

A General Picture of Cucurbit[8]uril Host-Guest Binding

Zhaoxi Sun^{1*}, Zhe Huai², Qiaole He³, Zhirong Liu¹

¹*Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China*

²*XtalPi - AI Research Center (XARC), 9F, Tower A, Dongsheng Building, No.8, Zhongguancun East Road, Haidian District, Beijing 100083, P.R. China*

³*Enzymaster (Ningbo) Bio-Engineering Co., Ltd., North Century Avenue 333, 315100 Ningbo, China*

*To whom correspondence should be addressed: z.sun@pku.edu.cn

Abstract

Describing, understanding, and designing complex interaction networks within macromolecular systems remain challenging in modern chemical research. Host-guest systems, despite their relative simplicity in both the structural feature and interaction patterns, still pose problems in theoretical modelling. The barrel-shaped supra-molecular container Cucurbit[8]uril (CB8) shows promising functionalities in various areas, e.g., catalysis and molecular recognition. It can stably coordinate a series of structurally diverse guests with high affinities. In this work, we examine the binding of 7 commonly abused drugs to the CB8 host, aiming at providing a general picture of CB8-guest binding. A thorough comparison of widely used fixed-charge models for drug-like molecules is presented. Extensive sampling of the configurational space of these host-guest systems is performed, and the binding pathway and interaction patterns of CB8-guest complexes are investigated in detail. Iterative refitting of the atomic charges suggests significant conformation-dependence of charge generation. The initial model generated at the original conformation could be inaccurate for new conformations explored during conformational search. Our investigations of the configurational space of CB8-drug complexes suggest that the host-guest interactions are more complex than expected. Despite the structural simplicities of these molecules, the conformational fluctuations of the host and the guest molecules and orientations of functional groups lead to the existence of an ensemble of binding modes (e.g., bracelet-like guest conformations and crescent free energy landscapes). Thus, understandings obtained from static calculations based on a single or several structures are limited for these host-guest interactions. The

investigation protocol provides useful guidelines for studying host-guest binding, and the insights of the binding thermodynamics, performance of fixed-charge models, and binding patterns of the CB8-guest systems are useful for elucidating the binding mechanism of other host-guest complexes.

Keywords: Cucurbit[8]uril, Host-guest Interaction, Abused Drugs, Binding Mode, Fixed-charge Models

1. Introduction

Drugs are small molecules triggering changes in the physiology or psychology of specific organism(s). They can be consumed in various ways (e.g., via injection or ingestion) for a limited duration or on a regular basis. They are small in size compared with biomacromolecules (e.g., proteins), but can play an important role in biological processes by interacting with specific targeted biomolecules. Therefore, protein-drug/ligand and nucleotide-ligand interactions are key research directions in drug discovery. Potential side effects are inevitable when consuming drugs. The abuse of psychoactive drugs could lead to psychological or physical addiction. Excess consumption of stimulants could result in stimulant psychosis.

The molecular container Cucurbit[8]uril (CB8 or CB[8], see Fig. 1a) belongs to a family of barrel-shaped supra-molecular containers named Cucurbit[n]uril (CB[n]) with wide application areas, e.g., catalysis and molecular recognition.¹⁻⁷ The partly enclosed hydrophobic cavity of CB8 is formed by 8 glycoluril monomers linked by 16 methylene bridges. Its hydrophilic carbonyl portals on glycoluril units provide stabilization effects for cationic species and act as hydrogen bond acceptors. As a result, CB8-related macrocycles can bind to a range of neutral and cationic guests with high association constants. Drug abuse has posed a societal problem. Due to the ability of the CB8 container to hold structurally diverse guests, it could also coordinate many commonly used drugs in a very stable way. Therefore, CB8-related molecules act as key units in many novel drug carriers.

Molecular simulations provide a feasible route to access the detailed motions of complex systems at atomic level. Thermodynamic and kinetic information could be extracted via statistical analyses. Biologically relevant processes generally happen at time scales of μ s, ms and even longer, which are inaccessible in modern computational modelling.⁸⁻¹⁶ Enhanced sampling techniques are often employed to deal with the time-scale problem.¹⁷⁻²³ Modifications of the Hamiltonian of the simulated system are added to accelerate the phase space exploration, and post-simulation analyses are used to recover the statistics in the original unperturbed ensemble. The Hamiltonian or model employed to describe the system under investigation determines the accuracy of the simulation outcome.²⁴⁻²⁸ Although detailed descriptions such as quantum mechanics (QM) calculations are accurate, their practical use is quite limited due to high computational costs.²⁹⁻³³ To obtain sufficient data for post-processing analyses, the approximated mean-field molecular mechanics (MM) force fields are widely employed in molecular dynamics (MD) simulations of biomolecular systems.³⁴⁻³⁹

In this work, we present a comprehensive computational investigation of the host-guest interaction in CB8-drugs systems. Seven commonly used and abused drugs including Metamphetamine, Fentanyl, Cocaine,

Phencyclidine, Ketamine, Morphine, and Hydromorphone are considered. Detailed interactions of these structurally diverse drugs/guests with the macrocyclic host CB8 would be investigated in detail via extensive atomistic simulations, aiming at providing a general picture of CB8-guest interactions at atomic level. The IUPAC name, chemical formula, the simplified molecular-input line-entry system (SMILES) identifier, and monatomic molecular weight for each molecule (e.g., the host and the guests) are summarized in Table 1, and the 2D chemical structures of these abused drugs are presented in Fig. 1a.

The first guest G1, Metamfetamine, is a potent stimulant for the central nervous system.^{40, 41} It has a high potential for addiction and abuse. Overdosage could lead to hallucinations, hyperpyrexia, rhabdomyolysis and so on.⁴² This scheduled drug is prescribed or dispensed sparingly. It is often employed when alternative therapy is tested to be ineffective. Structurally, the guest consists of a phenyl ring and an aliphatic chain containing an ionizable secondary amino group. It is relatively simple compared with the other guests under study.

The second guest G2, Fentanyl, is a potent synthetic lipophilic phenylpiperidine opioid agonist used for rapid analgesia.⁴³⁻⁴⁵ It is a controlled drug with a high potential for addiction, abuse and misuse. Fentanyl acts on the central nervous system and mimics the effects of endogenous opiates. Structurally, the monocarboxylic acid amid has three 6-membered rings (two aromatic and one heteroaliphatic). Considering our previous study of CB8-guest interactions,⁴⁶ this structurally complex guest is expected to have multiple binding patterns with the host.

The guests G3 Morphine and G4 Hydromorphone are structurally similar. Both of them are polycyclic compounds with a four-ring skeleton, the three condensed 6-membered rings of which form a partially hydrogenated phenanthrene moiety. One of the three 6-membered rings is aromatic and the other two are alicyclic. As the main alkaloid of opium, Morphine is a potent analgesic of limited use.⁴⁷ Excess consumption of Morphine and its derivatives could lead to various symptoms, such as respiratory depression, miosis, pulmonary edema and even death.⁴⁸⁻⁵⁰ The guest G4 Hydromorphone also belongs to the morphinans class and is a minor metabolite of Morphine.⁵¹⁻⁵⁴ This hydrogenated ketone derivative has a higher potency and a side effect profile comparable to the parent compound Morphine.⁵⁵ Hydromorphone is clinically used when patients do not tolerate the side effects of Morphine or suffer from renal failure or asthma.

The fifth abused drug under investigation, Ketamine, is a cyclohexanone derivative with rapid-acting analgesic and anesthetic properties.^{56, 57} It can rapidly produce a unique hypnotic state named dissociative anesthesia. The drug is clinically used for short-term diagnostic and surgical procedures, e.g., induction of

anesthesia prior to other general anesthetic agents.⁵⁸ It has been limited in use due to its significant psychological side effects such as hallucinations and agitation.⁵⁹⁻⁶¹ Long-term consumption could trigger inflammation and irritation to the urinary bladder, urethra and biliary tract, resulting in liver injuries.^{61, 62} Structurally, the drug/guest has two 6-membered rings with a limited number of functional groups and rotatable bonds and is simpler than complex guests such as G2 Fentanyl. However, the pKa of its ionizable secondary amino group is close to the physiological condition, which suggests the existence of acid-base/protonation-deprotonation equilibrium that should be properly accounted for in computational modelling.

The guest G6 Phencyclidine is structurally similar to G5 Ketamine, but its nitrogen atom is in a piperidinium moiety. As a result, this guest has three 6-membered rings. Similar to Ketamine, the drug could produce a dissociative state by inhibiting the ionotropic receptor named N-methyl-D-Aspartate receptor and thus interfering with functions of central nervous system.⁶³⁻⁶⁶ As a hallucinogen, Phencyclidine was formerly used as an anesthetic but was withdrawn due to severe side effects, e.g., visual disturbances, aggressive and violent behaviors, status epilepticus, and renal, respiratory and hepatic failure.^{63, 64, 66, 67}

The last abused drug under investigation, G7 Cocaine, is a topical anesthetic and vasoconstrictor.⁶⁸⁻⁷⁰ It is used to introduce local anesthesia of accessible mucous membranes of the oral, laryngeal and nasal cavities.^{68, 71-73} The drug could reversibly bind to sodium channels and act as a pore blocker, which blocks the sodium influx through the ion channels.⁷⁴ The addictive quality of Cocaine is highly related to its effects on dopamine levels. Cocaine could bind to dopamine transport proteins and thus directly prevent the re-uptake of dopamine.^{72, 73, 75, 76} Overdosage of Cocaine could trigger toxic effects such as rhabdomyolysis and acute liver injury.^{68, 69, 71, 77} Structurally, the drug is a benzoic acid ester with an ionizable tertiary amino group. Therefore, the host-guest interaction for this complex guest is expected to be complicated.

Intuitively, the CB8-guest binding and more generally host-guest interactions are simpler than the protein-ligand interactions. Macrocyclic hosts are structurally simpler than biomolecular receptors (e.g., proteins), in the heterogeneous environment of which complex interaction networks could be formed. The symmetric feature of the CB8 host further simplifies the possible patterns of host-guest interactions. Based on such assumptions, many computational investigations of similar host-guest systems in the CB[n] family ignored the possibility of the existence of multiple binding modes, and only a single bound conformation was considered in computational modelling.^{7, 78-81} However, our previous molecular modelling of CB8-guest systems revealed the existence of many unexpected binding patterns in some host-guest complexes. Specifically, we explored the configurational space in a series of CB8-related host-guest complexes and

observed a diverse set of interaction/binding patterns.^{46, 82} For instance, while many CB8-guest systems had only one stable binding pose, the CB8-G3 system studied in the previous work was observed to have multiple binding poses with similar thermodynamic stabilities.⁴⁶ This property poses a problem for the unbiased sampling technique. Enhanced sampling techniques in the configurational space are required to obtain a converged picture of the host-guest interaction. In order to provide a general picture of the CB8-guest interactions, in the current work, we selected the above seven commonly abused drugs with diverse structural features and investigate their interactions with the CB8 host. As CB8 has a medium internal cavity volume among the CB[*n*] family, understanding CB8-guest interactions would provide valuable insights into the host-guest binding patterns for the whole CB[*n*] family. The abused drugs under investigation have diverse structural features, including aromatic and aliphatic rings of different sizes, halogen substitutions, protonation-deprotonation equilibria and so on. Therefore, comprehensive investigations of these host-guest complexes would provide a general picture of the diverse CB8-guest interaction patterns. Further, two fixed-charge schemes widely used in drug discovery are employed and compared in great detail. Computationally efficient and widely used force fields in biomolecular simulations describe the system with bonded and non-bonded terms, the latter of which contribute directly and significantly to the inter-molecular interactions. Thus, comprehensive comparisons between fixed-charge models enable a thorough evaluation of these widely used force fields and provide hints on the selection of charge schemes in molecular simulations. In the following parts of the manuscript, we briefly outline the details of our modelling (i.e., model construction and sampling), provide a comprehensive comparison of fixed-charge models, and present a general picture of the CB8-guest binding.

