Comprehensive Evaluation of End-Point Free Energy Techniques in Carboxylated-Pillar[6]arene Host-guest Binding: I. Standard Procedure

Xiao Liu[1*], Lei Zheng[2], Chu Qin[1], John Z.H. Zhang[2,3,4,5], Zhaoxi Sun[6*]

[1]*School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China*

[2]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

[3]*School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, 200062, China*

[4]*Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China*

[5]*Department of Chemistry, New York University, NY, NY 10003, USA*

[6]*College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China*

*To whom correspondence should be addressed: liuxiaode2013@163.com, proszx@163.com

Abstract

Despite the massive application of end-point free energy methods in protein-ligand and protein-protein interactions, computational understandings about their performance in relatively simple and prototypical host-guest systems are limited. In this work, we present a comprehensive benchmark calculation with standard end-point free energy techniques in a recent host-guest dataset containing 13 host-guest pairs involving the carboxylated-pillar[6]arene host. We first assess the charge schemes for solutes by comparing the charge-produced electrostatics with many *ab initio* references, in order to obtain a preliminary albeit detailed view of the charge quality. Then, we focus on four modelling details of end-point free energy calculations, including the docking procedure for the generation of initial condition, the charge scheme for host and guest molecules, the water model used in explicit-solvent sampling, and the end-point methods for free energy estimation. The binding thermodynamics obtained with different modelling schemes are compared with experimental references, and some practical guidelines on maximizing the performance of end-point methods in practical host-guest systems are summarized. Finally, we compare our simulation outcome with predictions in the grand challenge and discuss further developments to improve the prediction quality of end-point free energy methods.

Keywords: Pillar[*n*]arenes, Host-guest Binding, End-point Free Energy Methods, Molecular Docking, Force Field

1. Introduction.

Pillar[*n*]arenes as a family of cylindrical macrocycles with symmetrical rims have exhibited great potentials in various laboratory and industrial applications, such as drug delivery, molecular machinery, catalyzed synthesis and fluorescence sensing.[1-8] Despite the guest-binding ability of pillar[*n*]arenes, the native form suffers from low solubility in aqueous environment and is unsuitable for biomedical applications, and chemical modifications are widely considered an effective solution to this problem.[9-12] Among modified pillar[*n*]arenes derivatives, carboxylated-pillar[*n*]arenes with *n*=6 or 7 are promising candidates in drug delivery due to their enhanced water-solubility, strong drug-binding ability, easy functionalization, and intermediate levels of cavity volume and entrance size.[13-16] An example of practical applications of functionalized pillar[*n*]arenes is the dramatic increase of the solubility and the significant pKa shift of sanguinarine upon the carboxylated-pillar[6]arene (WP6) encapsulation.[17]

Molecular simulations of protein-ligand and more generally host-guest systems accumulate time series data of atomic positions and energetics. These time-resolved information is then used to approximate ensemble averages under the ergodicity assumption.[18-22] However, due to the significant gap between the simulation-accessible time scale and that of the process of interest (e.g., host-guest binding/unbinding), it is generally impossible to observe multiple transitions between states of interest during the course of brute-force simulations.[23-27] Therefore, in simulation investigations of the binding/unbinding processes, designed modelling schemes are required. If direct observations of binding/unbinding events are pursued, free energy simulations in the physical configurational space or along the artificial alchemical pathway could be performed.[28-31] Although these enhanced sampling techniques sample the gradual change of the whole system during the binding/unbinding events and are considered rigorous free energy techniques, their high computational costs hinder large-scale applications in complex condensed matter systems.[32, 33] Relatively cheap but less accurate alternatives are end-point free energy methods including MM/PBSA and MM/GBSA,[34-37] the single-trajectory form of which only samples the well-defined bound state and thus is easier to converge. Currently, end-point free energy techniques are widely employed in drug discovery as a highly efficient tool for screening a large set of potential hits.[38-42] A computational technique cruder than end-point methods is molecular docking, which uses simpler energy functions (scoring function) and sampling techniques to ensure fast screening of an even larger set of molecules.[43-46] In the hierarchical screening with computational techniques, the docking procedure is often applied before end-point free energy calculation, and the latter is considered as a re-scoring procedure processing the docking outcome.[47] Aside from the sampling techniques, the model used in computation or Hamiltonian is also critical and

determines the ultimate accuracy level that a comprehensive simulation can achieve.[48-51] In complex systems such as protein-ligand and also host-guest complexes, it is often desirable to employ fixed-charge force fields due to their balanced efficiency and accuracy. A huge number of benchmark free energy calculations have revealed that different force-field parameters could lead to different conformational preferences, binding thermodynamics and also interaction networks.[52, 53]

Despite the importance of macrocyclic hosts and their interactions with drug-like guest molecules, the number of scientific reports on practical experiences of end-point free energy techniques in novel host-guest complexes is rather limited, not to mention bias-free evaluations covering a variety of critical modelling procedures. In this work, using an interesting dataset containing 13 WP6 host-guest systems provided in the recent SAMPL9 challenge,[54] we investigate four key modelling details of end-point free energy calculations extensively, aiming at providing a thorough evaluation of end-point methods in similar host-guest systems. The benchmarked four details include the method for the generation of initial condition, the charge scheme for host and guest molecules (i.e., solutes), the model for water molecules (solvents), and the parameters used in end-point free energy analysis. The accumulated comprehensive numerical experiences are summarized and interpreted to provide some general guidelines of maximizing the performance of standard end-point free energy calculations in host-guest systems similar to the current WP6 cases. Finally, we discuss about further directions (altered end-point schemes) to pursue improved performance.

2. Computational Setup.

The 3D chemical structures of all molecules under investigation are obtained from the GitHub site of the SAMPL9 challenge.[54] The determination of the protonation states of host and guest molecules follows a recent computational investigation.[55] The resulting net charges of all molecules under consideration can be interpreted from the chemical structures shown in Fig. 1. The aim of the current work is to provide a detailed and extensive assessment of end-point free energy techniques for WP6 host-guest binding. Thus, a set of parameter/force-field combinations are modelled in our simulations.

Charge Schemes.

First, for atomic charges, we consider two popular fixed-charge models. The first one is AM1-BCC,[56] which optimizes the molecule at the AM1 level and combines the Mulliken charges with a pre-fitted correction term named bond-charge correction. The aim of this charge model is to substitute computationally demanding *ab initio* calculations with computationally cheaper corrected semi-empirical charges, which is achieved mainly by training the bond-charge correction with the HF[57-59]/6-31G* electrostatic potential (ESP)

for a set of small molecules. The second charge model is the restrained electrostatic potential (RESP)[60] scheme. The structure of each molecule is first optimized at B3LYP[61-63]/6-31G*, which produces a high-quality low-energy configuration. This representative geometry is then used to generate ESP data with the Merz-Kollmann scheme. We consider two *ab initio* levels including the gas-phase HF/6-31G* and the implicit-solvent (IEFPCM for water) B3LYP/def2-TZVPP to scan the molecular ESP. The former *ab initio* level, the gas-phase HF/6-31G*, follows the tradition of AMBER-like force fields, while the latter B3LYP/def2-TZVPP with IEFPCM solvation serves as a representative and popular method employing a higher-level electronic structure treatment and taking solvent effects into consideration. The obtained ESP data are then used to fit atom-centered charges with a two-step regularization procedure, producing two sets of RESP charges named RESP-1 (HF/6-31G*) and RESP-2 (B3LYP/def2-TZVPP IEFPCM), respectively. All the other missing parameters are obtained from the transferable GAFF2 parameter set.[64] An exception for this GAFF2 extraction is the G4 guest due to the absence of Si parameters in the pre-fitted GAFF2. For this molecule, we use bonded and vdW parameters reported in a recent publication, where the parametrization of Si-involved species is the similar to the GAFF2 set.[65] It is worth noting that the Si parametrization in the reference is using AM1-BCC charges instead of the more rigorous HF/6-31G* RESP charges, but according to our detailed analysis of electrostatic data shown later, this corrected semi-empirical charge scheme works reasonably well for this molecule. Thus, the parametrization reported in the reference is generally usable and compatible with the charge schemes used in our modelling.

Water Models.

Second, although the force-field parameters for host and guest molecules would have a significant impact on the simulation outcome, the parameters of the other components involved in the simulation box, e.g., water molecules, would also influence the modelling results.[66-68] Therefore, in our extensive end-point free energy calculations, we consider two water models including the widely applied TIP3P[69, 70] and SPC/E models.[71] Although both water models have three interaction sites, all parameters (e.g., force constants and HOH angle) of them are different. As a result, various bulk properties of the pure solvent (water) obtained with the two water models differ, triggering differences in the solvation environments surrounding the host-guest solutes. This, ultimately, would alter the conformational/energetic preference of the complex structure and the resulting binding thermodynamics, leading to the solvent-model-dependent simulation outcome.

Initial Bound Conformation from Molecular Docking.

As only the structures of host and guest are given in the online server, a question that needs to be solved in model construction is securing the initial bound conformation. In our modelling, the initial guess of the

bound structure is obtained from docking calculations with the Autodock Vina program and the Autodock4 (AD4) scoring function.[72, 73] The top-1 (i.e., most stable) structure is extracted as the starting configuration for each host-guest complex. The simulation box is then constructed by parametrizing and solvating the host-guest complex with different parameter sets (e.g., solute charges and water) and adding non-polarizable monovalent spherical counter ions[74, 75] of Na^+ or Cl^- for neutralization. Periodic boundary conditions are always employed in our simulations. As the simulation is performed in an unbiased way in end-point free energy calculation, the whole sampling trajectory is often fluctuating in the neighborhood of the initial configuration. Thus, it is widely believed that the initial condition could have a significant impact on the simulation outcome. To investigate the impact of docking details on the simulation results, various parameters can be changed. Among these docking details, the scoring function itself is often considered the core of the problem. Therefore, this factor is investigated in our benchmark test of the docking details. Specifically, we repeat the docking calculation with another scoring function named Vina,[72] which differs from the AD4 scoring function in both the definition of scoring terms and computational complexity. These two scoring functions show different behaviors in many benchmark calculations,[72, 76, 77] which makes the current scoring-function variation indeed a meaningful test for the bound-conformation generation. Similar to the AD4 case, the box construction is repeated with the Vina-given top-1 bound structure.

End-point Free Energy Calculation.

With the constructed simulation box, we then turn to the molecular dynamics section. The SHAKE constraints on bonds involving hydrogen atoms are turned on.[78, 79] Langevin dynamics[80] with the collision frequency of 2 ps^{-1} are implemented for temperature regulation and the time step for integrating the equations of motion is set to 2 fs. The real-space cutoff for non-bonded interactions is set to 10 Å, and the PME method is used to treat long-range electrostatics.[81] Unbiased sampling is performed with the GPU version of the pmemd engine with the hybrid precision (SPFP) in the AMBER[82] suite.

To avoid perturbation of the docking-produced binding pose, for each host-guest complex, we first energy-minimize the system and perform 300 ps NVT heating with weak harmonic restraints on solutes. After this initial equilibration step, we turn to NPT equilibration for 1 ns to reach the 1 atm pressure. Then, we perform a long production run lasting 100 ns for data accumulation. The sampling interval of host-guest configurations is set to 10 ps, which is typical in end-point free energy analysis. After the sampling part, we process the trajectories with different end-point free energy methods to estimate binding affinities.

The free energy of binding in end-point free energy techniques can be generally expressed as the following equation,

$$\Delta G_{binding} = \Delta E_{\text{elec}} + \Delta E_{vdW} + \Delta G_{solv} - T\Delta S_{gas} \quad (1).$$

Here, the first two components are gas-phase electrostatic and vdW contributions to the binding enthalpy, the third term is the free energy contribution from solvation effects, and the last term is the gas-phase entropic contribution. The first two terms can be calculated directly with all-atom force fields, the third solvation term can be further divided into the polar and non-polar solvation parts that can be estimated with PBSA or GBSA implicit solvent models,[83-86] and the last term can be calculated with various methods such as normal mode analysis (NMA),[87] quasi-harmonic approximation[88] and so on. In our calculations, the GBSA solvation is performed with the GB^{OBC} model,[89, 90] which is quite popular in modern end-point free energy calculations. The normal mode approximation has been widely used in various cases such as static quantum mechanics calculations and protein-protein binding.[91-99] Due to the popularity of NMA in end-point free energy analysis of protein-ligand and host-guest systems, in the current benchmark calculation, we select this 'standard' method to estimate the entropic contribution to the binding free energy. As the computational cost of NMA calculations (specifically the calculation of the Hessian matrix) is quite high in complex systems and the entropic changes from NMA are similar for different configurations, often only several or tens of snapshots are used to compute this term.[100-102] Thus, in the current WP6 host-guest calculations, 50 frames equally spaced in each simulation trajectory (i.e., 2 ns per sample) are used for NMA calculations.

3. Results and Discussions.

In the above paragraphs of the previous section, four key influencing factors to be benchmarked in the current work have been discussed. The initial condition can be obtained with the AD4 or Vina scoring function; the charge scheme for solute molecules (i.e., host and guest) could be AM1-BCC or RESP (RESP-1 or RESP-2); the water model used for solvation could be TIP3P and SPC/E; and many details of the end-point free energy method could be varied. In the following discussion of the simulation outcome, we use the details of these four factors to indicate the modelling setup, e.g., Vina+AM1-BCC+TIP3P+MM/PBSA for the combination of modelling scheme using Vina-docked pose, AM1-BCC charges for solutes, TIP3P water, and MM/PBSA for free energy estimation. Note that the experimental binding affinities of the considered WP6 host-guest complexes are obtained from reference.[103]

Charge Quality.

Before performing any atomistic simulation, we first check probably the most influential factor in

molecular modelling, the charge scheme for solute molecules. Direct comparison of atomic charges produced by different charge schemes provides a pre-simulation evaluation of the charge quality, which could provide hints on possible behaviors of the binding thermodynamics produced by these charge schemes and also possible explanations of these observations. In charge fitting with ESP-related schemes, the loss function often includes ESP deviations and some user-defined restraints (regularization terms). In the current RESP fitting, ESP deviations and hyperbolic regularizations are included, and the quality of the regularized least-square fitting is often assessed by ESP relative root-mean-squared error (RRMSE), which is the percentage deviation of the charge-produced ESP from the reference *ab initio* data. Although we have two RESP charge sets (RESP-1 and RESP-2), in this preliminary evaluation of electrostatics we only consider the first one RESP-1 obtained from traditional AMBER-like RESP fitting, as it shares a goal similar to AM1-BCC (i.e., HF/6-31G* ESP) and a numerical procedure similar to RESP-2. The numerical results of ESP RRMSEs for all molecules are presented in Fig. 2a. It is clearly shown that the RESP charge scheme provides a much better fitting of the Coulombic ESP around each molecule than AM1-BCC, but it should be fair to point out that the ESP RRMSE of AM1-BCC is still acceptable, which indicates that both charge schemes do not have significant problems in reproducing the high-level electrostatics for WP6 host-guest systems. An observable that is not included in the loss function of RESP fitting, the dipole moment, is also checked in this preliminary evaluation of charge quality in Fig. 2b. We can see that the absolute deviation of the dipole moment is negligible for RESP-1 charges, but is a bit larger for AM1-BCC. This observation also suggests the superiority of the RESP charge scheme, but we should still point out that the AM1-BCC charge scheme is still not very problematic. A worth noting observation is the wide range of dipole moments of the guest molecules, which suggests the diversity of the investigated dataset. Thus, the computational results presented in the following parts of the manuscript, i.e., the comprehensive evaluation of binding thermodynamics with end-point free energy techniques, would provide useful hints on the practical performance of end-point free energy calculations in the coordination of pillar[*n*]arene derivatives and drug-like guests.

