Primary vs Secondary: Directionalized Guest Coordination in β-Cyclodextrin Derivatives

Zhaoxi Sun[1*], Lei Zheng[2], Kai Wang[3], Zhe Huai[4], and Zhirong Liu[1]

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

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

[3]*METiS Pharmaceuticals, Inc., Hangzhou 310052, China*

[4]*XtalPi - AI Research Center (XARC), 7F, Tower A, Dongsheng Building, No.8, Zhongguancun East Road, Haidian District, Beijing 100083, China*

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

Abstract

The cyclodextrin (CD) and its derivatives as useful drug carriers and reservoirs are widely used in pharmaceutical and chemical industries. The CD-guest (drug) and more generally host-guest coordinations are often considered prototypical cases that are much simpler than biological systems. As a result, in computational modelling even with dynamics propagation, only a single binding pose is considered. However, due to the asymmetric feature of the CD host, the guest can directionally bind to its primary 6' or secondary 3' face. Correct modelling of the primary-secondary equilibrium clearly poses a challenge, and any static or dynamic calculations modelling only a single binding preference would introduce systematic biases of unknown magnitude. In this work, using the β-CD host-guest set in a recent grand challenge, we present a comprehensive analysis of various aspects of fixed-charge modelling of β-CD host-guest complexes. The force field parameters (electrostatics and bonded terms) are evaluated in a detailed way, and the best parameter combination is employed in enhanced sampling simulations that accelerate the translational diffusion of the guest molecule, sample the binding/unbinding events extensively, and thus explore the space of possible binding modes. Finally, the binding affinities and the primary-secondary equilibrium (directionalized guest coordination preference) are computed and compared with the experimental reference. Possible reasons for deviations and further ways to improve the quality of calculation are discussed.

Keywords: β-Cyclodextrin, Host-guest Interaction, Directional Preference, Primary and Secondary Surfaces, Fixed-charge Force Field

1. Introduction

Macromolecular containers are receiving enormous attention due to their capability of tunable encapsulation of therapeutic agents and controllable release. Among so many macrocyclic receptors, e.g., cyclodextrin (CD), calixarene, cucurbit[*n*]uril and pillar[*n*]arene, the most commonly employed macrocycle is CD, due to their readily availability, low cost and high solubility.[1-4] Cyclodextrins (CDs) are a family of cyclic oligosaccharides widely used in pharmaceutical and chemical industries. The CD derivatives share the macrocyclic glucose backbone, which forms a hydrophobic cavity and a hydrophilic exterior. The unique structural feature makes the CD family nice candidates for drug carriers and reservoirs. Typically, bio-safe CD species have 6-8 glucose subunits, including α-CD (6 glucose units), β-CD (7 glucose units), and γ-CD (8 glucose units). As the portal size of β-CD is compatible with most pharmacologically active molecules, this member is often recognized as a promising host partner suitable for a number of drug-like molecules.[5-8] Molecular encapsulation of drugs in CD hosts can overcome various difficulties in the formulation of drug delivery systems, e.g., poor aqueous solubility of drug, poor bioavailability, and sensitivity to destructing factors.[9, 10] Sometimes the prototype CD (e.g., β-CD) does not fulfill the intended use. Thus, chemically modified CD molecules are often considered in the search of suitable hosts. On this aspect, computational modelling serves as an efficient screening tool due to its ability to cover a wide range of chemical space at a relatively low cost.

Compared with macromolecular species such as proteins of pharmaceutical interests, host molecules are more than an order of magnitude smaller in size (i.e., mass and volume). Further, their macrocyclic or polycyclic structures have only a small number of rotatable bonds, leading to relatively rigid dynamic features. As a result, in computational investigations of host-guest coordinations, there is a tendency to simplify the modelling procedure by neglecting conformational flexibility and considering only a single binding mode.[11-18] As a result, the computational inspection of host-guest binding is simplified to a single-state calculation that can be treated via single-point calculations with normal mode approximation or restrained sampling in the neighborhood of the initial bound configuration. However, recent modelling reports with comprehensive conformational sampling suggest the pivotal role of the conformational flexibility of both components involved in host-guest coordination and a general multi-modal binding behavior in host-guest complexes,[19-23] which suggests that the single-state approximation, regardless of the accuracy level of the employed Hamiltonian, could introduce systematic biases of unknown magnitude. This observation emphasizes the necessity of extensive exploration of the configurational space in host-guest modelling.

Correct modelling of host-guest systems requires accurate Hamiltonian and appropriate sampling technique.[24-29] In molecular simulations of β-CD-guest binding, it is common to employ general-purpose force fields such as the general AMBER force field (GAFF)[30] derivatives to treat bonded and vdW interactions,[31-33] while for electrostatics common selections are AM1-BCC[34, 35] and restrained electrostatic potential (RESP)[36]. However, assessments of these force fields for β-CD systems are mostly focused on binding thermodynamics, while detailed evaluations in other aspects (e.g., charge quality) are seldom performed. There is an AMBER-compatible force field parameterized for CD derivatives named q4MD,[37] which is also widely used in molecular simulations of CD host-guest systems. As for the sampling issue, it is well-known that the time-scale issue is a long-standing problem in molecular simulation of biomolecular and also host-guest systems. To overcome the significant gap between the integration time step of the equations of motion and the time scale of the processes of interest (e.g., conformational fluctuation and binding/unbinding events), enhanced sampling techniques are needed to improve the sampling efficiency and secure converged results in a reasonable simulation time. The multi-modal binding behavior in the β-CD-guest binding makes the situation more sophisticated.[38] According to the chemical structure of the β-CD host shown in Fig. 1, the top and bottom faces are asymmetric. The top surface with the hydroxymethyl rim is often called the primary surface, while the bottom surface with hydroxyl substitutions is the secondary surface. This asymmetric structural feature of the β-CD ring leads to two possible orientations for the guest molecule. Namely, the guest could directionally bind to the primary or secondary face of the host ring.[38] Computational perspective of β-CD host-guest binding thus concerns not just the correct reproduction of the experimental binding thermodynamics, but more importantly the correct description of this directionalized guest coordination (more specifically the primary-secondary equilibrium). Despite the existence of many machine learning tools trained to reproduce binding affinities,[39] these experience-based techniques would fail for newly encountered CD derivatives due to the absence of the training data or the dissimilarity of the training set and the systems under investigation. Further, simply getting the numbers (affinities) close to the experiment does not guarantee a correct description of β-CD host-guest binding but could involve fortuitous error cancellation, while the simultaneous predictions of binding thermodynamics and directionalized host-guest binding patterns would make the picture solid.

A recent grand challenge of computational modelling in drug discovery, SAMPL7,[40] involves a set of β-CD host-guest pairs with a variety of structural and chemical features. Aside from the prototype β-CD, the other host molecules are β-CD derivatives with chemical modifications on either the primary or the secondary surface. These modified β-CD rings can still stably coordinate guest molecules, but various physiochemical properties including the binding affinity and the energetic preference of the directional guest coordination are

altered. In this work, using this interesting β-CD host-guest dataset, we present a comprehensive modelling of β-CD host-guest binding, with a detailed evaluation of force-field accuracy and a correct treatment of the primary-secondary equilibrium. The obtained results would provide useful insights for both force-field development and practical considerations/guideline for molecular modelling of complex host-guest binding.

2. Methodology and Computational Details

2.1. System Preparation.

The 3D structures of hosts (i.e., β-CD derivatives) and the two guests are obtained from the online server of the grand challenge.[40] In Fig. 1, we present the 2D chemical structures of the host and guest molecules considered in this work. The host molecules include the prototype β-CD and eight modified β-CD derivatives. The backbone of β-CD remains untouched and chemical modifications are added on either the primary or the secondary surface. For the drug-like guests, we only consider two molecules with different net charges. The size of the guests is small, and exhaustive sampling of relevant conformational basins would be rather fast. Also, there is only one heteroatom in each guest molecule, which can be used to mark the primary or secondary coordination preference. The host-guest complexes formed by all hosts and the guest G1 are investigated, while for the host-G2 complex we only simulate the β-CD-G2 pair.

The force field parameters are obtained from GAFF derivatives[30] and system-specific charges derived from two schemes. The first charge scheme is AM1-BCC, which performs AM1 optimization in vacuo and then derives atomic charges by combining AM1 Mulliken charges and a correction term (i.e., BCC) trained from the HF/6-31G* electrostatic potential (ESP) for a large set of drug-like molecules. The second scheme is the RESP[36] fitting, which performs B3LYP/6-31G* optimization and HF/6-31G* ESP scan with the Merz-Kollman scheme in vacuo and then directly fits the atom-centered charges to reproduce the HF/6-31G* ESP in a two-step fashion with hyperbolic regularizations to avoid overfitting. The semi-empirical AM1 calculations are performed with the sqm engine in AMBER,[41] while for the calculations at *ab initio* levels we use the Gaussian 09 program.[42] Although in many cases the behaviors (e.g., binding thermodynamics and interaction patterns) of the two charge schemes are similar, there are also reports that they behave differently.[34, 35, 43-47] Thus, in the current β-CD host-guest modelling, we first assess their charge qualities before choosing one in extensive configurational sampling. The GAFF derivatives are used to obtain bonded and vdW parameters. Due to the high transferability of GAFF derivatives, their accuracies for specific systems could be relatively low compared with system-specified force fields. Therefore, we assess two GAFF versions including GAFF v1.81 (or simply GAFF) and the second generation of GAFF (GAFF

version 2.11, GAFF2), after which the best-performing one would be chosen in molecular simulation. Further, we consider reparametrizing GAFF2 in a system-specific manner with the generalized force-matching (FM) scheme[48, 49] to achieve a higher level of accuracy. Detailed discussion about the refitting procedure would be presented in the results part. Note that in force-field refitting and assessment, we only consider the host-guest pair of the prototypical β-CD and the guest G1.

As all host molecules share the same β-CD backbone and their difference lies in the substitution group, the β-CD backbone is used to define the binding pocket formed by hosts in all host-guest pairs. The starting structure of each host-guest complex is thus obtained by simply superposing the center of masses (COM) of the host backbone and that of the guest. The host-guest complex is solvated in TIP3P[50, 51] water and the minimum distance between the solute surface and the edge of the simulated box is set to 25 Å, which is sufficiently large to define a decoupled state. The truncated octahedron cell is replicated in whole space with periodic boundary conditions. Non-polarizable monovalent spherical counter ions[52, 53] of Na^+ or Cl^- parameterized for TIP3P water are added for neutralization.

2.2. Free Energy Simulations.

The interaction patterns between β-CD derivatives and guests are expected to be complex due to the asymmetric structural features of the hosts and the guests. As both guests considered in the current work have only one heteroatom, we use this heteroatom to illustrate the binding feature. The relative position of the heteroatom and the other parts of the guest inside the host cavity could determine at least two distinct binding poses, which feature the heteroatom at or closer to the primary or the secondary surface of the host. If the heteroatom tends to be closer to the primary surface than the secondary face in the bound state, then the host-guest binding prefers the primary orientation. If the primary and secondary binding poses are of similar thermodynamic stability, then the host-guest coordination shows an indifferent behavior. If the heteroatom is closer to the secondary surface than the primary one, then the host-guest coordination is secondary-preferred. To properly describe the thermodynamics in this primary-secondary equilibrium, exhaustive sampling of the binding/unbinding events is needed. Aside from the multi-modal binding behavior, as the binding affinities of host-guest complexes are much higher than the thermal fluctuation at the experimental condition (i.e., 298 K), some enhanced sampling techniques are needed to improve the sampling efficiency. Here, the protocol that has been applied to a series of similar host-guest systems in our previous works[19, 20, 54] is employed. The method employs well-tempered metadynamics[55-58] to accelerate the configurational exploration, and the biasing potential is imposed on the 3D spherical coordinates defined by the host and the guest. It should be

noted that the definition of the spherical coordinates (ρ,θ,φ) differs a little from previous host-guest modelling. In the previous works, the spherical coordinates are defined by the COMs of the two components involved in inter-molecular coordination (i.e., host and guest). As the host rings in the current work are chemically modified based on the prototype β-CD, the COM of the host backbone (i.e., without the substitution chain) is used to define the reference anchor point of the host. As for the guest part, the small size and the rigidity of the guest molecules enable the translational motion of the whole molecule to be described with one specific atom, which is selected as the only heteroatom in each guest (i.e., oxygen for G1 and nitrogen for G2). Therefore, we are actually biasing the host-heteroatom spherical coordinates in the current work. An illustration of such a biasing scheme is presented in Fig. S1. This sampling protocol enables a scan of the relative position of the host ring and the guest molecule and thus the exploration of possible binding poses. As the translational diffusion is accelerated in host-guest complexes, a series of binding/unbinding events would be observed, which makes the observation of guest binding to both the primary and secondary faces of the host possible. An upper wall is added on the first CV (i.e., distance/radius ρ) to limit the configurational space to explore. The maximum value of the radius ρ should be sufficiently large to define a fully decoupled state. In the current work, we put the upper wall on 22.5 Å, and the resulting entropic correction to recover the standard-state binding free energy is about 1.9 kcal/mol. For more detailed discussions about this spherical-coordinates-biased protocol in host-guest modelling, please refer to references.[19, 20, 54, 59]

The contact number between different groups of atoms $C = \sum_{i\in\text{group}_\text{A}} \sum_{j\in\text{group}_\text{B}} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^6}{1-\left(\frac{r_{ij}}{r_0}\right)^{12}}$ is used to analyze the detailed interactions inside the system. The variable r_{ij} is the distance between the *i*th atom in group A and the *j*th atom in group B, and the distance constant r_0 is set to 5 Å. The total contact number between the host and the guest $C_{\text{host-guest}} = \sum_{i\in\text{host}} \sum_{j\in\text{guest}} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^6}{1-\left(\frac{r_{ij}}{r_0}\right)^{12}}$ and its by-host-atom decomposition $C_i = \sum_{j\in\text{guest}} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^6}{1-\left(\frac{r_{ij}}{r_0}\right)^{12}}$ are used to identify the binding/unbinding events and the detailed host-guest coordination patterns, e.g., the host atoms directly coordinating the guest. The two components involved in the definition of the spherical

coordinates, the host backbone and the heteroatom of the guest, are also used to define the contact variable named the host-heteroatom contact $C_{\text{host-hetero}} = \sum_{i\in\text{host}} \sum_{j=\text{hetero}} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^6}{1-\left(\frac{r_{ij}}{r_0}\right)^{12}}$. In the analysis of the binding preference (i.e., binding at the primary or secondary surface of the host), the decomposed guest-primary-surface contacts are also analyzed, the details of which would be discussed later in the results part.

As mentioned in the previous system-preparation section, the starting configuration of each host-guest complex is obtained by simply superposing the COMs of the host backbone and the guest. As we are scanning the spherical coordinates covering the space of possible host-guest coordination patterns, this initial configuration would not introduce systematic bias to the simulation outcome. From this structure, we perform 10,000 steps energy minimization, 400 ps NVT equilibration, 2 ns NPT equilibration to obtain an equilibrated structure, from which the 1000 ns enhanced sampling simulation starts. The parameters for the metadynamics settings include 0.24 kcal/mol initial Gaussian height, the deposition interval of 0.5 ps, the bias factor of 10, and the Gaussian widths of 0.1 nm, $\frac{\pi}{16}$, and $\frac{\pi}{8}$ for the three polar coordinates, respectively. For temperature regulation we use the velocity rescaling algorithm[60] and for pressure regulation the Parrinello-Rahman barostat[61, 62] is used. A time step of 1 fs is used to propagate the dynamics. Long-range electrostatics are treated with the smooth Particle-mesh Ewald[63] method. GROMACS 2019.6 [64] patched with PLUMED 2.7.0[65] is used to perform simulations.

3. Result and Discussion.

3.1. Charge Generation.

Before computationally intense enhanced sampling simulations, we first evaluate the two charge schemes for solutes, which could be pivotal in determining the conformational preference and the inter-molecular coordination pattern in host-guest binding. As accurate reproduction of the *ab initio* electrostatics is achieved by accurate reproduction of the reference ESP (rather than the detailed values of atomic charges), we use the relative root-mean-squared error (RRMSE) of molecular ESP to assess the charge quality. The ESP RRMSE results for all solute molecules are presented in Fig. 2a, where the superiority of the RESP charge scheme is rather clear. For all solute molecules, the ESP RRMSE of the RESP scheme is smaller than AM1-BCC. For several systems such as the prototype β-CD, the ESP RRMSE under AM1-BCC is unacceptably large (~29%). This underperformance of AM1-BCC suggests the unsuitability of this

corrected semi-empirical charge scheme in the current β-CD host-guest set, and the inclusion of similar macrocyclic species in the training set of the correction term (BCC) could improve this situation. A similar phenomenon is also observed in our previous investigation of Cucurbit[8]uril host-guest binding.[19] Thus, the unsuitability of AM1-BCC in host-guest systems could be rather general, and care should be taken when employing this corrected semi-empirical charge scheme. This underperformance of the AM1-BCC charge scheme for β-CD derivatives could also be one of the reasons causing the unsatisfactory accuracy of computational results reported in many previous publications.[23, 31, 33, 66]

Although the ESP RRMSE could be informative to evaluate the consistency of the charge-produced ESP data and the *ab initio* reference, another interesting observable to check is the dipole moment, despite the fact that it is not included in the regularized charge fitting. The numerical results are presented in Fig. 2b, where we still observe a better performance of the RESP charge scheme. Interestingly, despite the huge difference between the ESP RRMSEs of RESP and AM1-BCC for the prototype β-CD, the dipole moments produced by the two charge schemes are extremely similar. By contrast, the difference between ESP RRMSEs of the two charge schemes is smaller for modified rings such as M34, but the dipole moments under the two charge schemes show a difference obviously larger than the β-CD case. Thus, the two quality-check metrics of ESP RRMSE and dipole moment are reflecting somehow different electrostatic properties, although in most cases their charge-scheme-dependent trends are similar. In Fig. 2c, we also present the number of ESP points used in charge fitting and assessment. We can see that a large number of ESP points are involved in our calculation, which ensures the coverage of relevant molecular surfaces and thus the reliability of our results.

