SAMPL6 host-guest binding affinities and binding poses from spherical-coordinates-biased simulations

Zhaoxi Sun[1,2*], Qiaole He[3], Xiao Li[4], and Zhengdan Zhu[5,6]

[1]*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

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

[3]*School of Biotechnology, East China University of Science and Technology, 200237 Shanghai, China*

[4]*Physics, engineering, earth, environment, mechanics (PhITEM), University Grenoble Alpes, 38000 Grenoble, France*

[5]*CAS Key Laboratory of Receptor Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*

[6]*University of Chinese Academy of Sciences, Beijing 100049, China*

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

Abstract

Host-guest binding is a challenging problem in computer simulation. The prediction of binding affinities between hosts and guests is an important part of the statistical assessment of the modeling of proteins and ligands (SAMPL) challenges. In this work, the volume-based variant of well-tempered metadynamics is employed to calculate the binding affinities of the host-guest systems in the SAMPL6 challenge. By biasing the spherical coordinates describing the relative position of the host and the guest, the initial-configuration-induced bias vanishes and all possible binding poses are explored. The agreement between the predictions and the experimental results and the observation of new binding poses indicate that the volume-based technique serves as a nice candidate for the calculation of binding free energies and the search of the binding poses.

Introduction

Accurate prediction of the free energy difference between states of interest is of great importance in computer modeling.[1-8] Due to the energy penalty in the Boltzmann factor and the numerical limitation in the integration of the equation of motion, long simulations are required to obtain converged statistics of the systems.[9-12] Fortunately, with the development of computer architecture and the optimization of simulation software, modern researchers could achieve sufficient level of convergence in many cases.[13-15] The sampling efficiency could be further increased with the enhanced sampling techniques, which greatly expand the applicability of computer simulations.[16-24] For instance, umbrella sampling[25] was employed to obtain converged thermodynamic profile in the protonation-dependent and sulfur-substitution-dependent base flipping cases with several μs simulations.[26-27] Nonphysical pathway was constructed to connect the states of interest and free energy calculation was performed along the alchemical reaction coordinate to determine the free energy difference between physical end states.[21, 28-38] Aside from the sampling issue, the accuracy of the description of the system or the so-called Hamiltonian is another source of error limiting the quality of prediction. Quantum mechanics Hamiltonians[39-43] provide accurate descriptions but are computationally demanding, while all-atom force fields[44-48] provide a faster alternative with moderate accuracy.

In computer-aided drug discovery, high throughput virtual screening provides preliminary hits, which are then refined with free energy calculations.[49-57] Hit-to-lead optimization is often performed with the so-called alchemical method, which features an alchemical pathway connecting the end states.[58-64] Direct simulation of the binding/unbinding event is also popular.[65-67] Although the simulation of the direct binding/unbinding event is often more time-consuming and computationally expensive, more physical pictures in the host-guest interactions could be obtained. All-atom force fields have often been employed due to efficiency considerations. Numerous biologically relevant systems reported by experimental investigators have been simulated.[37, 60] Recently, some challenging systems have been proposed by computational chemists to assess the sampling and Hamiltonian issues. The statistical assessment of the modeling (SAMPL) challenges are examples of such grand challenges.[68-73] Their expanding areas include solvation free energies, pKa, host-guest systems, and partition coefficients. In the 6th statistical assessment of the modeling (SAMPL6), the host-guest binding is one of the main challenge and many researchers submitted predictions obtained from various simulation techniques.[71, 74-75] For instance, equilibrium free energy simulation methods such as umbrella sampling and the double decoupling method were used to calculate the binding affinities.[70, 73] The nonequilibrium pulling technique in the alchemical space was employed to obtain the binding affinities.[72] The various mean errors reported are about 2 kcal/mol.[71]

In umbrella sampling, the distance between atoms in the host and the guest or its mass-weighted variants is often chosen as the collective variable (CV) describing the binding and unbinding events. However, this one-dimensional (1D) CV could only describe the distance between groups of atoms, but details about their relative position are absent. Recently, the spherical coordinates (ρ,θ,φ) have been emerged to enhance the sampling of the binding/unbinding event in protein-ligand systems.[76] The three-dimensional (3D) CV set enables us to scan all possible binding poses, which makes it a nice alternative for simulating the host-guest binding event in the SAMPL challenge. In the current work, with this set of CV, we explore all possible binding poses in the host-guest systems in SAMPL6 and reweight the statistics on the two-dimensional (2D) radius-contacts ($\rho - C$) surface to calculate the binding affinities. Two widely applied charge schemes will be employed. The illustrative calculations show that multiple binding poses could be identified and the quality of prediction is comparable to the existing publications.

Methodology and Computational Details

System preparation. The host molecules include two cavitands named octa acids (OA)[77] and tetra-endo-methyl octa-acids (TEMOA)[78] and one molecule from the cucurbituril family named cucurbit[8]uril (CB8).[79] OA and TEMOA are of low symmetry. Their basket-like binding pockets are accessible to the guest through a large entryway. The difference between these two hosts is the 4 methyl groups at the entryway, which alters the size of the acceptable guest. These two hosts share the same set of 8 guest molecules. The highly symmetric CB8 is composed of 8 identical glycoluril monomers, which form a ring. 13 guest molecules for CB8 are simulated in the current work. The structures of the hosts and guests are obtained from the online server of the SAMPL6 challenge.[80] The structures of the hosts and the guests are shown in Fig. 1.

Two widely accepted charge schemes are used in the current work. We construct the hosts and the guests with AM1-BCC[81] charges and the other parameters such as bonded terms and vdW radius are obtained from the general Amber force field (GAFF) force field.[82] The other charge scheme used is the restrained electrostatic potential (RESP) scheme. We construct the systems (hosts and guests) again with the traditional RESP fitting procedure, namely B3LYP/6-31G* optimization and HF/6-31G* electrostatic potential (ESP) scanning. A single conformation is used to generate the ESP data and restrained fitting gives the atomic charges. The other parameters are again obtained from GAFF. The systems are solvated with TIP3P[83-84] water molecules in octahedron boxes. The truncated octahedron cell is replicated in whole space by periodic

boundary conditions. The minimum distance between the box edge and the surface of the solute is set to 28 Å. This value is much larger than the one used in normal protein-ligand simulations. The reason for choosing such a value would be discussed later. Non-polarizable spherical counter ions of Na^+ and Cl^- parameterized for TIP3P water model by Joung and Cheatham[85-86] are added for neutralization.

MD simulations.

The well-tempered metadynamics method is used to enhance the sampling of the binding/unbinding event.[66, 87-88] Gaussian biasing potentials are added periodically and the overall biasing potential increases with time. The method defines the biasing potential with 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 V_n represents the overall biasing potential in the *n*th step, $\mathbf{s}$ is the CV matrix, $G(\mathbf{s}, \mathbf{s}_n)$ denotes the Gaussian kernels of biasing potentials, γ is the bias factor. The post-process reweighting is obtained by the time-independent algorithm.[89] The resulting free energy estimate can be obtained with the reweighting formula applicable for finite-time sampling estimates,

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

where the canonical bracket denotes ensemble average, O is the mechanical observable, β represents the reciprocal temperature, c is the offset of the biasing potential or the average biasing potential in the CV space (i.e. rct), and t is the time of simulation. This reweighting formula uses the rbias term $V_{rbias}(\mathbf{s}, t) = V(\mathbf{s}, t) - c(t)$, and requires to omit the first part of the simulation. Note that other enhanced sampling methods such as the Hamiltonian replica exchange method (H-REMD)[90-91] are also usable.