The fitting target, the ESP around each molecule, is influenced by the level of theory.[104] Thus, another interesting and valuable test performed in charge assessment is on the *ab initio* level. Although it is straightforward to change the level of theory used in RESP fitting, as AM1-BCC includes a pre-fitted term trained with HF/6-31G* data, its target level cannot be easily varied unless refitting the correction term. Thus, instead of directly varying the fitting target, we simply change the level of theory to generate the ESP data for evaluation. The deviations of the charge-produced ESP from different *ab initio* results are computed

to assess the quality of atomic charges. In this numerical test, aside from the AM1-BCC and RESP-1 target HF/6-31G*, we select four higher-level methods including B3LYP/def2-TZVPP, M06-2X[105]/def2-TZVPP, MN15[106]/def2-TZVPP and wB97X-D[107]/def2-TZVPP. These density functionals are more accurate than HF, and the basis sets are upgraded from the crude Pople series 6-31G* to the polarized triple-zeta def2-TZVPP. Aside from gas-phase calculations, we also include solvent effects with the IEFPCM implicit solvent model. The percentage errors of molecular ESP are shown in Fig. 3. The AM1-BCC ESP RRMSE using the gas-phase *ab initio* references is presented in Fig. 3a, where the error shows obvious system-dependence. In most cases, the deviation from HF/6-31G* is marginally smaller than those from the other higher-level results, which possibly arises from the fact that this level of theory (HF/6-31G*) is employed in the training of this semi-empirical charge model. The ESP percentage errors of AM1-BCC from the other *ab initio* levels are similar, which possibly suggests the similarity of ESP data generated with these levels of theory. The RESP-1 deviations from the gas-phase *ab initio* results are shown in Fig. 3b, where a similar behavior is observed. Namely, the magnitude of the deviation of charge-produced ESP from *ab initio* results shows system-dependence, the deviation from the HF/6-31G* reference is slightly smaller than the other, and the ESP RRMSEs computed with higher-level methods are similar. By comparing the AM1-BCC and RESP results in Fig. 3a and 3b, interestingly, the ESP deviations for RESP-1 charges are obviously smaller than AM1-BCC, regardless of the level of theory used to generate the reference ESP data. This phenomenon, along with the similar ESP deviations from different *ab initio* references, suggests the similarity of ESP data generated at different *ab initio* levels. As a result, the lowest level tested here and also the widely employed reference in AMBER-like force fields, HF/6-31G*, serves as a good yet efficient option to generate the reference ESP data. Consideration of higher-level techniques could be helpful, but the merits are of limited magnitudes and exhibit system-dependence. When incorporating solvent effects into the calculation, we reach the results presented in Fig. 3c-d. By comparing the ESP RRMSEs under the AM1-BCC charge scheme computed with gas-phase and IEFPCM *ab initio* references (i.e., Fig. 3a and 3c), interestingly, the ESP RRMSEs for many molecules (e.g., WP6, G1 and G8) decrease upon the inclusion of implicit solvent. However, the ESP RRMSE for many other molecules such as G3 increases. Thus, it is difficult to say whether the AM1-BCC charge scheme reproduces the solvated situation or the traditional gas-phase scan more accurately. The comparison between RESP-1-produced errors with gas-phase and implicit-solvent references gives a somehow similar behavior (c.f., Fig. 3b and 3d), but this time the ESP RRMSE increases for most molecules such as WP6 and G1 and only a small portion of molecules (e.g., G12) exhibit a decreasing behavior. This phenomenon is rather not unexpected, as the gas-phase ESP data are actually the

fitting target of RESP-1 charges. When we compare the ESP RRMSEs under the two charge schemes computed with IEFPCM references (i.e., Fig. 3c and 3d), interestingly, the RESP-1 errors are larger than AM1-BCC in many cases, which is also caused by the overfitting of the gas-phase ESP in the RESP-1 charge generation. Overall, with all levels of theory to generate the evaluation data, the AM1-BCC and RESP-1 ESP RRMSEs are not unreasonably huge for all molecules. When the gas-phase ESP data are targeted, RESP-1 outperforms AM1-BCC. When the aim is to reproduce the implicit-solvent ESP, RESP-1 is slightly worse than AM1-BCC. Although the ESP RRMSE of the RESP-2 charge set is not computed, according to the above observations about the RESP-1 charge set, we expect the RESP-2 charge set to reproduce the implicit-solvent ESP in a better way but exhibit slightly larger deviations from the gas-phase data. Due to the difference between fitting targets of the RESP-1 and RESP-2 charge sets, we consider both of them as realizations of the RESP charge scheme in molecular simulations of end-point free energy calculations. As the ESP errors of the two charge schemes (i.e., AM1-BCC and RESP) are within the reasonable range and are similar in magnitude, both of them ensure accurate calculation of electrostatic interactions with the Coulomb equation. Thus, their performances are similar.

The similar ESP reproductions of the two charge schemes are quite surprising in host-guest systems. In our previous investigation of other challenging host-guest coordinations involving the Cucurbit[8]uril (CB8) host, the ESP reproductions of these two charge schemes differ significantly.[108] Specifically, for the CB8 host, the RESP charge scheme produces an ESP RRMSE ~5%, which is similar to the current WP6 situation. By contrast, the AM1-BCC charge scheme fails to reproduce the *ab initio* ESP with a very huge ESP RRMSE ~33%. A similar observation is also obtained in β-cyclodextrin host-guest systems.[109] The major difference in ESP arises from the dissimilarity of the training set of AM1-BCC and the heterocyclic CB8 host.[108] The application of AM1-BCC to CB8 host-guest binding thus seems unsuitable due to its failure to reproduce the *ab initio* electrostatics. The binding affinities obtained through extensive sampling of binding/unbinding events under AM1-BCC deviate significantly from experimental values, which is in agreement with the detailed evaluation of electrostatic properties (e.g., ESP).[108] However, in the current WP6 situation, the AM1-BCC charge scheme performs quite well in reproducing the *ab initio* ESP, which suggests that the AM1-BCC training set already covers the features of WP6. Namely, the macrocyclic host WP6 shares similar structural and chemical features with many drug-like molecules in the AM1-BCC training set. Another possible consequence of the similarity of WP6 and the AM1-BCC training set composed of drug-like molecules is the perfect suitability of GAFF derivatives, as these general-purpose force fields are also fitted with a large set of drug-like molecules. Although we do not perform calculations

to evaluate the suitability of GAFF derivatives in the current WP6 host-guest systems like our previous works,[109-112] this speculation about the GAFF(2) suitability is highly probable. All these phenomena possibly suggest that the current SAMPL9 WP6 host-guest systems are 'easier' to model compared with the previous CB8 cases, as the broad range of highly efficient and accurate computational tools developed for drug-like molecules could be employed. However, it should be noted that these force-field similarities do not guarantee the satisfactory performance of end-point free energy methods, as various other factors including the intrinsic limitations of end-point methods (e.g., the suitability of normal mode approximation, the neglect of conformational change upon host-guest binding, and the accuracy of the implicit solvent model) and sampling problems still play a role.

Sampling Length.

We then turn to the simulation part and check the free energy estimates from end-point techniques. Before checking the consistency between computational results and experimental references, we first validate the convergence of the free energy calculation by monitoring the time-dependence of the free energy estimates in Fig. S1 and S2. The whole 100 ns sampling length in each simulation is divided into 10 groups, each of which lasts 10 ns. Starting from the first 0-10 ns point, we gradually add new trajectories to the existing one, leading to the 0-20 ns, 0-30 ns, and finally 0-100 ns results. When convergence is reached, the end-point estimates should not change with further sampling. Although the convergence can be reached within tens of ns (e.g., 40 ns) for many host-guest pairs under many simulation protocols, there are still cases where convergence cannot be reached even at the end of the long 100 ns unbiased sampling. This non-satisfactory convergence behavior is rather unexpected, as the sampling length of 100 ns employed in our simulation is already much longer than the commonly used ones such as 10 ns. The reasons causing this convergence problem could include the quality of the initial condition, the accuracy of Hamiltonian used in configurational sampling, and the compatibility of these two components (Hamiltonian and end-point free energy method). Specifically, if the docking procedure produces a good initial guess and the fixed-charge force field is good enough to stabilize this bound structure, the simulation employing the fixed-charge force field would achieve a good level of convergence. If the initial guess produced by docking is non-satisfactory, then the simulation would have trouble stabilizing the bound structure and the system keeps changing during the course of the simulation, leading to a not-so-good convergence behavior of the end-point calculation. In the case that the force field is not good enough but the initial guess is satisfactorily good, the simulation still could have trouble stabilizing the initial guess and thus lead to non-convergence of the end-point calculation.

The above sampling-length comparison suggests that converged end-point free energy estimates generally require tens of ns or longer sampling time, and the 100 ns sampling is already very long in modern end-point calculations. Thus, in the following parts of the manuscript, we use the free energy results obtained with this sampling length in further investigation of other influencing factors. The 100 ns end-point free energy estimates obtained with all combinations of modelling parameters (i.e., charge schemes, water models, scoring functions, and end-point calculation schemes) are summarized in Table S1-S4. Aside from RMSE mentioned previously, we also use the mean signed error (MSE), Kendall τ rank coefficient[113] and the Pearlman's predictive index (PI)[114] to evaluate the quality of computation for this WP6 host-guest dataset.

Changing the Parameter Set.

The parameter set or Hamiltonian used in explicit-water sampling is critical to the simulation outcome. As mentioned in the computational setup section, two charge schemes are employed for host and guest molecules (i.e., solutes), while for water we also test two models. In the following paragraphs, we would investigate the impact of the parameter set on the end-point free energy estimates.

The force-field/parameter set involves two components, the charge scheme for solutes and the water models. In Fig. 4, the quality metrics are grouped according to the water model and charge scheme. When comparing the error metrics of the three charge sets including AM1-BCC, RESP-1 and RESP-2, in most cases the AM1-BCC performs worst, the RESP-2 charge set exhibits an intermediate accuracy, and the RESP-1 set performs best (smallest RMSE), as shown in Fig. 4a. When checking the ranking coefficient in Fig. 4b, similar observations about charge sets could be obtained. Namely, the AM1-BCC charge set still performs worst, the RESP-2 set achieves an intermediate accuracy, and RESP-1 performs best in most cases. The consistent good performance of RESP-1 (traditional HF/6-31G* RESP fitting) in the reproduction of absolute binding affinities (RMSE) and the experimental rank (τ) suggests the superiority of this charge scheme in end-point free energy calculations. The AM1-BCC and RESP-2 charge sets are relatively less accurate. However, it should be fair to note that the apparent accuracies of these charge schemes rely heavily on error cancellation. Namely, other factors such as implicit solvent and the approximated calculation procedure of end-point methods also play a role here.

Then we check the solvent-model dependence of end-point results. When using the TIP3P water model, the RMSEs on the left part of Fig. 4a are somehow larger than the SPC/E ones on the right, regardless of the charge scheme and the docking procedure (scoring function). This observation suggests the superiority of the SPC/E model in reproducing the absolute values of binding affinities in end-point free energy

calculations. When comparing the ranking coefficients in Fig. 4b, we observe somehow different behaviors. When the solutes are described with the AM1-BCC charge model, TIP3P outperforms SPC/E. By contrast, when using either of the two RESP charge sets, SPC/E prevails. As the RESP charge scheme already shows higher accuracy in reproducing the absolute binding affinities (Fig. 4a) and achieves better ESP reproduction (c.f., Fig. 2 and 3), this charge scheme is considered more accurate and is recommended in the charge-scheme comparison presented in the previous paragraph. Thus, using this recommended charge scheme, SPC/E achieves better performance in end-point free energy calculations.

Overall, the above force-field comparison suggests that in end-point free energy calculations, the RESP-1 charge scheme performs best among charge schemes for solutes and the SPC/E water model achieves higher accuracy than the widely applied TIP3P.

Initial Condition and End-point Methods.

The initial condition can be pivotal in free energy calculations with the unbiased sampling technique, especially for systems with elevated (free-)energy barriers between relevant basins. If the initial bound structure is wrongly assigned, ns-length end-point free energy calculation could never really explore the true or optimal binding pose, introducing systematic bias of unknown magnitude. Thus, a detailed evaluation of this docking procedure for the generation of the initial condition seems indispensable in a thorough benchmark in the current WP6 host-guest binding.

As discussed in the computational setup section, the most stable (top-1) binding pose obtained with either of the two scoring functions including AD4 and Vina is selected as the starting bound configuration. Two end-point calculation methods including MM/PBSA and MM/GBSA are used to extract the free energy of binding. In Fig. 5, the quality metrics are grouped according to these two criteria. We first compare the generation procedure of the initial condition. As shown in Fig. 5a, the RMSEs of AD4- and Vina-produced initial conditions are somehow similar, and we cannot conclude which scoring function would lead to better reproduction of absolute binding affinities. When checking the ranking coefficient in Fig. 5b, still the relative performance of the two scoring functions shows combination-dependence, i.e., depending on the other modelling details such as the charge scheme.

We then turn to the comparison of the two end-point methods. For RMSE in Fig. 5a, the MM/PBSA results are larger than MM/GBSA in most cases, which suggests that the MM/GBSA method achieves a better performance than MM/PBSA in reproducing the absolute binding affinities. When checking the ranking coefficient in Fig. 5b, the qualities of predictions of the two end-point methods show parameter-

dependence, i.e., depending on other factors such as charge scheme and water model, which makes it difficult to conclude which method prevails.

As the combination of the RESP-1 charge set and the SPC/E water model achieves the best performance in the previous force-field comparison, we fix the force-field parameters to this recommended parameter set and re-visit the data in Fig. 5. Interestingly, when using this RESP-1+SPC/E parameter set, the free energy calculation achieves the highest performance among different force-field combinations for both the error and ranking metrics, regardless of the initial condition (scoring function) and end-point methods. Under this RESP-1+SPC/E parameter set, the AD4 scoring function achieves a better performance than Vina for both error and ranking metrics, which suggests the superiority of this scoring function. As for the end-point methods, the MM/GBSA method outperforms MM/PBSA for both error and ranking metrics. Therefore, when using the recommended RESP-1+SPC/E force-field combination, the AD4 scoring function and the MM/GBSA end-point method perform best. Overall, the above comparison of scoring functions used in molecular docking, charge schemes for solutes, water models, and end-point methods leads to the top-performing AD4+RESP-1+SPC/E+MM/GBSA combination in end-point free energy calculations in the current WP6 host-guest binding.

Despite the top performance of the AD4+RESP-1+SPC/E+MM/GBSA end-point free energy calculation scheme, the deviations of the computed values from the experimental reference are still very huge. The RMSE is as large as 12.2 kcal/mol, which is far from acceptable in free energy calculation. The MSE of this set is 9.4 kcal/mol, which suggests a significant level of systematic overestimation (more negative) of binding affinities. The Kendall τ and PI are only marginally larger than zero, which suggests the failure of the reproduction of the experimental rank. This is already the top-level accuracy of standard end-point free energy calculations, but the prediction quality is still non-satisfactory. To make the end-point free energy calculation really usable in host-guest binding, it is necessary to alter various details and pursue accuracy improvements.