2. Methodology and Computational Details

2.1. System preparation.

The structure of the pumpkin-like host CB8 is obtained from the PDBbank (RCSB PDB) and the structures of the seven commonly abused drugs are obtained from the GitHub repository of the 8th statistical assessment of the modelling of proteins and ligands challenge.⁸³ The 3D structure of the host CB8 and the 2D chemical structures of the guest molecules are shown in Fig. 1a. The first step in model construction is determining the protonation state of each molecule. The host could be confirmed to be net-neutral straightforwardly, while some care should be taken when dealing with the protonation forms of the guests. We summarize the experimental pKa of all guests in Table 1. The experimental pKa of G5 (Ketamine) is 7.5 and the predicted value from ChemAxon is 7.45. Both of these values are close to the experimental

condition that the binding affinities are measured (pH 7.4),⁸⁴ indicating the existence of acid-base equilibria under the experimental condition. Therefore, we consider both the protonated and deprotonated forms of the guest G5. The protonation forms of the other six guests could be easily determined according to the obvious differences between their pKa values and the experimental condition pH 7.4. The pKa values of these six guests are also predicted with ChemAxon, and the results are all higher than pH 8.5 (results not shown). Therefore, we have a total of 8 guests in the current calculation. Specifically, all 7 commonly abused drugs in their protonated forms and the deprotonated form of G5 are considered. The resulting net charges of the host and the guests and the experimental binding affinities⁸⁴ are summarized in Table 1.

Starting from the structures obtained from the online server, we derive atomic charges of the host and the guests with two widely accepted fixed-charge models, i.e., the AM1-BCC⁸⁵ model and the restrained electrostatic potential (RESP)⁸⁶ fitting. The former charge scheme performs AM1 optimization and derives atomic charges by combining the AM1 results with a correction term (bond charge correction, BCC) fitted to reproduce the HF/6-31G* electrostatic potential (ESP) for a set of small molecules,⁸⁵ while the latter performs B3LYP/6-31G* optimization, scans the ESP around the molecule with the Merz-Kollman scheme at the HF/6-31G* level, and fits the atom-centered charges with the two-step hyperbolic restraints/regularizations to avoid overfitting and provide a reasonable set of atomic charges.⁸⁶ In many cases, the predictions (e.g., binding thermodynamics and interaction patterns) obtained with these two charge models are similar, while in some cases obvious differences are observed.⁸⁷⁻⁹² The semi-empirical AM1 calculations are performed with the sqm engine in AMBER,⁹³ while for the ab initio QM calculations we use the Gaussian 09 program.⁹⁴ The comprehensive comparison of charge schemes in the current work would provide detailed evidences of their differences and the resulting impacts on the binding modes and thermodynamic profiles. The other missing parameters (e.g., vdW radius) are obtained from the general Amber force field (GAFF) force field.⁹⁵ TIP3P^{96, 97} water molecules are added and the truncated octahedron cell is replicated in whole space with periodic boundary conditions. Non-polarizable monovalent spherical counter ions^{98, 99} of Na⁺ or Cl⁻ parameterized for TIP3P water are added for neutralization. The starting configuration of the host-guest complex is obtained by simply superposing the center of masses (COM) of the host and that of the guest and letting the simulation equilibrate and relax the system.

2.2. Free Energy Simulations.

The CB8-guest interactions were found to be complex among host-guest systems investigated in our previous computational modelling.^{46, 82, 91} The guest molecules often have diverse structural features with many rotatable bonds that pose sampling problems for conventional MD simulations. Further, the binding free

energies of CB8-guest complexes are much higher than thermal fluctuations, which calls for the use of enhanced sampling techniques. More importantly, multiple binding poses are observed for CB8-guest complexes.⁴⁶ Therefore, in the current work, we use some enhanced sampling schemes to accelerate the exploration of the configurational space.

Here, we use the protocol that has been applied to similar host-guest cases in our previous work.^{46, 82} The enhanced sampling technique well-tempered metadynamics¹⁰⁰⁻¹⁰³ is used to enhance the sampling efficiency, and the spherical coordinates (ρ, θ, φ) defined by the COMs of the host and the guest are selected as the biasing coordinates. Illustrations of the sampled region and the 3D spherical coordinates reaction coordinate or collective variable (CV) are presented in Fig. 1b and Fig. S1, respectively. This protocol enables the scan of the relative position of the host and the guest molecules and thus the exploration of possible binding poses, which avoids the initial-configuration-induced systematic bias introduced in model building. An upper wall is added on the first CV (i.e., distance/radius ρ) to limit the configurational space explored during the enhanced sampling simulations. The maximum value of the radius ρ must be sufficiently large to define a fully decoupled state, and in the current work we put it on 23 Å. The entropic correction to recover the standard-state binding free energy is 2.0 kcal/mol. For more detailed discussions about this spherical-coordinates-biased protocol, please refer to our previous work on similar host-guest systems.^{46, 82} Although in the current work we bias the 3D spherical coordinates and sample a sphere around the host, other regimes are also usable. For instance, coupling the cylindrical restraint shown next to the spherical scheme in Fig. 1b with the 1D host-guest COM distance could also sample the binding/unbinding event in the physical space effectively. In that case, the explored region is limited to the poles of the spherical system and the guest cannot reach or bind to the outer sides of the host. Also, we can limit the sampling in a sub-region of the current sphere or the cylinder (e.g., the upper z^+ side). However, it should be noted that although the cylindrical restraint could save some computational resources due to the limitation of the sampled region, it may fail under some circumstances. For example, for chemically modified CB8 rings that stably coordinate guests at the outer face, the cylindrical restraint eliminates the possibility of the guest reaching the outer surface, making it impossible to explore side-binding poses. As a result, the explored center-binding pose may not be the global minimum, and the calculated binding affinity is wrong. Further, even in the current CB8 case, there is no guarantee that such side-binding poses are absent. Thus, we sample the spherical region in the current work to avoid such biases. A further altered scheme that couples the spherical and cylindrical restraints is shown next to the cylindrical scheme in Fig. 1b. In this case, both the side-binding and center-binding regions around the host could be

sampled, while the large- z regions are used to define the decoupled state. The resulting sampling results would be similar to the spherical scheme used in the current work. Another alternative is widening the cylindrical restraint to cover the host and leave some spaces for the guest to bind and using 2D distance CVs (i.e., the z -component and the in-plane xy -component of the host-guest COM distance), which would lead to similar results to the combined scheme. Another note about the sampling strategy is that we only bias the coordinates describing the relative position of the host and the guest molecules, while other CVs describing the internal degrees of freedom could also be added to the CV set in order to enhance the sampling efficiency of internal motion (e.g., dihedral flipping) in systems with complex intra-molecular interactions.

The contact numbers between different groups of atoms $C = \sum_{i \in \text{group}_A} \sum_{j \in \text{group}_B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}}$ with r_{ij} being the

distance between the i th atom in group A and the j th atom in group B and $r_0 = 6 \text{ \AA}$ are used to analyze the binding pattern. For instance, the total contact number between the host and the guest is given by

$C_{\text{host-guest}} = \sum_{i \in \text{host}} \sum_{j \in \text{guest}} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}}$. It can also be decomposed into the contribution of each host atom

$C_i = \sum_{j \in \text{guest}} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}}$ to investigate the detailed interaction pattern, e.g., the host atoms coordinating the guest.

When investigating the detailed motions of specific groups of the guest, such host-group contacts could be

calculated as $C_{\text{host-group}} = \sum_{i \in \text{host}} \sum_{j \in \text{group}} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}}$.

As mentioned in the system preparation section, the starting configuration is obtained by superposing the COMs of the host and the guest. From this structure, we perform 5000 steps minimization, 300 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 20, and the Gaussian widths of 0.1 nm, $\frac{\pi}{16}$, and $\frac{\pi}{8}$ for the three polar coordinates, respectively. The simulation is performed with GROMACS 2019.6¹⁰⁴ patched with PLUMED 2.7.0¹⁰⁵. The velocity rescaling algorithm¹⁰⁶ is employed for temperature regulation at 298 K and the Parrinello-Rahman barostat^{107, 108} is used for pressure regulation. A time step of 1 fs is used to propagate the dynamics. Long-range electrostatics are treated with the smooth Particle-mesh Ewald¹⁰⁹ method.

3. Result and discussion

3.1. Charge generation.

As discussed in the system-preparation section, the two charge models used in our work have many differences, e.g., the Hamiltonian for geometry optimization and the charge generation procedure. However, they share the same target, i.e., the HF/6-31G* ESP. Therefore, before comparing the thermodynamic profiles produced under the two charge schemes, we first check the differences between these generated atom-centered charges. As the atomic charges are configuration-dependent, we then compare the optimized structures in the two charge-generating procedures. The AM1 and B3LYP/6-31G* optimized structures are shown in Fig. S2. For the guests G4 and protonated G5, the AM1 and B3LYP/6-31G* optimized structures are virtually identical. For the guest G1 and G6, the optimized structures are extremely similar. For the other guest molecules, minor differences exist in some regions (e.g., rotation of some bonds). As for the host CB8, the structures optimized at different levels are extremely similar.

Then, we compare the qualities of the atomic charges produced by the two charge fitting procedures. The charge itself is not the target in charge generation, but the reproduction of the ESP around the molecule is the ultimate aim. Therefore, we then compare the deviation of the charge-generated ESP from the reference HF/6-31G* one to assess the quality of these atom-centered charges. The relative root-mean-squared error (RRMSE) of molecular ESP, the dipole moment and the number of points in the ESP scanning are given in Table 2. Overall, the statistics show that for all molecules under investigation, the RESP model provides a better description of the Coulombic molecular ESP and therefore should provide a more accurate description of intermolecular interactions. The AM1-BCC model is a bit worse in most cases, and significant deviations are observed for the large heterocyclic host CB8. For all the guest molecules, the RRMSEs of the two charge schemes are similar, and the situation is similar for the dipole moments. Below, we will provide detailed discussions about the fitting quality.