The above statistical assessment provides the general performance of the charge schemes. However, the detailed reason for their unusual behaviors remains unclear. Although the RESP charge scheme reproduces the *ab initio* ESP data in a more accurate manner than AM1-BCC, in several systems still obvious deviations could be identified. These deviations, according to detailed analyses (visualization) of the ESP surface, result from the intrinsic limitation of fixed-charge representations. Namely, the atom-centered charges cannot accurately reproduce the anisotropic ESP distribution caused by, e.g., lone electron pairs. We use the G1 guest as an illustrative example. Structurally, this guest molecule is quite simple. However, both charge schemes fail to reproduce the high-level ESP data with large deviations (~20%). The main reason for the fitting failure is the lone electron pairs at its heteroatom oxygen and the resulting anisotropic ESP distribution. An illustration of the fitting scheme for this molecule is presented in Fig. 2d. The carbon atoms of the 6-membered ring are colored in cyan and magenta. Atoms sharing the same color are constrained to

have the same parameter in RESP fitting, the underlying physical reason for which is the chemical equivalence of these atoms due to the flexible rotation in the -OH region. This equivalent-carbon constraint leads to an elevated ESP RRMSE of 19%. If we try to better the fitting quality by switching off this constraint, new degrees of freedom are added in the regularized parameter adjustment and the ESP RRMSE is improved significantly to 9%. However, this treatment overfits a single configuration, neglects the equivalence of chemically equivalent carbon atoms and hinders the flexible orientation in the -OH region. To avoid overfitting a single configuration and comply with the most widely applied RESP fitting scheme, we turn on the equivalence constraints in our modelling. Note that similar fitting problems exist in many compounds. For example, for methanol CH3OH, the chemical environments of the three hydrogen atoms linked to carbon are equivalent, which suggests that the three hydrogen atoms should share the same charge parameter. However, this behavior cannot be achieved via (R)ESP fitting with any low-energy structure of this molecule. A constraint on this chemical equivalence is necessary and is considered a standard treatment in modern charge fitting, but this term would obviously regularize the parameter space and worsen the ESP reproduction.

Overall, the above detailed evaluation of the two charge schemes for solutes suggests the superiority of the RESP charge scheme, which would thus be employed in the extensive exploration of the configurational space.

3.2. Bonded Parameters.

Inter-molecular interactions are mainly contributed by electrostatics, which have been considered in the previous charge-generation section. We then turn to the other terms in classical force fields. The vdW term contributes directly to the non-bonded interactions, but its magnitude is often not as significant as electrostatic interactions. Further, this term is often less system-dependent. Thus, vdW parameters are often grabbed from pre-fitted parameter sets such as GAFF derivatives. As for the remaining parameters in fixed-charge force fields, i.e., the bonded interactions including the bond stretching, angle bending, and dihedral flipping potentials, although they are not directly involved in the calculation of inter-molecular interactions, they do play a crucial role in molecular recognition by influencing the intra-molecular conformational preference. Therefore, in this section, we focus on the comparison of the bonded terms from GAFF derivatives and recalibrate these pre-fitted parameters to produce system-specific force fields that reproduce *ab initio* properties. In this force-field evaluation and refitting, we only consider the β-CD-G1 host-guest pair due to the similarities of chemical environments in all host molecules. In the following paragraphs of

this section, we first describe the refitting procedure and then compare the behaviors of all force fields.

We employ the generalized FM scheme[49] to refit force field parameters following the GAFF2 definition. From this force field, we refit all the bonded terms with a weak harmonic (L2) regularization term to avoid overfitting. The non-bonded terms (RESP charges and vdW parameters) and the scaling factors for 1-4 non-bonded interactions are untouched. Aside from the regularization term, the objective function further includes the atomic force and the system energy. The atom-type symmetry defined in GAFF2 is applied, and the dihedral periodicity is fixed. The configurational ensemble used to generate reference data is obtained by 110 ns gas-phase unbiased sampling with a sampling interval of 10 ps. A high temperature of 600 K is employed to enhance the sampling efficiency. The obtained 11000 structures are then fed to ORCA[67, 68] for *ab initio* calculations. Due to the large sample size and the unconstrained exploration of configurational space at the high temperature, this reference set covers relevant regions and further adding samples leads to negligible changes of the refitting results. The target level used for the large host molecule is BLYP[69-72]/def2-SVP with the latest D4 dispersion correction[73, 74] considering its good performance in modelling various properties (including non-covalent interactions) among pure functionals, while for the small guest molecules a higher-level method, the r^{2}SCAN-3c[75, 76] composite scheme, is selected. The resolution-of-the-identity approximation[77, 78] is always employed in *ab initio* calculations.

For quality assessment, we generate a 10 ns trajectory (1000 configurations) at 300 K for each molecule. These newly generated configurations are independent of the training set (11000 structures mentioned in the previous paragraph). With this testing set, we compute the deviations of both system energy and atomic force under the pre-fitted GAFF2 and the refitted FM parameter sets. Two error metrics including the root-mean-squared error (RMSE) and the mean absolute error (MAE) are used for quality assessment. The correlograms for energetics are presented in Fig. 3a (β-CD) and Fig. 3b (G1). For the host β-CD, the deviation of GAFF2 energetics is very large (RMSE ~7.2 kcal/mol and MAE ~5.7 kcal/mol), which suggests the inappropriateness of direct application of pre-fitted general-purpose force fields in β-CD derivatives. By contrast, the RMSE energy produced by GAFF2 is quite small for the guest G1, which suggests the suitability of transferable force fields in this situation. Upon force field refitting, improvements are observed for both the macrocyclic host β-CD and the drug-like guest G1, but the magnitudes of betterments show obvious differences. For the host β-CD, the regularized force-field refitting leads to pronounced improvements (~1.6 kcal/mol for RMSE and ~1.2 kcal/mol for MAE), while the RMSE improvement for the small guest molecule is only ~0.4 kcal/mol. The RMSE for the guest G1 is already small enough, but the deviation of the force-field energetics from the *ab initio* reference is still very large for the host β-CD. If the FM parameter set indeed reaches the optimal position

in the parameter space for each system, then the accuracy level of the FM parameter set is the limit of the employed functional form (i.e., AMBER-like force fields). If further improvements are pursued, then more complex forms of force field formulations should be considered.

We then check the force error in Fig. 3c-f. The time series of atom-specified force errors ($\|\Delta\mathbf{F}_i\|_2$ for the *i*th atom) for the β-CD host are shown in Fig. 3c-d. The repeating glucose units are numbered sequentially. The first 11 atoms of each repeating unit are heavy atoms, while the other (12-21) atoms are hydrogen atoms. By comparing the GAFF2 (Fig. 3c) and FM-BLYP-D4 (Fig. 3d) results, we know that the FM refitting effectively improves the force reproduction for all atoms, but the force errors on non-hydrogen atoms are still relatively large. This observation is somehow consistent with our previous Cucurbit[8]uril work[79] and thus is expected to be general for macrocyclic hosts. By averaging the contributions from all atoms, the RMSE of atomic forces is obtained. Upon FM refitting, the force RMSE is improved from ~18.9 kcal/(mol·Å·atom) to ~12.8 kcal/(mol·Å·atom). The time series of force errors for the drug-like guest G1 are shown in Fig. 3e-f, where we can see that the GAFF2 parameter set is generally not problematic for this molecule, and FM refitting only leads to minor improvements. The force RMSE is improved from the GAFF2 ~10.8 kcal/(mol·Å·atom) to the FM-r^{2}SCAN-3c ~7.9 kcal/(mol·Å·atom).

From both the analyses of energy and atomic forces, we know that the refitted parameter set reproduces the *ab initio* results in a better way than the transferable GAFF2. However, these numerical data provide no hints about the dynamic behaviors of the macrocyclic host produced by these parameter sets. Therefore, we then perform 20 ns explicit-solvent simulations to investigate the dynamics of the β-CD host. The GAFF v1.81 (GAFF) force field mentioned in Section 2.1 is also employed in this unbiased simulation. We superpose the snapshots accumulated from unbiased simulations under the three parameter sets in Fig. 4. It is clearly shown that the GAFF parameters produce a flexible β-CD cavity that can be squashed frequently in explicit water, which might pose a problem for the guest to enter the host cavity and form stable host-guest coordination. By contrast, the GAFF2 and FM-BLYP-D4 host cavities are rather stiff. The host cavity remains open during the course of unbiased simulations. The radius of gyration (Rg) could be used to monitor the open-to-close transformation of the host cavity, the numerical results of which are presented also in Fig. 4. The GAFF2 values are quite stable during the whole 20 ns unbiased simulation and fluctuate around a large value of 6.15 Å. The GAFF Rg fluctuates around a smaller value ~5.8 Å, and the magnitude of fluctuation is much more significant than the GAFF2 case. Thus, we can conclude that the host cavity under GAFF is more flexible and squashed than GAFF2. The FM-BLYP-D4 curve is generally similar to GAFF2, but a transition between the fully open and partially closed conformations is observed at about 13 ns, after which the host cavity remains

partially closed. As this conformational transition is obviously a rare event, it is difficult to conclude the relative stiffness of GAFF2 and FM-BLYP-D4 parameter sets or whether there is an obvious difference between the dynamic behaviors produced by GAFF2 and FM-BLYP-D4. However, we can safely say that the refitted parameter set produces a host cavity at least stiffer than GAFF. Note that similar conformational fluctuations between the open and closed forms of the β-CD cavity have also been observed in previous studies.[80]

Overall, the assessment and refitting of bonded parameters presented above suggest the superiority of the FM parameter set in reproducing high-level results. However, the dynamic behaviors of the host cavity produced by the pre-fitted GAFF2 and the refitted FM parameter sets are somehow similar. Due to this consistency and the better reproduction of *ab initio* results, in the following simulations of β-CD host-guest systems, we employ both GAFF2 and FM parameters to investigate the impact of force-field refitting on the binding thermodynamics and interaction patterns. Note that as the FM refitting is performed only on the prototype host β-CD and the guest G1, only this host-guest pair is tested in simulations using the FM parameter set.

3.3. Binding Affinities.

The above detailed evaluation of force field parameters provides an insightful picture of force field selection. We then perform spherical-coordinates-biased simulations with RESP charges and GAFF2/FM parameters to obtain the thermodynamics of β-CD host-guest binding. In Fig. 5 and Fig. S2, we present the time series of the total host-guest contact and its decomposition into the contribution of each host atom-guest pair. For the total host-guest contacts, their values fluctuate significantly and reach the system-dependent maximum and zero many times, which suggests the observations of a number of binding/unbinding events during the course of our simulations. Thus, a large number of binding poses would be explored, making the simulation outcome independent of the initial binding pose. The informative atom-guest decompositions provide a detailed picture of host-guest coordination. We can know which host atoms are directly coordinating the guest and their closeness to the guest. Similar to the previous by-atom force error shown in Fig. 3c-d, the atom-guest contacts are also numbered sequentially. Among the 21 atom-guest contacts in each of the 7 repeating units, the first 11 are the heavy-atom-guest contacts and the other (12-21) are hydrogen-guest contacts. For the prototype β-CD-G1 binding shown in Fig. 5, many center-binding configurations with large atom-guest contacts for all host atoms could be identified, and many host-guest coordination patterns with side-binding behaviors could be observed. Similar observations could be obtained in the β-CD-G2 case shown

in Fig. S2. For modified rings, the substitution groups are numbered after the β-CD backbone atoms. Still, for the binding cases of these β-CD derivatives and G1, a large number of binding/unbinding events and host-guest coordination patterns could be observed, suggesting the exhaustiveness of our conformational search.

We then start with the investigation of binding thermodynamics by projecting the free energy surfaces on the host-guest COM distance and the host-heteroatom contact, the results of which are presented in Fig. 6 and Fig. S3. All free energy surfaces share similar features and thus we only discuss the systems shown in Fig. 6. Only one free energy basin featuring the center-binding modes is observed for each system. For the β-CD-G1 binding case shown in Fig. 6a-b, the GAFF2 and FM parameter sets produce extremely similar free energy surfaces. We extract the most stable bound conformation for each parameter set with marks (vdW representation) for the oxygen atoms on the primary face of the host and the heteroatom (oxygen) of the guest. We can see that the binding poses obtained under the two parameter sets are also very similar. When the guest is changed from G1 to G2, the details of the free energy surface vary a little, but the general single-minimum center-binding picture remains unchanged, as shown in Fig. 6c. The situation is similar when the host is varied from the prototype β-CD to chemically modified species (c.f., Fig. 6d and Fig. S3) and thus would not be discussed further.

With the projected radius-contact surfaces, we calculate the binding free energies of these host-guest complexes, the results of which are summarized in Table 1. The experimental values of binding affinities are obtained from reference.[38] We first take a look at the absolute values of the computed binding affinities. For the complex formed by β-CD derivatives and G1, the results obtained under the RESP+GAFF2 combination are generally overestimated (more negative) than the experimental values. As the sampling has been proven to be sufficient in our modelling, this deviation should arise from the inaccuracy of the employed force field, the details of which have already been discussed in previous sections. An interesting observation for the β-CD-G1 complex is that when replacing the pre-fitted GAFF2 parameter set with the molecule-specific refitted FM parameter set (i.e., FM-BLYP-D4 for β-CD and FM-r^{2}SCAN-3c for G1), the β-CD-G1 affinity agrees with the previous GAFF2 result within statistical uncertainty. This phenomenon is rather not unexpected, as the GAFF2 parameters are already very accurate for the guest G1 (c.f. Fig. 3b) and the dynamic behavior of β-CD produced by the refitted parameter set is similar to the pre-fitted GAFF2 (c.f., Fig. 4). The perfect agreement also suggests the well-converged behavior of our simulation. Namely, the simulations under GAFF2 and FM parameter sets can be considered two independent trials due to the similarity of their behaviors, and the outcomes of the two repeats are in perfect agreement. Similar to the β-CD-derivatives-G1 complexes, the β-CD-G2 binding affinity is also overestimated. Thus, we can conclude

that our modelling protocol tends to give overestimated β-CD host-guest binding affinity. This trend is also clearly presented in the correlogram shown in Fig. 7.

To obtain some quantitative insights of the quality into the computed binding thermodynamics, we compute two error metrics including RMSE and mean signed error (MSE) and two ranking coefficients including Kendall τ and prediction index (PI). The error metrics evaluate the deviations of the computed values from the experimental reference, while the ranking coefficients assess the consistency of the computed rank of binding affinities and the experimental rank. The values of these quality metrics are also provided in Table 1. The RMSE of the calculated binding affinities is 1.5 kcal/mol. This error size is intermediate in computational modelling of protein-ligand and host-guest complexes. As the sampling has been proven sufficient in our simulation, this deviation is expected to arise from the force field inaccuracy discussed in previous sections. The positive MSE suggests the systematic overestimation of binding affinities (stronger than experiment) with the modelling protocol RESP+GAFF2/FM. Similar observations (overestimation) have been reported in previous studies that employ similar force field combinations but sample the β-CD host-guest binding in the alchemical space with nonequilibrium transformation.[23, 66] The values of both ranking coefficients are positive but close to zero, which suggests that the simulation outcomes agree with the experimental values to some extent, but the correlation is not very well. It should be noted that the poor experiment-calculation rank correlation in the current β-CD host-guest systems is not unexpected, as the range of binding affinities is quite narrow and is not obviously larger than the statistical uncertainty. Specifically, in the 1.4 kcal/mol range of G1-involved affinities from -4.6 kcal/mol to -3.2 kcal/mol, we have 9 host-guest pairs under investigation (or 10 systems considering the FM parameter set), while the statistical uncertainty for binding affinities is ~0.4 kcal/mol. As a result, the differences between binding affinities are close in value and similar in size to the statistical errors of the calculations. The β-CD-G2 affinity is much stronger than the β-CD-derivatives-G1 ones, and this trend is correctly computed from our simulation. Also, considering the system-dependent inaccuracy of force field parameters, we cannot expect to obtain extremely good ranks of binding affinities in the current β-CD host-guest systems.

3.4. Primary or Secondary? The Directionalized Binding Mode Preference.

Compared with other macrocyclic hosts such as Cucurbit[8]uril studied in our previous work,[19] the carbohydrate β-CD is structurally asymmetric with two surfaces of different portal sizes. This property leads to an extra asymmetric directional binding preference of external agents. Therefore, complex interaction patterns are expected to be observed in β-CD host-guest binding. However, compared with the structurally

diverse abused drugs investigated in our previous works,[19] the guest molecules in the current β-CD case are rather simple. They (i.e., G1 and G2) are formed by an aliphatic cycle with heteroatom substituted at specific site(s), creating a simple asymmetric feature of the guest itself. As a result, the guest is unable to a form complex binding pattern with the β-CD host and the main interaction possibility in the current β-CD systems is the asymmetric directionalized binding behavior.