Comparison with Other Methods.

In the SAMPL9 challenge, there are five ranked submissions for the current WP6 host-guest dataset. We access the results from the online GitHub website of the grand challenge[54] and compute the quality metrics (specifically RMSE and τ) in Fig. 6. When using the RMSE error metrics as the ranking criterion, the machine learning (ML) technique achieves the smallest RMSE ~2 kcal/mol and performs best. The second top-performing method is the extended linear interaction energy (ELIE) with ~2.5 kcal/mol RMSE. This

ELIE technique is a trained end-point method following the idea of the original linear interaction energy, which requires system-specific parametrization of the weights of different energy terms in free energy estimation.[115] The other three computational results are all obtained with alchemical free energy calculations, although their detailed simulation and modelling protocols differ.[55] The average RMSE of these rigorous free energy techniques in the WP6 host-guest systems is approximately 3 kcal/mol, which is a bit larger than the end-point ELIE result. When considering the ranking coefficient τ as the criterion, interestingly, the performance of alchemical free energy calculations improves significantly, and the averaged τ of the three alchemical calculations is similar to the end-point one. By contrast, the ML technique performs poorly in reproducing the experimental rank of binding affinities. This contrasting behavior of the error and ranking metrics, on the one hand, suggests that the ML method can predict absolute values similar to the experimental reference without many outliers, but the relative binding free energies of different systems cannot be accurately assigned. On the other hand, free energy calculations based on atomistic simulation (ELIE and alchemical calculation) can properly differentiate the relative binding affinities of different host-guest complexes, but the absolute values cannot be accurately computed.

The trained MM/PBSA method, ELIE,[115] is very similar to the end-point calculations performed in the current work. The difference lies in the system-specific training step that gives the relative weights of different energy terms defined in Eq. (1). Considering this parameterization step, this ELIE method also shares some similarities with ML and thus is not fully physics-based and is not directly transferable to other systems, especially for targets without known binding affinities. It should be noted that there are some other end-point methods sharing a similar parametrization procedure, e.g., the PBSA_E method.[116] Compared with the results reported in the current work that end-point free energy methods are straightforwardly applied without any system-specific parameter tuning, the trained MM/PBSA method exhibits a much better performance in both the computation of absolute values (RMSE) and the ranking of binding stabilities (τ). This improvement can be solely attributed to the system-specific parameterization step, and indeed can be considered a way (aside from the four key details benchmarked in the current work) to improve the end-point free energy results. However, it is fair to note that the involvement of the training step could hinder its application to newly encountered systems.

Inspired by a recent work reporting a surprisingly good estimation of binding affinities for host-guest systems with docking scores,[117] we also extract the docking results under the two scoring functions and compare them with the experimental affinities. The experiment-computation correlogram is shown in Fig. 7, where the systematic under-estimation can be observed under both scoring functions. Under both the AD4

and Vina scoring functions, the WP6 host-guest binding affinities are generally smaller than the experimental values, with MSE of -1.2 kcal/mol for AD4 and -2.1 kcal/mol for Vina. The RMSE values of the two scoring functions are 2.1 kcal/mol and 2.6 kcal/mol, which are plotted along with the quality metrics of SAMPL9 submissions in Fig. 6a. Interestingly, the docking-produced free energy estimates agree well with experimental values and perform better than computationally intense alchemical simulations and are comparable with ML-based techniques, which agrees with previous observations in other host-guest systems.[117] As for the ranking coefficient τ presented in Fig. 6b, the docking results are significantly better than ML but a bit worse than other simulation-based methods. Notably, for both error and ranking metrics, molecular docking outperforms the end-point results reported in the current work, regardless of the scoring function. Considering the widely acknowledged fact that post-processing docking results with atomistic simulations (normally with end-point free energy calculations) could lead to improved affinity ranks of compounds in protein-ligand binding, this underperformance of the current end-point calculations in WP6 host-guest binding is somehow counterintuitive.[118-123] However, this is actually not unexpected, as this hierarchical workflow of computational screening has not been well-validated in host-guest modelling. In these prototypical host-guest binding cases, error cancellations become less successful and the problems of various modelling details become exposed. Our calculations reveal these problematic issues, degrade the popularity of the end-point re-scoring procedure, and deepen the understanding of the hierarchical screening in host-guest binding.

Directions to Improve the Performance of End-point Methods.

The comprehensive benchmark presented above follows the widely applied standard workflow of end-point free energy calculations. We first obtain the initial condition with molecular docking and then perform molecular simulations to relax the structure and grab an ensemble of configurations in the bound state. Four key modelling details including the scoring function used in docking (AD4 and Vina), the charge scheme for solutes (AM1-BCC, RESP-1 parameterized with HF/6-31G* ESP and RESP-2 parameterized with B3LYP/def2-TZVPP ESP in IEFPCM solvation), the water model for solvents (TIP3P and SPC/E), and the end-point calculation details (MM/PBSA and MM/GBSA) are tested extensively. End-point free energy calculations with all possible combinations are performed and compared with experimental values. However, none of the combined schemes works well in reproducing the absolute binding affinities. The RMSE of WP6 host-guest binding free energies is generally as large as 12 kcal/mol, suggesting the inappropriateness of naively applying end-point methods in host-guest systems. Despite the failure of the

end-point free energy calculations, there are indeed ways to alter the standard procedure and thus improve the performance. For instance, fitting the weights of different terms in the end-point calculation of Eq. (1) with existing datasets could improve the accuracy. Methods following this spirit include PBSA_E,[116] ELIE[115] and so on. Aside from numerical fitting, jumping out of the fixed-charge model could also be attempted. Another direction is altering the dielectric constant, which has been proven useful in many protein-ligand and protein-protein complexes.[101, 124] Alteration of the implicit solvent model and the estimation of the entropic contribution could also be considered.[125-127] A further critical factor influencing the outcome is the exhaustiveness of the conformational sampling. As the end-point calculation is using configurations sampled with the unbiased technique, there could be several host-guest coordination patterns unexplored and thus not included in the ensemble average in Eq. (1), which could introduce systematic bias of unknown magnitude. This factor could be of great importance especially for host-guest binding, as such multi-modal binding behavior has been observed in several important host-guest coordinations such as CB8 and β-cyclodextrin.[108, 109, 111, 128, 129] On this aspect, enhanced sampling techniques could be employed to remove the sampling bias.

4. Concluding Remarks.

Molecular modelling of complex donor-acceptor systems such as protein-ligand and protein-protein complexes is of great importance in modern computational biophysics. Due to the large size and the conformational flexibility of these macromolecular systems, fixed-charge force fields are often incorporated with molecular simulation to probe and scope the spatial and temporal motions of different parts of the systems. Host-guest complexes are considered prototypical models for biomacromolecular systems due to the small size of the host and the relatively simple interaction network formed between host-guest pairs. However, despite the assumed simplicity of host-guest coordination, recent computational studies have revealed significant problems in the accurate modelling of host-guest binding. The problems lie both in the accuracy of employed Hamiltonian and the coverage of conformational space through configurational sampling. End-point free energy techniques are considered a choice with balanced accuracy and efficiency in free energy calculation and thus are widely used in drug discovery. A huge number of scientific reports on applications of end-point free energy methods in practical protein-ligand, protein-protein and protein-DNA systems could be accessed easily. However, numerical experiences of end-point free energy calculations in host-guest coordination are relatively limited. As these prototypical host-guest systems are simple in structure and small in size, fortuitous error cancellations can be avoided and potential issues in force-field

parameters and sampling techniques can be relatively easy to be identified. Thus, in the current work, using a dataset containing pillar[*n*]arene-based host-guest systems in a recent grand challenge, we benchmark four modelling details in end-point free energy calculations extensively. The focused four details include the docking procedure to generate the initial condition, the charge scheme for solutes, the water model for solvation, and the end-point method for free energy estimation.

The preliminary analysis of charge quality suggests the suitability of the AM1-BCC charge scheme for WP6 host-guest systems. The AM1-BCC and two RESP charge sets (RESP-1 targeting HF/6-31G* and RESP-2 targeting B3LYP/def2-TZVPP IEFPCM) produce molecular ESP similar to *ab initio* references. However, this does not guarantee the satisfactory performance of end-point free energy calculations employing these charge schemes, as other factors also influence the simulation outcome. For instance, the intrinsic limitations of end-point free energy techniques, e.g., the suitability of the normal mode approximation in entropy estimation and the neglect of conformational change in configurational sampling, could degrade the quality of calculation even with good charge schemes such as RESP. Thus, to obtain a reliable assessment of the practical performance of end-point methods, extensive calculation of binding thermodynamics is necessary.

We first investigate the impact of sampling length on the end-point free energy estimates in order to assess the convergence issue. Tens of ns are generally sufficient for convergence in the current WP6 host-guest dataset. However, there are still cases where convergence cannot be reached within the length of 100 ns. The reasons could include the quality of the initial condition produced by docking, the accuracy of the fixed-charge force field, and the compatibility of these two components. For example, if the initial bound structure produced by docking is good and the force field used in simulations is accurate, the simulation would explore the neighborhood of the starting configuration and the convergence is expected to be good. Even with a good initial guess, if the force field is inaccurate, the initial bound configuration cannot be stable and the inter- and/or intra-molecular interaction patterns changes during the course of the simulation, which would lead to convergence problems in end-point calculations. If the initial guess is non-satisfactory, then the force field cannot stabilize the initial structure and the system is kept moving in the configurational space, leading to non-satisfactory convergence behavior of the end-point free energy calculation.

Then, we benchmark two charge schemes (specifically, three charge sets including AM1-BCC, RESP-1 and RESP-2) and two water models (TIP3P and SPC/E) for binding thermodynamics. The RESP-1 charge set performs best among charge models, and the SPC/E water model achieves the highest level of accuracy. Thus, in the end-point free energy calculations of WP6 host-guest binding, the RESP-1 charge set and the

SPC/E water model are recommended. As for the investigation of the docking procedure to generate the initial condition/configuration, we perform docking calculations with two scoring functions (AD4 and Vina) and select the top-1 (most stable) binding mode for later simulation. AD4 is found to produce smaller errors compared with Vina, which suggests that the former scoring function could be more compatible with the end-point calculation procedure. As for the relative performance of the two end-point methods, we observe that MM/GBSA outperforms MM/PBSA in the reproduction of the absolute binding affinity and the experimental rank. Overall, the extensive benchmark test of end-point free energy calculations in WP6 host-guest interactions recommends the AD4+RESP-1+SPC/E+MM/GBSA combination of modelling parameters. We further compare our simulations with submissions in the SAMPL9 challenge and discuss potential problems and further directions to improve the end-point technique. The docking scores are also used to estimate the binding affinities, and the docking-produced estimates are found to be more successful in reproducing both the absolute binding affinities and their rank.

Even with this best combined end-point simulation procedure, the deviations from experiment are still very large with RMSE ~12 kcal/mol and τ ~0.08, suggesting the unsuitability of standard end-point calculations in host-guest binding. The underperformance of standard end-point methods in host-guest systems is not unexpected, as various issues could be identified in the calculation scheme. For instance, the suitability of normal mode approximation in the calculation of the entropic contribution, the accuracy of the implicit solvent model, the sampling convergence, the conformational change upon host-guest binding and so on can all influence the performance of end-point free energy estimates. To make end-point calculations practically useful in host-guest modelling, it is necessary to apply some modifications. In a following work, we would present another comprehensive perspective of further developments and modifications of end-point methods to improve the end-point estimates in host-guest binding, e.g., regression and parameter adjustments.

Acknowledgement

This work was supported by Beijing Natural Science Foundation (Grant No. 7224357) and the National Natural Science Foundation of China (Grant No.22107063). We thank the anonymous reviewers for valuable comments and critical reading.

Conflict of Interest Statement

There are no conflicts of interest to declare.

Supporting Information Description

Time series of binding affinities estimated with all possible combinations of the four modelling details, and the 100 ns MM/PBSA and MM/GBSA binding free energies obtained with three charge schemes, two water models, initial bound structures from AD4 and Vina docking are given in the supporting information.

References

1. Liu, L.; Cao, D.; Jin, Y.; Tao, H.; Kou, Y.; Meier, H., Efficient synthesis of copillar [5] arenes and their host–guest properties with dibromoalkanes. *Organic & Biomolecular Chemistry* **2011**, 9, 7007-7010.
2. Zhang, C.-C.; Li, S.-H.; Zhang, C.-F.; Liu, Y., Size switchable supramolecular nanoparticle based on azobenzene derivative within anionic pillar [5] arene. *Scientific reports* **2016**, 6, 1-9.
3. Peerannawar, S. R.; Gejji, S. P., Theoretical investigations on vibrational spectra of pillar [5] arene-bis (pyridinium) complexes. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2013**, 104, 368-376.
4. Xia, B.; He, J.; Abliz, Z.; Yu, Y.; Huang, F., Synthesis of a pillar [5] arene dimer by co-oligomerization and its complexation with n-octyltrimethyl ammonium hexafluorophosphate. *Tetrahedron letters* **2011**, 52, 4433-4436.
5. Li, S.-H.; Zhang, H.-Y.; Xu, X.; Liu, Y., Mechanically selflocked chiral gemini-catenanes. *Nature communications* **2015**, 6, 1-7.
6. Yu, G.; Han, C.; Zhang, Z.; Chen, J.; Yan, X.; Zheng, B.; Liu, S.; Huang, F., Pillar [6] arene-based photoresponsive host–guest complexation. *Journal of the American Chemical Society* **2012**, 134, 8711-8717.
7. Qin, S.; Xiong, S.; Han, Y.; Hu, X. Y.; Wang, L., Controllable fabrication of various supramolecular nanostructures based on nonamphiphilic azobenzene derivatives and pillar [6] arene. *Chin. J. Chem .* **2015**, 33, 107-111.
8. Ogoshi, T.; Yamafuji, D.; Akutsu, T.; Naito, M.; Yamagishi, T.-a., Achiral guest-induced chiroptical changes of a planar-chiral pillar [5] arene containing one π-conjugated unit. *Chem. Commun.* **2013**, 49, 8782-8784.
9. Strutt, N. L.; Zhang, H.; Schneebeli, S. T.; Stoddart, J. F., Amino-Functionalized Pillar [5] arene. *Chem. Eur. J.* **2014**, 20, 10996-11004.
10. Ma, Y.; Yang, J.; Li, J.; Chi, X.; Xue, M., A cationic water-soluble pillar [6] arene: synthesis, host–guest properties, and self-assembly with amphiphilic guests in water. *RSC advances* **2013**, 3, 23953-23956.
11. Yang, K.; Chang, Y.; Wen, J.; Lu, Y.; Pei, Y.; Cao, S.; Wang, F.; Pei, Z., Supramolecular vesicles based on complex of trp-modified pillar [5] arene and galactose derivative for synergistic and targeted drug delivery. *Chem. Mater.* **2016**, 28, 1990-1993.
12. Strutt, N. L.; Schneebeli, S. T.; Stoddart, J. F., Stereochemical inversion in difunctionalised pillar [5] arenes. *Supramol. Chem.* **2013**, 25, 596-608.
13. Dasgupta, S.; Mukherjee, P. S., Carboxylatopillar [n] arenes: a versatile class of water soluble synthetic receptors. *Organic & Biomolecular Chemistry* **2017**, 15, 762-772.
14. Gu, A.; Wheate, N. J., Macrocycles as drug-enhancing excipients in pharmaceutical formulations. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2021**, 100, 55-69.
15. Wheate, N. J.; Dickson, K.-A.; Kim, R. R.; Nematollahi, A.; Macquart, R. B.; Kayser, V.; Yu, G.; Church, W. B.; Marsh, D. J., Host-guest complexes of carboxylated pillar [n] arenes with drugs. *Journal of Pharmaceutical Sciences* **2016**, 105, 3615-3625.
16. Li, Z.; Yang, J.; Yu, G.; He, J.; Abliz, Z.; Huang, F., Water-soluble pillar [7] arene: synthesis, pH-controlled complexation with paraquat, and application in constructing supramolecular vesicles. *Org. Lett.* **2014**, 16, 2066-2069.
17. Ping, G.; Wang, Y.; Shen, L.; Wang, Y.; Hu, X.; Chen, J.; Hu, B.; Cui, L.; Meng, Q.; Li, C., Highly efficient complexation of sanguinarine alkaloid by carboxylatopillar [6] arene: p K a shift, increased solubility and enhanced antibacterial activity. *Chem. Commun.* **2017**, 53, 7381-7384.
18. And, S. E. T.; Smithrud, D. B., Carboxylates Stacked over Aromatic Rings Promote Salt Bridge Formation in Water. *Journal of the American Chemical Society* **2002**, 124, 442.
19. 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, 315-20.
20. Rani, P.; Biswas, P., Diffusion of hydration water around intrinsically disordered proteins. *J. Phys. Chem. B* **2015**, 119, 13262-13270.
21. 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, 14622-14630.

