SAMPL7 TrimerTrip host-guest binding poses and binding affinities from spherical-coordinates-biased simulations

Zhaoxi Sun[1*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

*To whom correspondence should be addressed: proszx@163.com

Abstract

Host-guest binding remains a major challenge in modern computational modelling. The newest 7th statistical assessment of the modeling of proteins and ligands (SAMPL) challenge contains a new series of host-guest systems. The TrimerTrip host binds to 16 structurally diverse guests. Previously, we have successfully employed the spherical coordinates as the collective variables coupled with the enhanced sampling technique metadynamics to enhance the sampling of the binding/unbinding event, search for possible binding poses and predict the binding affinities in all three host-guest binding cases of the 6th SAMPL challenge. In this work, we employed the same protocol to investigate the TrimerTrip host in the SAMPL7 challenge. As no binding pose is provided by the SAMPL7 host, our simulations initiate from randomly selected configurations and are proceeded long enough to obtain converged free energy estimates and search for possible binding poses. The predicted binding affinities are in good agreement with the experimental reference, and the obtained binding poses serve as a nice starting point for end-point or alchemical free energy calculations.

Introduction

The predictions of the free energy differences between different states are at the center of computational modelling.[1-13] There are various factors limiting the accuracy and precision of computer simulations.[14-21] Two main sources of error are the convergence of the simulation and the accuracy of the description. Molecular dynamics (MD) or Monte Carlo simulations estimate the ensemble averages via the ergodicity assumption, where the frame/time-averaged finite-sample estimates are used to estimate the expectation of observables. Thus, to get a time-invariant estimate, long simulation times are required. However, the integration time step limits the time scale accessible in MD simulations, and the Boltzmann weighting function hinders the sampling of high-energy regions and the exploration of the phase space. Smart sampling techniques enhance the sampling efficiency by introducing artificial biasing potentials, connecting states with higher flexibilities, or constructing non-physical but computationally feasible pathways.[5, 22-33] For instance, umbrella sampling[34] adds (harmonic) biasing potentials to enhance the sampling efficiency in specific regions of the phase space. The base flipping event, which happens at millisecond time scales, could be sampled extensively within several microseconds' umbrella sampling simulations.[13, 35-36] The replica exchange method enhances the flexibility and mobility of the system by attempting to exchange configurations with higher temperatures and Hamiltonians periodically.[37-42] The alchemical method constructs artificial and easy-to-converge pathways connecting the states of interest to avoid extensive free energy simulations in physically meaningful transformations.[43-48] These enhanced sampling techniques greatly extends the applicability of MD simulations. The description of the system is often called as Hamiltonian. Electronic structure calculations provide accurate descriptions but are computationally demanding in condensed phase simulations,[49-55] while all-atom force fields[56-60] or coarser models provide a faster alternative with moderate accuracy. Multiscale models combine different descriptions in the same simulation box, saving computational resources and extending the applicability of molecular simulations.[32-33, 52, 61-64] In biomolecular simulations, all-atom force fields are often employed due to efficiency considerations.

Drugs are small molecules targeting specific biomolecules of unique functionalities. Understanding the protein-ligand interactions is now one of the key research directions in the computer-aided drug discovery. Current machine learning techniques enable the large-scale screening and provide a set of preliminary hits.[65-74] Further refinement of the dataset could be performed with end-point and alchemical free energy calculations.[75-85] This workflow could be very efficient when the free energy difference is the only quantity of interest. However, the weakness of these free energy calculation methods is also obvious. As only

fluctuations around the starting configuration are sampled, the initial-configuration-induced bias may not be eliminated in finite-time simulations. Further, the details about the intermediates in the binding/unbinding pathway are absent. If more details about the binding event are pursued, direct simulations of the binding/unbinding event to construct physically meaningful transformation pathways are necessary.[81, 86-89] Initial-configuration-related bias could also be eliminated effectively in this way.

The statistical assessment of the modeling (SAMPL) challenges feature the assessments of the sampling and Hamiltonian issues in the computational modeling of solvation free energies, pKa, host-guest systems, partition coefficients, and protein-ligand binding.[90-95] The host-guest systems are analogues of protein-ligand complexes. They are smaller and simpler than proteins and ligands. The hosts are often macrocyclic and rigid molecules, and the guests are drug-like molecules. The binding/unbinding pathway of the host-guest complexes is often simple and the number of binding poses is limited. Also, their binding affinities are comparable to those of protein-ligand complexes. Therefore, they serve as nice candidates for calibrating computational approaches.[93, 96-98] Due to the similarity between host-guest systems and protein-ligand complexes, similar free energy methods are employed to investigate the host-guest binding. For instance, equilibrium free energy methods such as umbrella sampling and the double decoupling method were used to calculate the binding free energies in the SAMPL6 host-guest cases.[92, 95] Nonequilibrium free energy simulations in the alchemical space were performed to calculate the host-guest binding affinities.[94] Although the free energy methods are accurate, the mean deviations from the experimental reference are often 2 kcal/mol,[93] which in principle arise from the Hamiltonian issue.

In host-guest binding simulations, one-dimensional (1D) collective variable (CV) is often used. The alchemical parameter is employed in free energy simulations in the alchemical space, while the distance between non-hydrogen atoms of the host and those of the guest or its mass-weighted variants is chosen to describe the binding and unbinding events in the physical space. In our previous work employing the three-dimensional (3D) spherical-coordinate-(ρ,θ,φ)[99] CV set, the host-guest relative position was scanned.[100] Although the simulations were started from the bound configuration provided by the SAMPL6 online server,[101] more possible binding poses were explored and the initial-configuration-induced bias vanished. The statistics are reweighted on the two-dimensional (2D) radius-contacts ($\rho - C$) surface to calculate the binding affinities. Compared with the published reports on the SAMPL6 host-guest binding, our predictions of the binding free energies obtained with two widely applied charge schemes were of similar accuracy.[100]

In the newest 7th SAMPL challenge (SAMPL7), a new TrimerTrip host is synthesized and the

thermodynamic parameters of the host-guest binding systems are measured.[96] No binding pose is provided by the server of SAMPL7,[102] which indicates that the binding-pose prediction is also a challenge in the current case. Therefore, in this work, we employed the same spherical-coordinate-biased strategy to explore the space of binding poses and calculate the binding affinities of the TrimerTrip-guest systems.

Methodology and Computational Details

System preparation. The host molecule is TrimerTrip formed by a central glycoluril trimer and two triptycene caps. No significant self-association is observed for this host.[96] All of the 16 guest molecules for the host in the blinded dataset of the SAMPL7 challenge are simulated. The structures of the hosts and guests are obtained from the online server of the SAMPL7 challenge.[102] The structures of the host and the guests are shown in Fig. 1. The protonation states of the guests are adjusted to match the experimental reference,[96] and a summary of the net charges of the host and guests and the experimental binding affinities are given in Table S1. As in the host-guest and protein-ligand binding cases, the corrected semi-empirical charges and the restrained electrostatic potential (RESP) charges often give similar results, here we only use the AM1-BCC[103] charge scheme. The other parameters such as the bonded terms and the vdW radius are obtained from the general Amber force field (GAFF) force field.[104] Solvation is performed with TIP3P[105-106] water molecules and the truncated octahedron cell is replicated in whole space with periodic boundary conditions. As no bound conformation is provided on the online server,[102] we simply put the host and the guest together and let the simulation run to equilibrate the system and find stable binding poses. The minimum distance between the box edge and the surface of the solute is set to 28 Å, considering the radius of the spherical restraint, the fluctuations of the box size in NPT simulations and the sizes of the solute molecules. Non-polarizable spherical counter ions[107-108] of Na^+ or Cl^- parameterized for TIP3P water are added for neutralization.