In the analysis of the asymmetric binding behavior of the guests (i.e., binding at primary and secondary surfaces), we need some CVs to properly describe the orientation of the asymmetric guests inside the asymmetric host cavity. To this aim, we select the primary surface of the β-CD derivatives and the orientation of the heteroatom of the guest. In each repeating unit of β-CD, there are 2 oxygen atoms in -OH groups on the secondary surface and only 1 on the primary surface. In some substituted rings such as M35, the -OH oxygen on the primary surface is substituted in one unit, while in others one oxygen on the secondary surface is substituted. To provide a consistent definition in all β-CD derivatives, we only include the consecutive 5 of 7 oxygen atoms on the primary surfaces in the definition of the primary surface of each host. The 5 oxygen atoms already cover the main interacting regions of the primary surface, and using other numbers such as 4 or 6 of the 7 oxygen atoms does not change the results significantly. Then, we select two anchor atoms in each guest to investigate the orientation of the asymmetric guest. The first anchor atom is the only heteroatom in each guest, which is biased directly in enhanced sampling simulations. The second anchor atom is a carbon atom on the other side of the guest. The orientation of the asymmetric guest relative to the primary surface is thus described by the contacts between the anchor atoms and the primary surfaces $C_{\text{hetero-primary}} = \sum_{i\in\text{primary oxygen}} \frac{1-\left(\frac{r_{i\text{-hetero}}}{r_0}\right)^6}{1-\left(\frac{r_{i\text{-hetero}}}{r_0}\right)^{12}}$ and $C_{\text{carbon-primary}} = \sum_{i\in\text{primary oxygen}} \frac{1-\left(\frac{r_{i\text{-carbon}}}{r_0}\right)^6}{1-\left(\frac{r_{i\text{-carbon}}}{r_0}\right)^{12}}$. The free energy surfaces projects on these two new CVs are presented in Fig. 8 and Fig. S4. When the minimum with large $C_{\text{hetero-primary}}$ is more stable than that with large $C_{\text{carbon-primary}}$, the heteroatom is closer to the primary surface and thus the primary binding mode is preferred. When the two minima are of similar stabilities, the system is considered binding-mode indifferent. When the large-$C_{\text{carbon-primary}}$ minimum is more stable than the large-$C_{\text{hetero-primary}}$ one, the system prefers the secondary binding mode. The directionalized binding-mode preference for all host-guest pairs can be directly identified from these $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surfaces. For instance, the β-CD-G1 and β-CD-G2 complexes prefer the secondary binding mode, while the MG8-G1 and M36-G1

complexes exhibit a binding-mode indifferent behavior. An interesting comparison is performed still between the GAFF2 β-CD-G1 complex in Fig. 8 and the FM one in Fig. S4. Despite the differences between the bonded parameters in these two force fields, the $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surfaces under them are extremely similar. Also considering the binding-affinity accordance shown in Table 1 and the similarity of the $\rho - C_{\text{host-hetero}}$ surfaces in Fig. 6, we can indeed conclude that the (thermo-)dynamic behaviors produced by these two parameter sets are very similar and the transferable GAFF2 parameters are already very good for β-CD host-guest systems.

The predicted binding-mode preferences of all host-guest complexes under investigation are summarized along with the experimental reference[38] in Fig. 9. Among β-CD-derivatives-G1 complexes, the binding-mode preferences of only three β-CD-derivatives-G1 pairs are correctly predicted, possibly due to the force-field issue (electrostatics) probed in previous analyses (c.f. Fig. 2). For the other β-CD-derivatives-G1 pairs, the simulation-derived binding-pose preference is generally more secondary-mode preferred. Specifically, for systems with the indifferent experimental preference, the calculated binding-pose preference tends to be secondary or indifferent, and the experimentally primary-preferred systems tend to be predicted indifferent or secondary-preferred. By contrast, for the G2 guest the force-field problem is relatively small (c.f., Fig. 2a-b), and the simulation can correctly capture the secondary binding-pose preference of the β-CD-G2 complex and also reproduce the binding affinity shown in Table 1. Overall, it is difficult to correctly capture the primary-secondary preference in the current β-CD host-guest binding due to the force-field inaccuracy, and improvements could be expected upon the use of more accurate models. Aside from the problems of electrostatics, another possible reason for the deviations is the small free energy difference between primary and secondary binding modes (e.g., ~ 0.5 kcal/mol for β-CD-G1 shown in Fig. 8), which is similar to the statistical uncertainty (~0.4 kcal/mol) reported in Table 1.

4. Concluding Remarks.

Macrocyclic hosts as molecular containers are currently applied in various areas, e.g., catalyzed synthesis, molecular machinery, drug delivery, fluorescence sensing and molecular recognition. The CD host family is commonly employed due to their readily availability, low cost and high solubility. In computational modelling, the β-CD host-guest binding is often recognized as a relatively easy/simple case due to its carbohydrate composition and similarity to drug-like molecules. Often, a single host-guest coordination pattern is considered and the calculations are based on a single bound conformation or limited

sampling in the neighborhood of the starting configuration. However, the situation is actually complicated rather than simplified due to the asymmetric feature of the β-CD host. The external agent (guest) could bind to its primary 6’ or secondary 3’ face, and the unbiased modelling of the β-CD host-guest coordination requires proper treatment/averaging of the primary-secondary equilibrium. Otherwise, systematic biases of unknown magnitude would be introduced, and the computational insights would be unreliable.

Correct modelling of host-guest binding requires accurate Hamiltonians and appropriate techniques for configurational space exploration. The first term determines the ultimate accuracy obtained with infinite sampling, while the latter determines the computational cost or the time scale of modelling. A balanced treatment of the two issues is always the predominant issue in computational modelling. Due to the time scale of host-guest binding, μs-length sampling is often required even with accelerated sampling techniques. At this time scale, the model/description affordable for most researchers is all-atom fixed-charge model, which involves intra- and inter-molecular potentials. The former intra-molecular terms determine the conformational preference of each molecule in the absence of the system-environment coupling, while the latter non-bonded terms describe the strength and preference of inter-molecular interactions.

The non-bonded terms directly contribute to inter-molecular interactions, and the electrostatics part is often of higher importance than the vdW component. Accurate description of inter-molecular electrostatic interactions is often achieved by accurate reproduction of the Coulombic ESP around each molecule. In the current β-CD host-guest systems, the RESP charge scheme performs better than AM1-BCC in ESP reproduction. Similar observations are also reported in another heterocyclic host Cucurbit[8]uril, which suggests that this RESP-better-than-BCC phenomenon is quite general for host-guest systems. However, for both charge schemes, there are cases where ESP RRMSE is not satisfactorily small, mainly due to the anisotropic ESP distribution caused by lone electron pairs. A worth noting failure is the structurally simple guest G1. The lone electron pairs on its oxygen atom pose a problem, especially in the presence of the chemical equivalence constraint. Multi-conformer averaging and more detailed models (e.g., off-center fitting points/charges and polarizable force fields) are needed to accurately describe the electrostatics. However, such treatment significantly increases the computational burden and comprehensive sampling would soon become very expensive. Under the fixed-charge representation, the RESP charge scheme is superior and thus is recommended in host-guest modelling.

Intuitively, the non-bonded terms are directly involved in the calculation of inter-molecular interactions and thus are more important than intra-molecular terms. As a result, many computational investigations of host-guest and protein-ligand complexes vary the charge schemes, while less attention is attached to the

bonded terms. However, these stiffer bonded terms determine the intra-molecular conformational preference, the inaccurate description of which could lead to an incorrect/unphysical ensemble to sample. The intra-molecular terms include bond stretching, angle bending, and tortional terms, the parameters of which are often obtained with carefully fitted transferable force fields such as GAFF derivatives. However, their accuracy in specific cases could be relatively low compared with parameter sets calibrated for specific systems. Therefore, we evaluate and refit all bonded parameters of the β-CD host and the guest G1 to assess the suitability of GAFF derivatives in β-CD host-guest modelling. The refitted force fields produce smaller deviations from the *ab initio* reference for both system energy and atomic force, especially for the macrocyclic host. Explicit-solvent ns-length unbiased simulations are also performed to check the dynamic behavior produced by different parameter sets. For the β-CD host, GAFF produces a squashed host cavity in explicit solvent, which makes it hard for external agents (i.e., guests) to form stable host-guest coordinations. By contrast, the GAFF2 parameters provide a much stiffer host cavity and the ring remains open in tens of ns simulations in water, which makes the translational diffusion of the guest around the host cavity feasible and the formation of host-guest complex favorable. The refitted parameter set behaves similarly to GAFF2. This similarity, along with the better reproduction of high-level energetics, suggests that GAFF2 could be a choice of higher reliability than GAFF. Thus, the GAFF2 parameter set is recommended in molecular simulations of β-CD host-guest systems.

The RESP+GAFF2 combination as the best force field choice is employed in enhanced sampling simulations that sample the binding/unbinding events extensively. From the thermodynamic information projected on the radius-contact surface, the binding affinities are calculated. Compared with the experimental reference, the computed values exhibit systematic overestimation, which arises from the inaccuracy of force field parameters especially electrostatics. The RMSE of binding affinities achieves ~1.5 kcal/mol, which is of intermediate size in modern computational investigations of protein-ligand and host-guest binding. The β-CD-G1 binding affinities obtained under the pre-fitted transferable GAFF2 and the refitted molecule-specific FM parameter sets agree within statistical uncertainty, which agrees with the similarity of host dynamics produced by the two parameter sets and indicates the applicability of GAFF2. The directionalized binding-mode preference is also extracted by projecting the thermodynamics on the heteroatom-primary-surface and carbon-primary-surface CVs. Compared with the experimental results, the binding-mode preferences of four host-guest pairs are correctly reproduced, while the calculation-predicted binding mode preference is generally more secondary-preferred for the other systems. Namely, the experimentally primary-preferred systems are predicted to be indifferent or secondary-preferred, while the

indifferent host-guest pairs are prone to be indifferent or secondary-preferred. These deviations should also attribute to the force-field issues discussed previously. Interestingly, the binding-pose preferences for the β-CD-G1 complex under GAFF2 and the refitted parameter set are the same. This agreement along with the binding-affinity accordance suggests the molecule-specific force-field refitting for β-CD derivatives is generally unnecessary and the pre-fitted GAFF2 parameters are already good enough.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 21633001) and Beijing Natural Science Foundation (Grant No. 7224357). Part of the simulation was performed on the high-performance computing platform of the Center for Life Science (Peking University). We thank anonymous reviewers for valuable comments and critical reading.

Conflict of Interest Statement

There are no conflicts of interest to declare.

Supporting Information Description

An illustration of the spherical CV, the time series of the host-guest contacts and the by-host-atom decomposition during enhanced sampling simulations, the 2D $\rho - C_{\text{host-hetero}}$ free energy surfaces for MG9-G1, M19-G1, M23-G1, M24-G1, M34-G1, M35-G1 and M36-G1 complexes under GAFF2, the $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surfaces for MG9-G1, M19-G1, M23-G1, M24-G1, M34-G1, M35-G1 and M36-G1 pairs under GAFF2 and the $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surface for the β-CD-G1 complex under the refitted FM parameter set are given in the supporting information.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Jansook, P.; Ogawa, N.; Loftsson, T., Cyclodextrins: structure, physicochemical properties and pharmaceutical applications. *International journal of pharmaceutics* **2018**, 535, 272-284.
2. Carvalho, R. M.; Rosa, I. G.; Gomes, D. E.; Goliatt, P. V.; Goliatt, L., Gaussian processes regression for cyclodextrin host-guest binding prediction. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2021**, 101, 149-159.
3. Gebhardt, J.; Kleist, C.; Jakobtorweihen, S.; Hansen, N., Validation and comparison of force fields for native cyclodextrins in aqueous solution. *J. Phys. Chem. B* **2018**, 122, 1608-1626.
4. Vyas, A.; Saraf, S.; Saraf, S., Cyclodextrin based novel drug delivery systems. *Journal of inclusion phenomena and macrocyclic chemistry* **2008**, 62, 23-42.
5. Papezhuk, M. V.; Volynkin, V. A.; Stroganova, T. A.; Krapivin, G. D.; Usacheva, T. R.; Thi, L. P., Theoretical and experimental study of inclusion complex formation of β-cyclodextrin with some 1, 4-diazepine derivatives. *Macroheterocycles* **2020**, 13, 64-73.
6. Barbiric, D.; Castro, E.; De Rossi, R., A molecular mechanics study of 1: 1 complexes between azobenzene derivatives and β-cyclodextrin. *Journal of Molecular Structure: THEOCHEM* **2000**, 532, 171-181.
7. Paulino, P. H. S.; de Sousa, S. M. R.; Da Silva, H. C.; De Almeida, W. B.; Ferrari, J. L.; Guimaraes, L.; Nascimento Jr, C. S., A theoretical investigation on the encapsulation process of mepivacaine into β-cyclodextrin. *Chem. Phys. Lett.* **2020**, 740, 137060.
8. Chen, K.; Ye, R.; Liu, X.; Wong, C. F.; Xu, S.; Luo, J.; Gong, X.; Zhou, B., Why 2, 6-di-methyl-β-cyclodextrin can encapsulate OH-substituted naphthalenes better than β-cyclodextrin: Binding pose, non-covalent interaction and solvent effect. *Computational and Theoretical Chemistry* **2021**, 1206, 113496.
9. Otero-Espinar, F.; Torres-Labandeira, J.; Alvarez-Lorenzo, C.; Blanco-Méndez, J., Cyclodextrins in drug delivery systems. *J. Drug Deliv. Sci. Technol.* **2010**, 20, 289-301.
10. Rasheed, A., Cyclodextrins as drug carrier molecule: a review. *Scientia Pharmaceutica* **2008**, 76, 567-598.
11. Belica-Pacha, S.; Daśko, M.; Buko, V.; Zavodnik, I.; Miłowska, K.; Bryszewska, M., Thermodynamic Studies of Interactions between Sertraline Hydrochloride and Randomly Methylated β-Cyclodextrin Molecules Supported by Circular Dichroism Spectroscopy and Molecular Docking Results. *Int. J. Mol. Sci.* **2021**, 22, 12357.
12. Peerannawar, S. R.; Rao, S. S.; Gejji, S. P., Density functional investigations on 2-naphthalenecarbonitrile dimerization within cucurbit [8] uril cavitand. *J. Mol. Model.* **2014**, 20, 1-9.
13. Ahmadian, N.; Mehrnejad, F.; Amininasab, M., Molecular Insight into the Interaction between Camptothecin and Acyclic Cucurbit[4]urils as Efficient Nanocontainers in Comparison with Cucurbit[7]uril: Molecular Docking and Molecular Dynamics Simulation. *J. Chem. Inf. Model.* **2020**, 60, 1791-1803.
14. Ali, H. S.; Chakravorty, A.; Kalayan, J.; de Visser, S. P.; Henchman, R. H., Energy–entropy method using multiscale cell correlation to calculate binding free energies in the SAMPL8 host–guest challenge. *J. Comput.-Aided Mol. Des.* **2021**.
15. Grishaeva, T.; Masliy, A.; Bakovets, V.; Kuznetsov, A., Inclusion compound based on Bis (ethylenediamine) copper (II) complex and cucurbit [8] uril: Quantum chemical prediction for structure and formation thermodynamic parameters. *Russian Journal of Inorganic Chemistry* **2015**, 60, 1247-1252.
16. Guo, X.; Wang, Z.; Zuo, L.; Zhou, Z.; Guo, X.; Sun, T., Quantitative prediction of enantioseparation using β-cyclodextrin derivatives as chiral selectors in capillary electrophoresis. *Analyst* **2014**, 139, 6511-6519.
17. Assaba, I. M.; Rahali, S.; Belhocine, Y.; Allal, H., Inclusion complexation of chloroquine with α and β-cyclodextrin: Theoretical insights from the new B97-3c composite method. *Journal of Molecular Structure* **2021**, 1227, 129696.
18. Ferrero, R.; Pantaleone, S.; Delle Piane, M.; Caldera, F.; Corno, M.; Trotta, F.; Brunella, V., On the interactions of melatonin/β-cyclodextrin inclusion complex: A novel approach combining efficient semiempirical extended tight-binding (xTB) results with ab initio methods. *Molecules* **2021**, 26, 5881.
19. 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.
20. 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.