22. Huai, Z.; Sun, Z., Titration of Adenine in a GA mismatch with Grand Canonical Simulations. *Journal of Computational Biophysics and Chemistry* **2020**, 20, 165-173.
23. Marchi, M.; Procacci, P., Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the NPT ensemble. *J. Chem. Phys.* **1998**, 109, 5194-5202.
24. Jurasz, J.; Czub, J.; Baginski, M.; Wieczor, M., Molecular mechanism of proton-coupled ligand translocation by the bacterial efflux pump EmrE. *bioRxiv* **2021**.
25. Sun, Z.; Zhang, J. Z. H., Thermodynamic Insights of Base Flipping in TNA Duplex: Force Fields, Salt Concentrations, and Free-Energy Simulation Methods. *CCS Chemistry* **2021**, 3, 1026–1039.
26. 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, 15005-15020.
27. Wang, X., Conformational Fluctuations in GTP-Bound K-Ras: A Metadynamics Perspective with Harmonic Linear Discriminant Analysis. *J. Chem. Inf. Model.* **2021**.
28. Bruckner, S.; Boresch, S., Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, 32, 1320–1333.
29. Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammont, J. A., Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.* **2013**, 9, 4131-4139.
30. Lickert, B.; Wolf, S.; Stock, G., Data-Driven Langevin Modeling of Nonequilibrium Processes. *J. Phys. Chem. B* **2021**, 125, 8125-8136.
31. Sun, Z.; Gong, Z.; Xia, F.; He, X., Ion Dynamics and Selectivity of Nav channels from Molecular Dynamics Simulation. *Chem. Phys.* **2021**, 111245.
32. Huai, Z.; Yang, H.; Sun, Z., Binding thermodynamics and interaction patterns of human purine nucleoside phosphorylase-inhibitor complexes from extensive free energy calculations. *J. Comput.-Aided Mol. Des.* **2021**.
33. Anandakrishnan, R.; Drozdetski, A.; Walker, R.; Onufriev, A., Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysical Journal* **2015**, 108, 1153-64.
34. Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem. B* **2013**, 117, 8408-8421.
35. Miller, B. R.; Mcgee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: an efficient program for end-state free energy calculations. J Chem Theory Comput 8:3314-3321. *J. Chem. Theory Comput.* **2012**, 8, 3314-3321.
36. 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, 7865-7877.
37. 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, 2082-9.
38. Su, P. C.; Tsai, C. C.; Mehboob, S.; Hevener, K. E.; Johnson, M. E., Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (F abl). *J. Comput. Chem.* **2015**, 36, 1859-1873.
39. Sangpheak, W.; Khuntawee, W.; Wolschann, P.; Pongsawasdi, P.; Rungrotmongkol, T., Enhanced stability of a naringenin/2, 6-dimethyl β-cyclodextrin inclusion complex: Molecular dynamics and free energy calculations based on MM-and QM-PBSA/GBSA. *Journal of Molecular Graphics and Modelling* **2014**, 50, 10-15.
40. Tsitsanou, K. E.; Hayes, J. M.; Keramioti, M.; Mamais, M.; Oikonomakos, N. G.; Kato, A.; Leonidas, D. D.; Zographos, S. E., Sourcing the affinity of flavonoids for the glycogen phosphorylase inhibitor site via crystallography, kinetics and QM/MM-PBSA binding studies: comparison of chrysin and flavopiridol. *Food and chemical toxicology* **2013**, 61, 14-27.
41. Peng, C.; Zhu, Z.; Shi, Y.; Wang, X.; Mu, K.; Yang, Y.; Zhang, X.; Xu, Z.; Zhu, W., Computational Insights into the Conformational Accessibility and Binding Strength of SARS-CoV-2 Spike Protein to Human Angiotensin-Converting Enzyme 2. *J. Phys. Chem. Lett.* **2020**, 11, 10482-10488.
42. Abula, A.; Xu, Z.; Zhu, Z.; Peng, C.; Chen, Z.; Zhu, W.; Aisa, H. A., Substitution Effect of the Trifluoromethyl Group on the Bioactivity in Medicinal Chemistry: Statistical Analysis and Energy Calculations. *J. Chem. Inf. Model.* **2020**, 60, 6242-

6250.
43. Van Zundert, G.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P.; Karaca, E.; Melquiond, A.; van Dijk, M.; De Vries, S.; Bonvin, A., The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology* **2016**, 428, 720-725.
44. 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, 395-419.
45. Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C.; Waldman, M., LigScore: a novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling* **2005**, 23, 395-407.
46. Pecina, A.; Eyrilmez, S. M.; Köprülüoğlu, C.; Miriyala, V. M.; Lepšík, M.; Fanfrlík, J.; Řezáč, J.; Hobza, P., SQM/COSMO Scoring Function: Reliable Quantum-Mechanical Tool for Sampling and Ranking in Structure-Based Drug Design. *ChemPlusChem* **2020**, 85, 2362-2371.
47. Giordano, D.; Biancaniello, C.; Argenio, M. A.; Facchiano, A., Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals* **2022**, 15, 646.
48. Sun, Z.; He, Q., Seeding the Multi-dimensional Nonequilibrium Pulling for Hamiltonian Variation: Indirect Nonequilibrium Free Energy Simulations at QM levels. *Phys. Chem. Chem. Phys.* **2022**, 24, 8800-8819.
49. Sun, Z.; Liu, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of QM/MM Free Energy Landscapes: Varying the QM Region. *Adv. Theory Simul.* **2021**, 4, 2100185.
50. Haskopoulos, A.; Maroulis, G., Carbon dioxide interacting with rare gases: Insights from high-level ab initio calculations of polarizability and hyperpolarizability effects. *Chem. Phys.* **2016**, 475, 90-103.
51. Kang, D.; Sun, W.; Shi, H.; Lu, C.; Kuang, X.; Chen, B.; Xia, X.; Maroulis, G., probing the structure and electronic properties of beryllium doped boron clusters: A planar $BeB16^-$ cluster motif for metallo-borophene. *Scientific reports* **2019**, 9, 1-9.
52. Sun, Z.; Wang, X.; Song, J., Extensive Assessment of Various Computational Methods for Aspartate's pKa Shift. *J. Chem. Inf. Model.* **2017**, 57, 1621-1639.
53. 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**, 34, 589-600.
54. https://github.com/samplchallenges/SAMPL9.
55. Procacci, P.; Guarnieri, G., SAMPL9 blind predictions using nonequilibrium alchemical approaches. *J. Chem. Phys.* **2022**, 156, 164104.
56. 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, 1623-41.
57. Mcweeny, R.; Diercksen, G., Self-Consistent Perturbation Theory. II. Extension to Open Shells. *J. Chem. Phys.* **1968**, 49, 4852-4856.
58. Pople, J. A.; Nesbet, R. K., Self-Consistent Orbitals for Radicals. *J. Chem. Phys.* **1954**, 22, 571-572.
59. Roothaan, C. C. J., New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, 23, 69-89.
60. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1992**, 97, 10269-10280.
61. Hertwig, R. H.; Koch, W., On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chem. Phys. Lett.* **1997**, 268, 345-351.
62. Becke, A. D., Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.* **1996**, 104, 1040-1046.
63. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, 98, 11623-11627.
64. 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.
65. Dong, X.; Yuan, X.; Song, Z.; Wang, Q., The development of an Amber-compatible organosilane force field for drug-

like small molecules. *Phys. Chem. Chem. Phys.* **2021**, 23, 12582-12591.
66. Onufriev, A. V.; Izadi, S., Water models for biomolecular simulations. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2018**, 8, e1347.
67. Abascal, J. L.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, 123, 234505.
68. Vassetti, D.; Pagliai, M.; Procacci, P., Assessment of GAFF2 and OPLS-AA general force fields in combination with the water models TIP3P, SPCE, and OPC3 for the solvation free energy of druglike organic molecules. *J. Chem. Theory Comput.* **2019**, 15, 1983-1995.
69. 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, 926-935.
70. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, 121, 10096-10103.
71. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, 91, 6269-6271.
72. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S., AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, 61, 3891-3898.
73. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, 30, 2785-2791.
74. 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, 9020-9041.
75. 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, 13279-13290.
76. Nguyen, N. T.; Nguyen, T. H.; Pham, T. N. H.; Huy, N. T.; Bay, M. V.; Pham, M. Q.; Nam, P. C.; Vu, V. V.; Ngo, S. T., Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf. Model.* **2020**, 60, 204-211.
77. Gaillard, T., Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark. *J. Chem. Inf. Model.* **2018**, 58, 1697-1706.
78. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of The Cartesian Equations of Motion of A System with Constraints: Molecular Dynamics of n -alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.
79. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of The SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, 13, 952-962.
80. Pastor, R. W.; Brooks, B. R.; Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics* **1988**, 65, 1409-1419.
81. 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, 6811-6815.
82. Case, D. A.; Cheatham, T. E.; Tom, D.; Holger, G.; Luo, R.; Merz, K. M.; Alexey, O.; Carlos, S.; Bing, W.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, 26, 1668-1688.
83. Cai, Q.; Hsieh, M. J.; Wang, J.; Luo, R., Performance of Nonlinear Finite-Difference Poisson−Boltzmann Solvers. *J. Chem. Theory Comput.* **2010**, 6, 203.
84. Holst, M.; Baker, N.; Wang, F., Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples. *J. Comput. Chem.* **2000**, 21, 1319-1342.
85. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B., Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, 23, 128-137.
86. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, 101, 3005-3014.

87. Case, D. A., Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **2010**, 4, 285–290.
88. Karplus, M.; Kushick, J. N., Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, 14, 325-332.
89. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Structure Function & Bioinformatics* **2004**, 55, 383-94.
90. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A., Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, 25, 265-84.
91. Bahar, I.; Rader, A. J., Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, 15, 586-592.
92. Tama, F.; Sanejouand, Y. H., Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, 14, 1-6.
93. Levitt, M.; Sander, C.; Stern, P. S., Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology* **1985**, 181, 423-447.
94. Geick, R.; Perry, C. H.; Rupprecht, G., Normal Modes in Hexagonal Boron Nitride. *Phys. Rev.* **1966**, 146, 543-547.
95. 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, 1097-1115.
96. Litim, A.; Belhocine, Y.; Benlecheb, T.; Ghoniem, M. G.; Kabouche, Z.; Ali, F. A. M.; Abdulkhair, B. Y.; Seydou, M.; Rahali, S., DFT-D4 Insight into the Inclusion of Amphetamine and Methamphetamine in Cucurbit[7]uril: Energetic, Structural and Biosensing Properties. *Molecules* **2021**, 26, 7479.
97. Crean, R. M.; Pudney, C. R.; Cole, D. K.; van der Kamp, M. W., Reliable In Silico Ranking of Engineered Therapeutic TCR Binding Affinities with MMPB/GBSA. *J. Chem. Inf. Model.* **2022**.
98. Karaca, E.; Prévost, C.; Sacquin-Mora, S., Modeling the Dynamics of Protein–Protein Interfaces, How and Why? *Molecules* **2022**, 27, 1841.
99. Yau, M. Q.; Loo, J. S. E., Consensus scoring evaluated using the GPCR-Bench dataset: Reconsidering the role of MM/GBSA. *J. Comput.-Aided Mol. Des.* **2022**.
100. Liu, X.; Peng, L.; Zhang, J. Z., Accurate and Efficient Calculation of Protein–Protein Binding Free Energy-Interaction Entropy with Residue Type-Specific Dielectric Constants. *J. Chem. Inf. Model.* **2018**, 59, 272-281.
101. Zhou, Y.; Liu, X.; Zhang, Y.; Peng, L.; Zhang, J. Z., Residue-specific free energy analysis in ligand bindings to JAK2. *Molecular Physics* **2018**, 116, 2633-2641.
102. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T., End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chemical reviews* **2019**, 119, 9478-9508.
103. Deng, C.-L.; Cheng, M.; Zavalij, P. Y.; Isaacs, L., Thermodynamics of pillararene·guest complexation: blinded dataset for the SAMPL9 challenge. *New J. Chem.* **2022**, 46, 995-1002.
104. Zhu, Z.; Wang, G.; Xu, Z.; Chen, Z.; Wang, J.; Shi, J.; Zhu, W., Halogen bonding in differently charged complexes: basic profile, essential interaction terms and intrinsic σ-hole. *Phys. Chem. Chem. Phys.* **2019**, 21, 15106-15119.
105. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functiona. *Theor. Chem. Acc.* **2008**, 120, 215-241.
106. Haoyu, S. Y.; He, X.; Li, S. L.; Truhlar, D. G., MN15: A Kohn–Sham global-hybrid exchange–correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chemical science* **2016**, 7, 5032-5051.
107. Chai, J. D.; Head-Gordon, M., Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, 128, 57-63.
108. Sun, Z.; Huai, Z.; He, Q.; Liu, Z., A General Picture of Cucurbit[8]uril Host–Guest Binding. *J. Chem. Inf. Model.* **2021**, 61, 6107-6134.
109. Sun, Z.; Zheng, L.; Kai, W.; Huai, Z.; Liu, Z., Primary vs Secondary: Directionalized Guest Coordination in β-Cyclodextrin Derivatives. *chemrxiv* **2022**.