Free Energy Simulations.

In order to explore the phase space efficiently, we employ the well-tempered metadynamics method to enhance the sampling efficiency.[87, 109-110] Gaussian biasing potentials are added periodically and the overall biasing potential increases with time. The resulting biasing potential could be defined by the following equation,

$$V_{n+1}(\mathbf{s}) = V_n(\mathbf{s}) + G(\mathbf{s}, \mathbf{s}_{n+1}) e^{-\frac{1}{\gamma - 1} V_n(\mathbf{s}_{n+1})} \quad (1),$$

where the subscript n denotes the nth step, V is the time-dependent overall biasing potential, $\mathbf{s}$ is the CV

matrix, $G(\mathbf{s},\mathbf{s}_n)$ represents the Gaussian kernels of biasing potentials, and γ is the bias factor. The time-independent algorithm[111] is employed for post-process reweighting to recover the unbiased statistics in the original ensemble. The finite-time estimate of the expectation of mechanical observable O can be obtained by

$$\langle O(\mathbf{s})\rangle = \left\langle O(\mathbf{s}) e^{\beta(V(\mathbf{s},t)-c(t))} \right\rangle \quad (2),$$

where the canonical bracket denotes ensemble average, β represents the reciprocal temperature, c is the offset of the biasing potential, and t is the time of simulation.

In order to explore the space of possible binding poses, we bias the spherical coordinates (ρ,θ,φ) [99] defined by the relative position of the center of masses (COM) of the host and that of the guest. In our previous simulations of the SAMPL6 host-guest systems, this set of CV could differentiate different binding poses and enhance the sampling of the binding/unbinding event.[100] With this set of CVs, we could scan possible binding poses efficiently and get rid of the initial-configuration-induced bias. When the guest is sufficiently far away from the host, the host-guest interactions vanish and the unbound state is produced. The fully decoupled state could be defined by the zero or near-zero contact between the host and the guest. The contact number is given by the following switching function,

$$C = \sum_{i\in A}\sum_{j\in B} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^n}{1-\left(\frac{r_{ij}}{r_0}\right)^m} \quad (3),$$

where A and B denote two groups of atoms (i.e. the host and the guest), the subscripts i and j represent the *i*th and *j*th atoms in the groups, m and n are 12 and 6, respectively, r refers to the distance and the threshold for the contact $r_0 = 6$ Å. Only heavy atoms are included in the calculation. Consider the case that the contact number between the the *i*th atom in group A and the *j*th in group B is under calculation. When the distance r_{ij} becomes $2r_0$ (i.e. 12 Å), the switching function becomes $C = \frac{1}{1+2^6} \approx 0.015$. Therefore, if a configuration gives a near-zero contact, all heavy atoms in the group is far from the other group and the interactions between A and B groups are near-zero. Therefore, a near-zero contact could be used to define the decoupled state. As the simulation box is of finite size, an upper wall is added on the distance/radius ρ to limit the volume of phase space that the guest could explore. The upper

wall on the radius ρ is set at 28 Å, which is large enough to define a fully decoupled state with near-zero contacts between the host and the guest. An entropic correction defined in Eq. (4) is thus added to recover the unbiased free energy.

$$T\Delta S = -\frac{1}{\beta}\ln\left(\frac{V^0}{\frac{4}{3}\pi\rho_s^3 - V_{\text{host}}}\right) \quad (4),$$

where ρ_s is the upper wall on the radius ρ, $V^0 = 1660\,\text{Å}^3$ is the standard state volume, and V_{host} is the volume of the host.

For each host-guest system, the starting configuration is obtained by simply putting the host and the guest together, as mentioned previously. We perform minimization, 100 ps NVT equilibration and 5 ns NPT equilibration to equilibrate the system. After that, 1000 ns enhanced sampling simulation is performed. The parameters for the metadynamics simulation used in previous simulations of SAMPL6 host-guest systems are employed in the current work.[100] Namely, the initial Gaussian height is 0.24 kcal/mol, the deposition interval is 0.5 ps, and the bias factor used is 20. Gaussian widths are set as 0.1 Å, $\frac{\pi}{16}$, and $\frac{\pi}{8}$ for the three polar coordinates, respectively. The simulation is performed at 298 K (the experimental condition) with GROMACS 2018.6 [112] patched with PLUMED 2.6.0-dev[113]. The V-rescale algorithm[114] is employed for temperature regulation and the Parrinello-Rahman barostat[115-116] is used for pressure regulation. A time step of 1 fs is used to propagate the dynamics to avoid bond-length-constraint-related systematic errors. Long-range electrostatics are treated with the PME[117-118] method.

Result and discussion

Before analyzing the detailed results of enhanced sampling simulations, we check the convergence of the simulations. The height of Gaussian potentials decreases to very small values (e.g. 0.002 kcal/mol) at the end of simulations (data not shown), and after 400 ns the offset bias function $c(t)$ in Fig. S1 displays a linearly increasing behavior with the logarithm of the simulation time. Therefore, the quasi-stationary state is reached and we analyze the statistics obtained in 400-1000 ns. We reweight the statistics in the metadynamics simulations with the time-independent algorithm and perform the projection on the radius-contacts (ρ-C) surface. A typical 2D free energy surface is presented in Fig. 2. The free energy difference between the global free energy minimum and the zero-contact large-distance unbound state is used to

estimate the binding free energy. The time-evolution of the estimated binding affinities $\Delta G_{\mathrm{metad}}$ is presented in Fig. S2. The free energy difference $\Delta G_{\mathrm{metad}}$ presents the time-invariant behavior in the last part of the simulations, which indicates that the binding affinity has converged.

As no binding pose is provided in the SAMPL7 host-guest challenge, the spherical-coordinates-biased simulation is also used to obtain the stable host-guest binding pose. In Fig. 3, we present the representative structures of the bound states of the host-guest complexes. The top-6 stable structures visited during enhanced sampling simulations are provided in the online depository at https://github.com/proszxppp/SAMPL7_TTP. The binding poses presented in Fig. 3 are, of course, included. For each host-guest system, the top-6 structures are very similar, which indicates that they are from the same binding pose.

Compared with the previous cyclic hosts (e.g. CB8 in SAMPL6), the new TrimerTrip molecule is more flexible. It could close to form a ring-like pocket to coordinate the guest molecules. As there is no chemical bond restraining the 'ring', it could tolerate a high degree of fluctuations in the bound state. To illustrate the conformational fluctuations in the bound state, for the guest g2, we presented two structures extracted from the global free energy minimum in Fig. 3. Both of the binding poses represent the bound host-guest complex, and the difference mainly lies in the degree of closure of the TrimerTrip. We can also see the fluctuations in the 2D free energy surface. For instance, for the host-g5 complex in Fig. 2, the free energy basin in the bound state is quite wide, which indicates that the degree of local fluctuations is significant in the host-guest complex.