21. Boz, E.; Stein, M., Accurate Receptor-Ligand Binding Free Energies from Fast QM Conformational Chemical Space Sampling. *Int. J. Mol. Sci.* **2021**, 22, 3078.
22. 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**.
23. Procacci, P.; Guarnieri, G., SAMPL7 blind predictions using nonequilibrium alchemical approaches. *J. Comput.-Aided Mol. Des.* **2021**, 35, 37-47.
24. 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.
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.; 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**, 2100185.
27. 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.
28. Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019**, 21, 14923-14940.
29. Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R., Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput.-Aided Mol. Des.* **2013**, 27, 989-1007.
30. 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.
31. Slochower, D. R.; Henriksen, N. M.; Wang, L.-P.; Chodera, J. D.; Mobley, D. L.; Gilson, M. K., Binding Thermodynamics of Host–Guest Systems with SMIRNOFF99Frosst 1.0.5 from the Open Force Field Initiative. *J. Chem. Theory Comput.* **2019**, 15, 6225-6242.
32. Zhang, H.; Yin, C.; Yan, H.; van der Spoel, D., Evaluation of Generalized Born Models for Large Scale Affinity Prediction of Cyclodextrin Host–Guest Complexes. *J. Chem. Inf. Model.* **2016**, 56, 2080-2092.
33. Suárez, D.; Díaz, N., Affinity Calculations of Cyclodextrin Host–Guest Complexes: Assessment of Strengths and Weaknesses of End-Point Free Energy Methods. *J. Chem. Inf. Model.* **2019**, 59, 421-440.
34. 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, 1623-1641.
35. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, 21, 132-146.
36. 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.
37. Cézard, C.; Trivelli, X.; Aubry, F.; Djedaïni-Pilard, F.; Dupradeau, F.-Y., Molecular dynamics studies of native and substituted cyclodextrins in different media: 1. Charge derivation and force field performances. *Phys. Chem. Chem. Phys.* **2011**, 13, 15103-15121.
38. Kellett, K.; Slochower, D. R.; Schauperl, M.; Duggan, B. M.; Gilson, M. K., Experimental characterization of the association of β-cyclodextrin and eight novel cyclodextrin derivatives with two guest compounds. *J. Comput.-Aided Mol. Des.* **2020**.
39. Mizera, M.; Muratov, E. N.; Alves, V. M.; Tropsha, A.; Cielecka-Piontek, J., Computer-aided discovery of new solubility-enhancing drug delivery system. *Biomolecules* **2020**, 10, 913.
40. https://github.com/samplchallenges/SAMPL7/tree/master/host_guest/cyclodextrin_derivatives.
41. 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.
42. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., GAUSSIAN09, Gaussian, Inc., Wallingford, CT, USA,(2009). *Google Scholar* **2016**.
43. He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J., A fast and high-quality charge model for the next generation

general AMBER force field. *J. Chem. Phys.* **2020**, 153, 114502.
44. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A., Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, 111, 2242-2254.
45. Huai, Z.; Yang, H.; Li, X.; Sun, Z., SAMPL7 TrimerTrip host-guest binding affinities from extensive alchemical and end-point free energy calculations. *J. Comput.-Aided Mol. Des.* **2021**, 35, 117-129.
46. 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**.
47. 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.
48. Ercolessi, F.; Adams, J. B., Interatomic potentials from first-principles calculations: the force-matching method. *EPL (Europhysics Letters)* **1994**, 26, 583.
49. Morado, J.; Mortenson, P. N.; Verdonk, M. L.; Ward, R. A.; Essex, J. W.; Skylaris, C.-K., ParaMol: A Package for Automatic Parameterization of Molecular Mechanics Force Fields. *J. Chem. Inf. Model.* **2021**, 61, 2026-2047.
50. 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.
51. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, 121, 10096-10103.
52. 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.
53. 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.
54. Sun, Z., SAMPL7 TrimerTrip Host-Guest Binding Poses and Binding Affinities from Spherical-Coordinates-Biased Simulations. *J. Comput.-Aided Mol. Des.* **2021**, 35, 105-115.
55. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011**, 1, 826-843.
56. Valsson, O.; Tiwary, P.; Parrinello, M., Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annual Review of Physical Chemistry* **2016**, 67, 159.
57. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters* **2008**, 100, 020603.
58. Tiwary, P.; Parrinello, M., A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015**, 119, 736-42.
59. Capelli, R.; Carloni, P.; Parrinello, M., Exhaustive Search of Ligand Binding Pathways via Volume-based Metadynamics. *J. Phys. Chem. Lett.* **2019**, 10, 3495-3499.
60. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, 126, 2384.
61. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983**, 50, 1055-1076.
62. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, 52, 7182-7190.
63. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, 103, 8577-8593.
64. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1, 19-25.
65. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, 185, 604-613.
66. Khalak, Y.; Tresadern, G.; de Groot, B. L.; Gapsys, V., Non-equilibrium approach for binding free energies in cyclodextrins in SAMPL7: force fields and software. *J. Comput.-Aided Mol. Des.* **2021**, 35, 49-61.
67. Neese, F., The ORCA program system. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2012**, 2, 73-78.

68. Neese, F., Software update: the ORCA program system, version 4.0. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2018**, 8, e1327.
69. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, 38, 3098.
70. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, 37, 785.
71. Weigend, F., Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, 8, 1057-1065.
72. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, 7, 3297-3305.
73. Caldeweyher, E.; Bannwarth, C.; Grimme, S., Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, 147, 034112.
74. Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S., A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, 150, 154122.
75. Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M., r2SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, 154, 064103.
76. Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J., Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **2020**, 11, 8208-8215.
77. Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R., Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, 97, 119-124.
78. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R., Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, 240, 283-290.
79. 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**.
80. He, P.; Sarkar, S.; Gallicchio, E.; Kurtzman, T.; Wickstrom, L., Role of Displacing Confined Solvent in the Conformational Equilibrium of β-Cyclodextrin. *J. Phys. Chem. B* **2019**, 123, 8378-8386.

Table 1. The host-guest binding affinities in kcal/mol obtained under RESP charges combined with the GAFF2 or FM parameter set. ΔG_{exp} and ΔG_{calc} denote the experimental and calculated binding free energies, respectively. SD denotes the standard error of the free energy estimate obtained from block analysis. MSE, RMSE, τ, and PI serve as quality measurements. For β-CD-derivatives-G1 complexes, the binding affinities are generally overestimated (more negative) than experimental references, which should arise from the force-field problem analyzed in previous sections. This systematic overestimation is successfully captured by a large MSE (~1 kcal/mol). The binding affinities of the β-CD-G1 complex under the pre-fitted transferable GAFF2 and the refitted molecule-specific FM parameter sets agree within statistical uncertainty, which suggests that the bonded terms in GAFF2 are generally good enough for β-CD derivatives.

Host-Guest Pair	ΔG_{exp}	ΔG_{calc}	SD
β-CD G1	-4.5	-5.5	0.3
MG8 G1	-3.3	-4.3	0.3
MG9 G1	-3.2	-5.8	0.3
M19 G1	-3.2	-5.2	0.3
M23 G1	-3.3	-4.3	0.4
M24 G1	-3.3	-4.4	0.3
M34 G1	-4.6	-2.8	0.4
M35 G1	-3.8	-5.0	0.3
M36 G1	-3.3	-5.4	0.4
β-CD G2	-6.3	-7.1	0.3
FM β-CD G1	-4.5	-5.8	0.3
RMSE		1.5	
MSE		1.1	
τ		0.1	
PI		0.1	

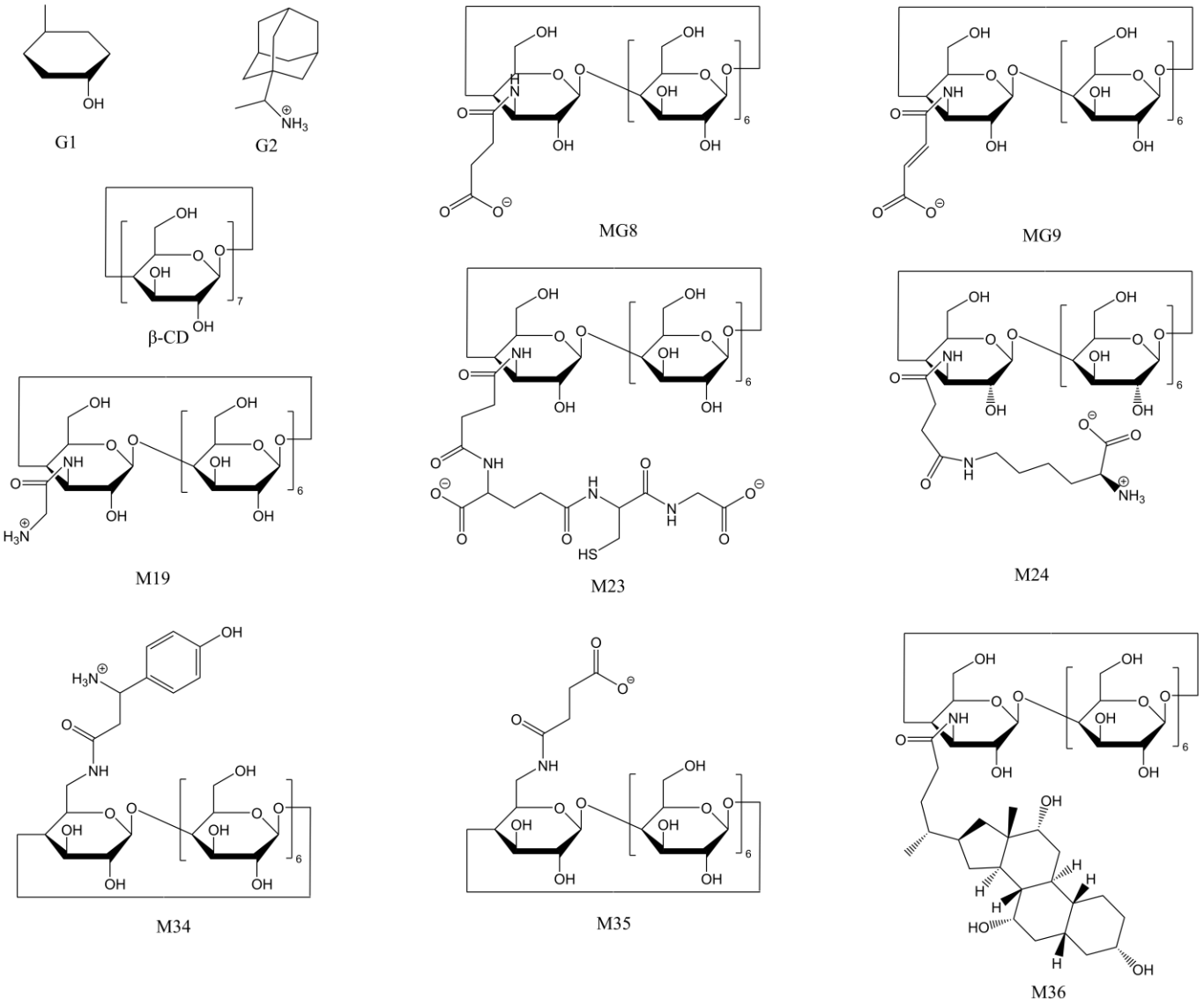


Fig. 1. The 2D chemical structures of the prototype β-CD, its derivatives and the two guest molecules.

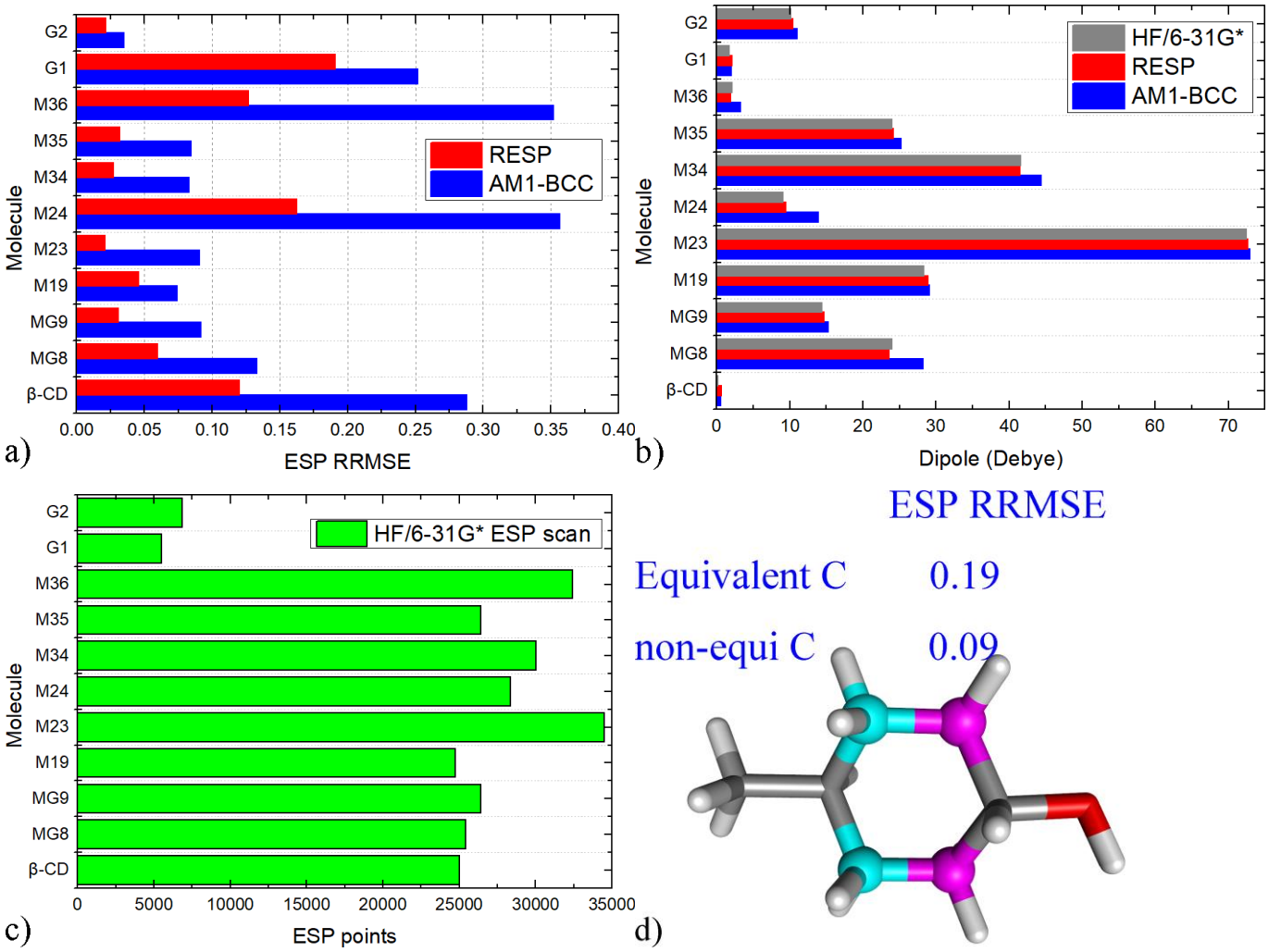
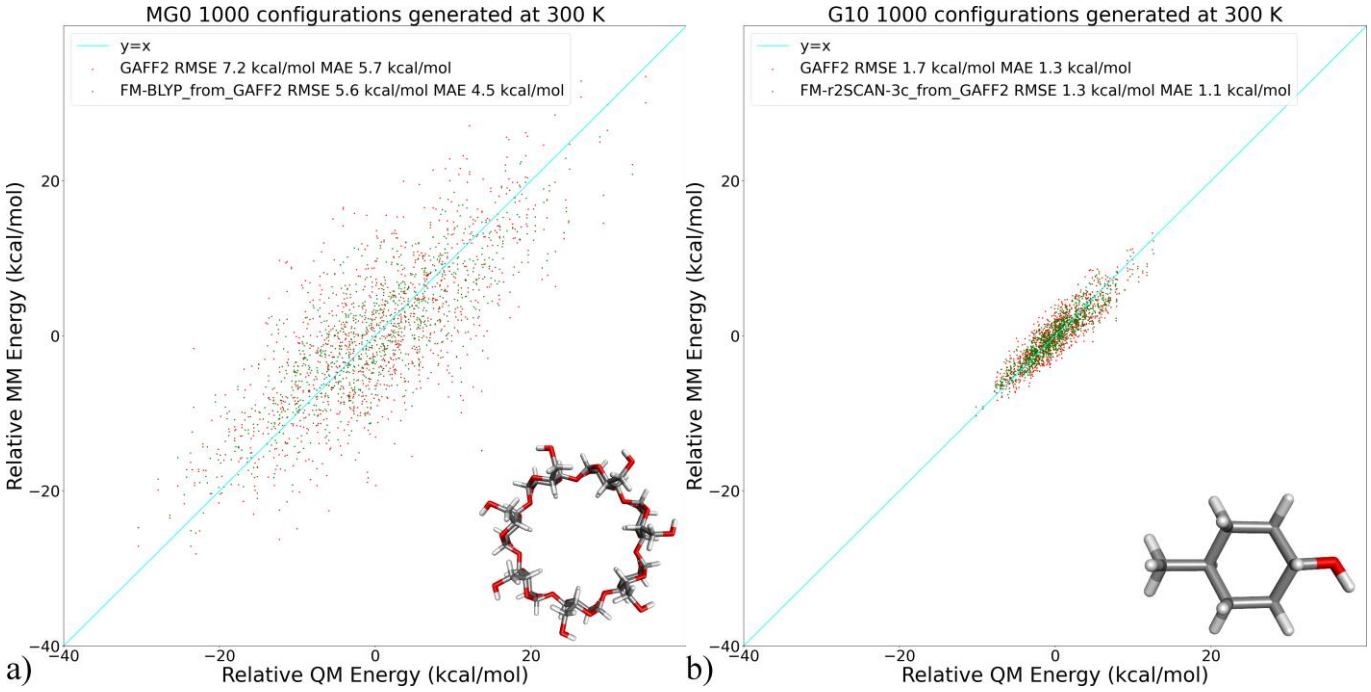
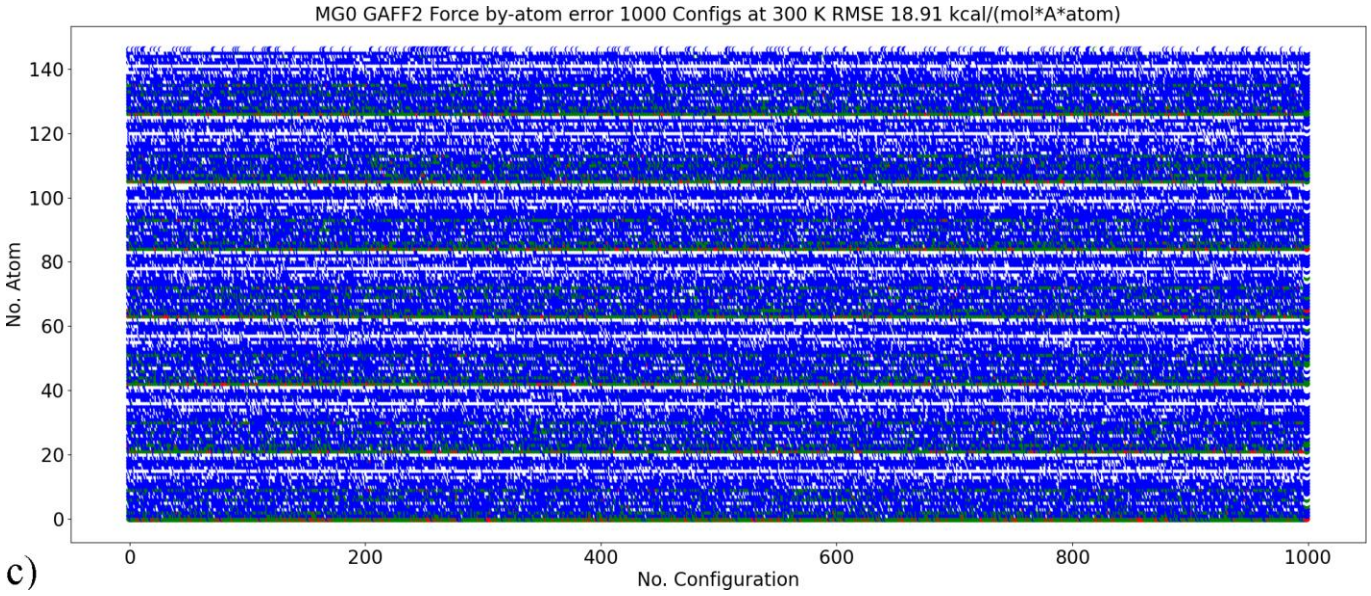


Fig. 2. Quality assessment for the generated atom-centered fixed-charge models: a) the ESP RRMSE, b) dipole moments obtained from the two charge schemes and the electronic structure calculations, c) the number of points in the reference HF/6-31G* ESP scan, d) an illustration of equivalence constraints used in charge fitting. Among the two guests under study, the guest G1 is a more challenging system for both charge schemes, which arises from the lone electron pairs at the oxygen atom that triggers anisotropic ESP distribution. The carbon atoms in cyan and magenta are considered equivalent in charge generation. Equivalence constraints are also imposed on the hydrogen atoms linked to them. Without the constraints on the two carbon pairs (note that the constraints on hydrogen atoms remain activated), the ESP RRMSE would be lower to ~9%. However, this leads to overfitting of the ESP of a single configuration and the neglect of the chemical equivalence of these atoms. Therefore, in our modelling, we use the RESP charges obtained with the equivalence constraints. Note that this equivalent-atom problems widely exist in many molecules, e.g., methanol.