110. Sun, Z.; Wang, M.; He, Q.; Liu, Z., Molecular Modelling of Ionic Liquids: Force-Field Validation and Thermodynamic Perspective from Large-Scale Fast-Growth Solvation Free Energy Calculations. *Adv. Theory Simul.* **2022**, 2200274.
111. Sun, Z.; He, Q.; Zhihao, G.; Payam, K.; Huai, Z.; Liu, Z., A General Picture of Cucurbit[8]uril Host-Guest Binding: Recalibrating Bonded Interactions. *chemrxiv* **2022**.
112. Sun, Z.; Gong, Z.; Zheng, L.; Payam, K.; Huai, Z.; Liu, Z., Molecular Modelling of Ionic Liquids: General Guidelines on Fixed-Charge Force Fields for Balanced Descriptions. *chemrxiv* **2022**.
113. Kendall, M. G., A New Measure of Rank Correlation. *Biometrika* **1938**, 30, 81-93.
114. Pearlman, D. A.; Charifson, P. S., Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *Journal of Medicinal Chemistry* **2001**, 44, 3417-3423.
115. He, X.; Man, V. H.; Ji, B.; Xie, X.-Q.; Wang, J., Calculate protein–ligand binding affinities with the extended linear interaction energy method: application on the Cathepsin S set in the D3R Grand Challenge 3. *J. Comput.-Aided Mol. Des.* **2019**, 33, 105-117.
116. Liu, X.; Liu, J.; Zhu, T.; Zhang, L.; He, X.; Zhang, J. Z., PBSA_E: A PBSA-Based Free Energy Estimator for Protein–Ligand Binding Affinity. *J. Chem. Inf. Model.* **2016**, 56, 854-861.
117. Casbarra, L.; Procacci, P., Binding free energy predictions in host-guest systems using Autodock4. A retrospective analysis on SAMPL6, SAMPL7 and SAMPL8 challenges. *J. Comput.-Aided Mol. Des.* **2021**.
118. Sun, Z.; Wang, X.; Zhao, Q.; Zhu, T., Understanding Aldose Reductase-Inhibitors interactions with free energy simulation. *Journal of Molecular Graphics and Modelling* **2019**, 91, 10-21.
119. Huai, Z.; Shen, Z.; Sun, Z., Binding Thermodynamics and Interaction Patterns of Inhibitor-Major Urinary Protein-I Binding from Extensive Free-Energy Calculations: Benchmarking AMBER Force Fields. *J. Chem. Inf. Model.* **2021**, 61, 284-297.
120. 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, 1511-1524.
121. Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **2018**, 20, 14450-14460.
122. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, 51, 69-82.
123. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, 32, 866-877.
124. Zheng, L.; Yang, Y.; Bao, J.; He, L.; Qi, Y.; Zhang, J. Z. H., Discovery of novel inhibitors of CDK2 using docking and physics-based binding free energy calculation. *Chemical Biology & Drug Design* **2022**, 99, 662-673.
125. Panday, S. K.; Alexov, E., Protein–Protein Binding Free Energy Predictions with the MM/PBSA Approach Complemented with the Gaussian-Based Method for Entropy Estimation. *ACS Omega* **2022**.
126. Basit, A.; Yadav, A. K.; Bandyopadhyay, P., Calcium Ion Binding to the Mutants of Calmodulin: A Structure-Based Computational Predictive Model of Binding Affinity Using a Charge Scaling Approach in Molecular Dynamics Simulation. *J. Chem. Inf. Model.* **2022**.
127. Kohut, G.; Liwo, A.; Bősze, S.; Beke-Somfai, T.; Samsonov, S. A., Protein–Ligand Interaction Energy-Based Entropy Calculations: Fundamental Challenges For Flexible Systems. *J. Phys. Chem. B* **2018**, 122, 7821-7827.
128. Bertazzo, M.; Gobbo, D.; Decherchi, S.; Cavalli, A., Machine Learning and Enhanced Sampling Simulations for Computing the Potential of Mean Force and Standard Binding Free Energy. *J. Chem. Theory Comput.* **2021**.
129. Boz, E.; Stein, M., Accurate Receptor-Ligand Binding Free Energies from Fast QM Conformational Chemical Space Sampling. *Int. J. Mol. Sci.* **2021**, 22, 3078.

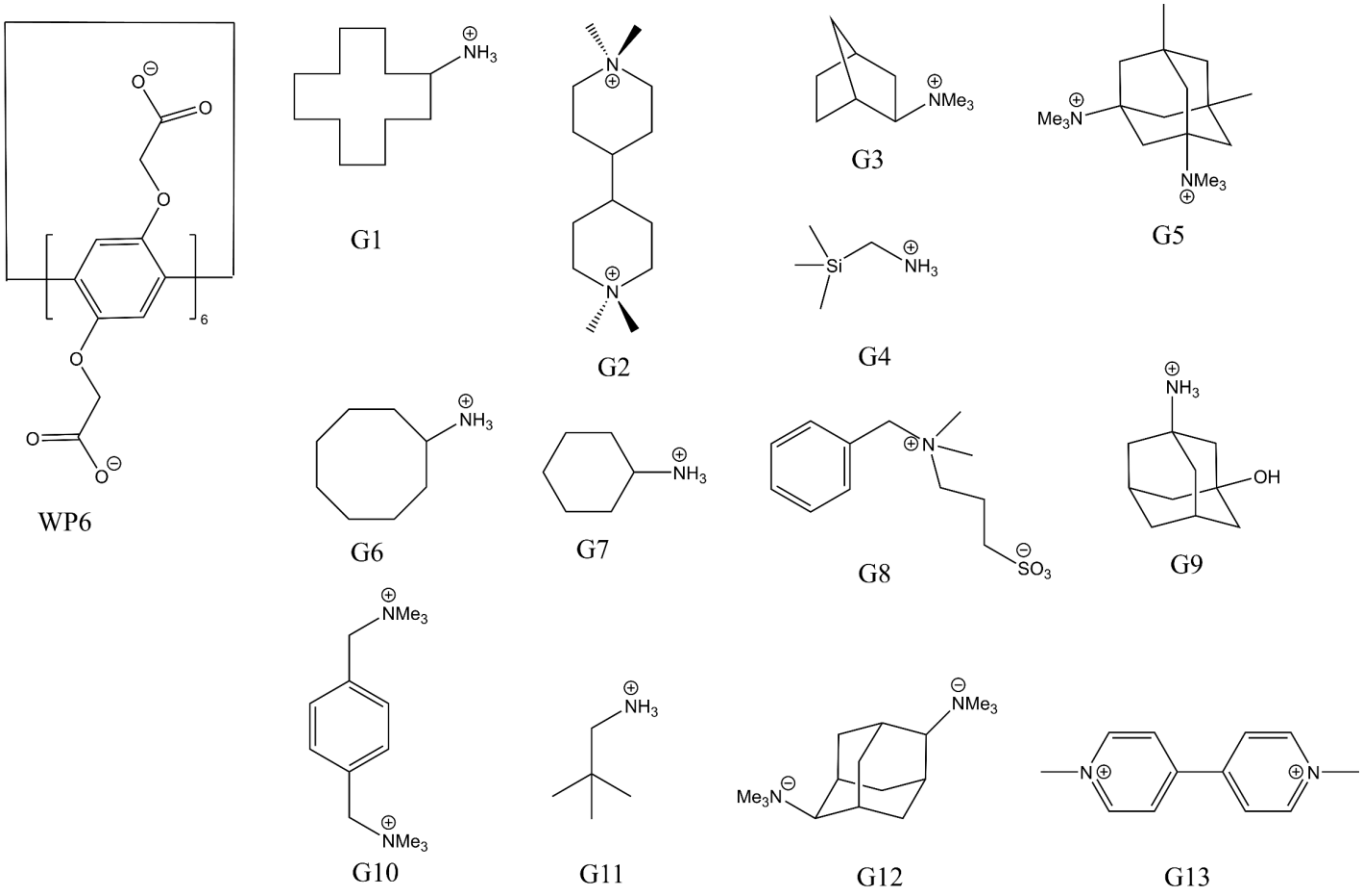


Fig. 1. The 2D chemical structures of WP6 host-guest systems investigated in the current work.

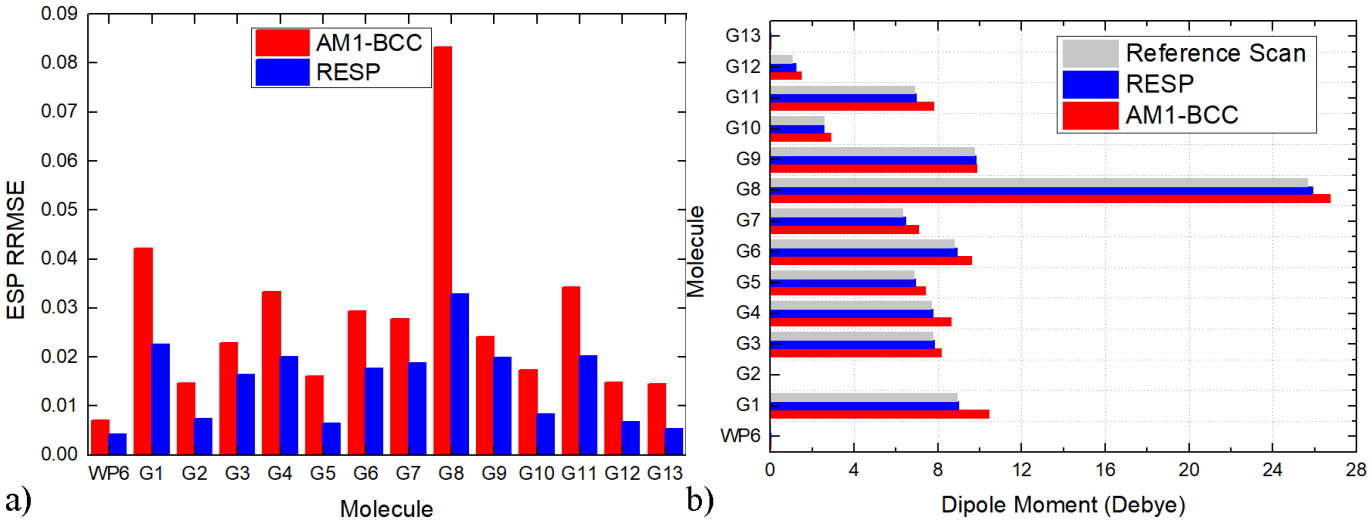


Fig. 2. Preliminary of charge quality for the two charge schemes (AM1-BCC and RESP-1 charge sets). a) The ESP RRMSE used to assess the percentage deviation of the charge-produced ESP from the *ab initio* reference. b) The dipole moments obtained from the two charge schemes and the electronic structure calculations. For both the Coulombic ESP and the dipole moments, RESP-1 charges reproduce the HF/6-31G* results obviously better than the AM1-BCC charge scheme. However, it is fair to note that the difference between the AM1-BCC results and the *ab initio* reference is also insignificant. This phenomenon is in stark contrast to the Cucurbit[8]uril and β-cyclodextrin host-guest systems investigated in our previous works, where the AM1-BCC charge scheme triggers significant problems and RESP charges are required to obtain accurate ESP and binding information. According to the satisfactory behavior of AM1-BCC, the current WP6 host-guest systems can be considered similar to the training set of the AM1-BCC charge scheme and thus the electrostatics behavior would be relatively easy to model compared with other host-guest systems.

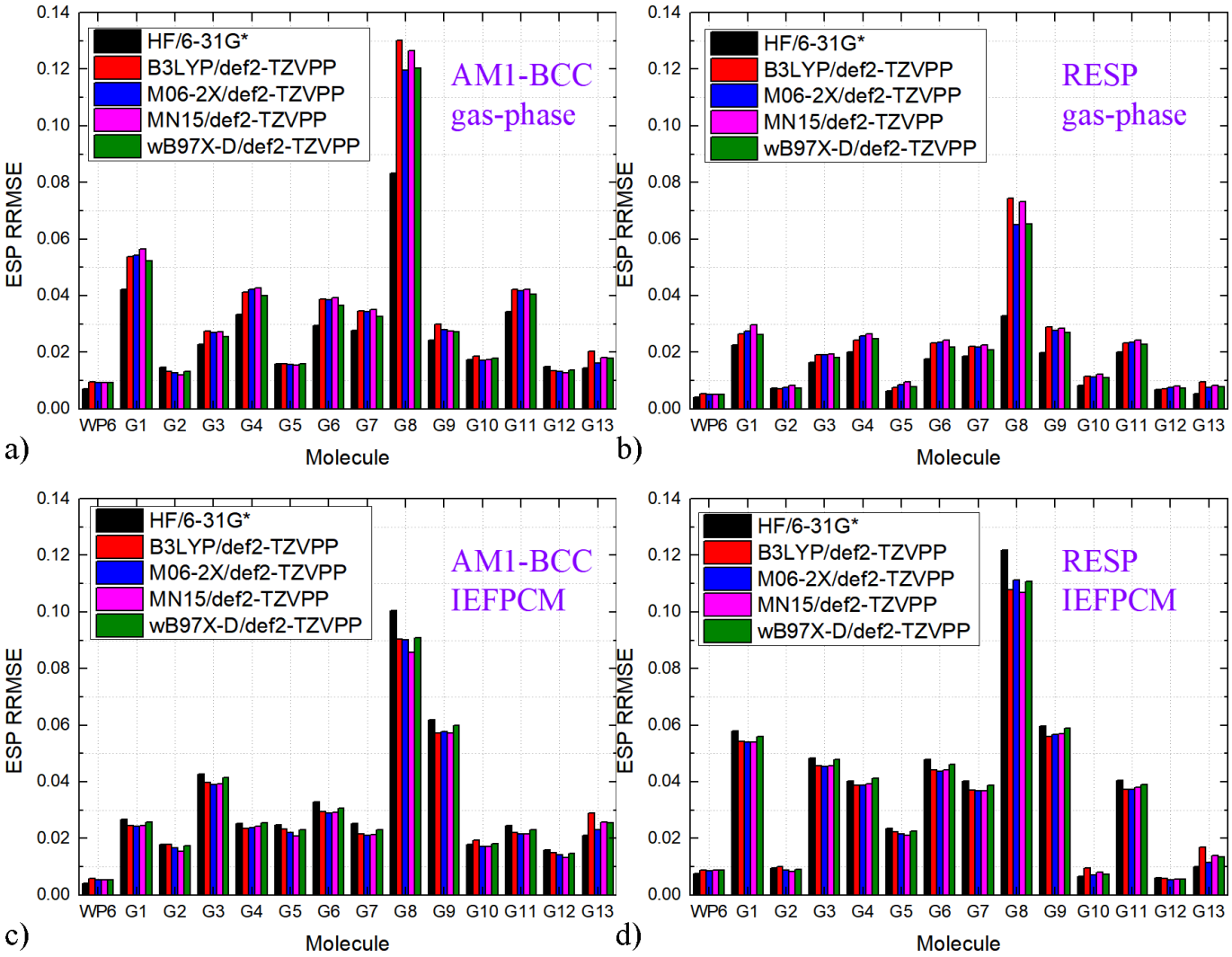


Fig. 3. The percentage deviation of the charge-produced ESP from different *ab initio* references. The ESP RRMSE of a) AM1-BCC and b) RESP-1 charges fitted targeting the HF/6-31G* ESP when the gas-phase data are used as the reference. The ESP RRMSE of c) AM1-BCC and d) RESP-1 when using implicit-solvent data as reference. The implicit solvent model of IEFPCM is used to mimic the solvent environment.