For the host molecule, the AM1 and B3LYP/6-31G* optimized structures are extremely similar. However, although the dipoles under the two charge schemes are almost identical, the ESP RRMSE under the AM1-BCC charge model is significantly larger than RESP, which indicates that the corrected semi-empirical charge scheme is unsuitable for macrocycles. As the training set of the AM1-BCC model does not include such large heterocyclic species, refitting some parameters for macrocycles could be useful to improve the general performance of the BCC model. For G4 and protonated G5, the AM1 and B3LYP/6-31G* optimized structures are virtually identical, and the AM1-BCC and RESP results are also extremely similar. For the guests G1 and G6, the optimized structures at the two levels are extremely similar, and the charge qualities of the two charge schemes are also very similar. For the guests G2, G3 and G7, the optimized structures at the two levels show some differences, but the fitted charges with the two schemes are still of similar qualities. These observations suggest that the AM1-BCC model is a useful tool for small drug-like molecules, and is able to produce high-quality atomic charges comparable to the RESP scheme. A worth noting failure of both charge schemes is the deprotonated form of G5. The RRMSEs of both charge schemes are very large, but the dipoles of the two charge models are very similar. However, the protonated G5 does not seem to cause problems in charge fitting. This phenomenon should arise from the inability of the atom-centered charge model to provide an accurate description of the lone electron pairs in the neighborhood of N and Cl atoms. Specifically, the lone pairs between N and Cl atoms cannot be properly described with atom-centered charges at these two atoms. Note that the lone pair of nitrogen is more problematic due to the anisotropic distribution and is the main accuracy-limiting factor in the current fitting. Upon the protonation of G5, the addition of one hydrogen atom on the N atom and the formation of the intramolecular N-H \cdots Cl hydrogen bond introduce another fitting point between N and Cl atoms, thus significantly improving the fitting quality. We have also tested a larger basis set 6-311G** and empirical dispersion corrections (e.g., D3) in geometry optimization and ESP scan, which do not change the optimized structure or improve the fitting quality for the deprotonated G5. Therefore, this ESP-reproduction failure indicates that more accurate charge models (e.g., polarizable force fields) could be used. Adding extra charges or fitting centers between N and Cl atoms to emulate the lone pairs could also help to improve the fitting quality. Another worth noting but commonly overlooked solution to the above lone pair problem is separating the charge and vdW centers. Specifically, the numbers of charges and vdW centers remain unchanged, but the charges of some atoms (e.g., N) and the corresponding vdW centers could be separated in the parametrization procedure, similar to the fitting procedures of optimal fixed-charge models for water molecules.¹¹⁰ In this way, more degrees of freedom are added and the molecular properties could be represented more accurately.

It should be noted that fixed-charge models have limitations and could be improved in many aspects. As the molecular ESP is conformation-dependent, taking multiple conformations into consideration in charge fitting could be important. Another worth noting point is that the ESP generated at the HF level with the 6-31G* basis set in vacuo may not be a good reference but relies on fortuitous error cancellation. On this aspect, generating the ESP in implicit solvent models and using a higher-level *ab initio* method and a larger basis set could be helpful. However, even with a perfect match of the charge-generated ESP and the reference, the polarization effect is not fully accounted for. Although it is not significant in many cases, more sophisticated models are needed when the environmental perturbation is large.¹¹¹⁻¹¹⁶ However, it should be fair to point out that polarizable models are relatively computationally demanding and the time scale affordable with modern computational resources may not be sufficient for converged phase space sampling.

3.2. Binding modes.

We then turn to the statistics obtained during enhanced sampling simulations. In Fig. 2, we present the time series of the total contact number and its decomposition into the contributions of each host atom (i.e., atom-guest contacts) for the CB8-G3 complex. Note that here we include all atoms of the host in the calculation of the contact number, while only heavy atoms are included in the later projection of free energy landscapes. The total contact number in Fig. 2a shows significant fluctuations, reaches its maximum around 2000 and approaches zero many times, which suggests that the host-guest complex experiences multiple binding and unbinding events. The by-atom decomposition in Fig. 2b is also informative. We can identify the host atoms directly involved in the coordination of the guest. We can further get some insights into the binding poses from the total and by-atom decomposition contacts. For instance, in the cyan oval on these plots, the total host-guest contact number approaches its maximum, and all host atoms are in close contact with the guest. These observations suggest a bound state, and the guest stays at the center of the CB8 ring. Another binding pattern could be observed in the magenta oval, where the total contact number fluctuates around an intermediate value and only some of the host atoms are coordinating the guest. These phenomena suggest a side-binding pose, where the guest interacts with the outer surface of the CB8 ring. Note that such side-binding configuration is not observable when the narrow cylindrical restraint is applied, as discussed in the previous section. The combination scheme (i.e., the green one in Fig. 1b), by contrast, could still access this point in the configurational space. From Fig. 2b, we can similarly identify many other binding positions that are sampled during the 1000 ns spherical-coordinates-biased simulation. The time series of the total host-guest contact numbers and the by-host-atom decompositions for other CB8-drugs systems under the two charge

schemes are presented in Fig. S3 and Fig. S4, where similar phenomena could be observed. Thus, our enhanced sampling simulations effectively sample the binding/unbinding event and explore the possible binding poses.

We strip the first 400 ns of the metadynamics simulation and use the 400-1000 ns trajectory to reweight the data. The statistical error or the standard deviation (SD) of the binding thermodynamics is obtained with block analysis. The free energy surfaces are projected onto the radius-contact ($\rho - C_{\text{host-guest}}$) CVs to investigate the binding pose and the binding pathway. The AM1-BCC surfaces are presented in Fig. S5, while the RESP ones are shown in Fig. S6. Representative structures are extracted from free energy minima to present detailed host-guest interactions. The spacing in the color bar is set as 4 kcal/mol in order to present the main features on the free energy landscapes of all host-guest systems clearly. This spacing could be a bit large to differentiate binding poses with similar thermodynamic stabilities inside the free energy basins. Therefore, only basic binding conformations are discussed here, while more detailed investigations of different binding modes would be provided later.

Before proceeding, we would spend some time discussing a useful check for sampling convergence. Often, the time-dependence of the binding affinity is estimated in convergence check. When the free energy of binding does not vary with further sampling, the free energy simulation is considered to be converged. Here, we present another useful option: extracting the most stable structures in specific states of interest from the trajectories and checking whether they are close in the time series. We use the bound state as an example to illustrate this convergence check procedure. For a well-converged sampling, the bound and unbound states should be visited multiple times, which indicates that the most stable bound structures should be reached many times. Therefore, the most stable binding poses should be observed many times, and these observations should be separated in the time series. A practical example is the CB8-G1 complex under the RESP charge scheme. The top-6 bound configurations with the highest thermodynamic stabilities extracted from the global free energy minimum are observed at 978.152 ns, 977.048 ns, 424.897 ns, 640.019 ns, 425.358 ns, and 642.897 ns, respectively. These observations are separated in the time series and could be divided into at least three binding/unbinding events, which indicates that the binding affinity is determined by multiple binding/unbinding observations rather than a specific shot. Similar conclusions could also be reached in the time series of the total host-guest contacts and their by-atom decompositions presented in Fig. S4. Therefore, the estimated binding affinity in our 1000 ns 3D spherical-coordinates-biased enhanced sampling simulation is statistically reliable.

Then, we discuss the AM1-BCC surfaces in Fig. S5. The structure of the guest G1 is the simplest one among all guests under study. As a result, the free energy surface of the host-guest system is simple. Only one

free energy minimum is presented. The most stable binding pose extracted from this global minimum features the whole guest staying at the center of the host. All guest atoms are in close contact with the host, and the host conformation becomes distorted to hold the guest tightly. As the distances between the hydrogen bond donor (i.e., -NH_2^+ of the guest) and the -C=O portals are larger than the normal 3.5 Å criterion for hydrogen bonds in this configuration, no hydrogen bond is formed. However, we should note that this configuration is just a low-energy snapshot extracted from the free energy basin. The structures in the bound ensemble could possibly include ones that form host-guest hydrogen bond(s). Thus, the formation of host-guest hydrogen bonds remains possible.

The guest G2 is structurally complex, which is also reflected on its free energy surface. Two minima are observed on this free energy surface. The free energy minimum that has large host-guest contacts features center-binding and the intra-molecular stacking between phenyl rings with two possible orientations. In the first orientation, the hydrogen bond donor -NH^+ directly interacts with the -C=O portal of the host and forms one inter-molecular hydrogen bond. In the other orientation, the hydrogen atom of the -NH^+ group stays in the pocket formed by two phenyl rings of the guest, and thus no inter-molecular hydrogen bond is formed. Obviously, the former orientation is more stable than the latter. The other free energy minimum on the $\rho\text{-C}$ surface has smaller host-guest contacts and larger host-guest COM distances. It is about 2 kcal/mol more stable than the previous center-binding minimum. Therefore, this free energy basin contains the most stable binding mode of the CB8-Fentanyl complex. The binding poses extracted from this free energy minimum contain bound structures with diverse features. For instance, instead of having the whole guest at the center of the host, two structures extracted there have the guest binding to the outer surface of the host, and one of the two phenyl rings of the guest is close to the center of the host. This phenomenon suggests the existence of an ensemble of interconverting conformational states. This also agrees with the great width of this free energy basin. Note that although this large-distance minimum features the extended guest clinging to the outer surface of the squashed host, this is not a side-binding mode but an altered center-binding one. In our later projection of the results on host-guest and host-aromatic contacts, we will show that there are several host-guest binding patterns in this bracelet conformation, where the guest forms a semi-closed pocket itself and holds one side of the CB8 ring.

The aromatic heteropolycyclic G3 and G4 are structurally similar. The ring strain in these polycyclic molecules limits intra-molecular orientations/rotations. The exploration of the configurational space is actually sampling the rigid-body motion of joined rings and the local rotations of functional groups. The resulting free energy surfaces of these two host-guest systems are similar. Their free energy basins are not

wide, and only one binding pose exists for each system. For G3 Morphine, the -NH^+ and two -OH groups form inter-molecular hydrogen bonds with -C=O portals of the host. For G4 Hydromorphone, the -NH^+ and only one -OH groups form hydrogen bonds with the -C=O portals. In the bound state, as the host cavity is larger than the guest, the host becomes distorted to be elliptical to hold the guest tightly.

The free energy surfaces of the protonated and deprotonated forms of G5 are similar. Only one free energy minimum exists for each system. For the protonated G5, its chlorophenyl ring is parallel to the other 6-membered ring of the guest, and its -NH_2^+ group stays at the center of the host, forming a typical center-binding mode. Inter-molecular hydrogen bonds are formed between -NH_2^+ of the guest and the -C=O portals of the host. As for the deprotonated G5, the -NH group forms a hydrogen bond with -C=O portal with two energetically favorable interaction patterns/orientations. Specifically, the hydrogen atom of the -NH group could be on the z-upper or z-lower side of the oxygen atom of the -C=O portal when forming the inter-molecular $\text{-N-H}\cdots\text{O=C-}$ hydrogen bond.

G6 Phencyclidine is structurally complex. Compared with G2 Fentanyl, the three 6-membered rings of Phencyclidine are separated by only one chemical bond and thus are spatially close to each other. Similar to the guest G2, there are two free energy basins on the surface. The structure in the center-binding minimum at the large-contact small-distance position features two of the three rings penetrating the cavity, while the large-distance small-contact minimum has more parts of the guest outside the host. Unlike G2, the host does not collapse in the bound state of the large-distance wide free energy basin, and the guest interacts with -C=O portals and does not reach the outer side of the host. This phenomenon arises from the fact that the three 6-membered rings of G2 are separated by several bonds that enables this longer guest to form a semi-closed pocket holding one side of the host, which is impossible for the relatively short Phencyclidine.