The CV set we bias should be able to describe the binding/unbinding between the host and the guest and differentiate different binding poses appropriately. The spherical coordinates (ρ, θ, φ) define the relative position of the center of mass (COM) of the host and that of the guest, and thus satisfy the above requirements. Biasing these coordinates enables us to scan all possible binding poses of the host-guest systems and thus find the most stable binding pose. Therefore, the initial configuration of the host-guest systems is irrelevant and can be chosen randomly. With a large radius ρ, the guest is fully decoupled from the host, which provides a nice definition of the unbound state. As the simulation box is of finite size, an upper wall is added on the radius ρ, which limits the volume of phase space that the ligand could explore. An entropic correction defined in Eq. (3) is thus added to recover the unbiased free energy.

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

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

The volume-based technique scans a sphere around the host. Therefore, a huge water box centered at the COM of the host is required to solvate the system. The minimum size of the box is determined by the minimum distance between the COM of the host and that of the guest, at which the interactions between the host and the guest is zero. This minimum distance is determined by the sizes of the host and the guest and the strength of interactions between the host and the guest. A fully decoupled state could be defined by the zero or near-zero contact between the host and the guest. The following switching function is used to calculate the contact number between the host and the guest.

$$C = \sum_{i\in A}\sum_{j\in B}\frac{1-\left(\frac{r_{ij}}{r_0}\right)^n}{1-\left(\frac{r_{ij}}{r_0}\right)^m} \quad (4),$$

where A and B denote two groups of atoms (i.e. the host and the guest), the subscripts i and j are the *i*th and *j*th atom in the groups, m and n are 6 and 12, respectively, r refers to the distance and the threshold for the contact r_0. In our calculation of the host-guest contacts, all heavy atoms in the host and the guest are included and $r_0 = 6$ Å. The resulting decoupled state defined by this definition has the guest at least 6 Å away from the host. The upper wall on the radius ρ is set at 26 Å in the current simulation, which is large enough to define a fully decoupled state with near-zero contacts between the host and the guest.

For each host-guest system, the reference bound structure provided by the SAMPL6 online server is used as the starting structure, from which 100 ps NVT equilibration and 5 ns NPT equilibration is performed. After that, we perform 400 ns enhanced sampling simulation. As all relative positions are scanned, the starting conformation is irrelevant to the simulation results and all possible binding poses could be explored. The simulation is performed at 298 K (the experimental condition) with GROMACS 2018.4 [92] patched with PLUMED 2.6.0-dev[93]. The V-rescale algorithm[94] is employed for temperature regulation and the Parrinello-Rahman barostat[95-96] is used for pressure regulation. A time step of 1 fs is used to propagate

the dynamics. Long-range electrostatics are treated with the PME[97-98] method. 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.

Result and discussion

We discuss the results obtained with AM1-BCC charges to provide some insights into the simulation. To illustrate the enhancement of the sampling efficiency for the relative position of the host and the guest, we calculate the number of contacts between the host and the guest and decompose it into the atomic sequence. All atoms of the host and the guest are included in the calculation, which differs from the later free energy projection case, where only heavy atoms are included. We choose the OA-G5 complex as an example. The time-dependence of the total number of contacts between the host and the guest is shown in Fig. 2a, from which we notice that the binding/unbinding event is observed frequently. Namely, the sampling of the association/dissociation of the host-guest complex is enhanced by the biasing potentials on the spherical coordinates. As the total number of contacts represents the overall outcome of the host-guest interactions, in order to investigate the details of the host-guest binding, e.g. identifying the parts of the host coordinating the guest, we decompose the total number of contacts into the contributions of each atom of OA. The resulting time series of the atom-specific contacts is given in Fig. 2b, from which we know that the guest explores the accessible areas around the host. Therefore, all possible binding poses could be identified, as long as the simulation converges.

Convergence check is indispensable in free energy calculation. It provides the evidence of the reliability of the ergodic assumption and thus should be checked in the first place. The time-evolution of the height of Gaussian potentials (cf. Fig. S1.) indicates that for all host-guest systems, at the end of the simulation, the height of new Gaussian potentials decreases to a very small value (e.g. 0.0002 kcal/mol). Therefore, the overall biasing potential in the whole CV space changes slowly. The average bias or the offset $c(t)$ (rct in the plot) is shown in Fig. S2, from which we know that the average biasing potential increases smoothly over the simulation time. The convergence is further checked with the time-dependence of the binding affinity. We project the free energy on the radius-contacts (ρ-C) surface and two examples are shown in Fig. 3. We choose the global free energy minimum as the most stable binding pose and calculate the free energy difference between that state and the zero-contact state. The time-dependence of the binding affinities is shown in Fig. S3. We can see that 350 ns is already long enough for convergence and the binding affinities

do not change with further simulation.

After checking the convergence of the simulation, we then focus on the binding affinities predicted by the volume-based technique. The volume and the resulting entropic correction for the host-guest systems are summarized in Table S1. The volumes of the hosts are estimated to be 5510 $Å^3$, 5360 $Å^3$, and 3840 $Å^3$, respectively. Then, as the radius of the sphere is 26 Å, the volume corrections are calculated to be 2.195 kcal/mol, 2.193 kcal/mol and 2.206 kcal/mol.

The corrected binding free energies for the host-guest systems are given in Table 1, Table 2 and Table 3. 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 used to assess the errors and the consistency of the ranks of binding affinities from our modeling and the experimental reference. The reported RMSE values are 1.3 kcal/mol, 2.4 kcal/mol and 3.5 kcal/mol for the three hosts, which are comparable to the reported statistics in the overview of SAMPL6 submission.[71] The case of the ranking coefficients is similar. A comparison between the predictions and the experimental references is given in Fig. 4. We can see that for the hosts of OA and TEMOA, the predicted binding affinities agree to the experimental results, while the agreement for the CB8 case is not very good. The results obtained with RESP charges are given in Table S2, Table S3 and Table S4. The RESP results are also shown in Fig. 4. We can see that the predicted binding affinities slightly differ from the AM1-BCC results, but the mean errors and the ranking coefficients are similar. Specifically, for the host of OA, that the binding affinities of the guests G1 and G4 are predicted to deviate from the experimental reference in both charge schemes, while for the guests of G3, G5, G6 and G7, both charge schemes provide results consistent with the experiments. For G0 and G2, the BCC results seem to agree with the experiments, while the deviations of the RESP ones are relatively large. As a result, the binding affinities calculated with the BCC charges are in better agreement with the experimental references. The three error estimates (including RMSE, MSE, and MAE) are all smaller in BCC. Note that the rank coefficients are very similar for the predictions obtained from the two charge schemes and are larger than 0.5, which indicates that both of them are able to correctly predict the ranks of the binding affinities for these guest systems targeting OA. For the TEMOA case, the predictions from the two charge schemes differ from the experimental values for G0, G1, G2 and G7. In the G3 and G5 cases, the agreements between the predictions and the experiments are good for both charge schemes. As for G4 and G6, the BCC results are in better agreement with the experiments, compared with the RESP ones. The three mean error estimates of the dataset from different charge schemes are similar. The BCC charge scheme provides very good predictions on the ranking coefficients, while the RESP ones are

worse. For the last host CB8, the agreements between the predicted binding affinities and the experimental references are worse than the previous two cases. Only for the guest of G0, the predictions from the two charge schemes agree with the experiments within the statistical uncertainty, while for the other guests, the deviations are large. As a result, the mean error estimates are larger than the previous two cases. Note that the three mean errors are similar for the two charge schemes, but the ranking coefficient from the RESP charges is better than that from the BCC charges. Note that the quality of our predictions is comparable to the existing publications. Therefore, the performance of biased simulations along spherical coordinates is similar to other techniques. The success of these predictions indicates the applicability of the spherical coordinates in the simulations of the host-guest binding.