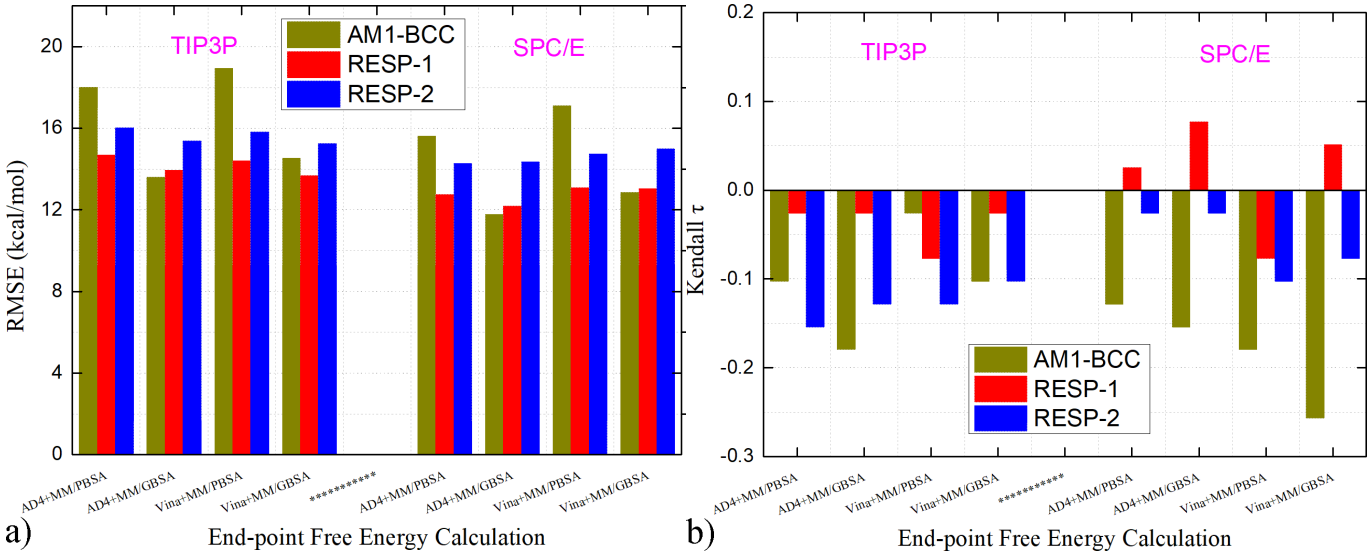


Fig. 4. Comparison of quality metrics obtained with different charge schemes and water models: a) RMSE and b) Kendall's ranking coefficient.

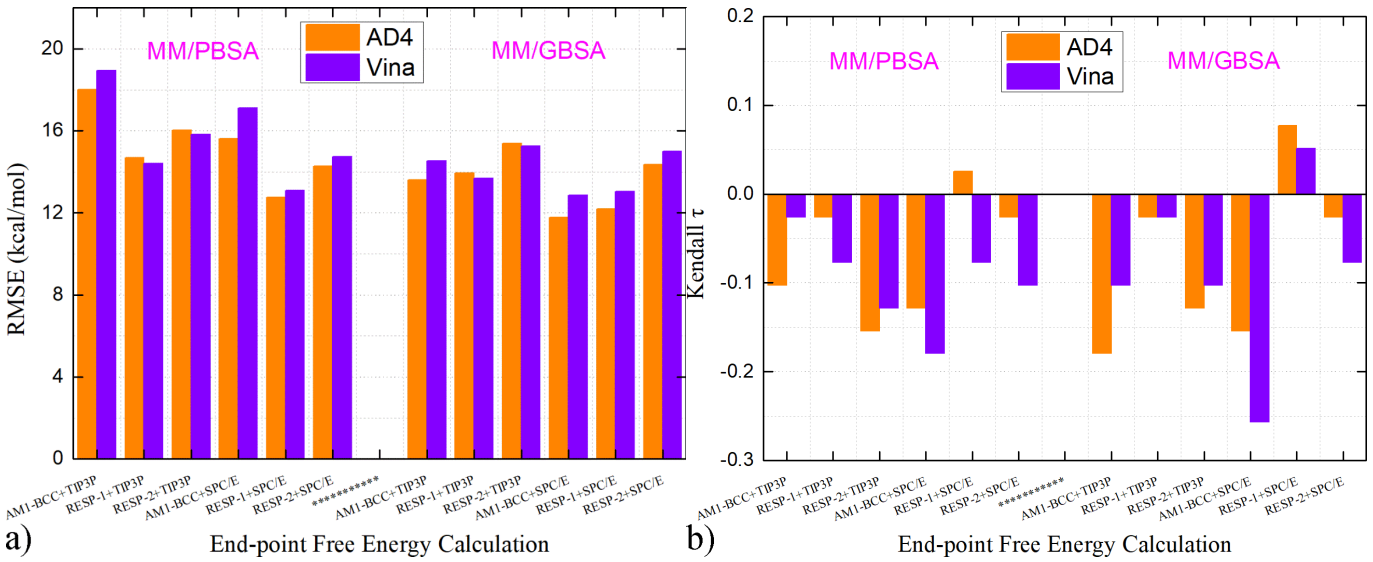


Fig. 5. Comparison of quality metrics obtained with different docking procedures (scoring functions) and free energy methods: a) RMSE and b) Kendall's ranking coefficient.

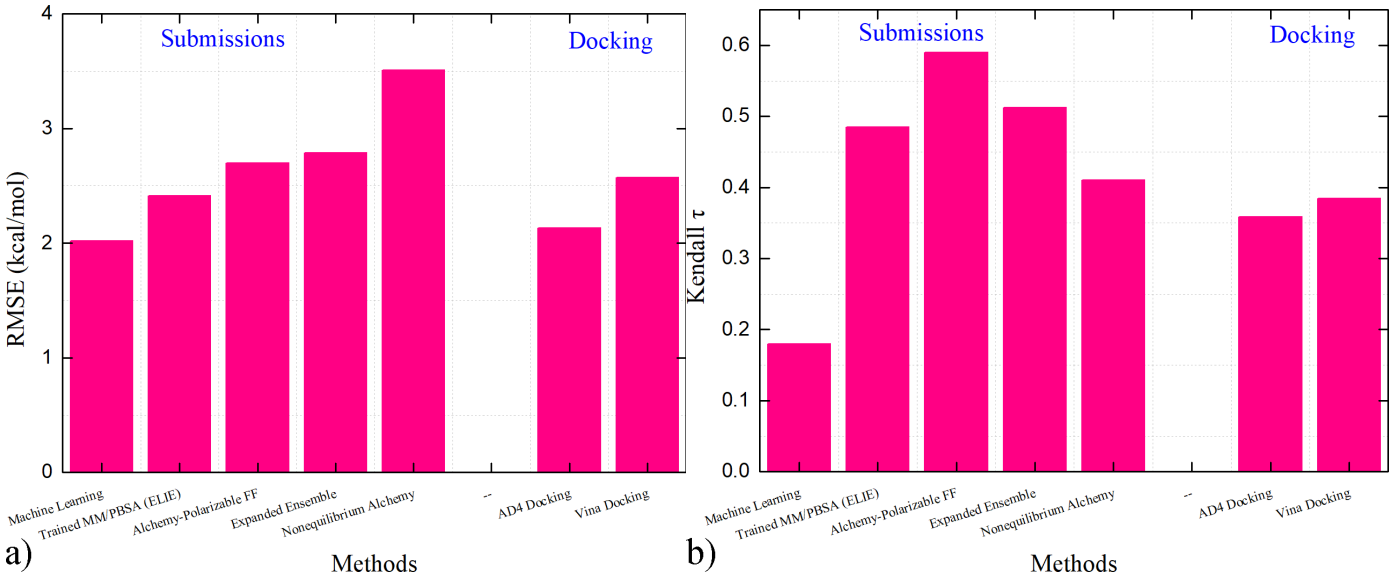


Fig. 6. Quality metrics for ranked submissions in the WP6 host-guest challenge and docking results obtained in the current work: a) RMSE and b) Kendall's ranking coefficient. Among submitted predictions, the top-performing techniques of machine learning and ELIE (trained MM/PBSA) are computationally cheap and both involve some level of training. By contrast, although alchemical free energy calculations require comprehensive sampling of the configurational space along the alchemical pathway, the alchemical results are less accurate. The ELIE technique involves an end-point free energy calculation procedure similar to the current work, and the difference is that the ELIE method uses pre-trained weighting factors for different components in the end-point free energy estimation. Interestingly, the docking-produced affinity estimates (scores) with both AD4 and Vina scoring functions perform quite well in reproducing the experimental values, despite the low computational costs of the docking procedures.

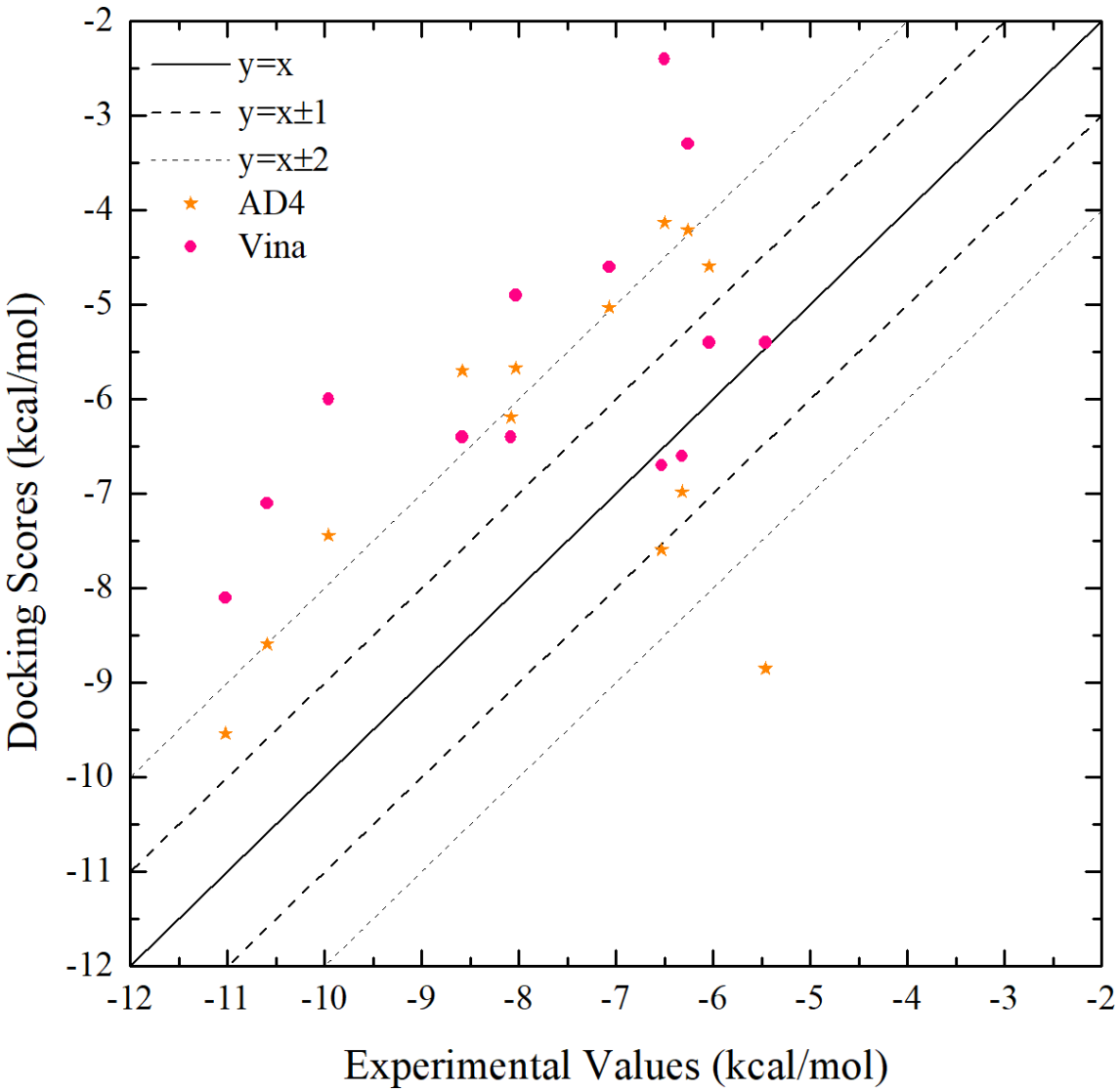


Fig. 7. Correlation between the docking scores obtained with the AD4 and Vina scoring functions and the experimental reference.

Supporting Information:

Comprehensive Evaluation of End-Point Free Energy Techniques in Carboxylated-Pillar[6]arene Host-guest Binding: I. Standard Procedure

Xiao Liu[1*], Lei Zheng[2], Chu Qin[1], John Z.H. Zhang[2,3,4,5], Zhaoxi Sun[6*]

[1]*School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China*

[2]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

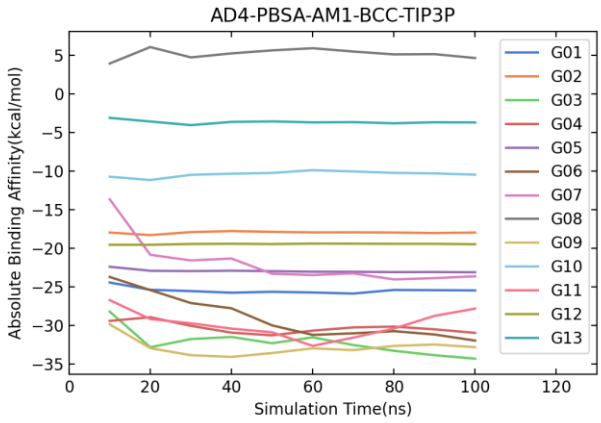
[3]*School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, 200062, China*

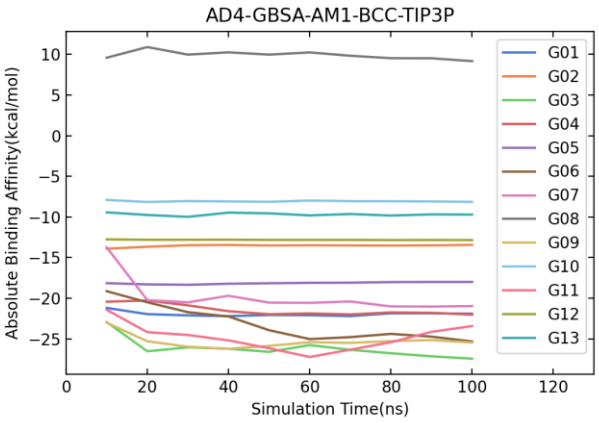
[4]*Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China*

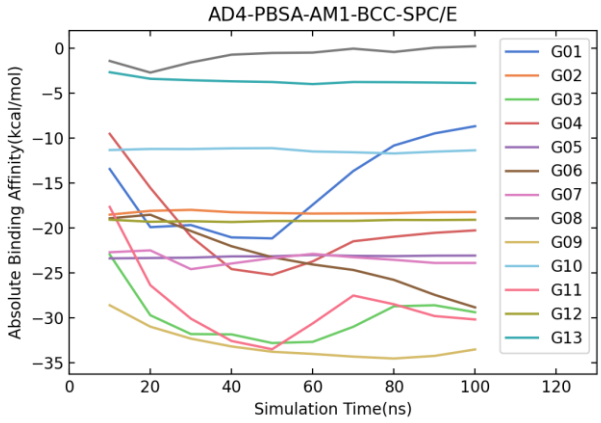
[5]*Department of Chemistry, New York University, NY, NY 10003, USA*