The free energy difference obtained from enhanced sampling simulations requires an entropic correction caused by the spherical restraint to recover the unbiased binding free energy. The volume and the resulting entropic correction are summarized in Table S2. The corrected binding free energies for the host-guest systems are given in Table 1. To assess the quality of prediction, several metrics including the mean signed error (MSE), the mean absolute error (MAE), the root-mean-squared error (RMSE), Kendall's rank correlation coefficient (τ), and Pearlman's predictive index (PI) are calculated. The first three errors are used to estimate the errors, while the last two estimators are used to assess the consistency of the predicted ranks of binding affinities and the experimental reference. In the current case, RMSE is 1.5 kcal/mol, MSE is -0.1 kcal/mol and MAE is 1.3 kcal/mol, which indicates that the predictions do not deviate too much from the experimental results and the quality of prediction is acceptable. The sizes of these errors are comparable to those of the SAMPL6 host-guest systems.[93] The ranking coefficients tell the same thing. The rank of

predicted binding affinities is similar to the experimental one. The linear correlation between the predictions and the experimental references is checked in Fig. 4, which also shows that the agreement between predictions and experimental results is good.

Conclusion

In this work, we employed the spherical coordinates (ρ,θ,φ) as the reaction coordinates to enhance the sampling of the binding/unbinding event and scan the space of binding poses in the SAMPL7 TrimerTrip-guest systems. Our simulation explores stable binding poses and estimates the binding affinities. The binding poses serve as a nice starting point for alchemical or end-point binding free energy calculations, and the predicted binding affinities are in good agreement with the experimental results. Compared with previous cyclic host molecules, the TrimerTrip host does not have chemical bonds for ring restraints. As a result, it is more flexible and could tolerate a higher degree of fluctuations in the bound state. The bound-state free energy basin is relatively wide.

Acknowledgement

This work was supported China Scholarship Council. We are grateful for many valuable and insightful comments from the anonymous reviewers.

Conflicts of interest

There are no conflicts of interest to declare.

Supporting information

The time-evolution of the bias offset function c(t), the time-evolution of the binding affinities from metadynamics simulations, the summary of the experimental binding free energies and the net charges of the host and the guest molecules, the volumes of the hosts and the resulting entropic corrections are given in the supporting information.

References

1. Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004,** *303* (5665), 1813-8.
2. Steinbrecher, T.; Labahn, A., Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* **2010,** *17* (8), 767-785.
3. Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W., Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **2010,** *6* (5), 1509-1519.
4. Gelman, A.; Meng, X.-L., Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science* **1998**, 163-185.
5. Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z., Interaction Entropy for Protein-Protein Binding. *J. Chem. Phys.* **2017,** *146* (12), 124124.
6. Pohorille, A.; Jarzynski, C.; Chipot, C., Good practices in free-energy calculations. *J. Phys. Chem. B* **2010,** *114* (32), 10235-53.
7. Kim, I.; Allen, T. W., Bennett's acceptance ratio and histogram analysis methods enhanced by umbrella sampling along a reaction coordinate in configurational space. *J. Chem. Phys.* **2012,** *136* (16), 164103-164103.
8. Shirts, M. R.; Pande, V. S., Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005,** *122* (14), 144107-144107.
9. Bruckner, S.; Boresch, S., Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011,** *32* (7), 1320–1333.
10. Kofke, D. A., On the sampling requirements for exponential-work free-energy calculations. *Mol. Phys.* **2006,** *104* (22), 3701-3708.
11. Hummer, G.; Pratt, L. R.; Garcia, A. E., Hydration free energy of water. *J. Phys. Chem.* **1995,** *99* (38), 14188-14194.
12. Hahn, A.; Then, H., Using bijective maps to improve free-energy estimates. *Physical Review E* **2009,** *79* (1), 011113.
13. Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chem. Phys. Lett.* **2017,** *684*, 239-244.
14. Villamaina, D.; Trizac, E., Thinking outside the box: fluctuations and finite size effects. *Eur. J. Phys.* **2014,** *35* (3), 035011.
15. Heidari, M.; Cortes-Huerto, R.; Kremer, K.; Potestio, R., Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations. *EUR PHYS J E* **2018,** *41* (5), 64.
16. Román, F.; White, J.; Velasco, S., Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.* **1997,** *107* (12), 4635-4641.
17. Lebowitz, J.; Percus, J., Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.* **1961,** *122* (6), 1675.
18. Salacuse, J.; Denton, A.; Egelstaff, P., Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. I. Theoretical method. *Phys. Rev. E* **1996,** *53* (3), 2382.
19. Brooks, B. R.; Brooks III, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614.
20. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.
21. Plimpton, S., Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995,** *117* (1), 1-19.
22. And, S. E. T.; Smithrud, D. B., Carboxylates Stacked over Aromatic Rings Promote Salt Bridge Formation in Water. *J. Am. Chem. Soc.* **2002,** *124* (3), 442.
23. Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C., Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (2), 315-20.
24. Rani, P.; Biswas, P., Diffusion of hydration water around intrinsically disordered proteins. *J. Phys. Chem. B* **2015,** *119* (42), 13262-13270.
25. Zerze, G. I. H.; Best, R. B.; Mittal, J., Sequence-and temperature-dependent properties of unfolded and disordered

proteins from atomistic simulations. *J. Phys. Chem. B* **2015,** *119* (46), 14622-14630.
26. Best, R. B.; de Sancho, D.; Mittal, J., Residue-specific α-helix propensities from molecular simulation. *Biophys. J.* **2012,** *102* (6), 1462-1467.
27. Makowski, M.; Liwo, A.; Scheraga, H. A., Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino Acid Side Chains in Water. VII. Charged–Hydrophobic/Polar and Polar–Hydrophobic/Polar Side Chains. *J. Phys. Chem. B* **2017,** *121* (2), 379-390.
28. Tobias, D. J.; Brooks III, C. L., Thermodynamics and mechanism of. alpha. helix initiation in alanine and valine peptides. *Biochemistry* **1991,** *30* (24), 6059-6070.
29. Hudson, P. S.; Woodcock, H. L.; Boresch, S., Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems. *J. Phys. Chem. Lett.* **2015,** *6* (23), 4850-4856.
30. Martínezveracoechea, F. J.; Escobedo, F. A., Variance minimization of free energy estimates from optimized expanded ensembles. *J. Phys. Chem. B* **2008,** *112* (27), 8120-8.
31. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Citeria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.
32. Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019,** *21* (12), 6672-6688
33. Sun, Z., BAR-based multi-dimensional nonequilibrium pulling for indirect construction of QM/MM free energy landscapes: from semi-empirical to ab initio. *Phys. Chem. Chem. Phys.* **2019,** *21* (39), 21942-21959
34. Kästner, J., Umbrella sampling. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011,** *1* (6), 932-942.
35. Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019,** *21* (27), 14923-14940.
36. Sun, Z.; Zhang, J. Z., Thermodynamic Insights of Base Flipping in TNA Duplex: Force Fields, Salt Concentrations, and Free Energy Simulation Methods. **2020**.
37. Fukunishi, H.; Watanabe, O.; Takada, S., On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002,** *116* (116), 9058-9067.
38. Itoh, S. G.; Damjanovic, A.; Brooks, B. R., pH replica-exchange method based on discrete protonation states. *Proteins* **2011,** *79* (12), 3420-36.
39. Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C., Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* **2006,** *2* (2), 420.
40. T, G.; CM, S., Mechanism of Amyloid-β Fibril Elongation. *Biochemistry* **2014,** *53* (44), 6981-91.
41. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999,** *314* (1-2), 141-151.
42. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000,** *113* (15), 6042-6051.
43. Sun, Z.; Wang, X.; Song, J., Extensive Assessment of Various Computational Methods for Aspartate's pKa Shift. *J. Chem. Inf. Model.* **2017,** *57* (7), 1621-1639.
44. Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R., Benchmarking pKa prediction. *Bmc Biochemistry* **2006,** *7* (1), 1-12.
45. Stanton, C. L.; Houk, K. N., Benchmarking pKa Prediction Methods for Residues in Proteins. *J. Chem. Theory Comput.* **2008,** *4* (6), 951-966.
46. Archontis, G.; Simonson, T., Proton binding to proteins: a free-energy component analysis using a dielectric continuum model. *Biophys. J.* **2005,** *88* (6), 3888-904.
47. Fowler, P. W.; Jha, S.; Coveney, P. V., Grid-based steered thermodynamic integration accelerates the calculation of binding free energies. *Philosophical Transactions of the Royal Society A Mathematical Physical & Engineering Sciences* **2005,** *363* (1833), 1999-2015.