>50 kcal/(mol·Å) >30 kcal/(mol·Å) >10 kcal/(mol·Å)



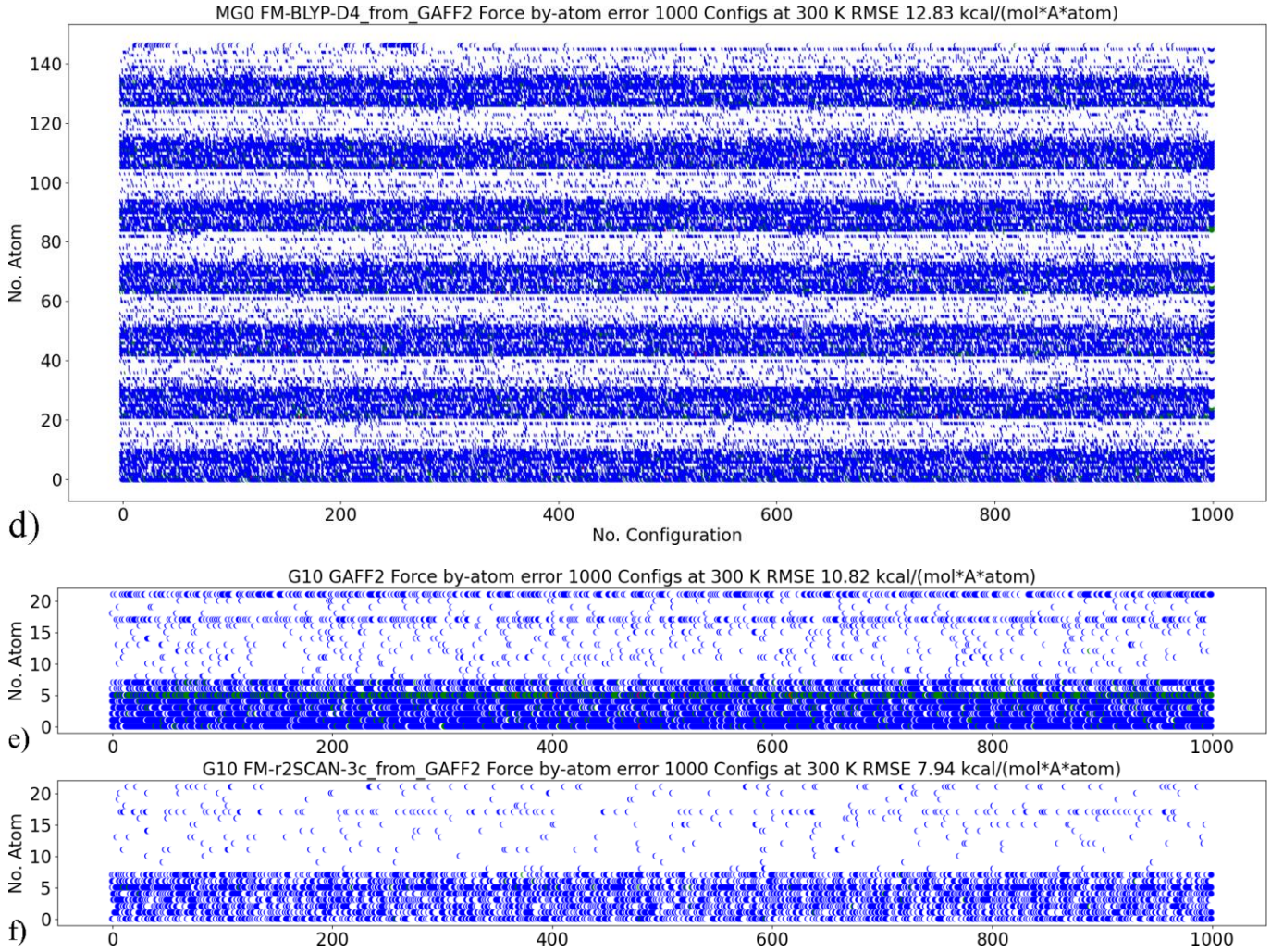


Fig. 3. Quality assessment for the generated models that combine the RESP charges and the pre-fitted GAFF2 or the refitted FM parameter set. Correlations of the force-field and *ab initio* energetics calculated from 10 ns trajectories with a sampling interval of 10 ps at 300 K in vacuo for a) the prototype host β-CD and b) the guest G1. The time series of errors/deviations of the atomic forces ($\|\Delta\mathbf{F}_i\|_2$ for the *i*th atom) produced by c) the original GAFF2 and d) the refitted FM-BLYP-D4 parameter sets for the host β-CD and the time series of force errors under e) GAFF2 and f) FM-r^{2}SCAN-3c for the guest G1. For the energy deviations, the errors produced by the pre-fitted GAFF2 are quite small for the guests that are small in size and drug-like and the GAFF2 refitting leads to marginal betterments, while more pronounced improvements are observed for the macrocyclic host β-CD. As for the atomic forces, when presenting the time series data of the atom-specified errors, the repeating units are numbered sequentially, and each glucose has 21 atoms. The first 11 atoms are non-hydrogen heavy atoms, while the other atoms are hydrogen atoms. Red dots for the force errors larger than 50 kcal/(mol·Å), green for force errors larger than 30 kcal/(mol·Å), blue for errors larger than 10 kcal/(mol·Å), and white for the other small-error points. A worth noting observation from the by-atom decomposition of force errors is that the forces on heavy atoms are more problematic than

those on hydrogen, which is similar to our previous refitting of the heterocyclic Cucurbit[8]uril.

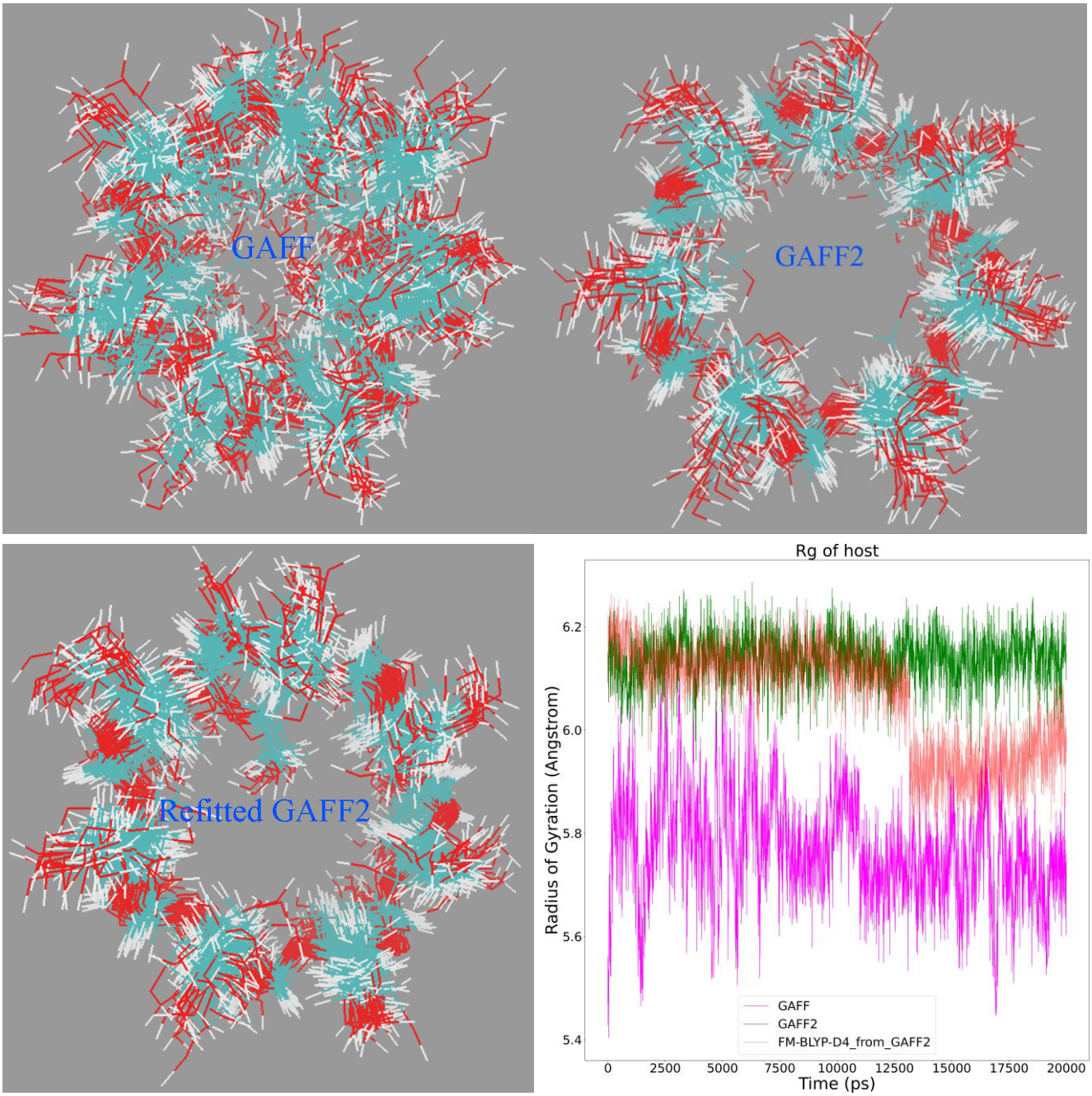


Fig. 4. The superposition of the host configurations during 20 ns unbiased simulations in explicit solvent with bonded parameters from GAFF, GAFF2, and the refitted parameter set targeting BLYP-D4/def2-SVP initiated from GAFF2 (FM-BLYP-D4_from_GAFF2) and the time series of the radius of gyration of the host ring. The GAFF parameters describe a squashed β-CD cavity in explicit solvent, GAFF2 provides a stiff host ring maintaining a large central cavity, and the host ring under the refitted GAFF2 parameter set is more flexible than the original GAFF2 but the cavity still remains open and thus is stiffer than the GAFF parameter set.

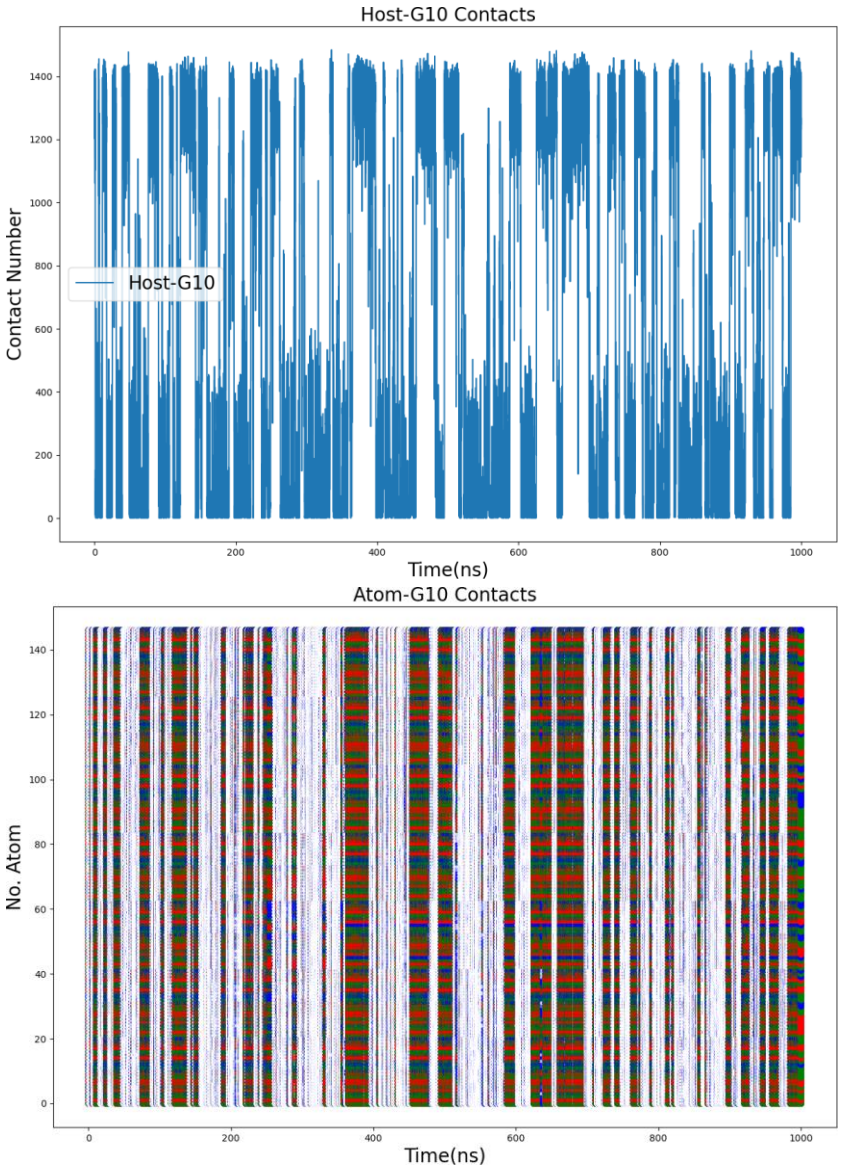


Fig. 5. The number of contacts between all atoms of the prototype β-CD host and the guest G1 and its decomposition by each atom of the host during enhanced sampling simulations under the GAFF2 force field. The y-axis represents the serial number of host atom. The repeating units are numbered sequentially, and each glucose has 21 atoms. The first 11 atoms are non-hydrogen heavy atoms, while the other atoms are hydrogen atoms. All atoms of the host and the guest are included in the calculation, which enables the identification of host atoms that directly coordinate the guest. Red dots denote contacts larger than 10, green dots represent contact number between 5 and 10, blue ones are those larger than 1, and the other are represented by white dots. We can see that a huge number of binding/unbinding events are observed, and the center-binding and side-binding modes are explored during enhanced sampling simulations. The results of all host-guest complexes are provided in Fig. S2.

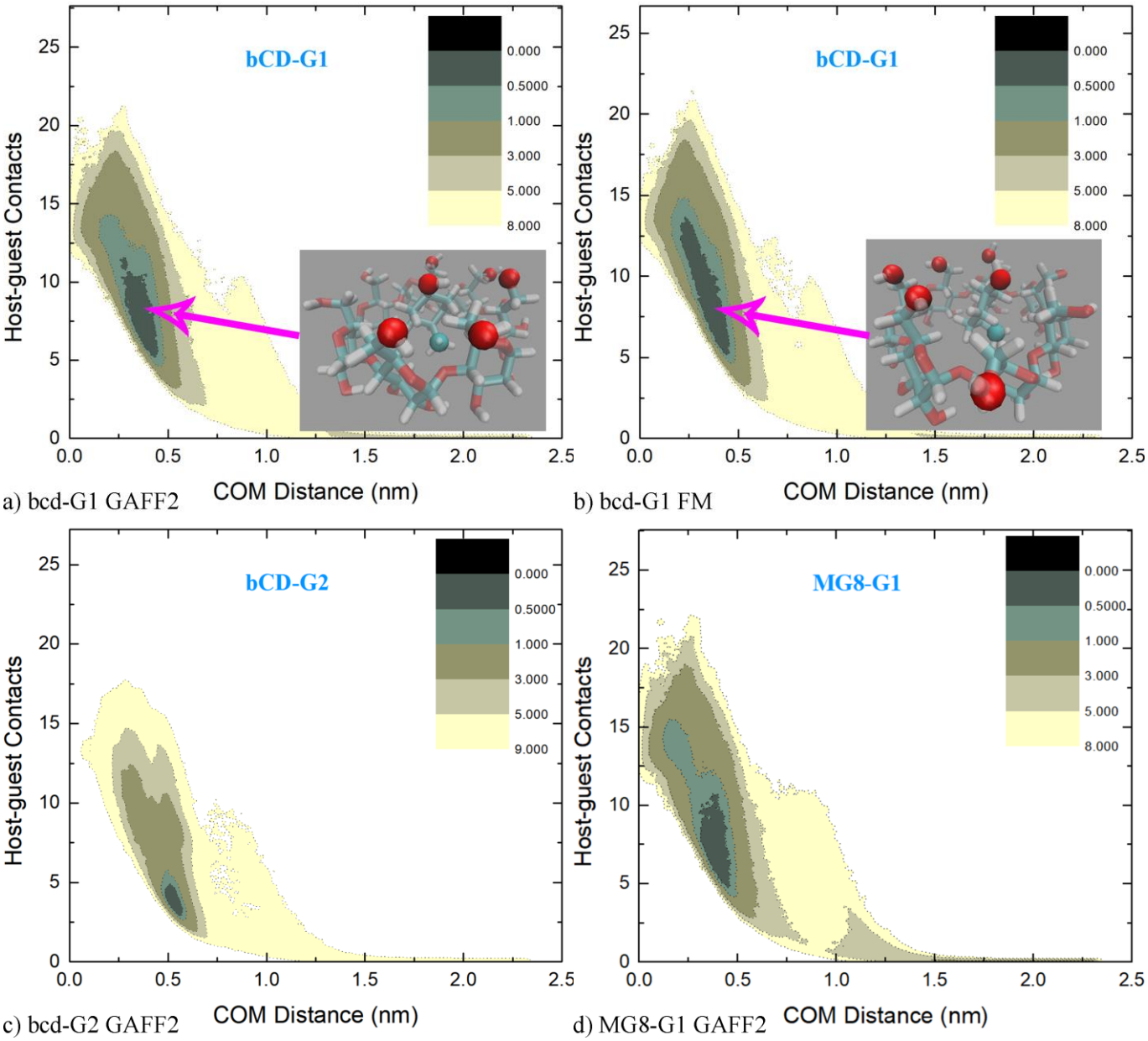


Fig. 6. 2D $\rho - C_{\text{host-hetero}}$ free energy surfaces in kcal/mol for a-b) the β-CD-G1 complex, c) the β-CD-G2 complex and d) the MG8-G1 complex. In a), c) and d), the pre-fitted GAFF2 parameter set is used to describe the system, while in b), the refitted FM parameter set (i.e., FM-BLYP-D4 for host and FM-r^2-SCAN-3c for guest) is employed. The results for the other host-guest complexes under the GAFF2 force field are given in Fig. S3.