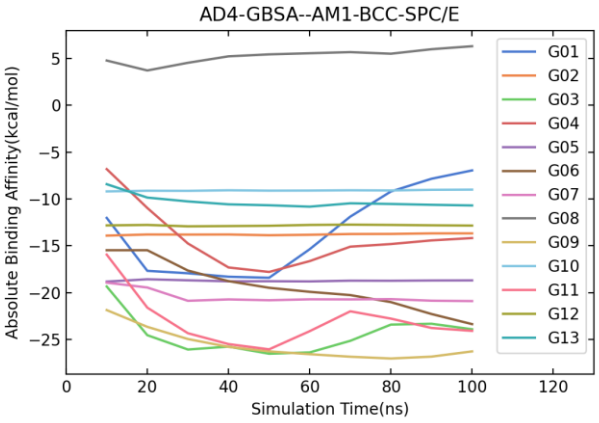
[6]*College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China*

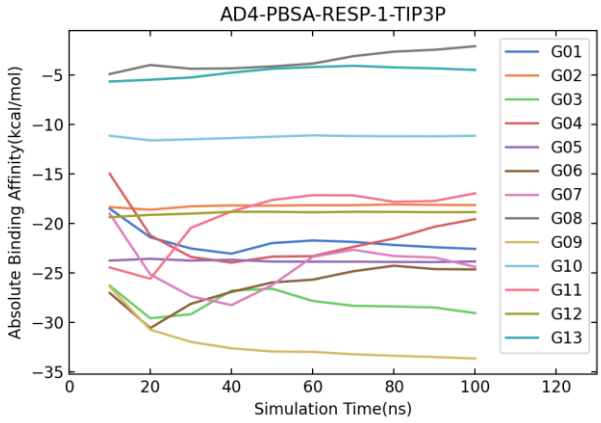
*To whom correspondence should be addressed: liuxiaode2013@163.com, proszx@163.com

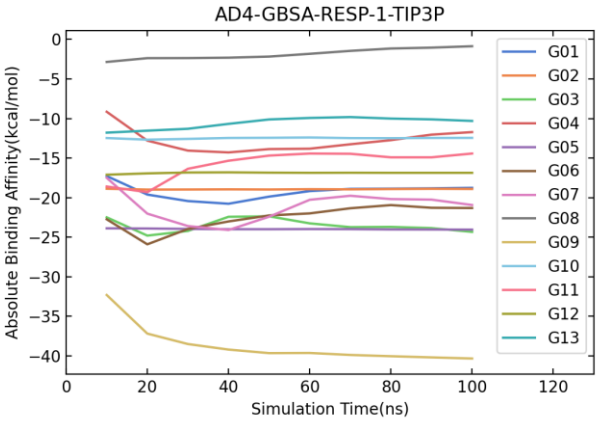

AD4-PBSA-AM1-BCC-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

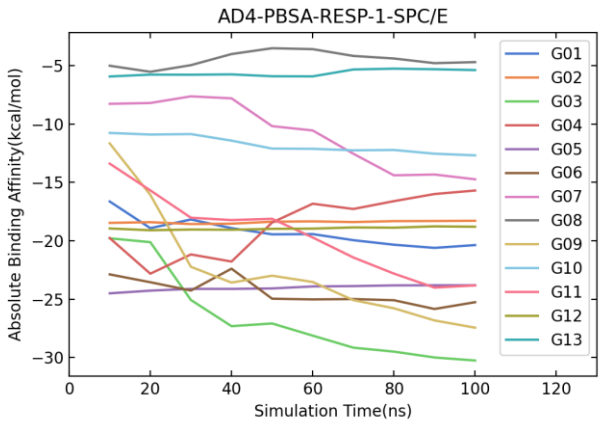

AD4-GBSA-AM1-BCC-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

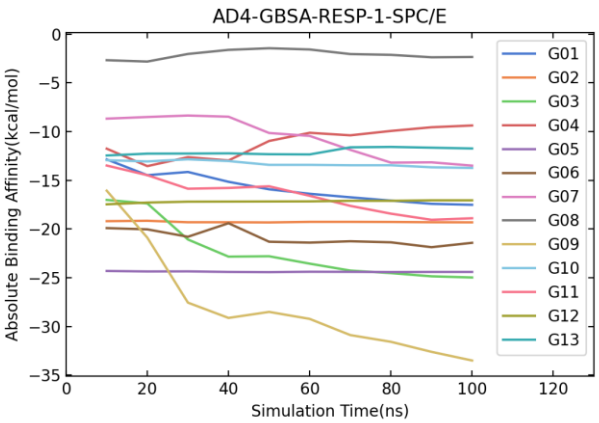

AD4-PBSA-AM1-BCC-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


AD4-GBSA--AM1-BCC-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


AD4-PBSA-RESP-1-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


AD4-GBSA-RESP-1-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


AD4-PBSA-RESP-1-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


AD4-GBSA-RESP-1-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

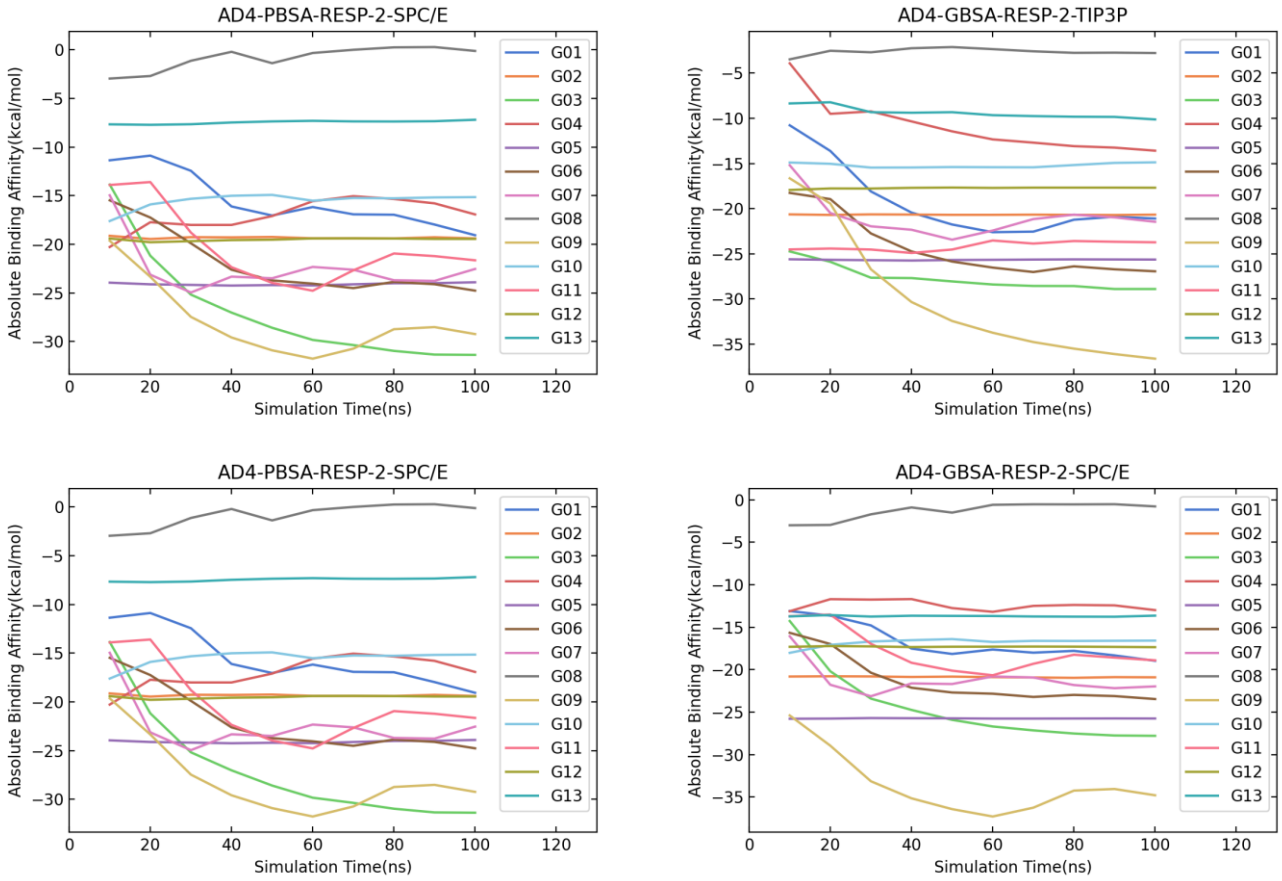
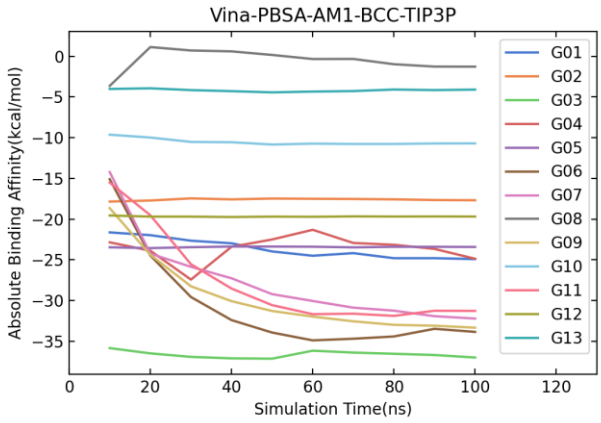
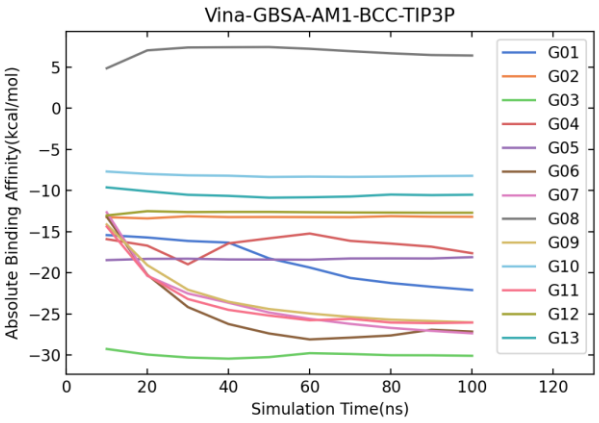
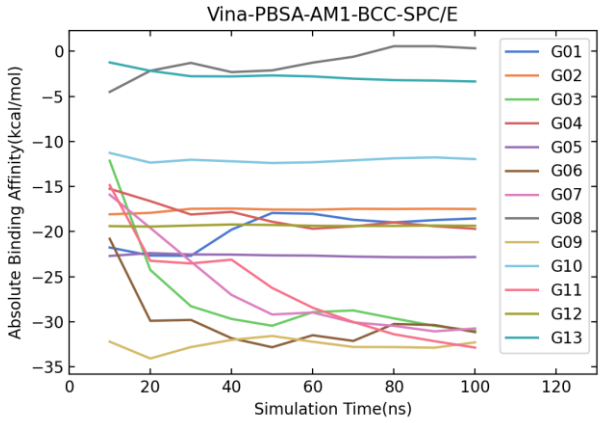
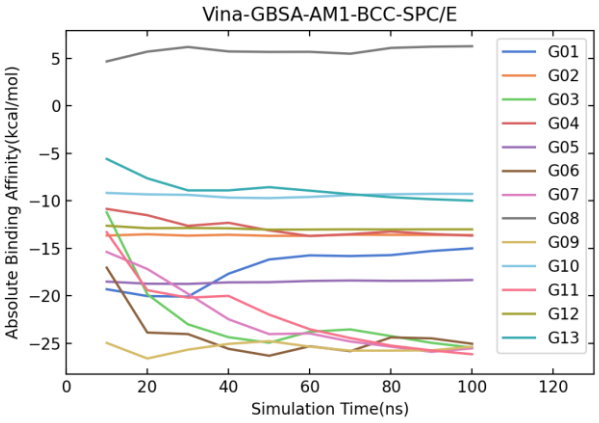


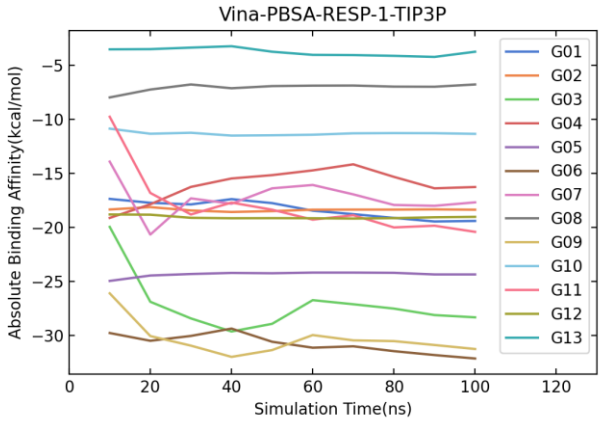
Fig. S1. Time-dependence of end-point free energy estimates using AD4-produced initial bound structures, two charge schemes for solutes, two water models and two end-point methods.

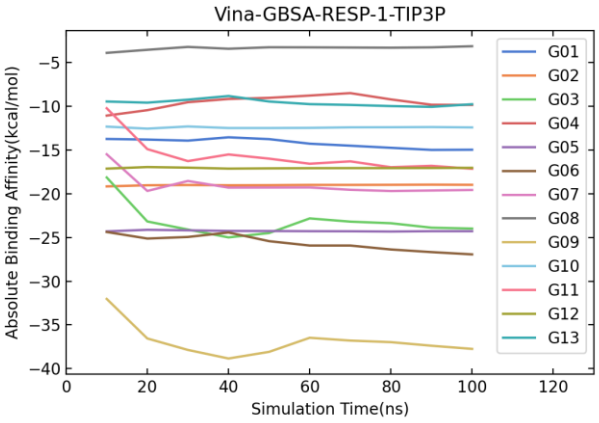

Vina-PBSA-AM1-BCC-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

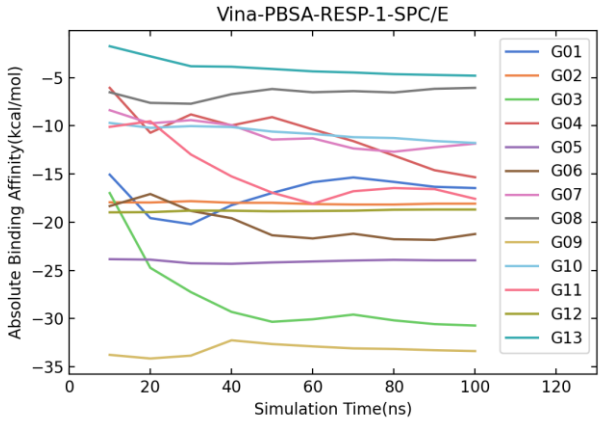

Vina-GBSA-AM1-BCC-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

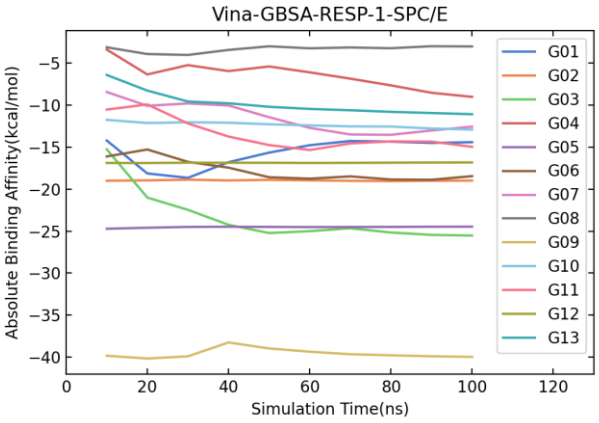

Vina-PBSA-AM1-BCC-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


Vina-GBSA-AM1-BCC-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


Vina-PBSA-RESP-1-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


Vina-GBSA-RESP-1-TIP3P
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


Vina-PBSA-RESP-1-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13


Vina-GBSA-RESP-1-SPC/E
Absolute Binding Affinity(kcal/mol)
Simulation Time(ns)
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
G13