Aside from the binding affinity, an interesting phenomenon we observed is the existence of multiple stable binding poses in the host-guest systems. Although we use the reference bound state given by the SAMPL6 online server[80] as the starting structure, all possible binding poses are scanned and the most stable binding pose could be found. We choose two examples to discuss the features of different binding cases in the following paragraph.

In Fig. 3, two typical free energy surfaces are given. The first one has only one free energy minimum, while the other has two minima, which indicates the existence of multiple stable binding poses. In Fig. 3a, starting from the free energy minimum, the free energy increases monotonically with the distance ρ, and the number of contacts decreases with ρ. This indicates that when the guest leaves the bound state, the contacts between the host and the guest become less and the host-guest interactions become weaker. When they are separated by about 15 Å, the contact number becomes near-zero and the guest is fully decoupled from the host. However, in Fig. 3b, such a monotonic behavior changes to some extent and multiple free energy minima exist. This suggests that there are multiple (meta-) stable binding poses. It is worth noting that the CB8-G3 complex is reported to have difficulty in convergence.[80] The bias due to the initial configuration could be significant and the correlation time is long. We visualize the structures of the binding poses to dig deeper into the problem. As two free energy minima exist, we extract one configuration from each free energy basin. Aside from the binding pose provided from the SAMPL6 online server, the binding poses observed in our simulation include one with the aromatic ring being coordinated at the center of the host. The free energy landscape indicates that these two binding poses are of similar binding affinities. Therefore, it is clear that by biasing the spherical coordinates, the whole space of binding poses could be explored efficiently and the simulation results are irrelevant to the starting configuration.

Conclusion

The current work provides illustrative calculations of the binding affinities in host-guest systems with the recently proposed volume-based metadynamics. Specifically, we simulate SAMPL6 host-guest systems by biasing the spherical coordinates (ρ,θ,φ), which scan all possible binding poses. For some systems, only one stable binding pose is observed, while for the other multiple stable poses exist. For the CB8-G3 complex, we observe a new binding pose that differs from the starting configuration provided by the SAMPL6 online server, which indicates that the initial-configuration bias is effectively avoided. The quality of the predicted binding affinities of the host-guest systems is comparable to existing publications. Note that the success of the current simulation is mainly determined by the choice of CV set, while the enhanced sampling method could be changed to alternatives such as H-REMD. Therefore, our results indicate that the spherical coordinates could be a nice candidate for the CV chosen in the computational modeling of the host-guest binding.

Acknowledgement

This work was supported China Scholarship Council. Computer access to the CLAIX cluster of RWTH Aachen University and clusters of Forschungszentrum Juelich is gratefully acknowledged. 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 height of Gaussian potentials during simulations, the average biasing potential in the CV space, the time-evolution of the binding affinities from metadynamics simulations, the volumes of the hosts and the resulting entropic corrections, the binding affinities of the host-guest systems with RESP charges are given in the supporting information.

References

1. Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019,** *40* (12), 1270-1289.
2. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann, M., Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **2001,** *93* (2-3), 215-230.
3. Makhatadze, G. I.; Privalov, P. L., Hydration effects in protein unfolding. *Biophys. Chem.* **1994,** *51* (2-3), 291-309.
4. Chang, C.-E.; Gilson, M. K., Free energy, entropy, and induced fit in host− guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004,** *126* (40), 13156-13164.
5. Chodera, J. D.; Mobley, D. L., Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Ann. Rev. Biopys.* **2013,** *42*, 121-142.
6. Reynolds, C. H.; Holloway, M. K., Thermodynamics of ligand binding and efficiency. *ACS Med. Chem. Lett.* **2011,** *2* (6), 433-437.
7. Böhm, H. J.; Klebe, G., What can we learn from molecular recognition in protein–ligand complexes for the design of new drugs? *Angewandte Chemie International Edition in English* **1996,** *35* (22), 2588-2614.
8. Wang, X.; Deng, B.; Sun, Z., Thermodynamics of helix formation in small peptides of varying length in vacuo, in implicit solvent, and in explicit solvent. *J. Mol. Model.* **2019,** *25* (1), 3.
9. 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.
10. 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.
11. Rani, P.; Biswas, P., Diffusion of hydration water around intrinsically disordered proteins. *J. Phys. Chem. B* **2015,** *119* (42), 13262-13270.
12. 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.
13. 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.
14. 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.
15. Plimpton, S., Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995,** *117* (1), 1-19.
16. Best, R. B.; de Sancho, D.; Mittal, J., Residue-specific α-helix propensities from molecular simulation. *Biophys. J.* **2012,** *102* (6), 1462-1467.
17. 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.
18. 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.
19. 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.
20. 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.
21. Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z., Interaction Entropy for Protein-Protein Binding. *J. Chem. Phys.* **2017,** *146* (12), 124124.
22. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Citeria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.

23. 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
24. 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
25. Kästner, J., Umbrella sampling. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011,** *1* (6), 932-942.
26. 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.
27. 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*, 14923-14940.
28. Wang, X.; Tu, X.; Zhang, J. Z. H.; Sun, Z., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation: The Nonequilibrium Stratification. *Phys. Chem. Chem. Phys.* **2018,** *20* (3), 2009-2021.
29. Shirts, M. R.; Pande, V. S., Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005,** *122* (13), 134508.
30. Hummer, G.; Pratt, L. R.; Garcia, A. E., Hydration free energy of water. *J. Phys. Chem.* **1995,** *99* (38), 14188-14194.
31. 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.
32. Sun, Z. X.; Wang, X. H.; Zhang, J. Z. H., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation. *Phys. Chem. Chem. Phys.* **2017,** *19* (23), 15005-15020.
33. Gallicchio, E.; Levy, R. M., Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011,** *21* (2), 161-6.
34. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S., Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011,** *21* (2), 150-60.
35. Parenti, M. D.; Rastelli, G., Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol. Adv.* **2012,** *30* (1), 244-250.
36. Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K., Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009,** *394* (4), 747-763.
37. 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.
38. Qiu, L.; Yan, Y.; Sun, Z.; Song, J.; Zhang, J. Z., Interaction entropy for computational alanine scanning in protein–protein binding. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2018,** *8* (2), e1342.
39. Dewar, M. J. S.; Thiel, W., Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977,** *99* (15), 4899-4907.
40. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985,** *107* (13), 3902-3909.
41. Roothaan, C. C. J., New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951,** *23* (23), 69-89.
42. Pople, J. A.; Nesbet, R. K., Self-Consistent Orbitals for Radicals. *J. Chem. Phys.* **1954,** *22* (3), 571-572.
43. Mcweeny, R.; Diercksen, G., Self-Consistent Perturbation Theory. II. Extension to Open Shells. *J. Chem. Phys.* **1968,** *49* (11), 4852-4856.
44. 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.
45. 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.
46. 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.
47. 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.