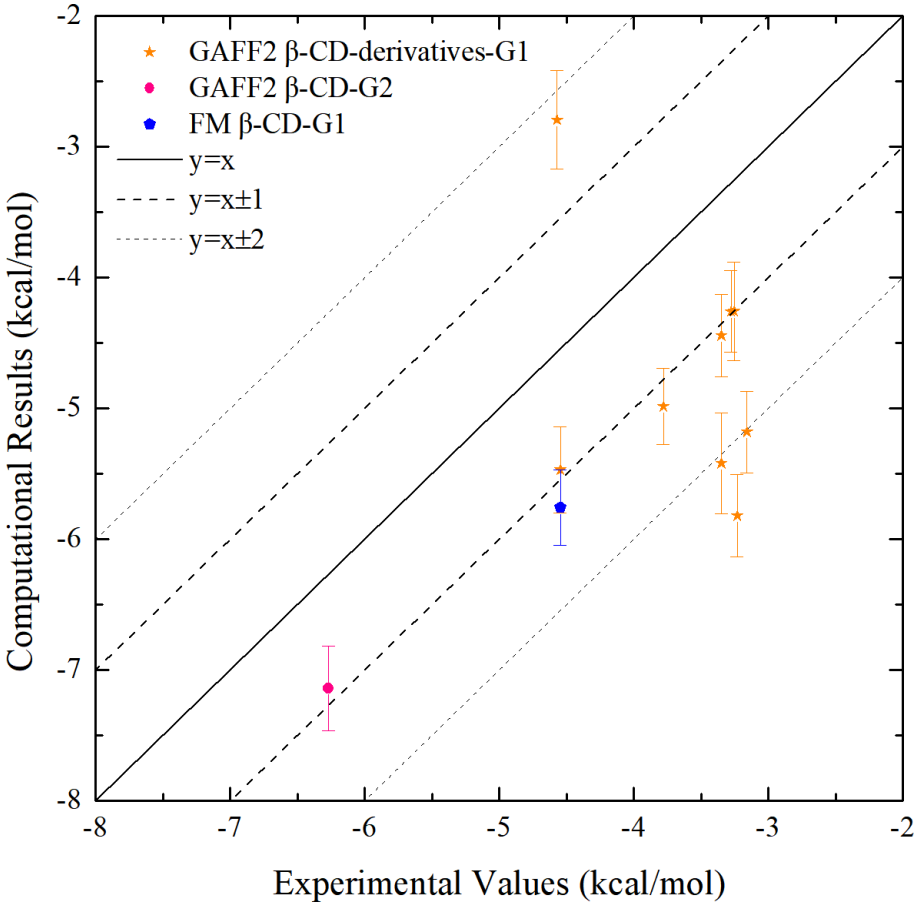


Fig. 7. Correlation between the binding affinities obtained from our computational modeling and the experimental reference for host-guest systems under RESP charges combined with the GAFF2 or FM parameter set. The exact values of the binding affinities are presented in Table 1.

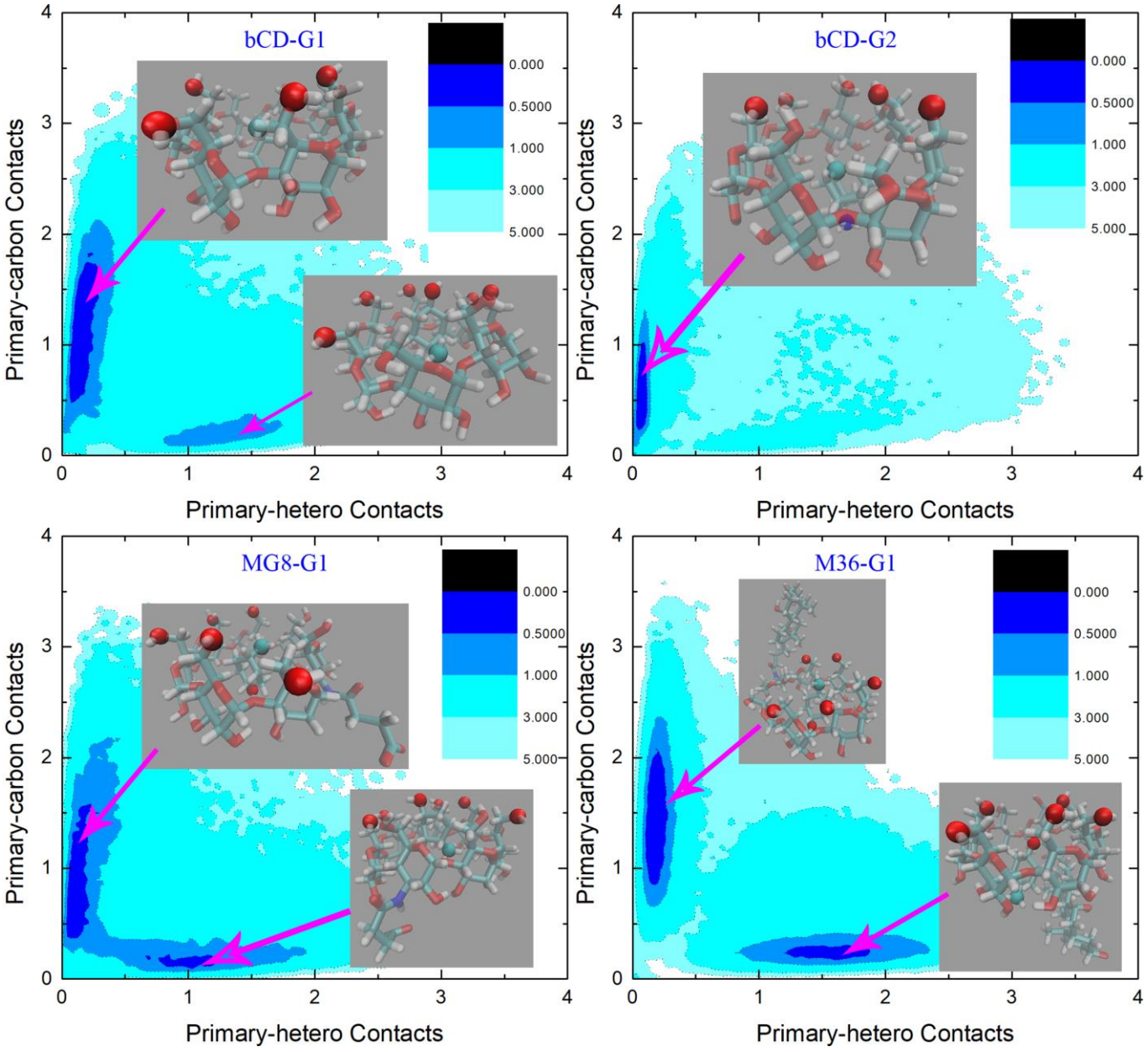


Fig. 8. 2D $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surfaces in kcal/mol for β-CD-G1, β-CD-G2, MG8-G1 and M36-G1 complexes obtained with the RESP charges and the GAFF2 force field. The results for the other host-guest systems are given in Fig. S4. Explicit marks (vdW representations) are added on the oxygen atoms defining the primary face of the host and the two anchor atoms of the guest.

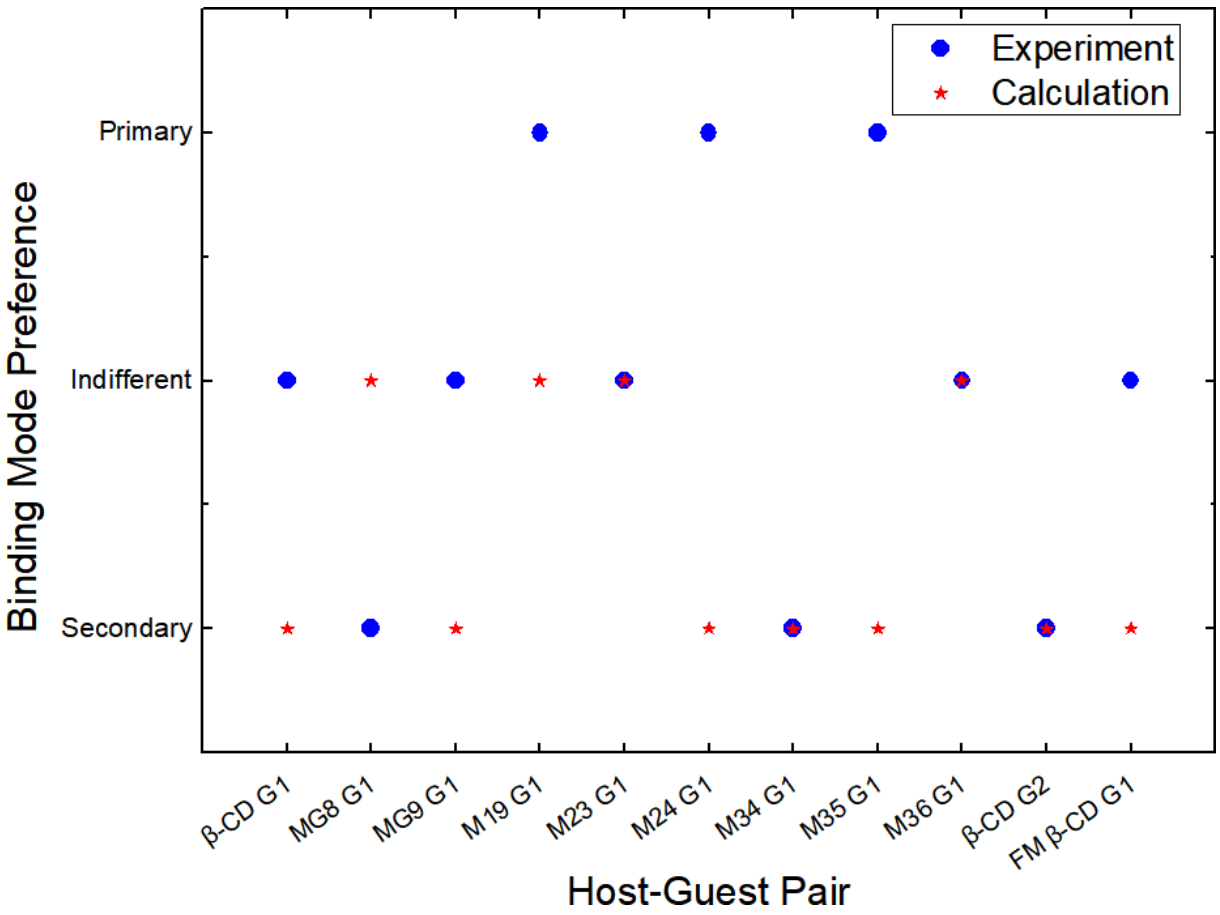


Fig. 9. Comparison between the experimental and computed directional preference for β-CD-derivatives-guest complexes. Primary represents the host-guest complex prefers a primary binding mode, secondary denotes the secondary binding mode preference, and indifferent represents no binding mode preference. It is generally observed that the calculation-predicted binding mode preference is more 'secondary' than the experimental reference. Namely, in computational modelling, the primary-preferred systems tend to be indifferent or secondary-preferred, while the indifferent systems are prone to be indifferent or secondary-preferred.

Supporting Information:

Primary vs Secondary: Directionalized Guest Coordination in β-Cyclodextrin Derivatives

Zhaoxi Sun[1*], Lei Zheng[2], Kai Wang[3], Zhe Huai[4], and Zhirong Liu[1]

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

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

[3]*METiS Pharmaceuticals, Inc., Hangzhou 310052, China*

[4]*XtalPi - AI Research Center (XARC), 7F, Tower A, Dongsheng Building, No.8, Zhongguancun East Road, Haidian District, Beijing 100083, China*

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

Fig. S1. The spherical coordinates defined by the COM of the host backbone and the only heteroatom of the guest.