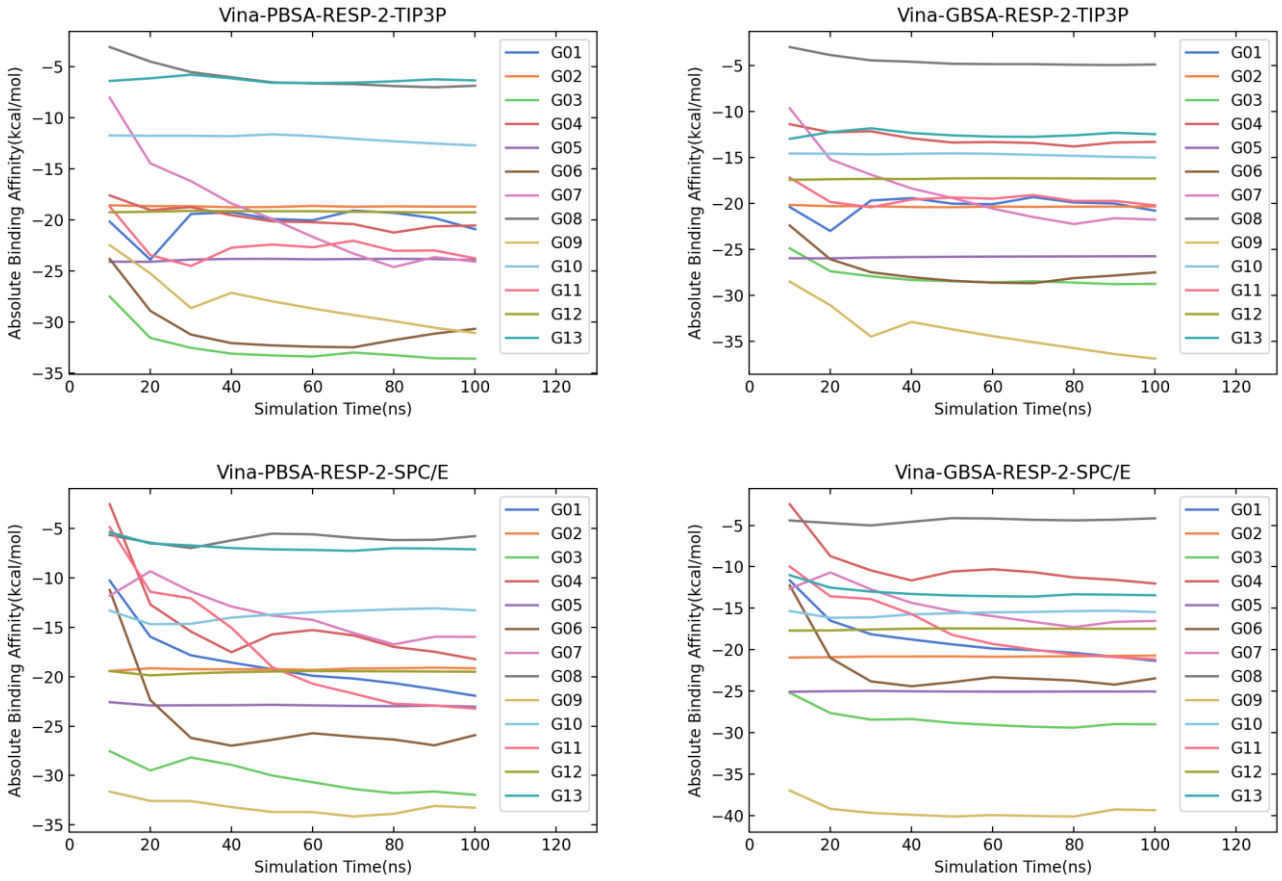


Fig. S2. Time-dependence of end-point free energy estimates using Vina-produced initial bound structures, two charge schemes for solutes, two water models and two end-point methods.

Table S1. MM/PBSA ($\Delta G_{MM/PBSA}$) binding free energies of WP6 host-guest systems obtained with the three charge schemes, two water models (TIP3P and SPC/E) and 100 ns unbiased sampling initiated from AD4-produced docking poses. MSE, RMSE, τ, and PI serve as quality measurements.

Host	Guest	Experiment	AM1-BCC				RESP-1				RESP-2			
			TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD
WP6	G1	-6.53	-25.4	0.2	-8.7	0.5	-22.6	0.3	-20.4	0.2	-20.5	0.2	-19.1	0.3
	G2	-10.59	-17.9	0.2	-18.2	0.2	-18.1	0.3	-18.3	0.3	-19.1	0.3	-19.4	0.3
	G3	-8.03	-34.3	0.2	-29.4	0.2	-29.1	0.2	-30.3	0.2	-32.1	0.2	-31.4	0.2
	G4	-6.50	-30.9	0.2	-20.3	0.3	-19.6	0.3	-15.7	0.3	-20.6	0.2	-16.9	0.3
	G5	-5.46	-23.1	0.2	-23.1	0.2	-23.8	0.3	-23.8	0.3	-23.9	0.2	-23.9	0.2
	G6	-8.08	-31.9	0.2	-28.8	0.2	-24.6	0.2	-25.3	0.2	-30.5	0.2	-24.8	0.3
	G7	-7.07	-23.6	0.3	-23.9	0.2	-24.4	0.3	-14.7	0.3	-22.6	0.3	-22.5	0.2
	G8	-6.04	4.7	0.3	0.2	0.3	-2.1	0.3	-4.7	0.3	-3.0	0.3	-0.1	0.3
	G9	-6.32	-32.8	0.2	-33.5	0.2	-33.7	0.2	-27.4	0.2	-31.0	0.2	-29.2	0.2
	G10	-9.96	-10.4	0.2	-11.4	0.2	-11.1	0.3	-12.7	0.4	-12.9	0.3	-15.2	0.3
	G11	-6.26	-27.8	0.3	-30.2	0.2	-17.0	0.3	-14.6	0.3	-28.5	0.3	-21.7	0.3
	G12	-11.02	-19.5	0.2	-19.1	0.2	-18.9	0.2	-18.8	0.2	-19.6	0.2	-19.5	0.2
	G13	-8.58	-3.7	0.2	-3.9	0.2	-4.5	0.3	-5.4	0.3	-4.6	0.2	-7.2	0.2
RMSE			18.0		15.6		14.7		12.8		16.0		14.3	
MSE			13.6		11.5		11.5		10.1		13.0		11.6	
τ			-0.1		-0.1		0.0		0.0		-0.2		0.0	
PI			-0.1		-0.1		-0.1		0.0		-0.1		0.0	

Table S2. MM/GBSA ($\Delta G_{\mathrm{MM/GBSA}}$) binding free energies of WP6 host-guest systems obtained with the three charge schemes, two water models (TIP3P and SPC/E) and 100 ns unbiased sampling initiated from AD4-produced docking poses. MSE, RMSE, τ, and PI serve as quality measurements.

Host	Guest	Experiment	AM1-BCC				RESP-1				RESP-2			
			TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD
WP6	G1	-6.53	-21.9	0.2	-7.0	0.5	-18.8	0.3	-17.5	0.2	-21.1	0.2	-19.0	0.3
	G2	-10.59	-13.4	0.2	-13.7	0.2	-18.9	0.3	-19.3	0.3	-20.7	0.3	-20.9	0.3
	G3	-8.03	-27.4	0.2	-23.9	0.2	-24.3	0.2	-25.0	0.2	-28.9	0.2	-27.8	0.2
	G4	-6.50	-22.0	0.2	-14.2	0.3	-11.7	0.3	-9.4	0.3	-13.6	0.2	-13.0	0.2
	G5	-5.46	-18.0	0.2	-18.7	0.2	-24.1	0.3	-24.4	0.3	-25.6	0.2	-25.8	0.2
	G6	-8.08	-25.3	0.2	-23.4	0.2	-21.3	0.2	-21.4	0.2	-26.9	0.2	-23.5	0.3
	G7	-7.07	-21.0	0.2	-20.9	0.2	-20.9	0.3	-13.5	0.3	-21.5	0.3	-22.0	0.2
	G8	-6.04	9.2	0.2	6.3	0.3	-0.9	0.3	-2.3	0.3	-2.8	0.3	-0.8	0.3
	G9	-6.32	-25.5	0.2	-26.3	0.2	-40.4	0.2	-33.5	0.2	-36.6	0.2	-34.8	0.2
	G10	-9.96	-8.2	0.2	-9.0	0.2	-12.5	0.3	-13.7	0.4	-14.9	0.3	-16.6	0.3
	G11	-6.26	-23.4	0.2	-24.1	0.2	-14.4	0.3	-13.4	0.2	-23.7	0.3	-18.9	0.3
	G12	-11.02	-12.9	0.2	-12.8	0.2	-16.9	0.2	-17.1	0.2	-17.7	0.2	-17.4	0.2
	G13	-8.58	-9.7	0.2	-10.7	0.2	-10.3	0.3	-11.7	0.3	-10.1	0.2	-13.6	0.2
RMSE			13.6		11.8		13.9		12.2		15.4		14.4	
MSE			9.2		7.5		10.4		9.4		12.6		11.8	
τ			-0.2		-0.2		0.0		0.1		-0.1		0.0	
PI			0.0		0.0		-0.1		0.0		-0.1		0.0	

Table S3. MM/PBSA ($\Delta G_{MM/PBSA}$) binding free energies of WP6 host-guest systems obtained with the three charge schemes, two water models (TIP3P and SPC/E) and 100 ns unbiased sampling initiated from Vina-produced docking poses. MSE, RMSE, τ, and PI serve as quality measurements.

Host	Guest	Experiment	AM1-BCC				RESP-1				RESP-2			
			TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD
WP6	G1	-6.53	-24.9	0.2	-18.5	0.2	-19.4	0.3	-16.5	0.3	-20.9	0.2	-21.9	0.2
	G2	-10.59	-17.7	0.2	-17.5	0.2	-18.4	0.3	-18.1	0.3	-18.7	0.3	-19.1	0.3
	G3	-8.03	-37.0	0.2	-31.0	0.2	-28.3	0.3	-30.7	0.3	-33.6	0.2	-32.0	0.2
	G4	-6.50	-24.9	0.2	-19.7	0.3	-16.3	0.2	-15.3	0.3	-20.6	0.3	-18.2	0.3
	G5	-5.46	-23.4	0.2	-22.8	0.2	-24.4	0.3	-24.0	0.2	-24.0	0.3	-23.0	0.2
	G6	-8.08	-33.9	0.2	-31.1	0.2	-32.2	0.2	-21.2	0.3	-30.7	0.2	-25.9	0.2
	G7	-7.07	-32.2	0.3	-30.7	0.2	-17.7	0.3	-11.9	0.3	-24.1	0.3	-16.0	0.3
	G8	-6.04	-1.3	0.3	0.3	0.3	-6.8	0.2	-6.1	0.2	-6.9	0.2	-5.8	0.3
	G9	-6.32	-33.4	0.2	-32.3	0.2	-31.3	0.2	-33.4	0.2	-31.1	0.2	-33.3	0.2
	G10	-9.96	-10.7	0.2	-12.0	0.3	-11.4	0.3	-11.8	0.3	-12.7	0.3	-13.3	0.3
	G11	-6.26	-31.3	0.3	-32.9	0.3	-20.4	0.3	-17.6	0.3	-23.2	0.2	-23.2	0.2
	G12	-11.02	-19.7	0.2	-19.4	0.1	-19.0	0.2	-18.7	0.2	-19.3	0.2	-19.5	0.2
	G13	-8.58	-4.1	0.2	-3.4	0.2	-3.8	0.3	-4.8	0.2	-6.4	0.2	-7.1	0.2
RMSE			18.9		17.1		14.4		13.1		15.8		14.7	
MSE			14.9		13.1		11.5		10.0		13.2		12.1	
τ			0.0		-0.2		-0.1		-0.1		-0.1		-0.1	
PI			-0.2		-0.1		-0.1		-0.1		-0.1		-0.1	

Table S4. MM/GBSA ($\Delta G_{\mathrm{MM/GBSA}}$) binding free energies of WP6 host-guest systems obtained with the three charge schemes, two water models (TIP3P and SPC/E) and 100 ns unbiased sampling initiated from Vina-produced docking poses. MSE, RMSE, τ, and PI serve as quality measurements.

Host	Guest	Experiment	AM1-BCC				RESP-1				RESP-2			
			TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD	TIP3P	SD	SPC/E	SD
WP6	G1	-6.53	-22.1	0.2	-15.0	0.2	-15.0	0.3	-14.4	0.3	-20.8	0.2	-21.4	0.2
	G2	-10.59	-13.2	0.2	-13.6	0.2	-19.0	0.3	-19.0	0.3	-20.3	0.3	-20.7	0.3
	G3	-8.03	-30.1	0.2	-25.5	0.2	-24.0	0.3	-25.5	0.3	-28.8	0.2	-29.0	0.2
	G4	-6.50	-17.6	0.2	-13.7	0.3	-9.9	0.2	-9.0	0.3	-13.3	0.3	-12.0	0.3
	G5	-5.46	-18.1	0.2	-18.3	0.2	-24.3	0.3	-24.5	0.2	-25.8	0.3	-25.1	0.2
	G6	-8.08	-27.2	0.2	-25.1	0.2	-26.9	0.2	-18.5	0.2	-27.5	0.2	-23.5	0.2
	G7	-7.07	-27.4	0.2	-25.5	0.2	-19.6	0.3	-12.6	0.3	-21.8	0.3	-16.5	0.3
	G8	-6.04	6.4	0.2	6.3	0.3	-3.2	0.2	-3.0	0.2	-4.9	0.2	-4.2	0.3
	G9	-6.32	-26.0	0.2	-25.4	0.2	-37.7	0.2	-40.0	0.2	-36.9	0.2	-39.4	0.2
	G10	-9.96	-8.2	0.2	-9.3	0.3	-12.4	0.3	-12.9	0.3	-15.0	0.3	-15.5	0.3
	G11	-6.26	-26.1	0.2	-26.2	0.3	-17.2	0.3	-15.0	0.3	-21.2	0.2	-21.2	0.2
	G12	-11.02	-12.7	0.2	-13.0	0.1	-17.1	0.2	-16.8	0.2	-17.3	0.2	-17.5	0.2
	G13	-8.58	-10.5	0.2	-10.0	0.2	-9.8	0.3	-11.1	0.2	-12.5	0.2	-13.4	0.2
RMSE			14.5		12.9		13.7		13.1		15.2		15.0	
MSE			10.2		8.8		10.4		9.4		12.7		12.2	
τ			-0.1		-0.3		0.0		0.1		-0.1		-0.1	
PI			-0.1		0.0		-0.1		-0.1		-0.1		-0.1	