48. Best, R. B.; Buchete, N.-V.; Hummer, G., Are current molecular dynamics force fields too helical? *Biophys. J.* **2008,** *95* (1), L07-L09.
49. 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.
50. Kapetanovic, I., Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem. Biol. Interact.* **2008,** *171* (2), 165-176.
51. 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.
52. 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.
53. Shoichet, B. K.; Leach, A. R.; Kuntz, I. D., Ligand solvation in molecular docking. *Proteins-structure Function & Bioinformatics* **1999,** *34* (1), 4.
54. 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.
55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. Wang, X.; Sun, Z., Understanding PIM-1 kinase inhibitor interactions with free energy simulation. *Phys. Chem. Chem. Phys.* **2019,** *21*, 7544-7558.
62. 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.
63. 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.
64. 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**.
65. 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.
66. 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.
67. 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.
68. 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.
69. 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.
70. 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.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. 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.
77. Gibb, C. L. D.; Gibb, B. C., Well-defined, organic nanoenvironments in water: the hydrophobic effect drives a capsular assembly. *J. Am. Chem. Soc.* **2004,** *126* (37), 11408-9.
78. Haiying, G.; Benjamin, C. J.; Gibb, B. C., Nonmonotonic assembly of a deep-cavity cavitand. *J. Am. Chem. Soc.* **2011,** *133* (13), 4770-3.
79. Murkli, S.; Mcneill, J. N.; Isaacs, L., Cucurbit[8]uril•guest complexes: blinded dataset for the SAMPL6 challenge. *Supramol. Chem.*
80. https://github.com/samplchallenges/SAMPL6.
81. 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.
82. 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.
83. 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.
84. 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.
85. 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.
86. 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.
87. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011,** *1* (5), 826-843.
88. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008,** *100* (2), 020603.
89. Tiwary, P.; Parrinello, M., A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015,** *119* (3), 736-42.
90. Sabri, D. D.; Roitberg, A. E., Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning. *J. Chem. Theory Comput.* **2013,** *9* (9), 4692–4699.
91. Park, S.; Kim, T.; Im, W., Transmembrane helix assembly by window exchange umbrella sampling. *Phys. Rev. Lett.* **2012,** *108* (10), 207-212.
92. 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.
93. 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.

94. Giovanni, B.; Davide, D.; Michele, P., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 2384.
95. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50* (5), 1055-1076.
96. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
97. 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.
98. 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 OA-guest binding affinities in kcal/mol obtained from metadynamics simulations with AM1-BCC charges. 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
OA	G0	-5.68	-3.4	0.6	2.2	-5.6	0.6
	G1	-4.65	-4.3	0.6	2.2	-6.5	0.6
	G2	-8.38	-6.7	0.6	2.2	-8.9	0.6
	G3	-5.18	-3.4	0.6	2.2	-5.6	0.6
	G4	-7.11	-7.8	0.5	2.2	-10.0	0.5
	G5	-4.59	-2.8	0.5	2.2	-5.0	0.5
	G6	-4.97	-2.1	0.6	2.2	-4.3	0.6
	G7	-6.22	-4.1	0.5	2.2	-6.3	0.5
RMSE						1.3	
MSE						0.7	
MAE						0.9	
τ						0.6	
PI						0.8	

Table 2. The TEMOA-guest binding affinities in kcal/mol obtained from metadynamics simulations with AM1-BCC charges. Error estimates are obtained from block averaging. 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.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
TEMOA	G0	-6.06	-7.5	0.5	2.2	-9.7	0.5
	G1	-5.97	-8.0	0.6	2.2	-10.2	0.6
	G2	-6.81	-7.7	0.6	2.2	-9.9	0.6
	G3	-5.6	-2.6	0.6	2.2	-4.8	0.6
	G4	-7.79	-6.9	0.6	2.2	-9.1	0.6
	G5	-4.16	-1.4	0.6	2.2	-3.6	0.6
	G6	-5.4	-4.4	0.5	2.2	-6.6	0.5
	G7	-4.13	-0.1	0.6	2.2	-2.3	0.6
RMSE						2.4	
MSE						1.3	
MAE						2.0	
τ						0.6	
PI						0.9	

Table 3. The CB8-guest binding affinities in kcal/mol obtained from metadynamics simulations with AM1-BCC charges. Error estimates are obtained from block averaging. 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.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
CB8	G0	-6.69	-4.3	0.6	2.2	-6.5	0.6
	G1	-7.65	-4.3	0.6	2.2	-6.5	0.6
	G2	-7.66	-11.0	0.6	2.2	-13.2	0.6
	G3	-6.45	-9.5	0.6	2.2	-11.7	0.6
	G4	-7.8	-11.0	0.5	2.2	-13.2	0.5
	G5	-8.18	-12.2	0.6	2.2	-14.4	0.6
	G6	-8.34	-8.6	0.6	2.2	-10.9	0.6
	G7	-10	-8.0	0.5	2.2	-10.2	0.5
	G8	-13.5	-11.2	0.6	2.2	-13.4	0.6
	G9	-8.68	-9.3	0.5	2.2	-11.5	0.5
	G10	-8.22	-2.0	0.6	2.2	-4.2	0.6
	G11	-7.77	-7.8	0.6	2.2	-10.0	0.6
	G12	-9.86	-7.6	0.6	2.2	-9.9	0.6
RMSE						3.5	
MSE						1.9	
MAE						2.7	
τ						0.1	
PI						0.2	