The last guest Cocaine G7 is formed by two rings. Only one free energy minimum is observed on the $\rho\text{-C}$ free energy surface. The structure extracted from this minimum features center-binding, the extended guest with solvent-exposed phenyl and -COOMe groups at different sides of the host, and the distorted CB8 ring coordinating the guest tightly. However, it should be noted that the free energy basin of the bound state for Cocaine is much wider than the previous cases, which suggests that the bound state of CB8-G7 could include many substates/conformations.

The above AM1-BCC results provide a picture that the host-guest complex has an ensemble of interconverting conformations in the bound state. The interaction pattern is related to the host flexibility, the structural features of the guest, the orientations of functional groups, and the guest fluctuations inside the bound state. The electrostatics of the molecules influence the inter-molecular interactions. As discussed in the

charge-generation section, the Coulombic molecular ESPs produced by RESP charges differ from the AM1-BCC ones, and the differences are significant for the host CB8 and the guest deprotonated G5. As the host CB8 is involved in all host-guest pairs, variations of its atomic charges would have dramatic impacts on host-guest interactions. Now, we would discuss the RESP results and compare them with the AM1-BCC ones. For the structurally simple G1, the $\rho-C$ free energy surface under the RESP charge scheme is similar to the AM1-BCC one. Only one minimum exists, and this center-binding pose has the whole guest penetrating the host cavity.

Similar to the AM1-BCC results, the structurally complex G2 has two conformational states in general. The large-contact state is still a typical center-binding mode with the whole guest penetrating the host cavity, which is similar to the AM1-BCC case. However, under the RESP charge scheme, this center-binding pose is of lower stability (i.e., higher relative free energy). The other conformational state has one end of the guest close to the center of the host, and the other parts of the guest clinging to the external side of the host. The central cavity of the host becomes squashed to maximize the host-guest contacts/interactions. Still, the bound structures extracted from this wide free energy basin have diverse features, which suggests the existence of an ensemble of conformations here. The great width of this free energy basin tells a similar thing.

The situations of guests G3 and G4 with RESP charges are similar to the AM1-BCC cases. The limitation of their cyclic structures leads to the rigid-body motion of the guest backbone. Their limited conformational flexibilities only include orientations of functional groups (e.g., $-NH^+$). For the guest G3, its $-NH^+$ and two $-OH$ groups form inter-molecular hydrogen bonds with the $-C=O$ portals of the host, but the distances between hydrogen bond donors and acceptors are slightly larger than the AM1-BCC case, leading to weaker hydrogen-bonding interactions. For the guest G4, its $-NH^+$ and $-OH$ groups form inter-molecular hydrogen bonds with the $-C=O$ portals of the host, and the host becomes elliptical to hold the guest tightly.

The protonated and deprotonated forms of G5 under the RESP charge scheme also have similar behaviors to the AM1-BCC cases. Only one free energy minimum is observed on their radius-contact surfaces. The detailed structures extracted from the minimum show some differences compared with the AM1-BCC ones. Specifically, the representative structure extracted for protonated G5 has the chlorophenyl ring perpendicular to the other 6-membered ring, and the $-NH_2^+$ group stays at the center of the host. For deprotonated G5, the top-6 stable configurations extracted from the global minimum only have one energetically favorable orientation for $-NH$. Specifically, the $N \rightarrow H$ and the $C \rightarrow Cl$ vectors share the same z-direction.

The free energy surface of the guest G6 Phencyclidine shows somehow different behaviors compared with the previous AM1-BCC result. Only one free energy minimum exists on the radius-contact surface, and

this minimum is also observed under the AM1-BCC charge scheme. In this bound state, two of the three 6-membered rings penetrate the host cavity. Two representative structures are extracted from this minimum. In the first bound structure, the phenyl and piperidinium moieties stay at the center of the host, and the cyclohexyl ring remains solvent-exposed. In the second bound structure, both phenyl and cyclohexyl moieties penetrate the central cavity, and the piperidinium ring remains in solvent. The -NH^+ group forms inter-molecular hydrogen bond with the -C=O portal. The thermodynamic stability of the second pose is slightly higher than the first one, and is extremely similar to the bound structure extracted from this minimum under the AM1-BCC charge scheme. Therefore, the two charge schemes provide similar host-guest interactions for this bound state. The phenomenon that bound structures with high thermodynamic stabilities extracted from the same conformational state have different structural features suggests that this bound state could include many microstates, which would be investigated in the next section.

For the last guest G7, only one free energy minimum is observed on the free energy surface. The conformational state features center-binding, the phenyl and -COOMe groups staying at the center of the host, and the distorted host coordinating the guest. It should be noted that the guest conformation is very similar to the structure used in charge generation (i.e., ESP scan), which suggests that the RESP charge scheme has a tendency to limit the conformational change of the guest and only the neighborhood of the initial configuration is energetically favorable.

Overall, the observations in these host-guest complexes under the two fixed-charge schemes show a picture that the distorted host holds the guest in different ways. For relatively simple and rigid molecules, only one free energy minimum exists on the $\rho-C$ free energy surface. For structurally complex guests, aside from the typical center-binding mode, other interaction patterns could exist. For instance, the guest could cling to the host and forms a half-closed pocket holding one side of the host. An ensemble of interconverting conformations could exist in these bound states, which emphasizes the dynamic feature of host-guest complexes. In the following section, we investigate the conformational ensemble in the bound states in great detail to understand the binding modes of CB8-guest complexes and provide a general picture of CB8-guest interactions.

3.3. Conformational fluctuations in the bound state.

The radius-contact surfaces presented above provide thermodynamic descriptions of the binding/unbinding event. Structural features of the bound conformation(s) are also extracted from the global and local minima to investigate the detailed interaction network within CB8-drug complexes. However, the

free energy basins defining the bound states are wide in some host-guest systems, which suggests the existence of an ensemble of bound conformations. Further, multiple binding modes are observed in some cases. The intrinsic complex structural features of the guests under investigation also make the interaction pattern complicated. Therefore, the total host-guest contacts may not differentiate all interaction patterns, and a more detailed description of host-guest interactions is needed. As all guests under investigation have at least one aromatic ring formed by 6 carbon atoms, we select the number of contacts between the aromatic 6-membered ring of the guest and the host as another CV to analyze detailed binding patterns of these host-guest complexes. Still, only heavy atoms are included in this calculation. There are two phenyl rings for Fentanyl G2, and the selected phenyl ring is connected to the amide bond. The resulting free energy surfaces projected on $C_{\text{host-guest}} - C_{\text{aromatic}}$ or $C - C_{\text{Ph}}$ are presented in Fig. 3 (AM1-BCC) and Fig. 4 (RESP charges).

We first discuss the AM1-BCC results shown in Fig. 3. Even for the structurally simple G1, the $C - C_{\text{Ph}}$ surface has a wide free energy minimum. If we use 2 kcal/mol increments for the color bar, there would be at least two minima in the bound state. Representative structures with high thermodynamic stabilities are extracted from these minima. The minimum with larger host-phenyl contacts is less stable (about 1 kcal/mol) than the other minimum. The main difference between the large- C_{Ph} and small- C_{Ph} states lies in the degree of distortion of the CB8 ring. When the CB8 ring is not distorted significantly, the host-phenyl contact is relatively small, and more guest atoms could penetrate the cavity. As a result, this small- C_{Ph} basin is wider on the C -axis and the fluctuation of the guest in this bound state is larger. For instance, the representative structure extracted from this minimum has the whole guest in the central cavity. Some other snapshots from this minimum (not shown) suggest that parts of the guest (except phenyl) could leave from the cavity. If only the phenyl group is in the cavity, the total host-guest contact would decrease but the host-phenyl contact remains almost unchanged. By contrast, in the large- C_{Ph} state, the CB8 ring is distorted significantly to hold the phenyl ring tightly. As a result, the host-phenyl contact is maximized, the volume of the cavity is decreased, and only phenyl could penetrate the host cavity. Thermodynamically, the small- C_{Ph} less distorted state is more stable, which suggests that the backbone torsion of the host could also influence the host-guest conformation. The above behaviors of this host-guest system indicate that the host-guest complex could have a diverse set of conformational states, even for structurally simple guest molecules. The conformational fluctuations of both the host and the guest play an important role and should be properly accounted for in host-guest modelling.

The structurally complex G2 Fentanyl is a challenging system in computational modelling, and multiple free energy minima are observed on the $C - C_{ph}$ surface. We could generally divide the conformational states into 2 groups. The first group has larger host-phenyl contacts, which suggests that the phenyl ring directly connected to the amide bond stays at the center of the host. Three substates exist in this macrostate. The state with the largest host-guest contacts is the typical center-binding mode, where the whole guest stays at the center of the host. The two phenyl rings form intra-molecular stacking interactions inside the host cavity. This center-binding bound conformation differs from the other conformational states, where only parts of the guest penetrate the host cavity. The other two states with the smaller host-guest contacts among the large- C macrostate are similar. In both states, the phenyl ring connected to the amide bond stays at the center of the host, and the other parts of the guest nestle against the host. The state with larger host-guest contacts features the other phenyl ring in direct interactions with the host, while in the conformational state with the smallest host-guest contacts, the 6-membered ring in the middle of the guest is in direct interactions with the host and the phenyl ring outside the cavity is relatively far from the host. For the macrostate with smaller host-phenyl contacts, two microstates could be identified. Both of them have the phenyl ring connected to the amide bond outside the cavity, but the other phenyl group stays at the center of the host. Namely, the two phenyl rings swap their positions compared with the large- C_{ph} states. Despite the similarities of the small- C_{ph} microstates and the large- C_{ph} ones, differences exist due to the asymmetric structural feature of the guest. Specifically, the guest in the small- C_{ph} microstates tends to form a half-closed bracelet-like pocket and hold the host tightly, which leads to a higher degree of distortion of the CB8 ring compared with the large- C_{ph} states. The difference between the two small- C_{ph} microstates lies in the extension of the guest. For a more extended guest, the total host-guest contact is smaller. The above complex binding patterns in the CB8-Fentanyl complex emphasize another factor that plays a crucial role in host-guest interactions, i.e., the guest structural feature. Aside from the typical center-binding mode commonly observed in host-guest systems, the guest itself could form a pocket and coordinate the guest, forming a bracelet-like bound conformation. This factor, along with the conformational fluctuations of the host and the guest, depicts a complex host-guest interaction picture that cannot be properly described with a single or several bound configurations.

The guests G3 Morphine and G4 Hydromorphone are structurally similar, as discussed in the previous section. Their joined rings limit the conformational flexibility, and only local fluctuations/orientations of

functional groups are possible. As a result, the $C-C_{ph}$ surfaces of these systems are rather simple. Only one minimum is observed for each system and corresponds to the typical center-binding pose that the whole guest stays at the center of the host and forms stable inter-molecular interactions (e.g., hydrogen bonds).