48. Pitera, J. W.; van Gunsteren, W. F., A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Molecular Simulation* **2002,** *28* (1-2), 45-65.
49. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., A direct MP2 gradient method. *Chem. Phys. Lett.* **1990,** *166* (3), 275-280.
50. Head-Gordon, M.; Pople, J. A.; Frisch, M. J., MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988,** *153* (6), 503-506.
51. Hertwig, R. H.; Koch, W., On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chem. Phys. Lett.* **1997,** *268* (5), 345-351.
52. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M., Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999,** *313* (3-4), 701-706.
53. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995,** *16* (11), 1357–1377.
54. Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J., Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.* **2003,** *24* (6), 760-769.
55. Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007,** *13* (12), 1173-1213.
56. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007,** *92* (11), 3817-3829.
57. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-25.
58. 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.* **2015,** *11* (8), 3696-3713.
59. Mahoney, M. W.; Jorgensen, W. L., A five-site model for liquid water and the reproduction of the density anomaly by rigid. *J. Chem. Phys.* **2000,** *112* (20), 8910-8922.
60. Best, R. B.; Buchete, N.-V.; Hummer, G., Are current molecular dynamics force fields too helical? *Biophys. J.* **2008,** *95* (1), L07-L09.
61. Sun, Z.; Zhu, T.; Wang, X.; Mei, Y.; Zhang, J. Z., Optimization of convergence criteria for fragmentation methods. *Chem. Phys. Lett.* **2017,** *687*, 163-170.
62. Raghavachari, K.; Saha, A., Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.* **2015,** *115* (12), 5643-5677.
63. Collins, M. A.; Bettens, R. P., Energy-based molecular fragmentation methods. *Chem. Rev.* **2015,** *115* (12), 5607-5642.
64. Sahu, N.; Gadre, S. R., Molecular tailoring approach: a route for ab initio treatment of large clusters. *Accounts of chemical research* **2014,** *47* (9), 2739-2747.
65. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015,** *36* (15), 1132-1156.
66. Kapetanovic, I., Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem. Biol. Interact.* **2008,** *171* (2), 165-176.
67. Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M., Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* **2008,** *51* (20), 6237-6255.
68. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R., Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins-structure Function & Bioinformatics* **2002,** *47* (4), 409-443.
69. Shoichet, B. K.; Leach, A. R.; Kuntz, I. D., Ligand solvation in molecular docking. *Proteins-structure Function & Bioinformatics* **1999,** *34* (1), 4.
70. Sotriffer, C. A.; Sanschagrin, P.; Matter, H.; Klebe, G., SFCscore: scoring functions for affinity prediction of protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **2008,** *73* (2), 395-419.
71. Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C.; Waldman, M., LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.* **2005,** *23* (5), 395-407.

72. Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B., Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002,** *20* (4), 281-295.
73. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S., A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006,** *49* (20), 5912-5931.
74. Yan, Y.; Wang, W.; Sun, Z.; Zhang, J. Z.; Ji, C., Protein–ligand empirical interaction components for virtual screening. *J. Chem. Inf. Model.* **2017,** *57* (8), 1793-1806.
75. Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G., Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Biorg. Med. Chem.* **2007,** *15* (24), 7865-7877.
76. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004,** *25* (2), 238-250.
77. Case, D. A., Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **2010,** *4* (2), 285–290.
78. Rapp, C.; Kalyanaraman, C.; Schiffmiller, A.; Schoenbrun, E. L.; Jacobson, M. P., A Molecular Mechanics Approach to Modeling Protein–Ligand Interactions: Relative Binding Affinities in Congeneric Series. *J. Chem. Inf. Model.* **2011,** *51* (9), 2082-9.
79. Swope, W. C., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982,** *76* (1), 637.
80. Pham, T. T.; Shirts, M. R., Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *J. Chem. Phys.* **2011,** *135* (3), 034114.
81. Procacci, P.; Chelli, R., Statistical Mechanics of Ligand–Receptor Noncovalent Association, Revisited: Binding Site and Standard State Volumes in Modern Alchemical Theories. *J. Chem. Theory Comput.* **2017,** *13* (5), 1924-1933.
82. Wang, X.; Sun, Z., Understanding PIM-1 kinase inhibitor interactions with free energy simulation. *Phys. Chem. Chem. Phys.* **2019,** *21* (14), 7544-7558.
83. Sun, Z.; Wang, X.; Zhao, Q.; Zhu, T., Understanding Aldose Reductase-Inhibitors interactions with free energy simulation. *J. Mol. Graphics Modell.* **2019,** *91*, 10-21.
84. Sun, Z.; Wang, X.; Zhang, J. Z., Determination of Binding Affinities of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Inhibitors from Free Energy calculation. *Chem. Phys. Lett.* **2019,** *723*, 1-10.
85. Sun, Z.; Wang, X.; Zhang, J. Z., Theoretical understanding of the thermodynamics and interactions in transcriptional regulator TtgR-ligand binding. *Phys. Chem. Chem. Phys.* **2020,** *22* (3), 1511-1524.
86. Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V., Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114* (11), E2136.
87. Valsson, O.; Tiwary, P.; Parrinello, M., Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annual Review of Physical Chemistry* **2016,** *67* (1), 159.
88. Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M., Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (5), 386-91.
89. Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P., Molecular mechanics methods for predicting protein–ligand binding. *Phys. Chem. Chem. Phys.* **2006,** *8* (44), 5166-77.
90. Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P., Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput. Aided Mol. Des.* **2014,** *28* (3), 135-50.
91. Monroe, J. I.; Shirts, M. R., Converging free energies of binding in cucurbit[7]uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. *J. Comput. Aided Mol. Des.* **2014,** *28* (4), 401-15.
92. Song, L. F.; Bansal, N.; Zheng, Z.; Merz, K. M., Detailed potential of mean force studies on host–guest systems from the SAMPL6 challenge. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 1013-1026.
93. Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L., Overview of the SAMPL6 host–guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 937-963.
94. Procacci, P.; Guarrasi, M.; Guarnieri, G., SAMPL6 host–guest blind predictions using a non equilibrium alchemical