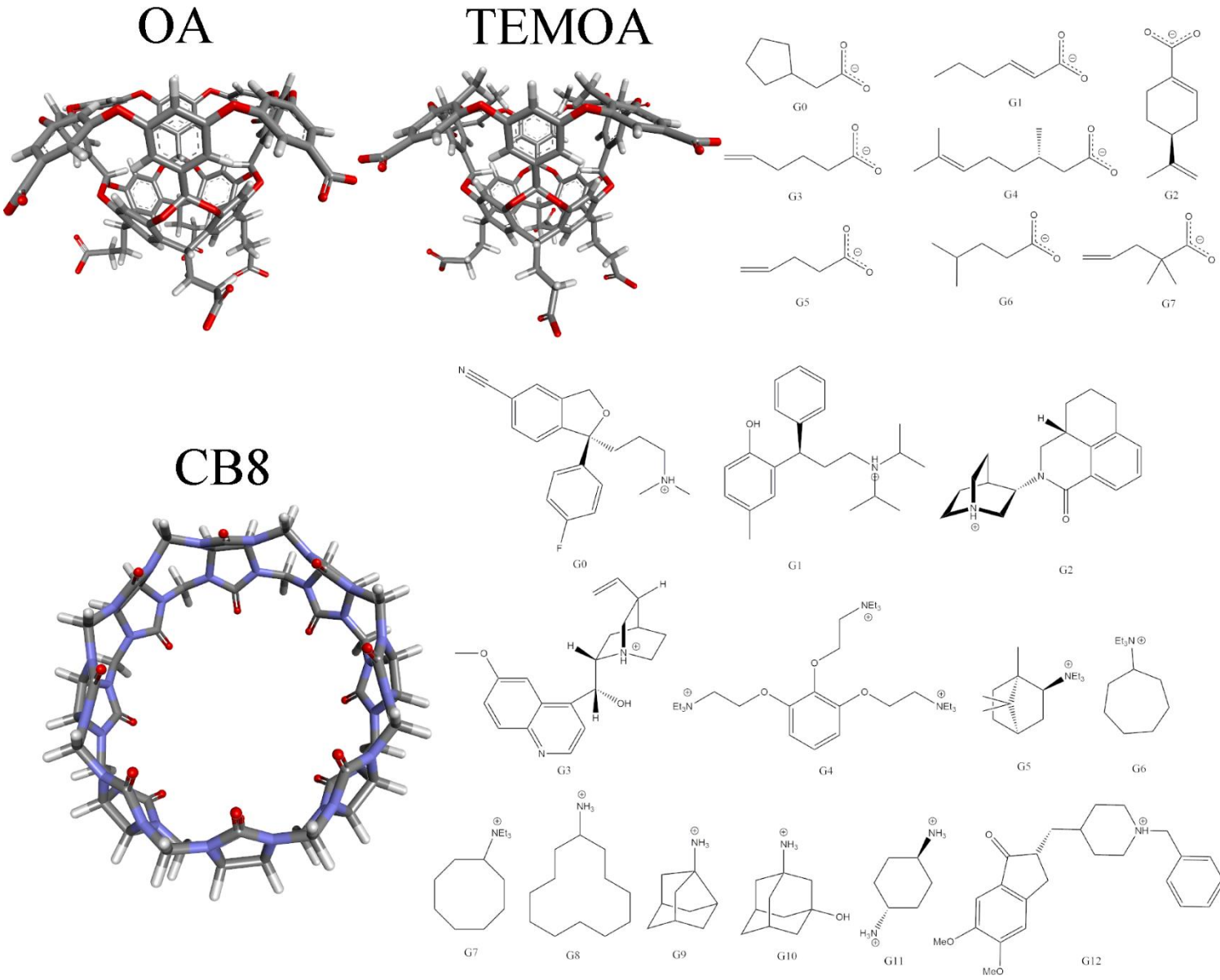

Fig. 1. 2D chemical structures of the hosts and guests. Top: OA, TEMOA and their guests from G0 to G7. Bottom: CB8 and its guests from G0 to G12.

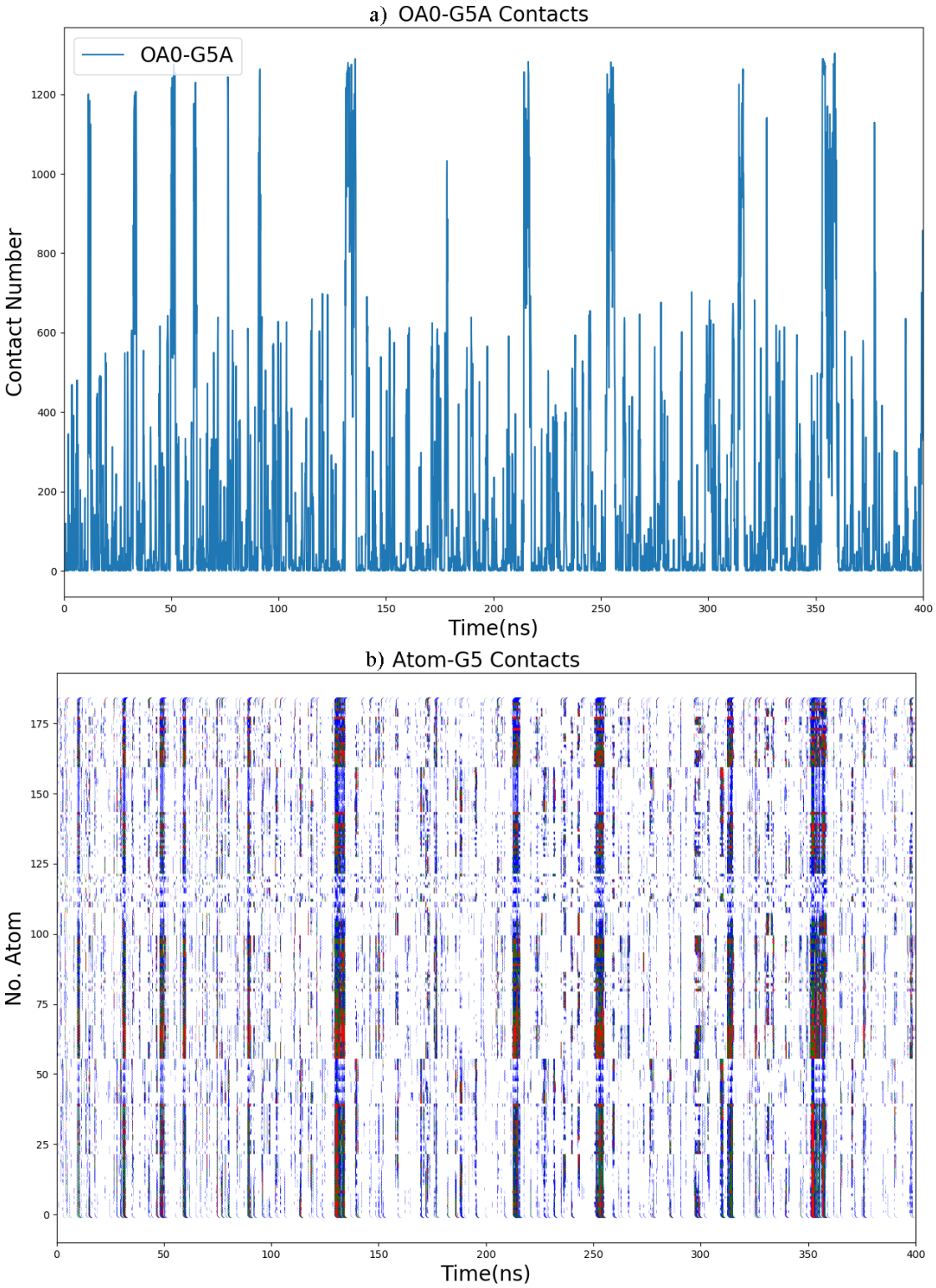

Fig. 2. a) The number of contacts between OA and G5 and b) its decomposition by each atom of OA during metadynamics simulations. Therefore, the y-axis represents the serial number of atom in the host. All atoms of the host and the guest are included in the calculation. Thus, the number of contacts differ from the later projection of the metadynamics results shown in other figures. Red dots denote contacts larger than 10, green dots represent contact number between 5 and 10, blue ones are those larger than 1, and the other are represented by white dots.