The guest G5 Ketamine has two 6-membered rings. Its protonated and deprotonated forms show similar behaviors. We first discuss the protonated G5. On the $C-C_{ph}$ surface, the low-energy regions form a crescent-shaped macrostate and three local minima could be identified. The representative structures extracted from these minima suggest the existence of significant conformational fluctuations in the crescent macrostate. In the microstate with the largest host-guest and host-phenyl contacts, the whole guest is in the cavity, which corresponds to the typical center-binding pose. However, such host-guest coordination is flexible, and one of the two 6-membered rings could exit the host cavity and becomes solvent accessible. The free energy surface of the deprotonated G5 shows some differences, but the general crescent landscape remains unchanged. The protonation/deprotonation alters the conformational fluctuations of the guest inside the host cavity to some extent.

The guest G6 Phencyclidine is another structurally complex guest formed by three 6-membered rings in the current study. Three free energy minima are observed on its free energy surface. The state with the largest host-guest and host-phenyl contacts is the typical center-binding one, where the whole guest penetrates the host cavity and the CB8 ring is distorted to hold the guest tightly. Among the three 6-membered rings, the phenyl and piperidinium moieties are closer to the center of the cavity, while the cyclohexyl ring is relatively solvent-exposed. Both of the other two states have smaller host-guest and host-phenyl contacts, suggesting looser host-guest coordinations. The piperidinium ring of the guest stays at the center of the host in these two states, and their differences mainly lie in the orientation of the guest molecule. In the state with smaller host-guest contacts, the phenyl moiety is close to the center of CB8, and the $-NH^+$ group is in the right position to form inter-molecular hydrogen bond with the $-C=O$ portal of the host. By contrast, in the other state, the cyclohexyl moiety is closer to the central cavity, and $-NH^+$ is impossible to form hydrogen bonds with the host. Note that there is an obvious free energy barrier between the small- C states and the typical center-binding one, which arises from the distortion of the CB8 ring to fit the guest. Such ring strain involves relaxation of internal stiff degrees of freedom (e.g., internal angles) and thus could be difficult to achieve thermodynamically.

Multiple minima are also observed on the free energy surface of the last guest G7 Cocaine. The four conformational states differ in the penetrating groups and the orientations of functional groups. The states

with larger host-phenyl contacts have the phenyl group at the center of the host, while in the other two states the phenyl group is outside the cavity and the other parts of the guest penetrate the host. In the state with the largest host-phenyl contacts, both the phenyl and the -COOMe groups penetrate the host cavity, and -NH⁺ is close to the -C=O portal of the host, forming inter-molecular hydrogen bond. In its neighboring conformational state, another orientation of the guest is taken. The penetrating groups are still phenyl and -COOMe, but the -NH⁺ is close to the center of the host cavity and does not form inter-molecular hydrogen bond with the host. The structures extracted from these two large-*C* states are similar to the charge-generation configuration. The other two states with smaller host-phenyl contacts feature the phenyl group outside the cavity. Among them, the guest conformation in the state with smaller host-guest contact is similar to the charge-generation configuration. Namely, both phenyl and -COOMe are outside the cavity and are close to each other, and the other part of the guest penetrates the host cavity. In the other state, only phenyl stays outside the cavity and the other parts of the guest lie at the center of the host. This guest conformation differs from that used in charge generation, and thus is explored during enhanced sampling simulations.

Overall, the above observations tell us that host-guest complexes have rather complicated bound conformations. Even simple guest molecules could form diverse binding poses with the symmetric host. Conformational fluctuations of both the host and the guest play crucial roles here. For structurally complex guests, further complicated binding modes could be formed. Modelling results show significant dependences on the employed Hamiltonian/model. As the AM1-BCC ESP reproduction is less well compared with RESP charges especially for the host CB8 that is involved in all host-guest pairs under investigation, whether these observations are still valid under the RESP charge scheme should be investigated. However, it should be noted that the RESP charges do not necessarily outperform AM1-BCC charges, as the initial configuration used in charge generation could differ from the finally observed guest conformation in the bound state. Therefore, we focus on whether the multi-state and highly dynamical behaviors of the host-guest systems remain when the RESP charge scheme is applied.

We then turn to the RESP surfaces shown in Fig. 4. Compared with the previous AM1-BCC case, the free energy surface of the CB8-G1 complex under the RESP charge scheme has similar low-energy regions, but the detailed patterns show some differences. Two free energy basins could be identified in general, and their positions differ from the AM1-BCC case. The state with smaller host-guest and host-phenyl contacts has a relative free energy of about 1 kcal/mol higher than the other state. In this state, only the phenyl ring penetrates the host cavity, and the distortion of the CB8 ring is low. As the other minimum is quite wide, we

extract two structures from it. The first structure has a similar C but a larger C_{ph} compared with the previous small- C small- C_{ph} state. The main structural features of this state include the phenyl penetrating the central cavity and a more distorted host molecule. The thermodynamic stability of this conformation is very similar to that of the previous state, which suggests that the distortion of the host does not have significant impacts on the host-guest binding thermodynamics. The second structure extracted from this state has larger host-guest and host-phenyl contacts. This conformation is observed to have a largely distorted CB8 ring and the whole guest penetrating the cavity. Considering the above detailed comparison of various aspects of the CB8-G1 binding under the two charge schemes, the host-guest interactions with these two charge sets are similar and only minor differences exist. The general picture of the CB8-G1 binding is the distorted CB8 ring holding parts of or the whole guest with significant conformational fluctuations of both the host and the guest molecules.

Significant differences between the surfaces under the two charge schemes are observed for the structurally complex G2 Fentanyl. Four conformational states could be identified with RESP charges, and the locations of these minima differ from the previous AM1-BCC results to some extent. Three macrostates could be identified according to the degree of penetration of the selected phenyl group. The macrostate with the largest host-phenyl contact could be further divided into two microstates. One of them is the typical center-binding mode, where the whole guest stays at the center of the host. Intra-molecular stacking and intermolecular hydrogen bonds are observed in this state. Compared with the AM1-BCC case, this center-binding minimum is much narrower, which indicates a lower thermodynamic stability and a lower level of conformational fluctuations. The other microstate has smaller host-guest contacts and is wider, suggesting the existence of significant conformational fluctuations here. This conformational state is also observed under the AM1-BCC charge scheme. Similar to the AM1-BCC case, this microstate has the phenyl group connected to the amide bond penetrating the central cavity and the other parts of the guest clinging to the host. The location of the macrostate with intermediate values of host-phenyl contacts corresponds to two local minima under the AM1-BCC charge scheme, but only one wide free energy minimum is observed with RESP charges. Similar to the AM1-BCC case, in this macrostate, the selected phenyl group binds to the outer surface of the host, while the other phenyl group penetrates the cavity. The guest generally nestles against the host. The macrostate with the smallest host-guest and host-phenyl contacts is a new state that is unstable under the AM1-BCC charge scheme. In this conformation, only the phenyl group at the $-NH^+$ side penetrates the cavity, the $-NH^+$ group forms inter-molecular hydrogen bond with $-C=O$ portal, and the other

parts of the guest stay away from the host. The above comparison of the free energy surfaces under the two charge schemes tells us that the atomic charges could have a significant impact on the binding pattern/thermodynamics. Under different charge schemes, the same binding mode could be of different thermodynamics stabilities, and different binding modes could be preferred. The above observations of the complexities of host-guest interactions, similar to the AM1-BCC case, suggest that complex binding patterns could be formed for structurally complex guest molecules. Computationally modelling of host-guest complexes should properly treat such an ensemble of bound conformations in order to obtain a practical and accurate description of host-guest interactions.

The structurally similar G3 Morphine and G4 Hydromorphone have similar free energy surfaces under the AM1-BCC charge scheme and only one minimum is observed for each system due to their limited conformational flexibility. In the current RESP case, G3 Morphine still has only one free energy minimum, but G4 Hydromorphone is observed to have two stable bound conformations. The center-binding mode is still the most stable binding pose, where the whole guest stays at the center of the host, the -OH group of the guest forms inter-molecular hydrogen bond with -C=O portal, and -NH⁺ is close to one -C=O portal with the donor-acceptor distance larger than the normal 3.5 Å criterion for hydrogen bonds. The other newly observed binding pose of the CB8-G4 complex is similar to the global minimum, but the -NH⁺ group of the guest is perpendicular to the CB8 ring and far from the host. These observations suggest that even for host-guest complexes formed by quite rigid molecules with limited conformational flexibilities, the differences between atomic charges could lead to different interaction patterns. Therefore, the atomic charge issue needs to be treated appropriately in host-guest interactions.

We then turn to the protonated and deprotonated forms of G5 under the RESP charge scheme. Compared with the AM1-BCC results, the crescent low-energy region on the $C - C_{ph}$ surface for both forms of G5 remains unchanged generally. For the protonated G5, the most stable bound conformation still has the two 6-membered rings penetrating the cavity. However, the low-energy region is narrower, which indicates that the conformational fluctuation is smaller and the bound conformation is locked in the typical center-binding pose. This phenomenon suggests that although the two charge schemes reproduce the ESP around the guest molecule in the configuration used in charge generation, they could produce obviously different descriptions of host-guest interactions. The RESP charges stabilize the typical center-binding mode, while AM1-BCC charges describe a host-guest complex with a higher level of conformational flexibility. The free energy surface for the deprotonated G5 is also similar to the AM1-BCC one. We extract two structures with obviously different host-phenyl contacts to illustrate the structural feature of the bound state.

The chlorophenyl ring stays at the central cavity and -NH forms inter-molecular hydrogen bond with -C=O portal. The difference between the two structures lies in the degree of penetration of the phenyl ring. In the configuration with larger host-phenyl contacts, phenyl is very close to the center of the cavity and is less solvent-exposed compared with the other parts of the guest.

The structurally complex G6 Phencyclidine is another guest that has significantly different free energy landscapes under different charge schemes. Unlike the previous AM1-BCC case, only two minima are observed on the RESP surface. These two states have similar host-guest contacts, but their host-phenyl contacts are significantly different. As the guest is formed by three 6-membered rings, the two states should differ mainly by groups/rings penetrating the host cavity. Extraction of the representative structures with the highest thermodynamic stabilities is performed for each of the two conformational states. In the state with larger host-phenyl contacts, typical center-binding behaviors are observed. Specifically, the whole guest is in the cavity, with phenyl and piperidinium closer to the cavity center than the cyclohexyl moiety. Note that this binding pose is also observed under the AM1-BCC charge scheme. By contrast, the other conformational state with smaller host-phenyl contacts is newly observed and unstable under the AM1-BCC charge scheme. In this conformation, the piperidinium and cyclohexyl moieties penetrate the cavity and the phenyl ring is solvent-exposed. Intra-molecular hydrogen bond is not formed in this pose, as -NH⁺ is at the center of the cavity and the N-H bond is almost perpendicular to the CB8 ring and far from -C=O portals. The above observations of the structurally complex G6 Phencyclidine again suggest that the charge scheme could have a significant impact on host-guest interactions and the phenomenon is more pronounced for structurally complex molecules.