approach. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 965-982.
95. Nishikawa, N.; Han, K.; Wu, X.; Tofoleanu, F.; Brooks, B. R., Comparison of the umbrella sampling and the double decoupling method in binding free energy predictions for SAMPL6 octa-acid host–guest challenges. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 1075-1086.
96. Ndendjio, S. Z.; Liu, W.; Yvanez, N.; Meng, Z.; Zavalij, P. Y.; Isaacs, L., Triptycene walled glycoluril trimer: synthesis and recognition properties. *New J. Chem.* **2020,** *44* (2), 338-345.
97. Caldararu, O.; Olsson, M. A.; Ignjatović, M. M.; Wang, M.; Ryde, U., Binding free energies in the SAMPL6 octa-acid host–guest challenge calculated with MM and QM methods. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 1027-1046.
98. Eken, Y.; Patel, P.; Díaz, T.; Jones, M. R.; Wilson, A. K., SAMPL6 host–guest challenge: binding free energies via a multistep approach. *J. Comput. Aided Mol. Des.* **2018,** *32* (10), 1097-1115.
99. Capelli, R.; Carloni, P.; Parrinello, M., Exhaustive Search of Ligand Binding Pathways via Volume-based Metadynamics. *J. Phys. Chem. Lett.* **2019,** *10* (12), 3495-3499.
100. Sun, Z.; He, Q.; Li, X.; Zhu, Z., SAMPL6 host–guest binding affinities and binding poses from spherical-coordinates-biased simulations. *J. Comput. Aided Mol. Des.* **2020**, 1-12.
101. https://github.com/samplchallenges/SAMPL6.
102. https://github.com/samplchallenges/SAMPL7/tree/master/host_guest/Isaacs_clip.
103. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23* (16), 1623-41.
104. 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.* **2004,** *25*, 1157–1173.
105. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
106. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004,** *121* (20), 10096-10103.
107. Joung, I. S.; Cheatham III, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008,** *112* (30), 9020-9041.
108. Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009,** *113* (40), 13279-13290.
109. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011,** *1* (5), 826-843.
110. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008,** *100* (2), 020603.
111. Tiwary, P.; Parrinello, M., A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015,** *119* (3), 736-42.
112. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.
113. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014,** *185* (2), 604-613.
114. Giovanni, B.; Davide, D.; Michele, P., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 2384.
115. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50* (5), 1055-1076.
116. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
117. York, D. M.; Darden, T. A.; Pedersen, L. G., The Effect of Long-range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of The Ewald and Truncated List Methods. *J. Chem. Phys.* **1993,** *99* (10), 8345-8348.
118. Tuckerman, M. E.; Berne, B. J.; Martyna, G. J., Molecular dynamics algorithm for multiple time scales: Systems with

long range forces. *J. Chem. Phys.* **1991,** *94* (10), 6811-6815.

Table 1. The TrimerTrip-guest binding affinities in kcal/mol. ΔG_{exp} is the experimental value, ΔG_{metad} denotes the free energy difference between the bound and unbound states, ΔG_{VC} represents the volume correction, and ΔG_{calc} is the predicted binding affinity. MSE, MAE, RMSE, τ, and PI serve as quality measurements. SD denotes the standard error of the free energy estimate, which is obtained from block averaging.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
TTP	g1	-6.1	-4.7	0.5	2.3	-7.0	0.5
	g2	-8.3	-5.3	0.6	2.3	-7.6	0.6
	g3	-10.1	-7.6	0.5	2.3	-10.0	0.5
	g5	-11.1	-9.4	0.6	2.3	-11.8	0.6
	g6	-9.6	-8.1	0.5	2.3	-10.5	0.5
	g7	-6.5	-4.1	0.5	2.3	-6.4	0.5
	g8	-9.5	-4.4	0.5	2.3	-6.8	0.5
	g9	-7.6	-6.7	0.5	2.3	-9.0	0.5
	g10	-8.2	-7.4	0.5	2.3	-9.7	0.5
	g11	-9.0	-6.3	0.5	2.3	-8.7	0.5
	g12	-8.3	-3.8	0.4	2.3	-6.1	0.4
	g15	-10.5	-4.9	0.5	2.3	-7.3	0.5
	g16	-11.5	-11.0	0.5	2.3	-13.3	0.5
	g17	-11.8	-7.9	0.5	2.3	-10.2	0.5
	g18	-10.6	-9.7	0.5	2.3	-12.0	0.5
	g19	-11.7	-10.0	0.5	2.3	-12.4	0.5
RMSE						1.5	
MSE						-0.1	
MAE						1.3	
τ						0.5	
PI						0.7	

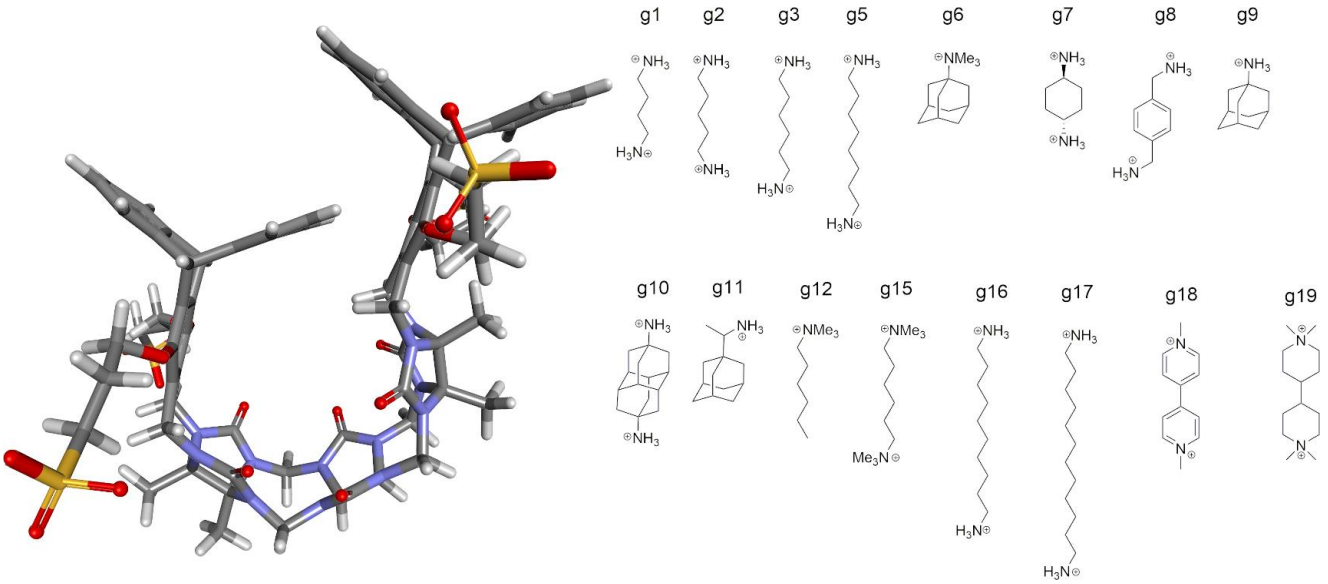

Fig. 1. The 3D structure of the host TrimerTrip (left) and the 2D chemical structures of its guests (right).