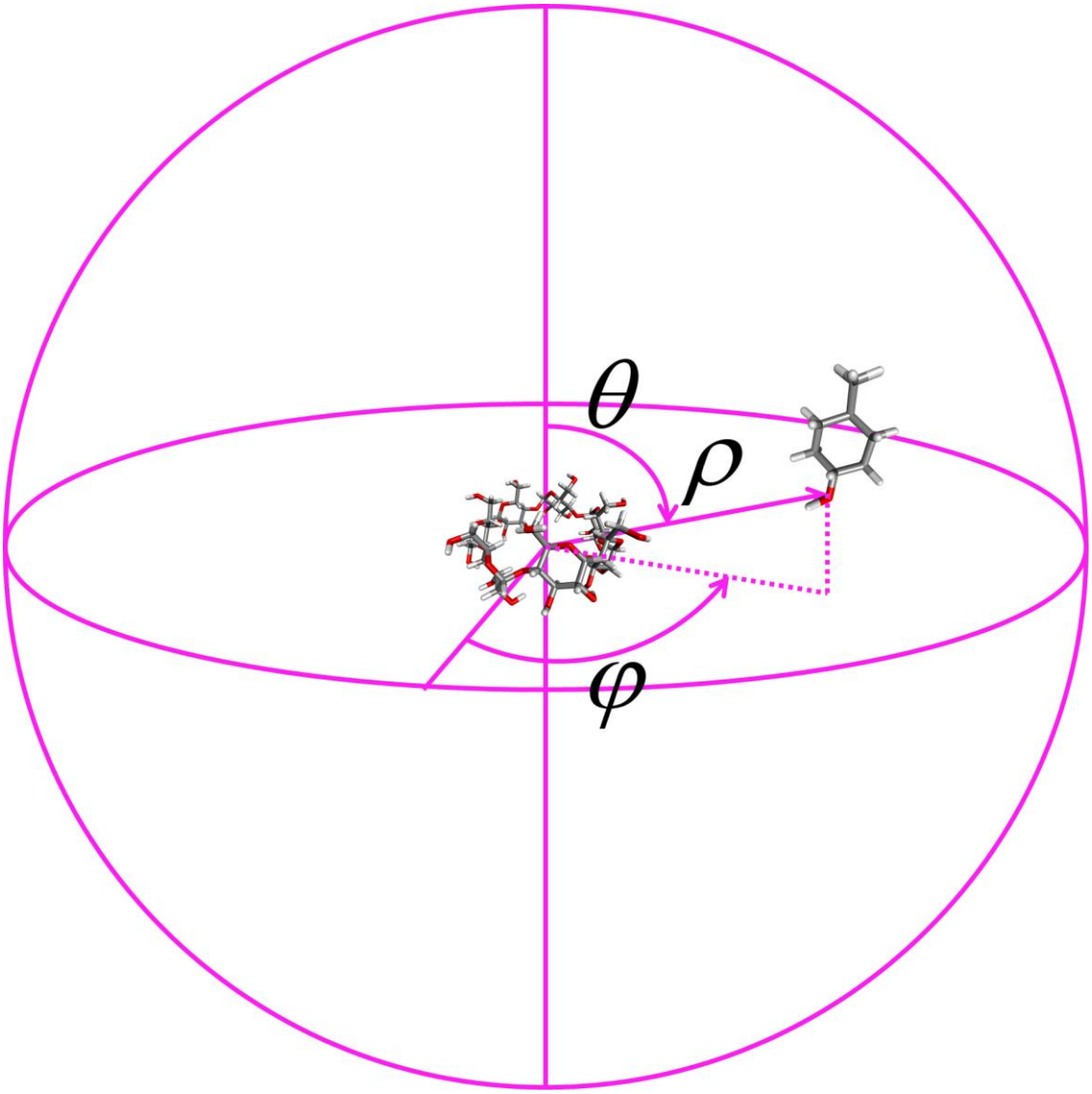
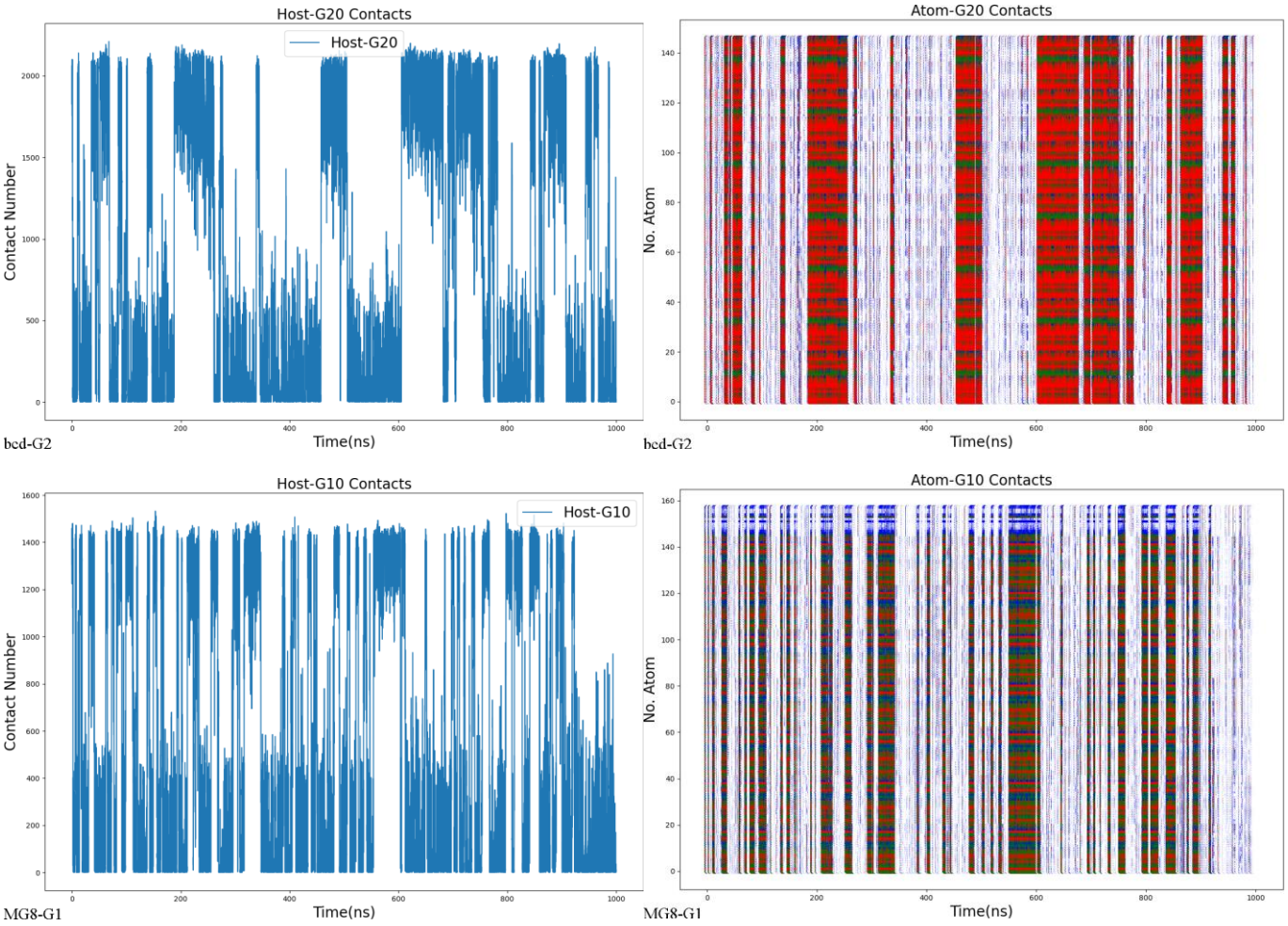
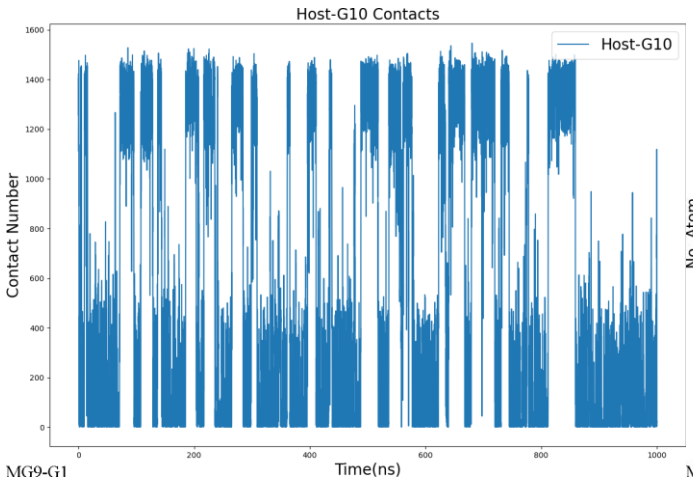
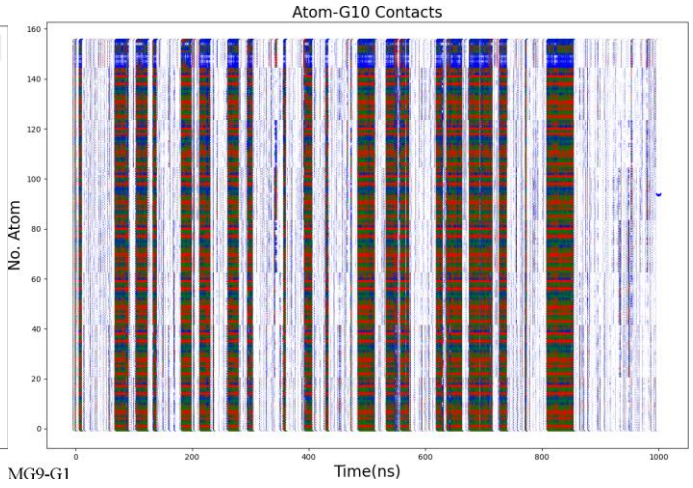


Fig. S2. The number of contacts between all atoms of the host β-CD derivatives and the guests and the by-host-atom decomposition during metadynamics simulations under the RESP charge scheme and the GAFF2 force field. The y-axis represents the serial number of host atom. All atoms of the host and the guest are included in the calculation. Each glucose has 21 atoms. The first 11 atoms are non-hydrogen heavy atoms, while the other atoms are hydrogen atoms. The repeating units are numbered sequentially and the substitution groups are put after the backbone atoms. Red dots denote contacts larger than 10, green dots represent contact number between 5 and 10, blue ones are those larger than 1, and the other are represented by white dots.