The last guest G7 Cocaine, despite its complex structural features, has a rather simple free energy landscape with RESP charges. This phenomenon is in sharp contrast to the AM1-BCC case, where four conformational states with different penetrating groups and different orientations of functional groups are observed. In this only stable bound conformation, both phenyl and -COOMe groups penetrate the cavity, and -NH⁺ forms inter-molecular hydrogen bond with -C=O portal. This pose is also observed under the AM1-BCC charge scheme and is very similar to the configuration used in charge generation, which suggests that the RESP charge scheme may have a tendency to limit the conformational flexibility and thus the sampling in the neighborhood of the initial configuration. These observations indicate that all parts of the charge-generation procedure (i.e., configuration and charge scheme) are crucial and could lead to significantly different thermodynamic landscapes in host-guest interactions.

The above observations of host-guest bound conformations provide a general picture of CB8-guest

interactions. Although the host CB8 is structurally simpler than biomolecular receptors such as proteins and nucleotides, the host-guest interactions that it could form are rather complex. Many (meta-)stable conformational states could be observed on the free energy landscapes describing the host-guest binding. The conformational fluctuations of both the host and the guest contribute to the dynamical behavior of host-guest complexes. Even for relatively simple and rigid molecules, multiple bound conformations of similar thermodynamic stabilities could be observed. Although these binding poses often have the center-binding feature, their detailed interaction patterns differ due to different degrees of guest penetration and different levels of distortions to the host. For example, the simplest guest considered in the current work, Metamfetamine, is observed to form multiple binding modes with the CB8 host. The main differences between these binding poses include the degree of penetration and the level of distortion of the CB8 ring. For structurally complex guests, more complex binding landscapes are formed and various types of binding poses are observed. For instance, the guest Fentanyl could form a half-closed pocket itself and nestle against the CB8 host, forming a bracelet-like bound conformation. It could also insert its two phenyl rings inside the central cavity of the host and form a typical center-binding mode. Different orientations of specific functional groups could also lead to different binding conformations. A case in point is the guest Phencyclidine, where two binding poses on the $C - C_{ph}$ free energy surface differ mainly in the orientation of the $-NH^+$ group. Further, the bound state could endure significant fluctuations. The guest is not restrained in the central cavity tightly but experiences obvious fluctuations. Some parts of the guests thus become more solvent-exposed. Such fluctuation also helps the transformations between different conformational states. The above behaviors suggest that the CB8-guest interactions involve a dynamical ensemble that cannot be properly described with a subset of conformations. Therefore, static calculations based on a single or several structures cannot properly describe host-guest binding. Another important issue that is highly related to the dynamic feature of host-guest binding is the Hamiltonian/model used in computational modelling. Different models could lead to significantly different thermodynamic profiles in host-guest binding. Further, as the configuration used in parameterization could significantly differ from the other configurations explored during extensive sampling, the parameter set generated initially could be inaccurate for the newly observed conformations. Therefore, iterative refitting of these parameters could be useful. The initial exploration of host-guest interactions could serve as a preliminary guess, and the stable host-guest conformation could be extracted and used to refit a new parameter set that provides more accurate descriptions of host-guest interactions. Such a trial will be presented in Section 3.5 after the analysis of the predicted binding thermodynamics.

3.4. Binding thermodynamics.

After finalizing the binding mode investigation, we analyze the calculated binding thermodynamics and compare them with the experimental references.⁸⁴ The free energy of binding is calculated as the sum of the free energy difference between the global minimum and the zero-contact unbound state and the volume correction, as discussed in the computational details section.

The predicted binding affinities and the corresponding statistical errors (standard deviation, SD) under the two charge schemes are summarized in Table 3. Three error metrics including the mean signed error (MSE), the mean absolute error (MAE), and the root-mean-squared error (RMSE) are used to assess the deviations of the predictions from the experimental reference, and two ranking metrics including Kendall's rank correlation coefficient (τ) and Pearlman's predictive index (PI) are used to assess the consistency of the predicted rank of binding affinities and the experimental rank. As the experimental and ChemAxon predicted pKas (i.e., pH 7.5 and pH 7.45, respectively) for G5 are very close to the pH condition (i.e., pH 7.4) that the experimental binding affinities are measured, the protonated and deprotonated forms of G5 are considered to be approximately equally populated. Thus, the final calculated binding affinity of the CB8-G5 complex is the average value of the protonated and deprotonated G5 results. This final value is used in the calculation of error metrics and ranking coefficients, but in Table 3 we still present protonation-dependent estimates for detailed analyses of the simulated results.

Overall, the AM1-BCC charge scheme systematically overestimates host-guest binding affinities, while RESP provides more accurate results than AM1-BCC in error estimates and ranking predictions. The RMSE and MAE of AM1-BCC predictions are much larger than RESP, which indicates that the RESP predictions are closer to the experimental values. The large and positive MSE for AM1-BCC suggests overestimations of binding affinities, while that for RESP is close to zero, suggesting the absence of systematic deviations. Although RESP is better than AM1-BCC for these host-guest systems, its RMSE is still very large (~ 3.6 kcal/mol), which indicates the inaccuracies of fixed-charge models. The ranking coefficients show similar behaviors. Two ranking coefficients for AM1-BCC are very close to zero, while the RESP ones are larger (~ 0.2) but are still not satisfactory. The correlation between the experimental values and the modelling results in the current work is shown in Fig. 5. We can see that 3 affinities of the RESP predictions locate outside the ± 2 kcal/mol line of the experimental results, while the number is 6 for AM1-BCC predictions, which suggests low accuracies of AM1-BCC results.

The reasons for the overestimation of binding affinities with AM1-BCC charges and the RESP-better-than-BCC phenomenon include the differences between atomic charges of the host CB8, those of the guests, and the differences between the most stable binding pose obtained under different charge schemes. The most stable binding poses obtained under the two charge schemes are similar and thus do not pose a problem here. Thus, the atomic charges play the most important role in the accuracy of binding thermodynamics. For the macrocyclic host CB8, the AM1-BCC charges cannot reproduce the molecular ESP, while the ESP RRMSE is small for RESP charges. For deprotonated G5, the atomic charges of both charge schemes fail to reproduce the target ESP, as discussed in the previous charge-generation comparison. Interestingly, the binding affinities obtained under the two charge schemes are extremely similar (within statistical uncertainty). For other guests, both of the two charge schemes provide similar ESP around the guests, and the binding affinities obtained with AM1-BCC charges are much stronger than the RESP results. The dipole does not pose a problem for both charge schemes and we will not discuss them here. Therefore, the overall overestimation of the binding affinities is closely related to the poor ESP reproduction around the CB8 host. The low-quality ESP around the CB8 host leads to inaccurate CB8-guest interactions and thus inaccurate binding thermodynamics. The accord between the binding affinities of the CB8-deprotonated-G5 complex under the two charge schemes arises from error cancellation. The host-guest interactions in the complex formed by the relatively accurate RESP CB8 and the poorly described RESP deprotonated G5 are similar to those in the complex formed by the inaccurate AM1-BCC CB8 and the poorly described AM1-BCC deprotonated G5. A possible way to improve the AM1-BCC results is substituting the atomic charges of CB8 with the RESP charges but still use the AM1-BCC charges for the guest molecules. Namely, we only treat the most problematic molecule with more accurate descriptions. In this way, the binding thermodynamics are expected to be improved and close to the RESP-CB8-RESP-guests results.

3.5. Iterative Refitting of Atomic charges.

Although the sampling results are independent of the initial configuration, the fitted atomic charges do depend on the configurations used to generate them. The configurations used in the first step of molecular modelling, i.e., parametrization, could differ significantly from the finally obtained bound states with the highest thermodynamic stability on the free energy surface. Therefore, we then choose guests that have obvious differences between the most stable bound conformations explored during enhanced sampling simulations and the initial configurations used in charge generation, re-generate the atomic charges with both charge schemes, and investigate the impact of the initial configuration on the fitted charges, the

resulting binding thermodynamics and the binding patterns. In this way, the initial fixed-charge model is considered as a guess to explore the configurational space, and the stable conformation obtained from configurational space exploration serves as a more appropriate model for the parameterization of host-guest complexes. The resulting refitted model, in principle, should provide a more accurate description of host-guest interactions and thus binding information. Such an iterative refitting procedure could be repeated multiple times to calibrate the sampling and fitting results. Here, we only perform one round of such refitting calculations as illustrative examples.

Refitting protonated G7 Cocaine. The bound conformation under the AM1-BCC charge scheme differs from the initial configuration used in charge generation, while the RESP one is similar to the initial configuration. As AM1-BCC charges enable more flexible configurational exploration, we employ the most stable configuration extracted from the bound state under the AM1-BCC charge scheme in charge refitting. The AM1 and B3LYP/6-31G* optimized structures are superposed in Fig. S7, where differences could be observed in some regions, e.g., the orientation at -COOMe. The qualities of the ESP reproduction are similar under the two charge schemes and are also similar to the previous charge fitting. The dipole moments produced by the two charge schemes are still similar.

Refitting deprotonated G5 Ketamine. As discussed in the previous charge-generation section, the deprotonated form of G5 has the largest problem. The failure of the ESP reproduction arises from the anisotropic lone-pair distribution at nitrogen and cannot be solved by atom-centered fixed-charge schemes. However, improvements of the fitting quality could be obtained by varying the configuration used in charge generation. The binding poses obtained with both charge schemes are different from the initial conformation used in charge generation. Thus, we consider refitting the atomic charges of this guest and resampling the whole configurational space with this newly generated model. The most stable binding pose extracted from the previous 3D enhanced sampling simulation is used as the starting structure in the new charge fitting procedure. Namely, the most stable configuration among the center-binding conformations is selected. We repeat the AM1-BCC and RESP charge generation procedures, and detailed comparisons are presented in Fig. S7. The AM1 and B3LYP/6-31G* optimized structures of deprotonated G5 in Fig. S7 are extremely similar. The main difference between this new conformation and the previously used one is the closeness of the N and Cl atoms. The N atom is far from Cl in this new conformation, and their lone pairs of these atoms are separated, which would improve the fitting quality. As expected, the ESP RRMSEs of both AM1-BCC and RESP schemes are smaller than the previous case, which suggests better reproductions of the ESP around the molecule and thus better fitting qualities. However, these RRMSEs are still much larger than the other cases,

which indicates that the lone-pair problem for atom-centered fixed-charge models cannot be fully solved by changing the conformation used in charge generation, but needs other techniques (e.g., multipoles and off-center charges) to get an accurate representation. The dipole moments produced by the two charge schemes are similar, which is similar to the case using the previous conformation.