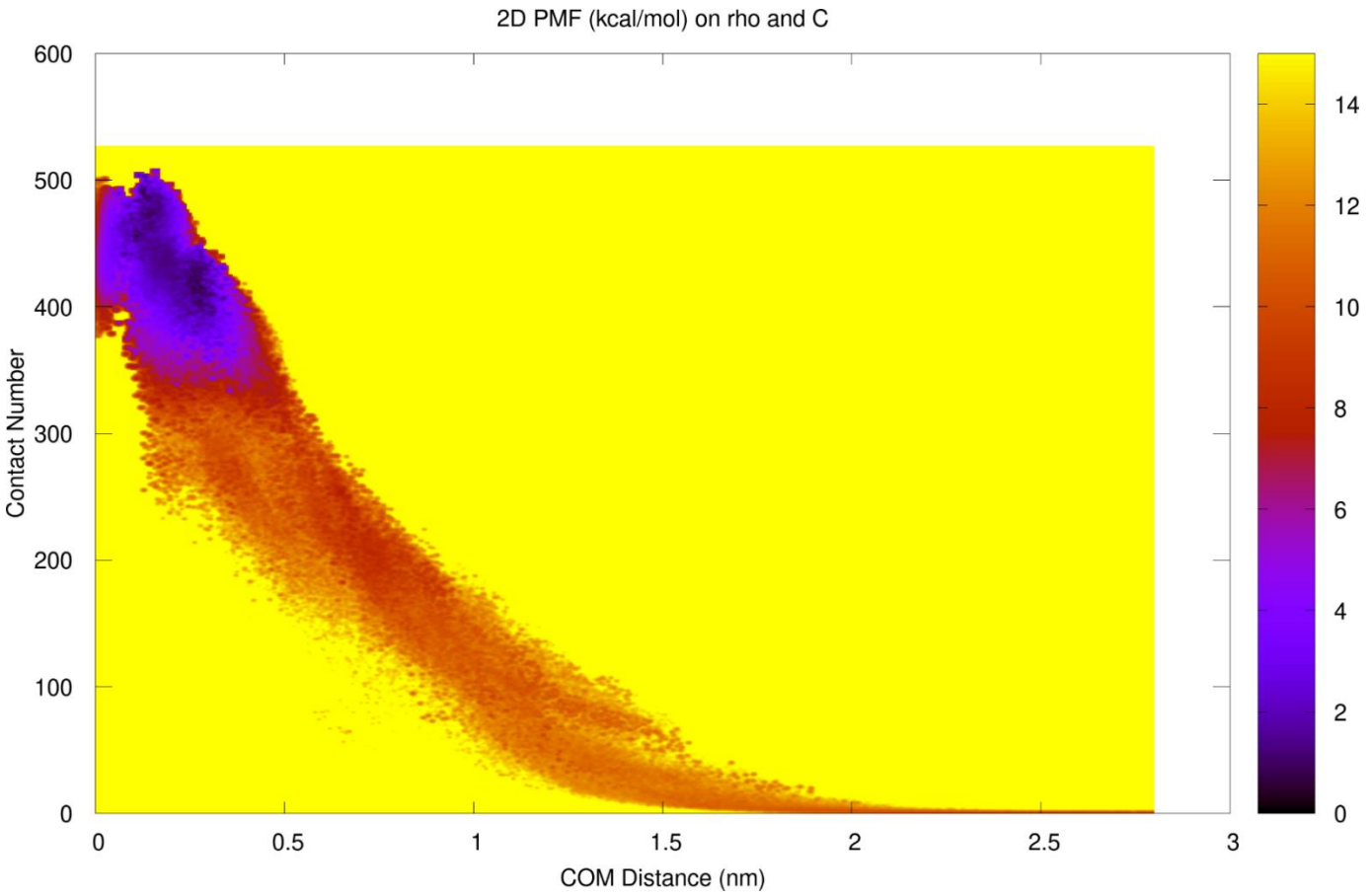

Fig. 2. Typical 2D $\rho - C$ free energy surface in kcal/mol. Here, the host-g5 complex is used to generate the plot.

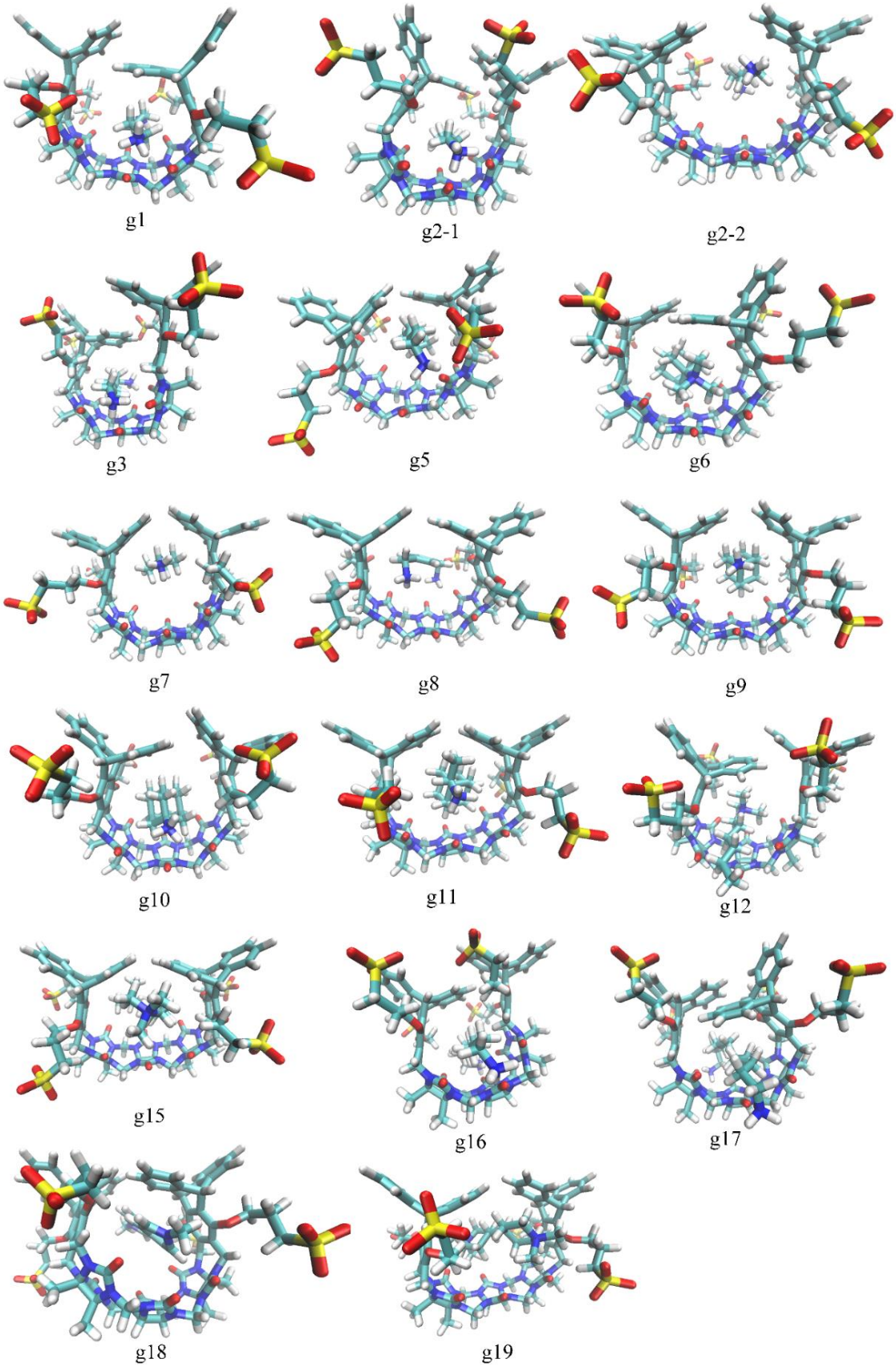

Fig. 3. The representative structures of the binding poses for all host-guest systems. For the guest g2, two structures from the global free energy minimum are extracted. The TrimerTrip ring could tolerate conformational fluctuations in the bound state.