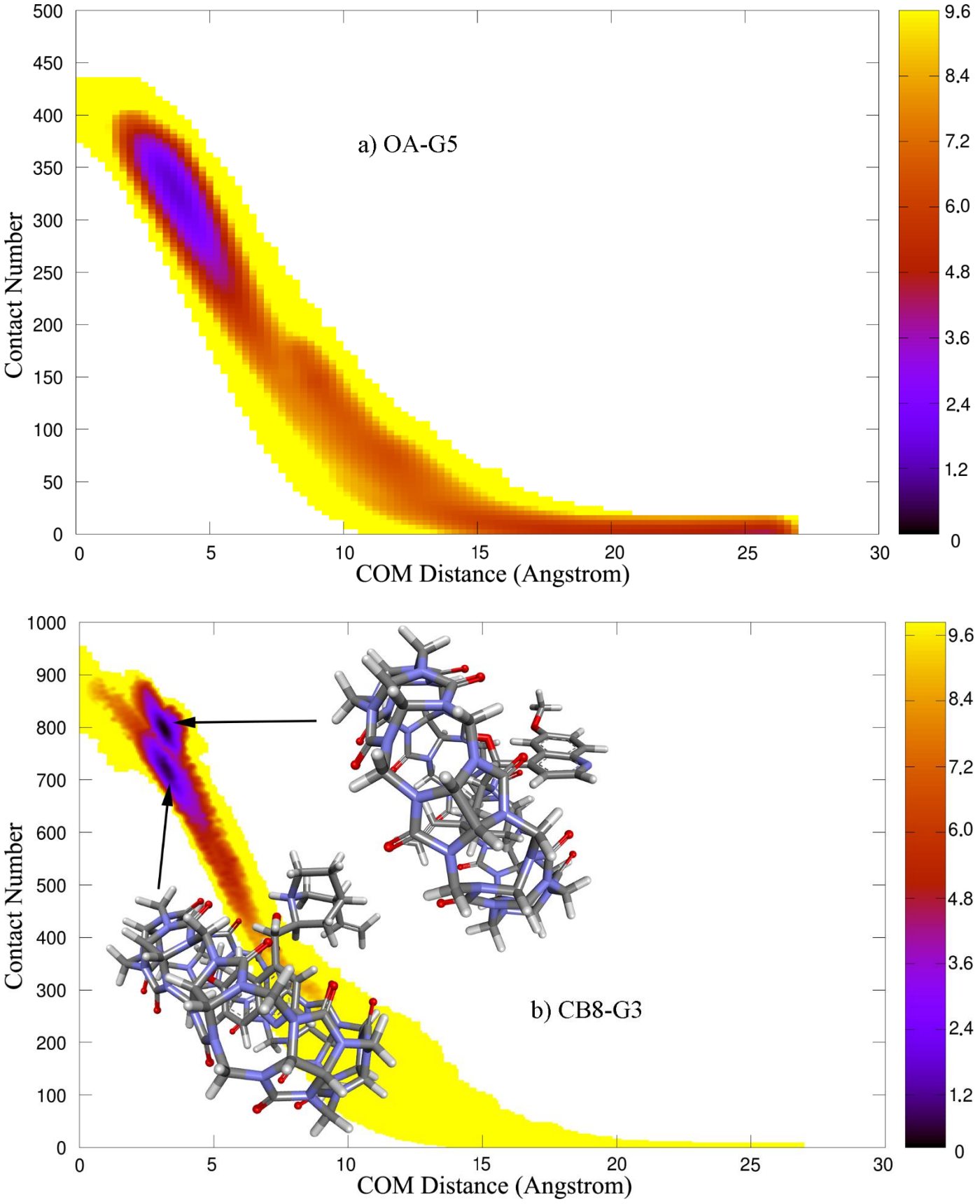

Fig. 3. Typical 2D ρ-C free energy surface in kcal/mol with a) a single free energy minimum (the OA-G5 system), and b) multiple free energy minima (CB8-G3).

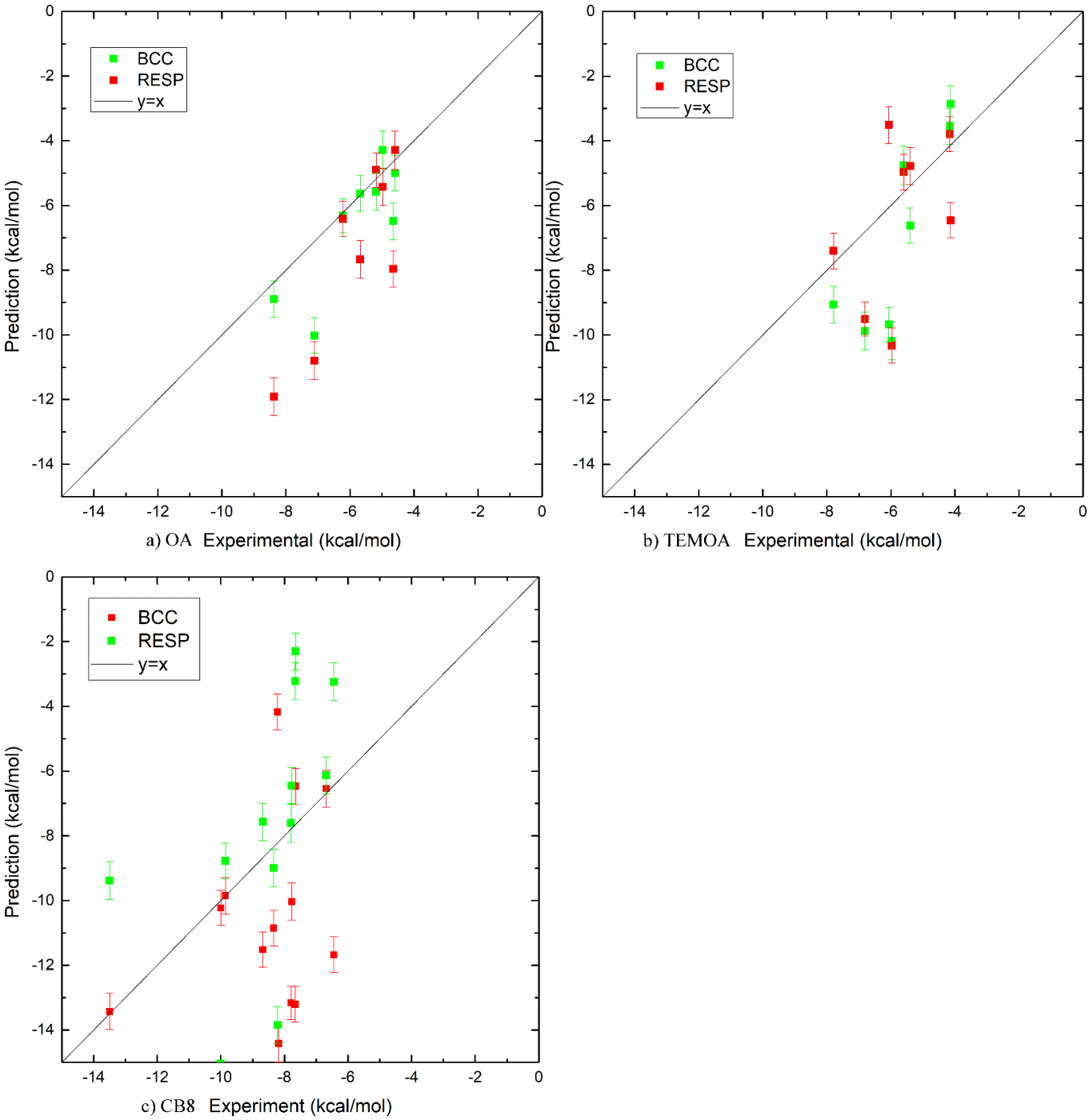

Fig. 4. Correlation between the predicted binding affinities and the experimental reference for a) OA-guest systems, b) TEMOA-guest systems, and c) CB8-guest systems. The exact values of the binding affinities are presented in Table 1-3 and Table S2-S4.

Supporting Information: SAMPL6 host-guest binding affinities and binding poses from spherical-coordinates-biased simulations

Zhaoxi Sun[1,2*], Qiaole He[3], Xiao Li[4], and Zhengdan Zhu[5,6]

[1]*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

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

[3]*School of Biotechnology, East China University of Science and Technology, 200237 Shanghai, China*

[4]*Physics, engineering, earth, environment, mechanics (PhITEM), University Grenoble Alpes, 38000 Grenoble, France*

[5]*CAS Key Laboratory of Receptor Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*

[6]*University of Chinese Academy of Sciences, Beijing 100049, China*

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

Fig. S1. Time evolution of the height of Gaussian potentials under the AM1-BCC charge scheme for a) OA-guest systems, b) TEMOA-guest systems, and c) CB8-guest systems.