Refitting protonated G5 Ketamine. The protonated Ketamine, despite with good fitting quality under both charge schemes, is observed to have multiple binding conformations in the bound state under the two charge schemes. These conformational states are mainly caused by the conformational fluctuations of the guest inside the host cavity. As the deprotonated form of Ketamine is included in the refitting calculations, we also include the protonated form. The most stable bound conformation under the AM1-BCC charge scheme differs from the initial conformation used in charge generation to some extent, while the RESP charges keep the guest in the neighborhood of the initial conformation. This could possibly result from the over-stabilization caused by polarization-induced conformation-dependent charges, which locks the guest in the initial conformation and makes it difficult to explore new guest conformations. By contrast, the more mean-field-like AM1-BCC charges provide a more flexible model for conformational exploration. Thus, in the iterative fitting of atomic charges, we select the most stable bound conformation obtained with AM1-BCC charges to re-generate new atomic charges with the two charge schemes. Another reason that this system is selected is to validate the low binding affinity produced by RESP charges. The AM1 and B3LYP/6-31G* optimized structures from the bound conformation are compared in Fig. S7. Still, these optimized structures are similar. The ESP RRMSEs of the two charge schemes are both very low, and their magnitudes are also close to the previous fitting results. The dipole moments provided by the two charge schemes are still similar. This set of refitted atomic charges is named as G5 prot 1 or protonated G5-1 in our presentation, and the structure after geometry optimization has the intra-molecular hydrogen bond formed by Cl and -NH_2^+ and thus is similar to that in the initial charge generation. Due to the similarity of this new configuration and the previous one, we expect the refitting to have minor impact on the binding thermodynamics. To provide a guest conformation obviously different from the previous one, we modify the configuration used to refit the deprotonated G5 by adding one hydrogen atom (i.e., protonating G5). The resulting structures after optimization at the AM1 and B3LYP/6-31G* levels have Cl and -NH_2^+ separated and no intra-molecular hydrogen bond, as shown in Fig. S7. We fit the atomic charges with the two charge schemes based on the optimized structures, and this model is named as G5 prot 2 or protonated G5-2. By comparing the refitting results of the deprotonated G5 and the G5 prot 2 model, we can see that the ESP RRMSE is significantly lowered upon the protonation of G5 for both charge schemes. As Cl

and -NH_2^+ are separated in this model, we know that the lone electron pairs of nitrogen are the most problematic part for fixed-charge models, while the Cl ones do not really pose a problem here.

More detailed insights into the conformational dependence. To provide a clearer picture of the ESP reproduction issue, we calculate the ESP RRMSEs of all 4 sets of atomic charges (i.e., AM1-BCC and RESP charges generated from the initial/old and new conformations) with references to the two HF/6-31G* ESP scans for these 3 guests, as shown in Fig. 6. For Cocaine G7, the ESP RRMSEs of the old and new AM1-BCC charges are larger than the RESP ones, which indicates that the RESP charge scheme provides more accurate descriptions of inter-molecular interactions than AM1-BCC for both old and new configurations. The charges generated with one conformation could still describe the other conformation in a better way than AM1-BCC charges. However, the conformational dependence of the ESP RRMSE is larger for the RESP charge scheme, which indicates that the conformational dependence of the RESP charges is more severe. This RESP-better-than-BCC phenomenon provides a good example that RESP could be more accurate than AM1-BCC.

The second example, the deprotonated form of Ketamine G5, shows somehow different behaviors. Among the 4 charge sets, the RESP charge set always provides the smallest ESP RRMSE when the reference is the HF ESP used to generate it. However, when a different conformation is explored, the ESP RRMSE of the original RESP charge set becomes extremely large. For instance, the old RESP charge set provides about 40% ESP RRMSE when the reference ESP is scanned at the new conformation. This phenomenon suggests that the RESP charge depends significantly on the conformation used to fit it, and the charge set generated at one conformation could be very inaccurate for another. By contrast, the AM1-BCC charge scheme shows smaller conformation-dependence, and the differences between the ESP RRMSEs of the old and new AM1-BCC charge sets are relatively modest for the two reference ESP scans. For the new reference ESP, the ESP RRMSEs of the two AM1-BCC charge sets are even smaller than the old RESP one. Therefore, this example suggests that the significant conformational dependence of the RESP charge scheme sometimes makes it a poor solution, while the more mean-field-like AM1-BCC charge scheme could be relatively accurate to describe all relevant conformations.

The last example of the protonated G5 serves as an intermediate of the above two examples. In the G5 prot 1 model, for a given reference ESP, the RESP charge set generated at this conformation is always more accurate than the AM1-BCC one. The charges generated at one conformation provide larger errors when a new conformation is explored, and the ESP RRMSEs of the two charge schemes show similar conformational dependences. Compared with the deprotonated G5, the fitting problem triggered by lone pairs is absent, and the ESP RRMSEs of all 4 charge sets are smaller for the two reference ESP scans. However, unlike the first

example of Cocaine, where RESP charges are always better than AM1-BCC irrespective of conformations, the ESP RRMSE of RESP charges could be larger than AM1-BCC when the reference conformation differs from that used in charge generation. Specifically, when the reference ESP is obtained at the new conformation, the AM1-BCC charge set generated at this conformation has a smaller ESP RRMSE than the old RESP charge set. In the G5 prot 2 model, the charges generated at one configuration produce obviously larger ESP RRMSEs at the other configuration, which is triggered by the difference between the old and new structures, especially the presence/absence of the intra-molecular hydrogen bond.

Overall, for all 3 cases under refitting, the new RESP charges reproduce the ESP around the molecule in a better way. Therefore, if the binding pose or the global minimum under the new charge set is similar to the bound conformation extracted from the previous free energy surface, the inter-molecular interactions should be more accurately described. As a result, the binding thermodynamics and interaction patterns are expected to be more accurate. Note that the configurational dependence of atomic charges is valid not only for host-guest systems, but is also of great importance for other complex systems such as protein-ligand complexes. Therefore, the iterative refitting procedure is useful for both host-guest and biomolecular systems.

Binding information. We then perform 1000 ns spherical-coordinates-biased metadynamics simulation for each host-guest system with the newly fitted AM1-BCC and RESP charges and reweight the statistics. The free energy surfaces projected on the host-guest COM distance and host-guest contacts are shown in Fig. S8, while the $C - C_{ph}$ free energy surfaces are provided in Fig. 7.

Similar to the previous discussions about binding modes, we first discuss the AM1-BCC results. The radius-contact surfaces for the three host-guest systems are presented in Fig. S8a-d. These free energy surfaces are similar to the previous AM1-BCC results shown in Fig. S5. Only one minimum corresponding to the center-binding mode exists on the $\rho - C$ free energy surface for each host-guest system. The low-energy regions have been explored in previous simulations with old atomic charges, and we do not repeat the discussion about the structural features here. Thus, we turn to the $C - C_{ph}$ surfaces shown in Fig. 7a-d to check the detailed binding modes. The first example, protonated Cocaine, has a free energy surface with obviously different features compared with the previous AM1-BCC result. The new AM1-BCC charges provide a picture that one conformational state has much higher thermodynamic stabilities than all other conformations. This low-energy conformational state is observed in the previous simulation with old AM1-BCC charges, and it features the selected phenyl ring remaining solvent-exposed and the other parts of the guest penetrating the host cavity. The other binding poses explored with the old AM1-BCC charge set are no

longer thermodynamically stable upon charge refitting. This phenomenon suggests that the thermodynamics and interaction pattern produced by AM1-BCC charges could have significant dependence on the configuration used in charge generation. The other examples, i.e., the deprotonated G5 and the two protonated G5 models, have similar $C - C_{ph}$ free energy surfaces to the previous AM1-BCC results. Minor differences could be observed for detailed patterns of the free energy surfaces, but the general features remain unchanged. The conformational fluctuations in the bound state still lead to flexible host-guest complexes. Therefore, the configuration used in charge generation does not have a significant impact on the inter-molecular interaction patterns in CB8-G5 complexes.

We also calculate the binding affinities with the newly fitted charges. Under the AM1-BCC charge scheme, the binding free energies are -13.5 ± 0.5 kcal/mol, -12.1 ± 0.5 kcal/mol, -12.8 ± 0.5 kcal/mol, and -10.6 ± 0.5 kcal/mol for protonated G7, deprotonated G5, protonated G5-1 and protonated G5-2, respectively. For the protonated G7, interestingly, although significant changes are observed on the free energy surfaces upon charge refitting, the binding affinity of this guest remains almost unchanged. This phenomenon suggests that the variation of atomic charges could trigger some differences in interaction patterns, but the strength of inter-molecular interaction could be almost unaltered. By contrast, the deprotonated G5 has a statistically significant change (~ 1.7 kcal/mol). The magnitude of this variation of binding thermodynamics is in agreement with the variations of ESP RRMSE shown in Fig. 6, where the change of ESP RRMSE is significant for deprotonated G5. For the protonated G5-1, minor changes (similar to the statistical uncertainty) are observed for the binding thermodynamics compared with the previous AM1-BCC estimate, which suggests that the strengths of inter-molecular interactions in the bound state are similar with the old and new AM1-BCC charge sets. This phenomenon is in agreement with the similar ESP RRMSEs produced by the old and new AM1-BCC charge sets and the similar conformational features of the two structures (i.e., old configuration and the new protonated G5-1) used in charge generation. By contrast, the binding affinity for the protonated G5-2 differs significantly from the previous result, which is expected due to the change of guest conformation in charge generation, especially the absence of intra-molecular hydrogen bond. Averaging the binding affinities of the protonated G5-1 and the deprotonated G5 gives the final estimate of the CB8-G5 binding affinity, which is -12.4 kcal/mol. This value is almost identical to the experimental result, which suggests that the iterative refitting procedure could improve the binding thermodynamics. The experiment-computation correlation is presented in Fig. 8 along with the error metrics (RMSE and MAE) and the ranking coefficients (τ and PI). We can see that the iterative refitting procedure does improve the quality of prediction. Specifically, MAE is lowered from 4.7 kcal/mol to 4.5 kcal/mol, and PI increases from

0.0 to 0.1. RMSE and Kendall τ remain almost unchanged. The quality metrics when replacing the protonated G5-1 result with the protonated G5-2 one are similar and thus will not be discussed further. The above observations indicate that refitting atomic charges with the AM1-BCC charge scheme could improve the accuracy of modelling.

We then turn to the RESP results. The radius-contact surfaces are given in Fig. S8e-h, which are still similar to the previous results shown in Fig. S6. We thus only discuss the $C-C_{ph}$ free energy surfaces shown in Fig. 7e-h to investigate the detailed binding patterns. Similar to the old RESP result, there is only one minimum on the $C-C_{ph}$ free energy surface for G7 Cocaine. However, the position of this minimum differs from the previous result. In this newly observed state, the host-guest contact number is smaller with new RESP charges, but the host-phenyl contact number remains unchanged. This phenomenon suggests that the phenyl ring is still inside the cavity, but some parts of the guest are relatively far away from the host. The representative structure extracted from this minimum features the phenyl penetrating the cavity and the other parts of the guest outside the host cavity. The main difference between this new conformational state and that obtained with the old RESP charge set is the location of the -COOMe group. This group is thermodynamically stable outside the host cavity with the new charge set, but is penetrating the central cavity with the old RESP charge set. As for Ketamine in its protonated (G5-1 and G5-2) and deprotonated forms, the conformational fluctuations and the resulting crescent landscapes remain with the new RESP charges. Thus, charge refitting does not alter the inter-molecular interaction patterns significantly in these two cases.

The binding affinities obtained with the new RESP charges are -8.1 ± 0.5 kcal/mol, -10.4 ± 0.5 kcal/mol, -3.9 ± 0.5 kcal/mol, and -3.5 ± 0.5 kcal/mol for protonated G7, deprotonated G5, protonated G5-1, and protonated G5-2, respectively. For the first example of protonated G7, as discussed in the previous paragraph, the binding mode is altered upon charge refitting. The binding affinity with the old RESP charges is significantly overestimated compared with the experimental value, while the newly fitted RESP charge set provides a binding affinity very close to the experimental result. Thus, the iterative refitting procedure, i.e., using the initial charge set as an initial guess to explore the configurational space and refitting atomic charges with newly obtained conformation(s), could significantly improve the binding thermodynamics. Some hints about inter-molecular interactions could also be obtained from the ESP RRMSE comparison in Fig. 6. The new RESP charge set has a relatively stable ESP RRMSE for different configurations, while the ESP RRMSE produced by the old RESP charge set is significantly different for different configurations.