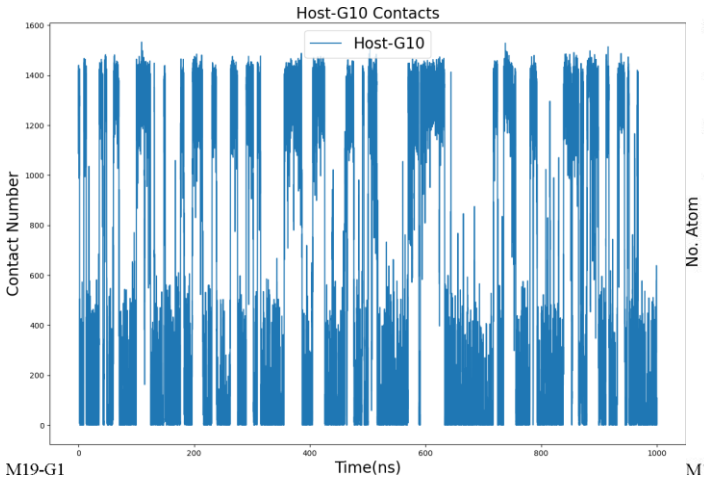




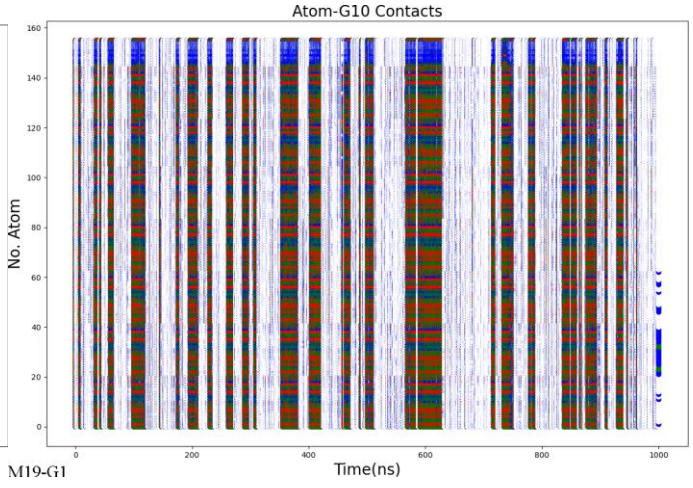
MG9-G1



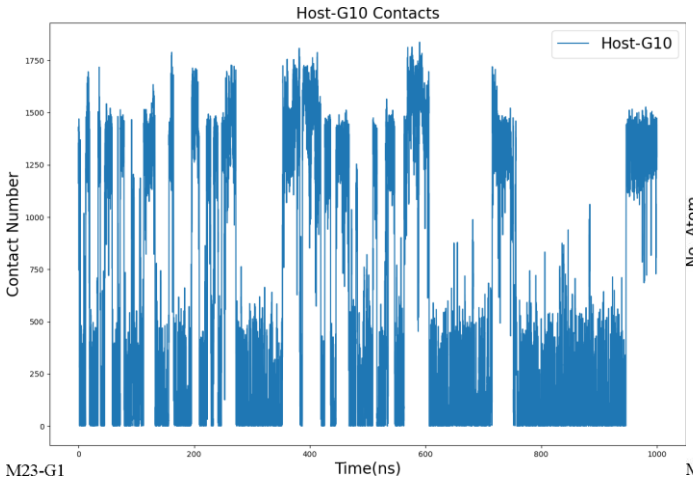
MG9-G1



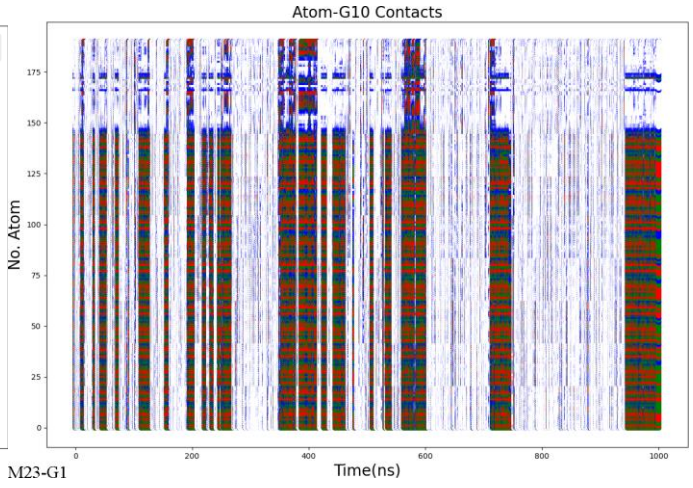
M19-G1



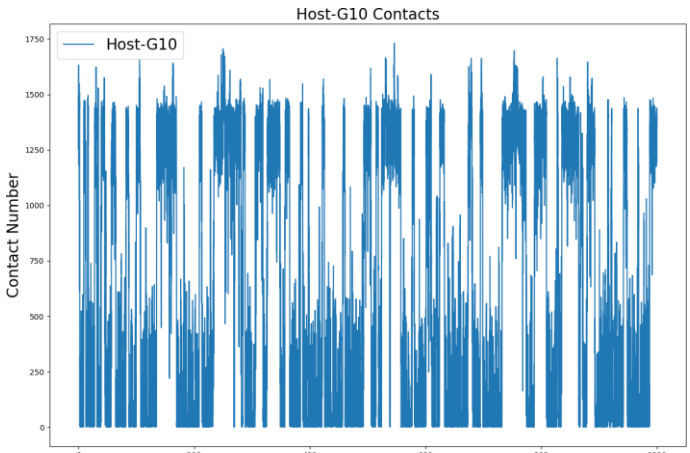
M19-G1



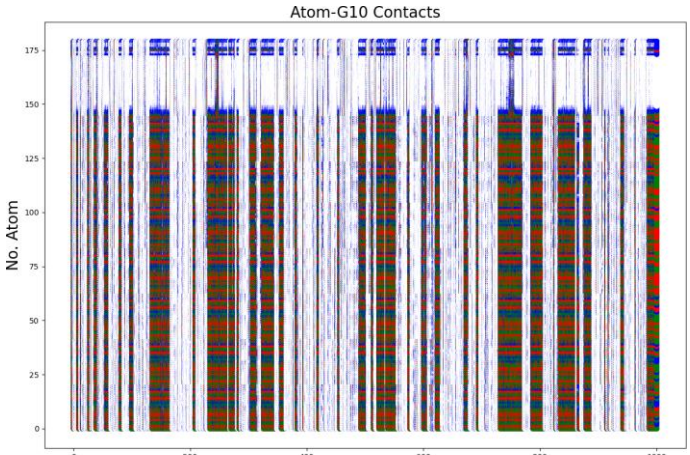
M23-G1



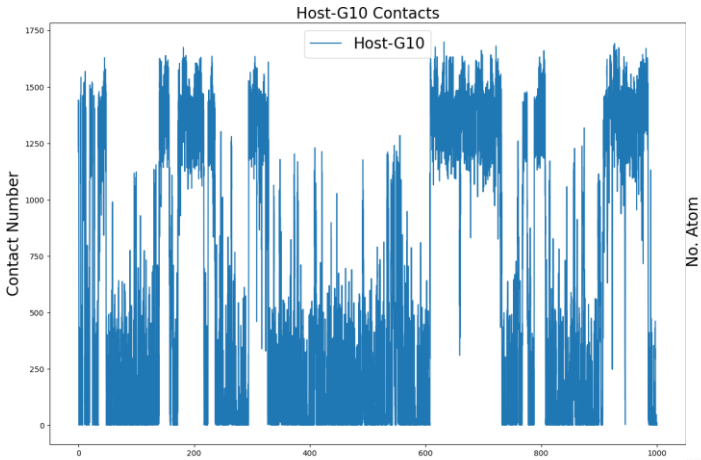
M23-G1



M24-G1

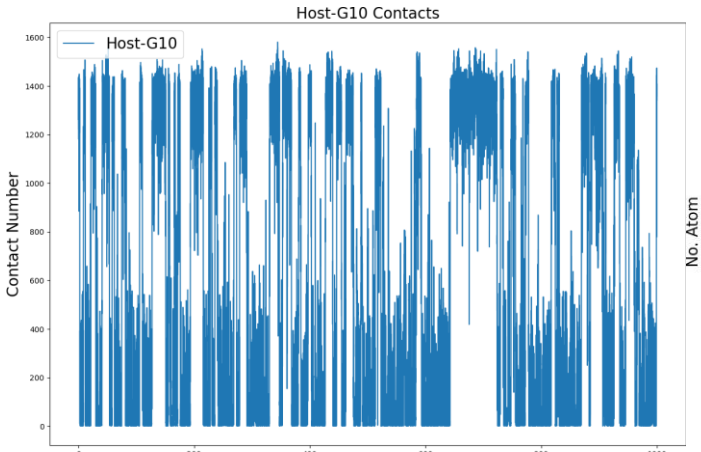


M24-G1

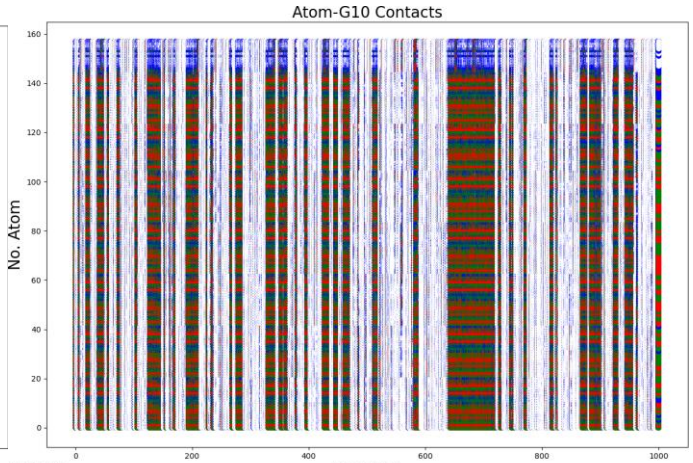


M34-G1

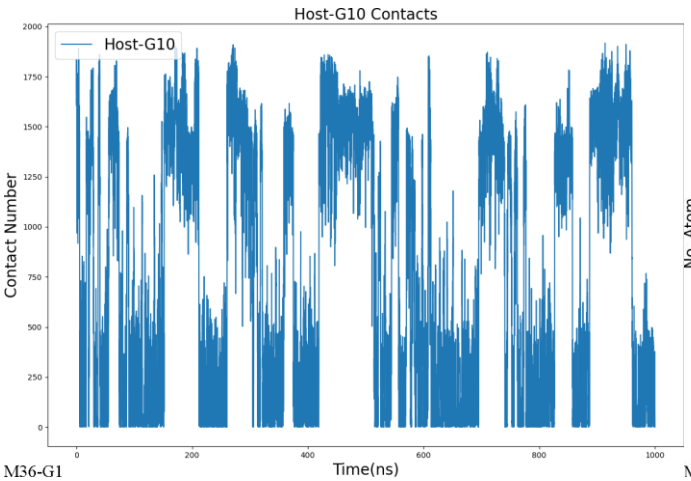
M34-G1



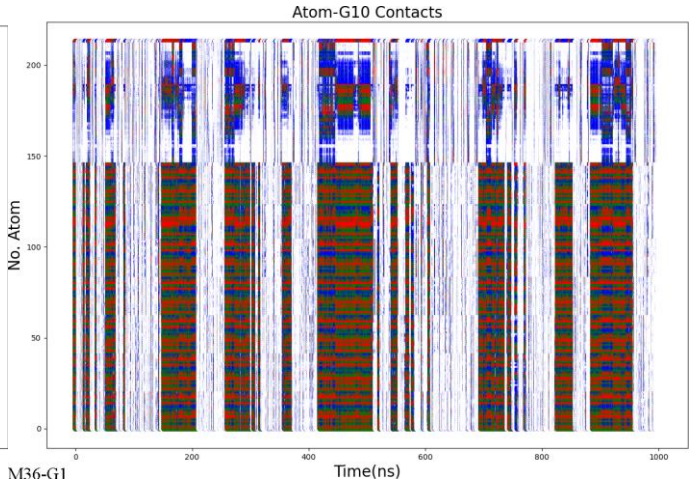
M35-G1



M35-G1

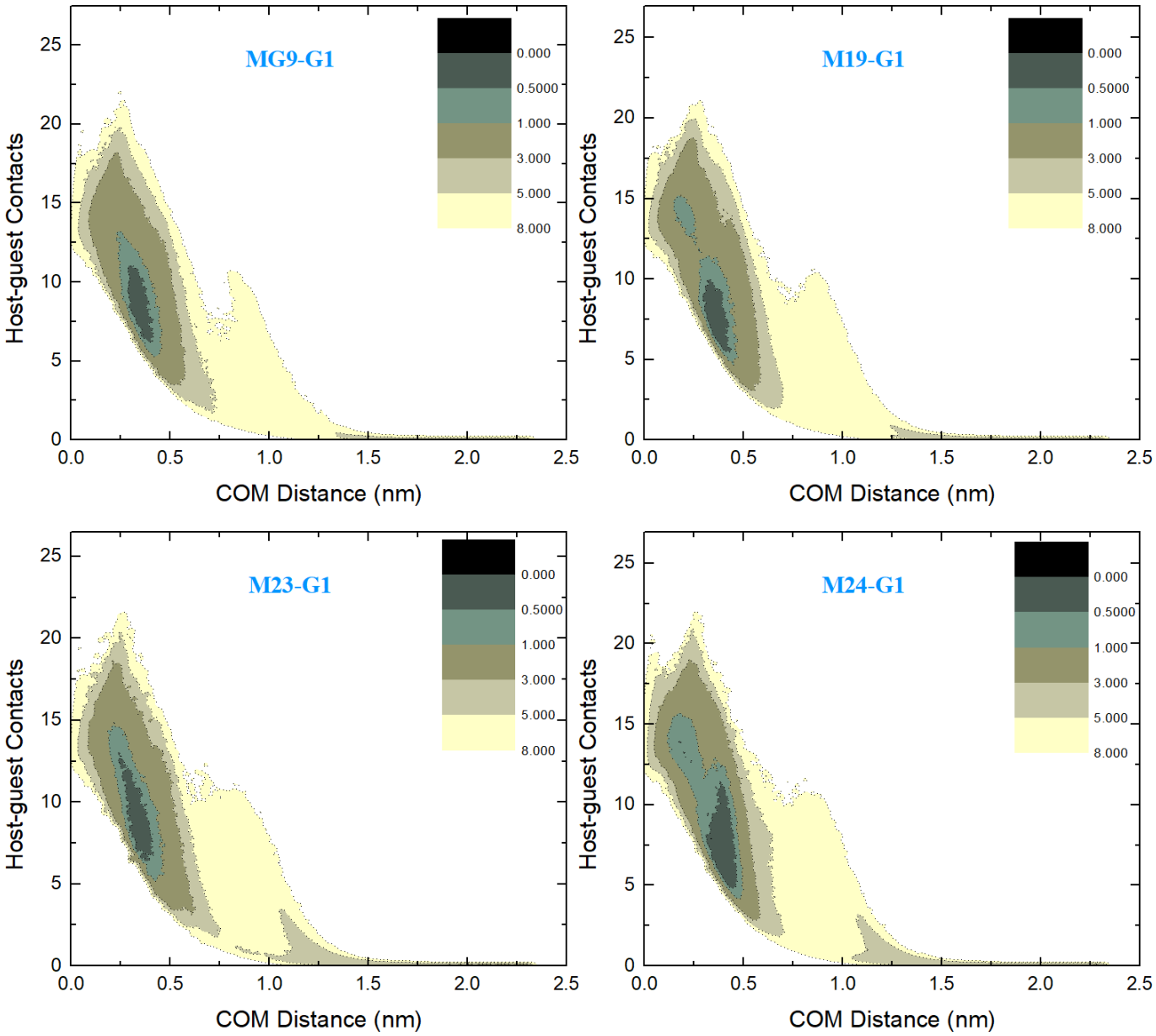

Host-G10 Contacts
Host-G10
Contact Number
Time(ns)

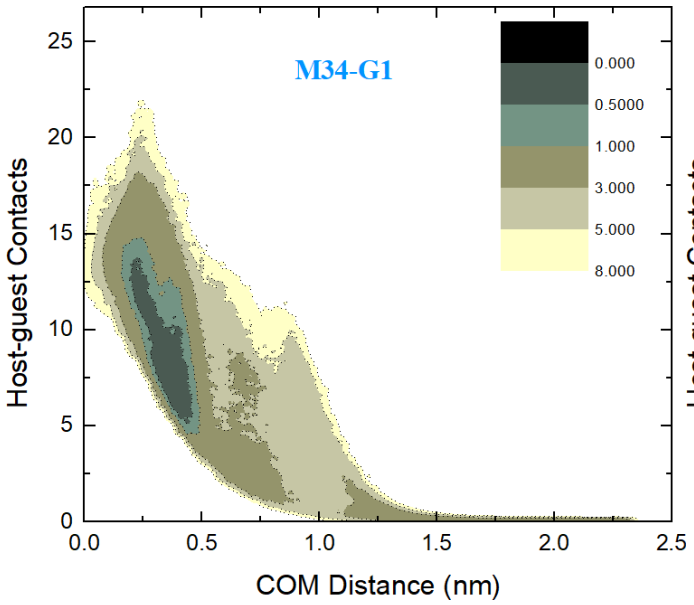
M36-G1

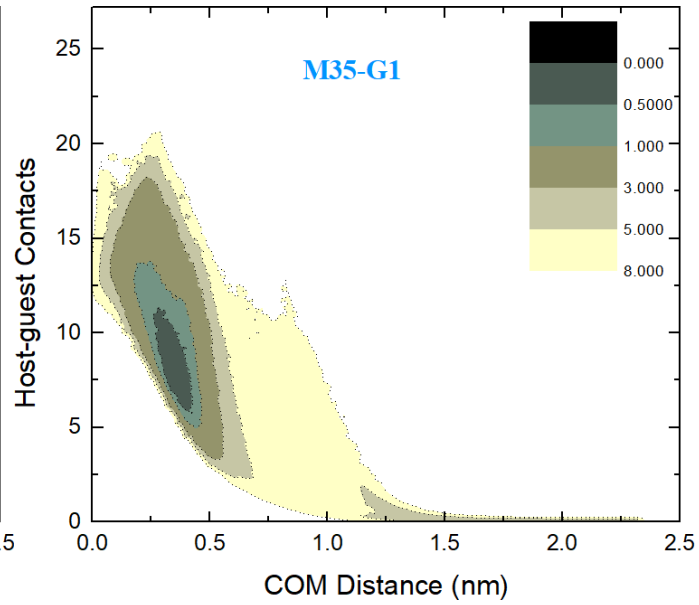

Atom-G10 Contacts
No. Atom
Time(ns)

M36-G1

Fig. S3. 2D $\rho - C_{\text{host-hetero}}$ free energy surfaces in kcal/mol for MG9-G1, M19-G1, M23-G1, M24-G1, M34-G1, M35-G1 and M36-G1 complexes obtained with the RESP charges and the GAFF2 force field.




M34-G1
Host-guest Contacts
COM Distance (nm)
0.000
0.5000
1.000
3.000
5.000
8.000


M35-G1
Host-guest Contacts
COM Distance (nm)
0.000
0.5000
1.000
3.000
5.000
8.000

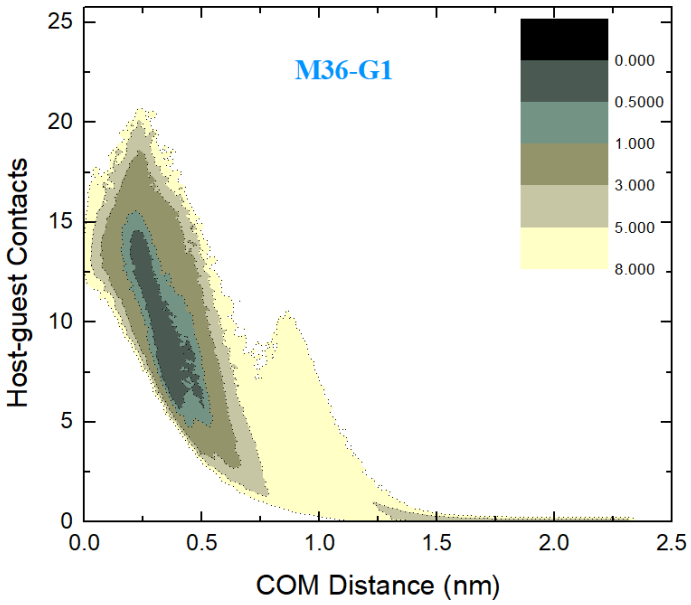
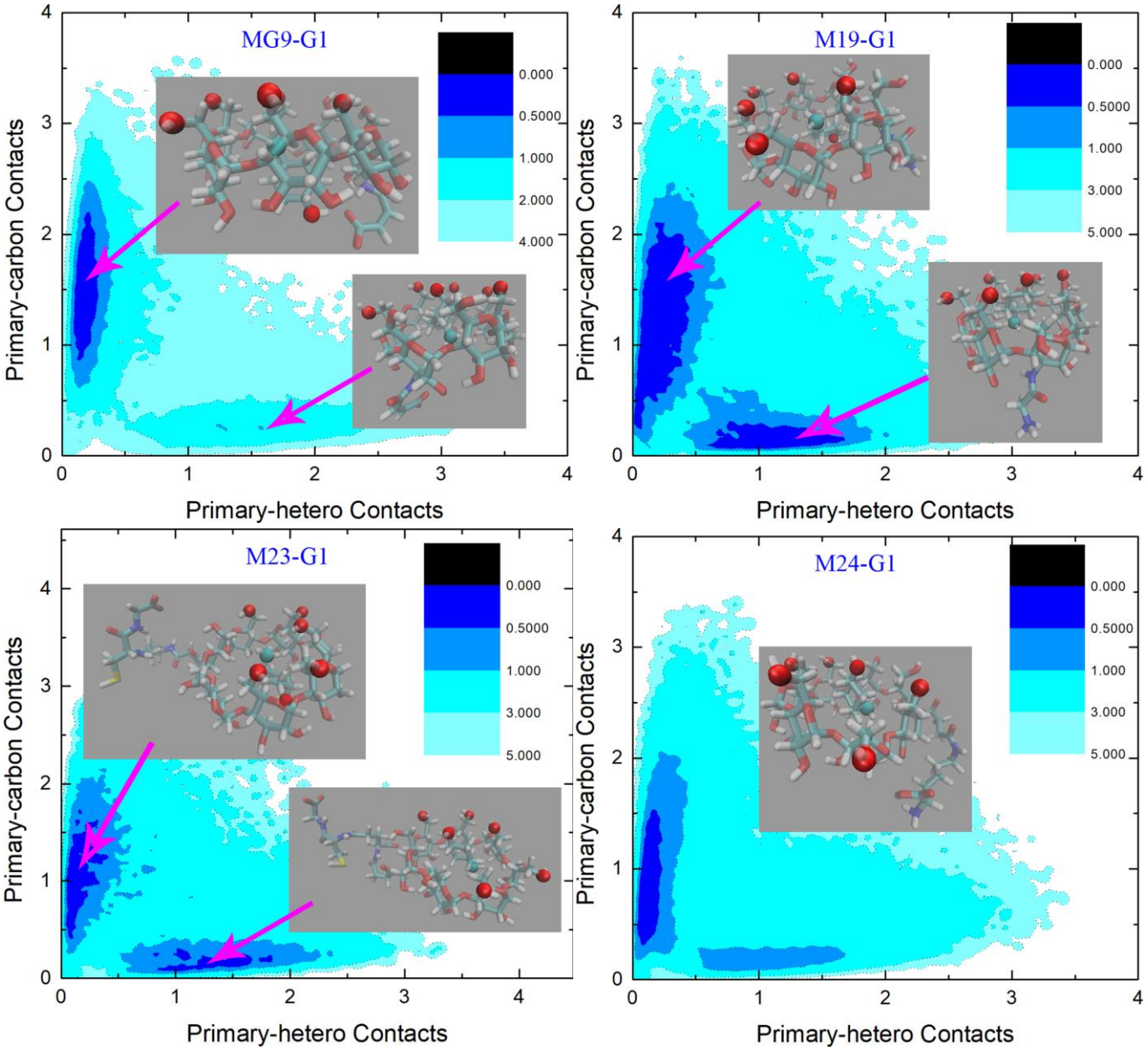
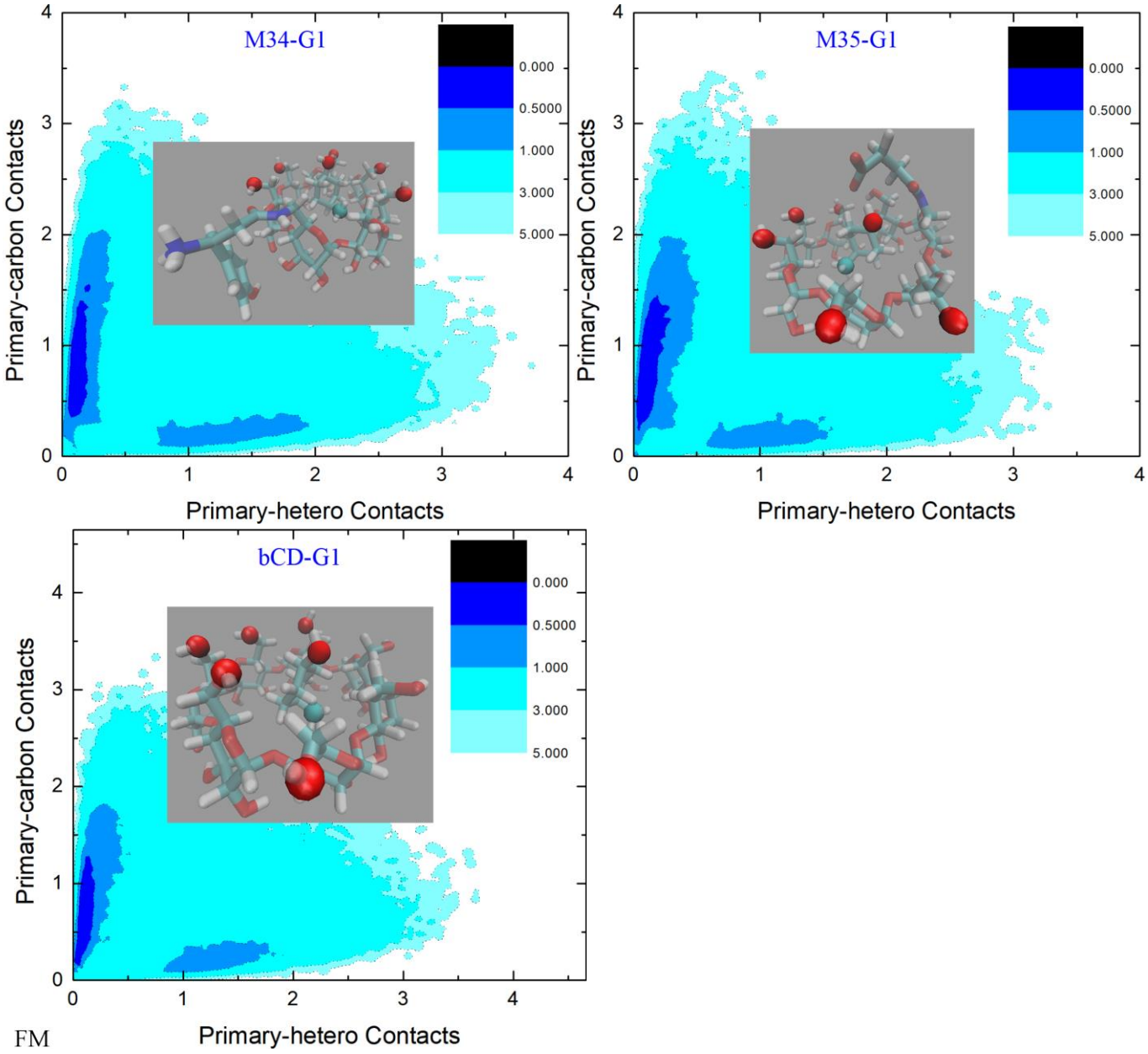

M36-G1
Host-guest Contacts
COM Distance (nm)
0.000
0.5000
1.000
3.000
5.000
8.000

Fig. S4. 2D $C_{\text{hetero-primary}} - C_{\text{carbon-primary}}$ free energy surfaces in kcal/mol for MG9-G1, M19-G1, M23-G1, M24-G1, M34-G1, M35-G1 and M36-G1 complexes obtained with the GAFF2 force field and that for the β-CD-G1 complex obtained with the refitted FM parameter set. For the first three systems, illustrative structures of binding poses are extracted from both minima (i.e., primary and secondary), while in the other host-guest cases we only present a single illustrative bound structure for each system. Explicit marks (vdW representations) are added on the oxygen atoms defining the primary face of the host and the two anchor atoms of the guest.




M34-G1
M35-G1
bCD-G1
0.000
0.5000
1.000
3.000
5.000
Primary-carbon Contacts
Primary-hetero Contacts
0
1
2
3
4

FM