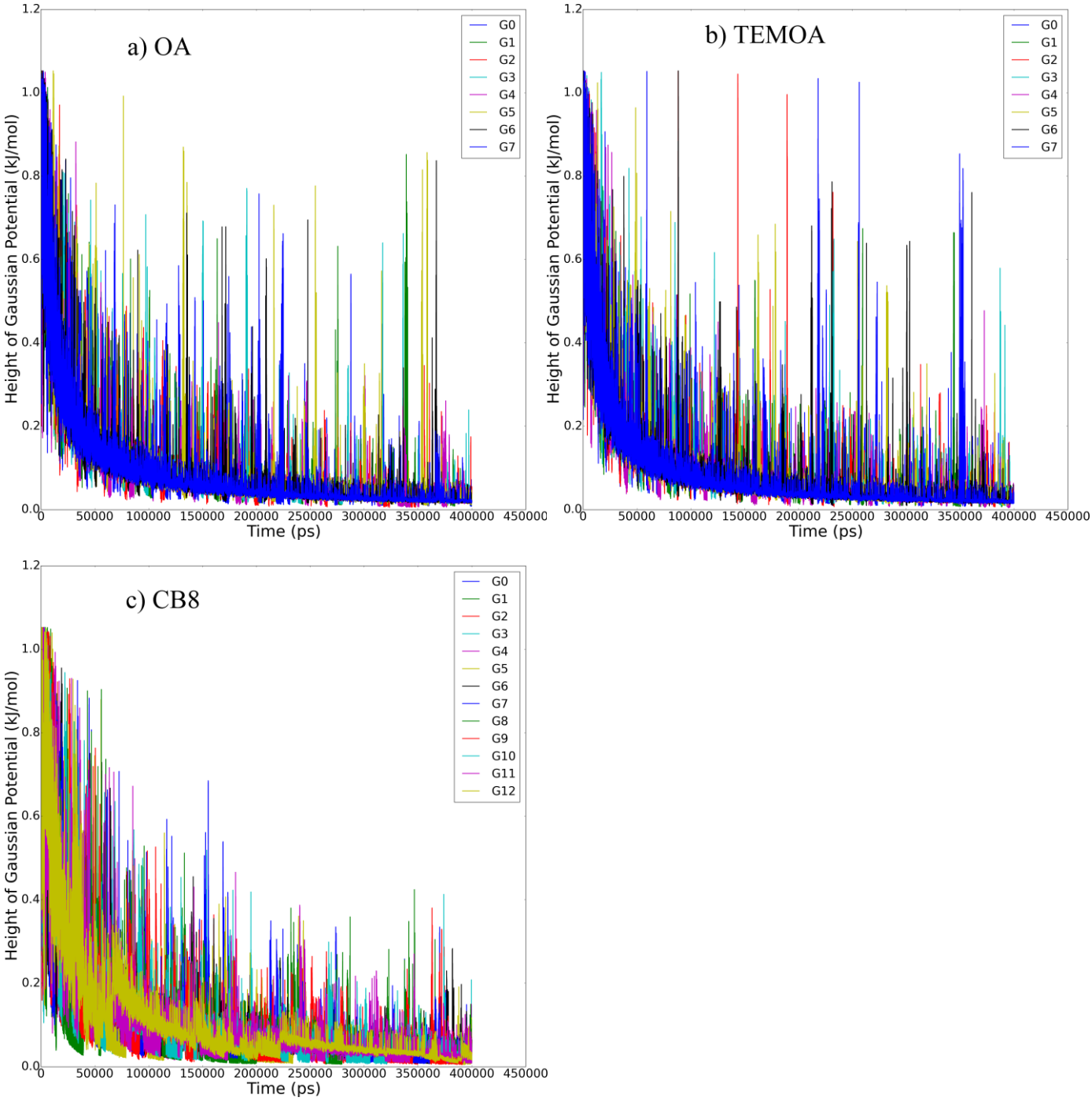

Fig. S2. Time evolution of the average bias in the CV space (i.e. rct) under the AM1-BCC charge scheme for a) OA-guest systems, b) TEMOA-guest systems, and c) CB8-guest systems.

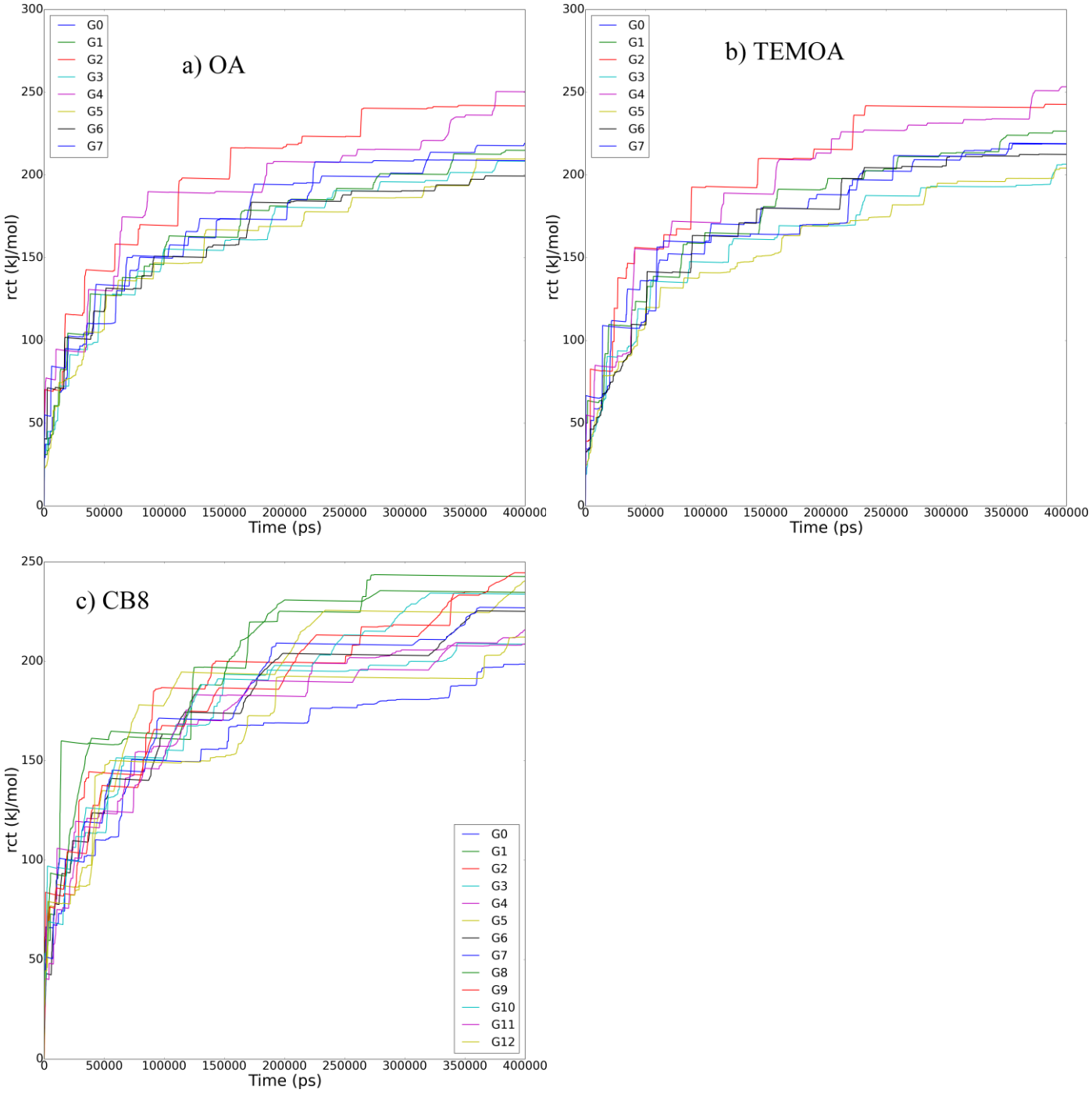

Fig. S3. Binding affinities from metadynamics simulations with AM1-BCC charges as a function of simulation time. The length of simulation to omit is set to 100 ns, which are chosen according to the average bias in the CV space (i.e. rct). The binding affinity is zero at the beginning as no free energy surface is reweighted.