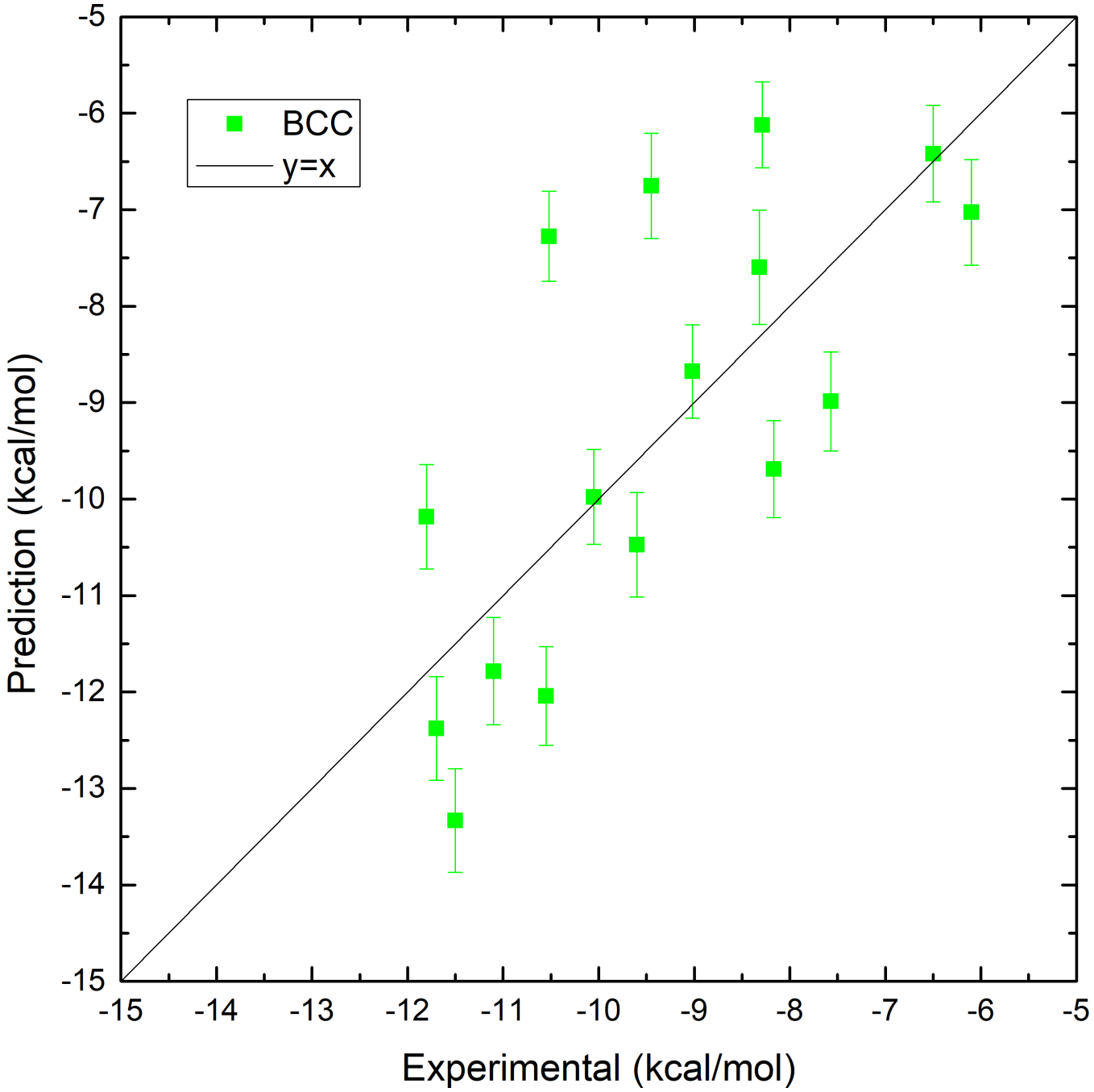

Fig. 4. Correlation between the predicted binding affinities and the experimental reference for TrimerTrip-guest systems. The exact values of the binding affinities are presented in Table 1.

Supporting Information: SAMPL7 TrimerTrip host-guest binding poses and binding affinities from spherical-coordinates-biased simulations

Zhaoxi Sun[1*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

*To whom correspondence should be addressed: proszx@163.com

Fig. S1. Time evolution of the offset bias c(t). After 400 ns (the green-yellow line), the bias offset function c(t) increases linearly with the logarithm of the simulation time, which indicates that the quasi-stationary state is reached.

