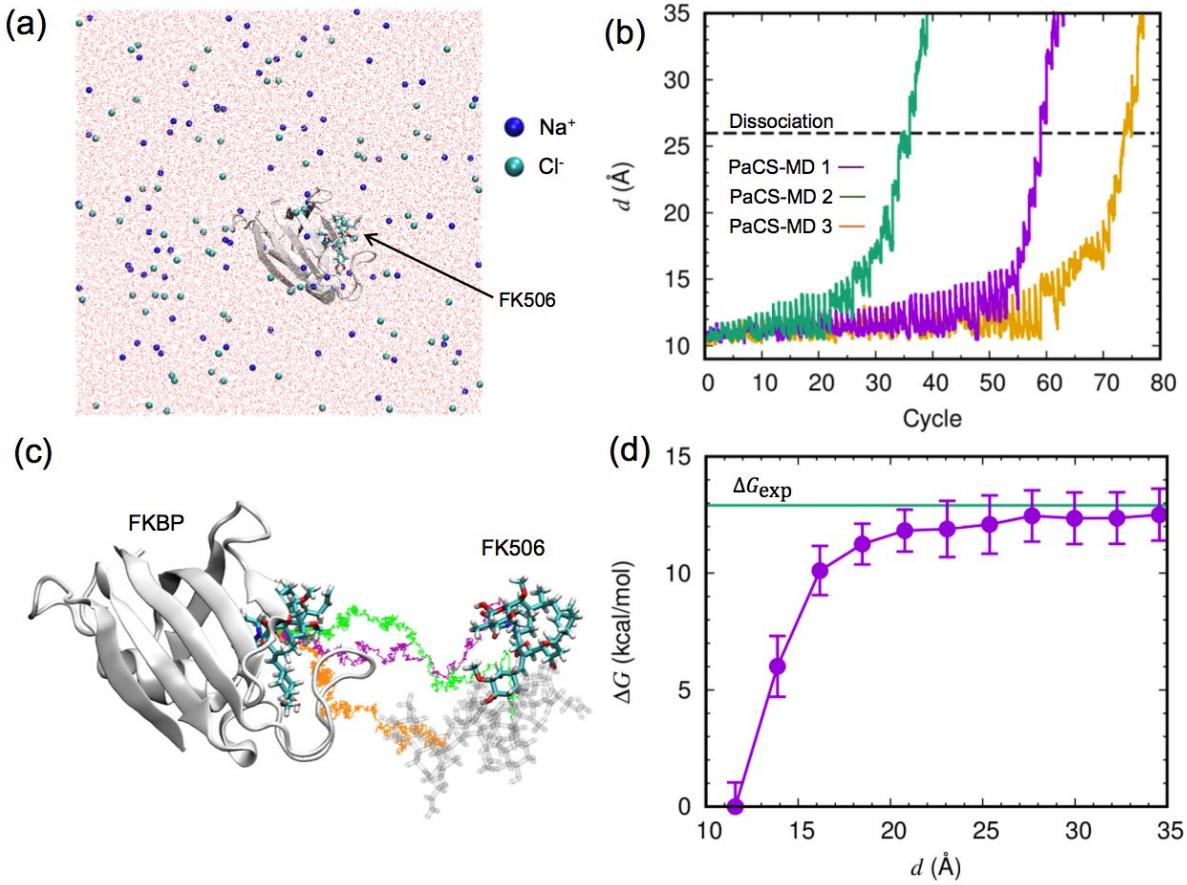
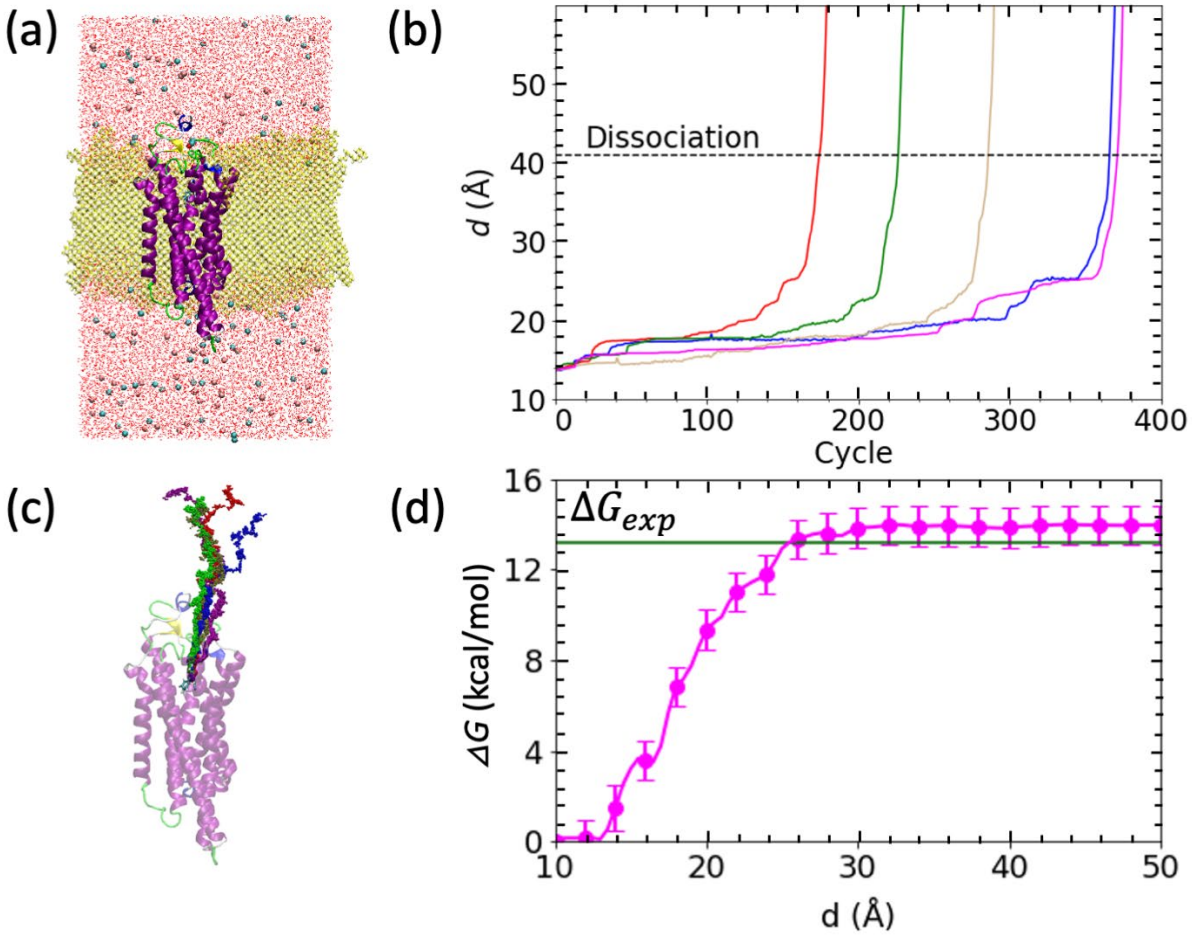


Figure 2.


(a)
Na+
Cl-
FK506
(b)
Dissociation
PaCS-MD 1
PaCS-MD 2
PaCS-MD 3
d (Å)
Cycle
(c)
FKBP
FK506
(d)
ΔG_exp
ΔG (kcal/mol)
d (Å)

589 **Figure 3.**



590