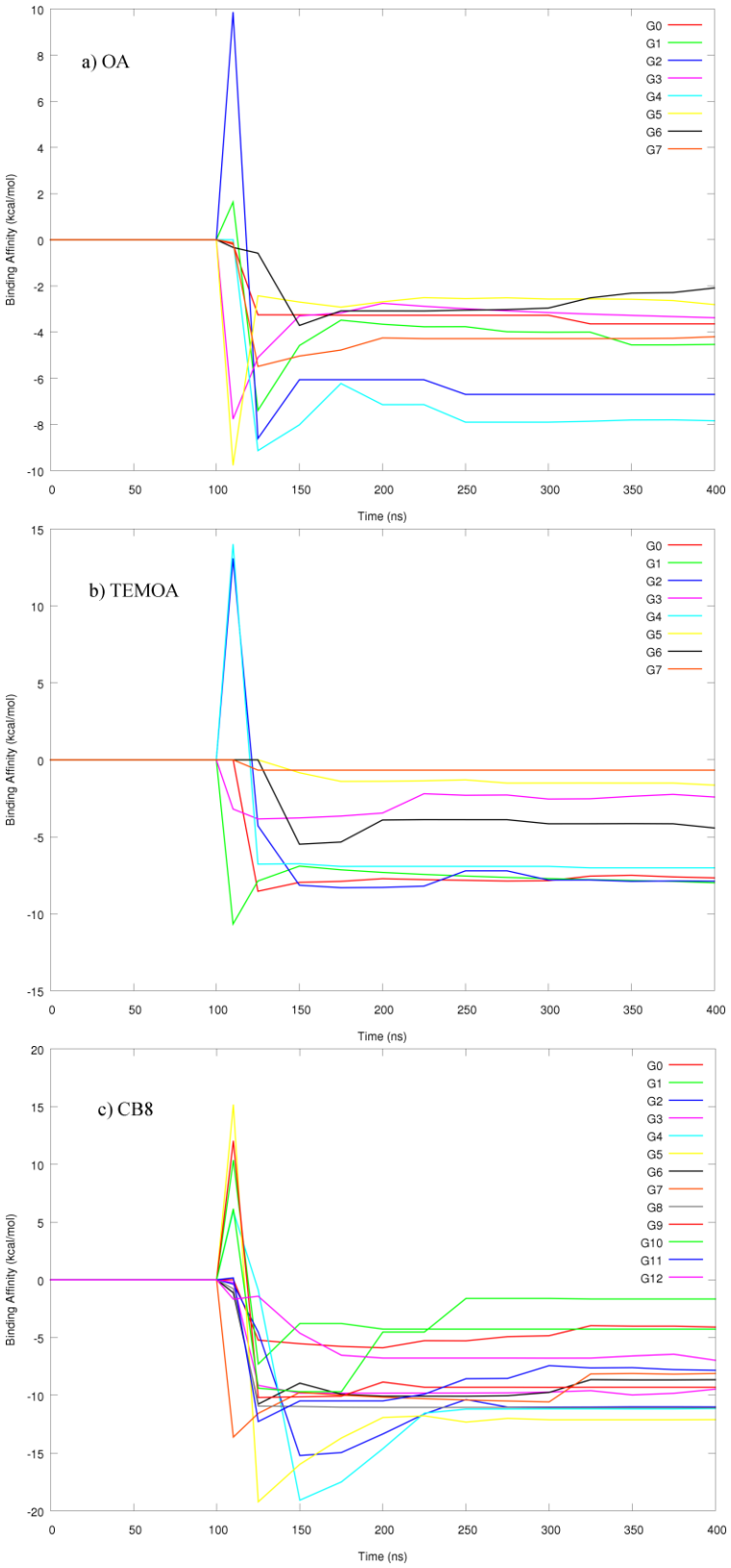

Table S1. The volume of the protein and the resulting entropic corrections. The probe radius used is 2.0 Å, and the grid step is set to 0.5 Å. The only statistical quantity in the equation for the entropic correction is the volume of the host molecule V_{host}. As the hosts are quite rigid and the fluctuation 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 / Systems	V^0 (Å^3)	V_{host} (Å^3)	ρ_s (Å)	V_s (Å^3)	entropic correction (kcal/mol)
OA	1660.0	5110.0	26.0	73622.2	2.195
TEMOA	1660.0	5360.0	26.0	73622.2	2.193
CB8	1660.0	3840.0	26.0	73622.2	2.206

Table S2. The OA-guest binding affinities in kcal/mol obtained from metadynamics simulations with RESP charges. Error estimates are obtained from block averaging. 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. 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) serve as quality measurements.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
OA	G0	-5.68	-5.5	0.6	2.2	-7.7	0.6
	G1	-4.65	-5.8	0.6	2.2	-8.0	0.6
	G2	-8.38	-9.7	0.6	2.2	-11.9	0.6
	G3	-5.18	-2.7	0.5	2.2	-4.9	0.5
	G4	-7.11	-8.6	0.6	2.2	-10.8	0.6
	G5	-4.59	-2.1	0.6	2.2	-4.3	0.6
	G6	-4.97	-3.2	0.6	2.2	-5.4	0.6
	G7	-6.22	-4.2	0.5	2.2	-6.4	0.5
RMSE						2.3	
MSE						1.6	
MAE						1.7	
τ						0.6	
PI						0.8	

Table S3. The TEMOA-guest binding affinities in kcal/mol obtained from metadynamics simulations with RESP charges. Error estimates are obtained from block averaging. 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.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
TEMOA	G0	-6.06	-1.3	0.6	2.2	-3.5	0.6
	G1	-5.97	-8.1	0.5	2.2	-10.3	0.5
	G2	-6.81	-7.3	0.5	2.2	-9.5	0.5
	G3	-5.60	-2.8	0.6	2.2	-5.0	0.6
	G4	-7.79	-5.2	0.6	2.2	-7.4	0.6
	G5	-4.16	-1.6	0.5	2.2	-3.8	0.5
	G6	-5.40	-2.6	0.6	2.2	-4.8	0.6
	G7	-4.13	-4.3	0.5	2.2	-6.4	0.5
RMSE						2.2	
MSE						0.6	
MAE						1.7	
τ						0.2	
PI						0.4	

Table S4. The CB8-guest binding affinities in kcal/mol obtained from metadynamics simulations with RESP charges. Error estimates are obtained from block averaging. 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.

Host	Guest	ΔG_{exp}	ΔG_{metad}	SD	ΔG_{VC}	ΔG_{calc}	SD
CB8	G0	-6.69	-3.9	0.6	2.2	-6.1	0.6
	G1	-7.65	-0.1	0.6	2.2	-2.3	0.6
	G2	-7.66	-1.0	0.6	2.2	-3.2	0.6
	G3	-6.45	-1.0	0.6	2.2	-3.2	0.6
	G4	-7.8	-5.4	0.6	2.2	-7.6	0.6
	G5	-8.18	-14.4	0.6	2.2	-16.6	0.6
	G6	-8.34	-6.8	0.6	2.2	-9.0	0.6
	G7	-10	-12.8	0.6	2.2	-15.0	0.6
	G8	-13.5	-7.2	0.6	2.2	-9.4	0.6
	G9	-8.68	-5.4	0.6	2.2	-7.6	0.6
	G10	-8.22	-11.6	0.6	2.2	-13.8	0.6
	G11	-7.77	-4.2	0.6	2.2	-6.4	0.6
	G12	-9.86	-6.6	0.6	2.2	-8.8	0.6
RMSE						4.0	
MSE						-0.1	
MAE						3.2	
τ						0.5	
PI						0.6	