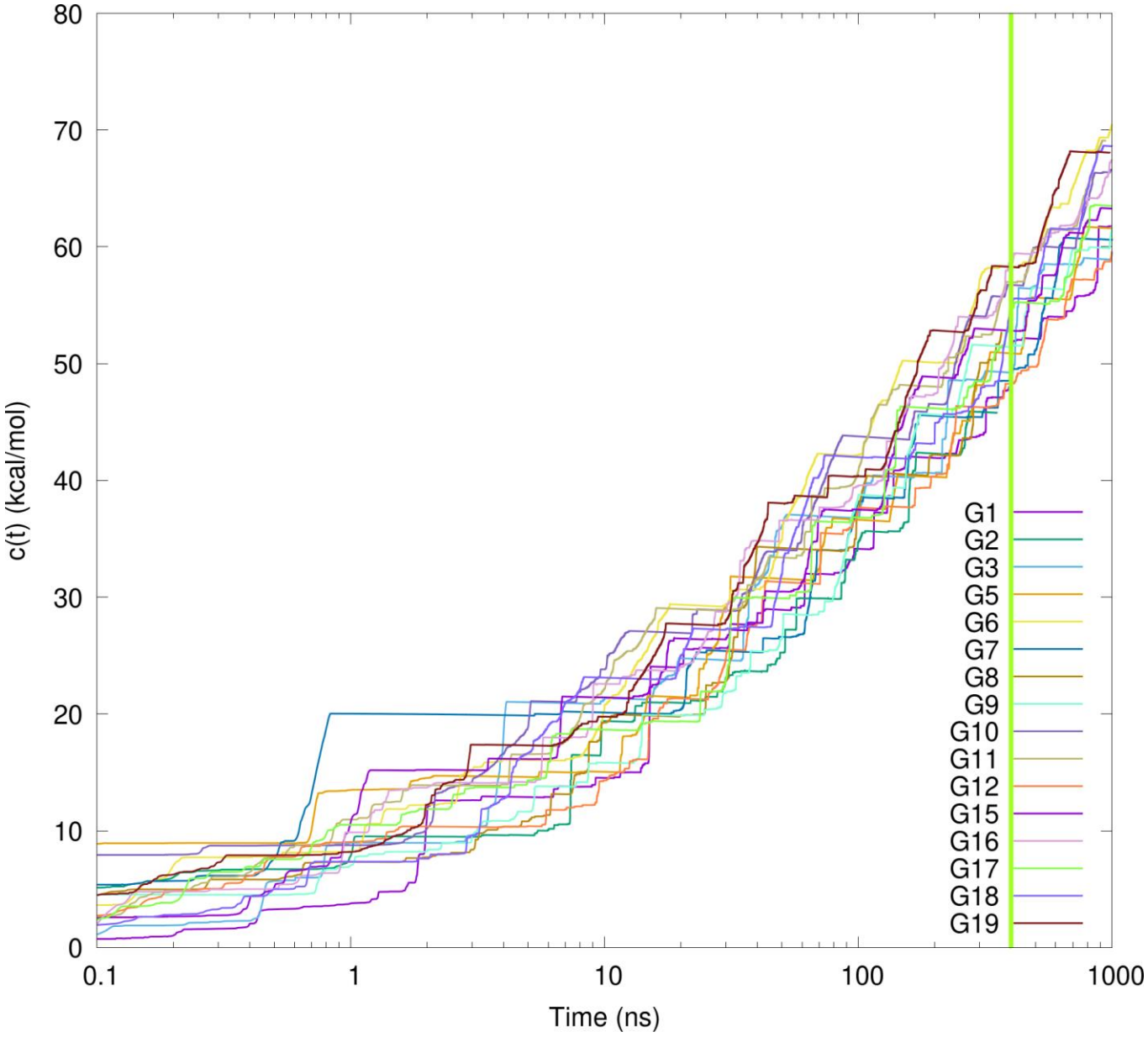

Fig. S2. Binding affinities from metadynamics simulations as a function of simulation time. The length of simulation to omit is set to 400 ns, which are chosen according to the offset c(t). The binding affinity is zero at the beginning as no free energy surface is reweighted.

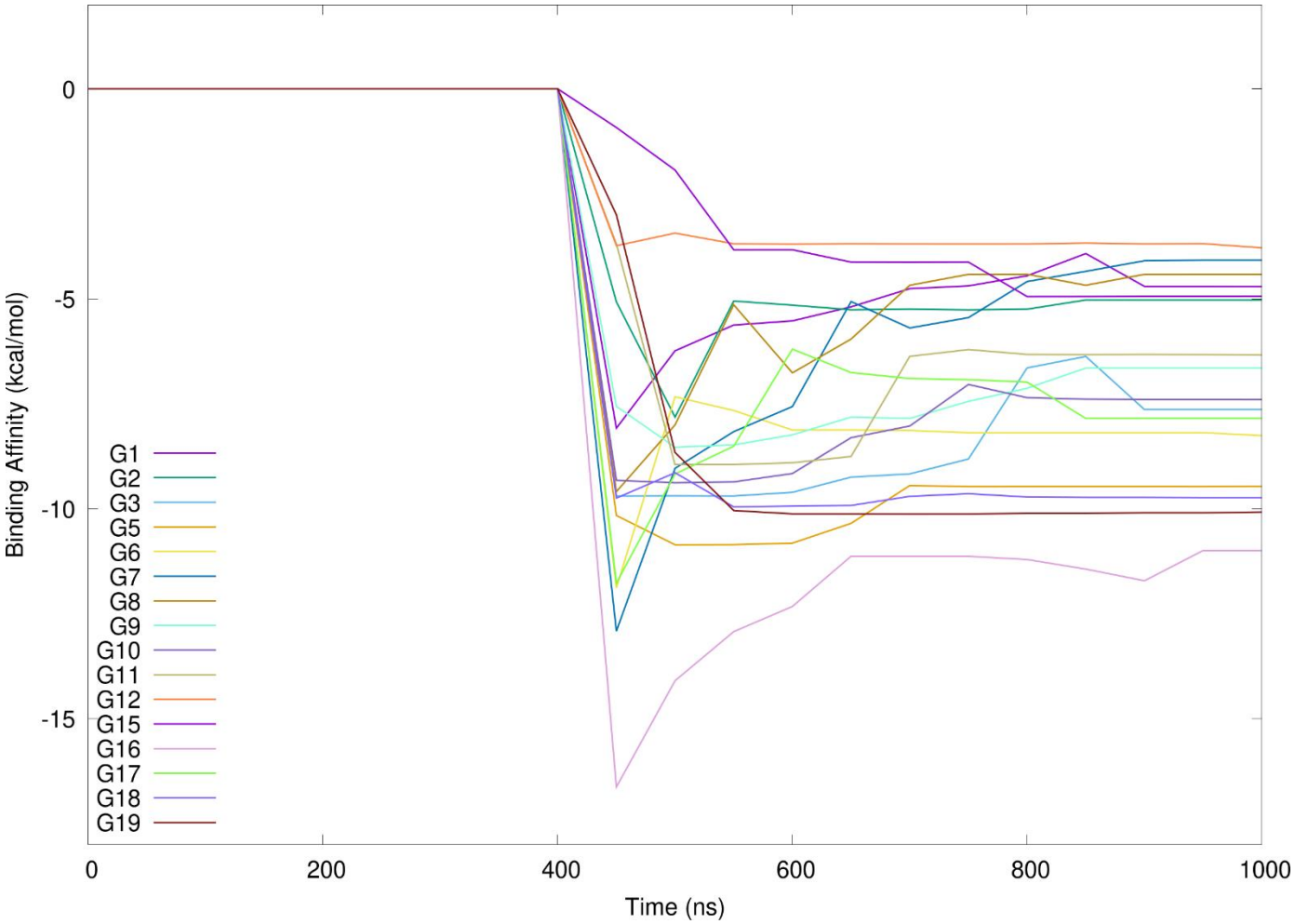

Table S1. The summary of the charges of the host TrimerTrip and its guests and the experimental binding affinities in kcal/mol.

Molecule	ΔG_{exp}	charge
TrimerTrip	-	-4
g1	-6.1	2
g2	-8.32	2
g3	-10.05	2
g5	-11.1	2
g6	-9.6	1
g7	-6.5	2
g8	-9.45	2
g9	-7.57	1
g10	-8.17	2
g11	-9.02	1
g12	-8.29	1
g15	-10.52	2
g16	-11.5	2
g17	-11.8	2
g18	-10.55	2
g19	-11.7	2

Table S2. The volume of the host TrimerTrip and the resulting entropic corrections. The probe radius used is 2.0 Å, and the grid step is set to 0.5 Å. The only statistical quantity in the equation for the entropic correction is the volume of the host molecule V_{host}. As the hosts are quite rigid and the fluctuation of their sizes is very small, the statistical error of V_{host} is negligible. Therefore, we do not give any statistical error about the entropic correction.

Terms / Host	V_0 (A^3)	V_{host} (A^3)	ρ_s (A)	V_s (A^3)	entropic correction (kcal/mol)
TrimerTrip	1660.0	5270.0	28.0	91952.3	2.334