Thus, the performance of the new RESP charge set is more stable for different configurations. The computed binding affinity of protonated G7 with the new RESP charge set is significantly different from the previous RESP result, which suggests that the conformational dependence of the RESP scheme is quite significant. The binding thermodynamics obtained with atomic charges generated at one configuration could differ dramatically from the other. Thus, the configuration used in charge generation could be the key determinant in computational modelling of host-guest complexes. As for the protonated and deprotonated forms of G5, the binding thermodynamics remain almost unchanged compared with the previous RESP results. This phenomenon suggests that the electrostatic interactions produced by the initial charge sets are similar to those produced by the new charges, even for newly explored conformations. Namely, the charge set obtained from a single configuration is already sufficient to describe the inter-molecular interactions for different guest conformations. Thus, iterative refitting the atomic charges is not very helpful in this case. The experiment-computation correlation for the new RESP charge set is shown in Fig. 8 along with the error metrics and the ranking coefficients. The quality metrics of the two protonated G5 models are similar and thus only the protonated G5-1 model is presented. For the updated charge set, RMSE is lowered from 3.6 kcal/mol to 2.6 kcal/mol and PI increases from 0.2 to 0.6. Therefore, the iterative refitting procedure does improve the quality of prediction and is useful when using the RESP charge scheme.

An interesting phenomenon observed here is that the binding thermodynamics obtained with the two protonated G5 models (i.e., G5 prot 1 and G5 prot 2) are extremely similar under the RESP charge scheme. This behavior is obviously different from the previous AM1-BCC case, where a statistically significant difference is observed between the binding affinities. In that case (i.e., AM1-BCC), as the charge correction (i.e., BCC) is obtained according to atom type and connectivity information, the differences between the two AM1-BCC charge sets arise from the differences between the AM1 Mulliken charges. Namely, although the detailed conformations differ in the two configurations used in charge generation, the additional BCC terms are exactly the same, which does not seem reasonable. The presence and absence of the intra-molecular N-H \cdots Cl interaction perturb the AM1 Mulliken charges and should also be properly accounted for in the BCC term. By contrast, the atomic charges under the RESP scheme are derived from restrained ESP fitting, which differs from the AM1-BCC procedure theoretically. As the strengths of inter-molecular interactions are similar for the two RESP models (i.e., G5 prot1 and G5 prot 2), it seems that the restrained fitting procedure leads to somehow less conformation-dependent behavior of the RESP charge scheme.

The above observations suggest that the iterative refitting procedure could change both the binding modes and the binding thermodynamics. Therefore, choosing a proper configuration used for model

parameterization plays an important role in molecular modelling. This impact is observed to be significant in host-guest complexes. As only the guest/ligand configuration is altered and the parameters of the host/receptor remain unchanged in the current CB8-drug modelling, this guest-conformation-dependent behavior would also be significant for other complex systems, e.g., protein-ligand complexes.

3.6. Limitations of the current work.

Although the RESP charges could reproduce the ESP around the molecules in a better way compared with AM1-BCC, the predicted binding affinities still have obvious deviations from the experimental references. The sampling is not a problem in the current work. Thus, we expect the errors/deviations to arise mainly from the accuracy of the model. The force field used in the current modelling consists of bonded and non-bonded terms. Among the non-bonded terms which contribute significant parts to the binding thermodynamics, the electrostatics term is often found to have larger problems. Although the RESP charges perform well in reproducing the *ab initio* ESP, the target gas-phase HF/6-31G* ESP itself may not be of good quality. Higher-level methods (e.g., post-HF), larger basis sets, and solvation models could be considered to provide more accurate reference ESP. Even with a good target level/ESP, atom-centered fixed-charge models are still unable to accurately reproduce the ESP around the molecule. We observe that the fitting of lone-pair species could be improved by changing conformations used in the charge generation, but the fitting quality is still obviously worse than the other situations. This observation indicates that more accurate models (e.g., polarizable models and off-center charges) are needed. Further, polarization effects are not fully included in the fixed-charge model. Therefore, polarizable force fields and even multi-scale QM/MM treatments could be useful to improve the accuracy of the modelling.^{111, 112, 117-124} The bonded terms also play an important role in the host-guest/protein-ligand recognition. On this aspect, instead of using the pre-fitted GAFF parameters, refitting the bonded terms (e.g., dihedral) could be helpful.

3.7. Alternative simulation approach—a perturbative solution to host-guest sampling.

Instead of directly simulating all host-guest complexes in the configurational space, an alternative approach to obtain thermodynamic profiles under different parameter sets (e.g., force fields) is performing alchemical transformations between different models/Hamiltonians to extrapolate the results from a reference calculation, e.g., one selected host-guest system. The indirect free energy scheme has been widely applied in many cases.^{123, 125-136} However, we should note that the indirect method relies heavily on the overlap between end states (i.e., similarities of different models). It fails to converge when the differences

between models are large, the introduced bias may be hard to quantify exactly, and the statistical fluctuations are much larger than the direct exploration of the configurational space. As a result, the perturbation-based estimates are not as reliable as direct simulations of the binding/unbinding event. Therefore, in the current work, we directly simulate the binding/unbinding of all host-guest complexes to obtain a reliable and general picture of the CB8-drug interactions. However, it should be fairly to note that the perturbative solution could be useful when we iteratively refit the atomic charges. In this case, the difference between atomic charges after several iterations would be small, and the perturbation-based method could be efficient, as the direct exploration of the configurational space is totally avoided.

4. Conclusion

The partly enclosed macrocycles CB[n] are a family of pumpkin-shaped molecular containers formed by glycoluril monomers connected with methylene bridges. Their unique structural features make them suitable hosts for an array of neutral and cationic guests. They are widely used in many research areas, e.g., drug delivery, supramolecular assembly and asymmetric synthesis. Among the CB[n] family of molecules, CB8 has intermediate internal cavity volume. Thus, it acts as a good example in the investigation of CB[n]-guest interactions. Despite the relative simplicity of the structural features of CB8, the interaction pattern of CB8-guest systems could be rather complex. Multiple binding patterns were observed in our previous modelling of a series of CB8-guest complexes, which posed a problem in the sampling strategy in MD simulations. In this work, we selected a series of structurally diverse and commonly abused drugs with experimental binding affinities available, explored the configurational space with spherical-coordinates-biased sampling strategy, and investigated the binding/interaction patterns that the CB8 ring could form extensively. We further compared fixed-charge models widely used in molecular simulations systematically, and provided a general picture of CB8-guest interactions.

Specifically, the host-guest binding requires advanced sampling strategies and accurate models/Hamiltonians to describe the intra- and inter-molecular interactions. The sampling strategy used in the current work samples a spherical region around the host, where all possible host-guest relative positions and thus binding poses are explored. Alternative schemes (e.g., the cylindrical scheme and the combination of the cylindrical and spherical ones shown in Fig. 1b) could also be used, as long as all energetically favored regions are included in the sampled region. The time series of the total host-guest contact number and its decomposition into the contributions of each host atom provide details about the coordination pattern. Multiple binding/unbinding events are sampled over the course of the 3D enhanced sampling simulation,

and center- and side-binding patterns are observed, which indicates that the possible binding poses are explored.

The fixed-charge models as a mean-field approximation of electronic structure calculations provide a fast and efficient option for long timescale simulations. However, they are often criticized due to the extent that the polarization effects are represented. We build fixed-charge models for abused drugs and the CB8 host and analyze the performance in detail (e.g., their abilities to reproduce the ESP around the molecule). The AM1 and B3LYP/6-31G* optimized structures are similar in most cases, and minor differences are observed for Fentanyl, Morphine, and Cocaine. The AM1-BCC derived charges provide similar results for the molecular ESP and the dipole in most cases, and significant deviations are observed for the CB8 host and the deprotonated Ketamine. The problems of the host should arise from the fact that the training set of the BCC model does not include such large macrocyclic species, while the deprotonated form of Ketamine has lone pairs that cannot be properly described with atom-centered fixed-charge models. We further extract the most stable bound conformation of the CB8-Ketamine (deprotonated) complex obtained with the initial poorly fitted charge models, refit the atomic charges of the deprotonated Ketamine, and resample the whole configurational space of the host-guest complex. Although some improvements of the fitting quality are observed, the lone electron pairs (especially those on nitrogen) still hinder the accurate reproduction of the ESP around the molecule with fixed-charge models, which indicates that the lone-pair problem cannot be fully solved with atom-centered fixed-charge models. Widely used solutions include adding off-center dummy points in charge fitting and employing polarizable force fields (e.g., adding multipoles) to represent the electrostatics. An alternative strategy that is often overlooked is separating vdW and charge centers in order to have more degrees of freedom to represent the molecular properties.

Despite the relative simplicity of the structures of the host CB8 and drugs, the binding mechanism of CB8-drug complexes is more complicated than expected. Even for structurally simple guests, the bound conformation endures large fluctuations and an ensemble of conformations could be observed in the bound state. The whole guest could stay at the central cavity of the host, or parts of the guest could fluctuate along the CB8 normal and become more solvent-exposed. The conformational fluctuations of the host make the bound conformations further complicated. Different levels of the distortion of the host result in different binding modes. For structurally complex guests, multiple binding modes are observed on the free energy surfaces. The binding poses observed in the current work are all center-binding ones, and the bound conformations vary with the structural details of the guest. For instance, Fentanyl could cling to the outer surface of the CB8 host and form a bracelet-like pocket coordinating the host, or insert its two aromatic

rings inside the central cavity of the host and form a typical center-binding mode. Orientations of specific functional groups of the guest molecules could also lead to different binding poses. These observations indicate that the host-guest interactions are more complex than commonly expected and static calculations based on quantum mechanics Hamiltonians are insufficient to obtain a general and accurate understanding of host-guest binding.

Comparison of the binding patterns obtained under the two charge schemes suggests that the atomic charge could have a significant impact on host-guest interactions. Different atomic charges could lead to different binding modes, and the same binding mode observed under different charge schemes could differ in thermodynamic stability. Even for host-guest complexes formed by quite rigid molecules, the interaction patterns could show obvious differences. Such charge-scheme-dependence is more significant for structurally complex molecules. Therefore, the appropriate treatment of molecular electrostatics is necessary for the computational modelling of host-guest systems.

The inter-molecular interactions are mostly affected by atomic charges. The configurations used in the initial charge fitting could differ from the stable poses that the small molecules or drugs tend to stay in the bound state. This difference would generally lead to inaccurate descriptions of inter-molecular interaction patterns, including both the binding thermodynamics and the binding modes. We thus iteratively refit the atomic charges of guest molecules that present obvious differences between the initial configuration and the stable binding poses explored during enhanced sampling simulations. The errors of the ESP around the molecules are observed to have significant conformational dependence. This problem is more severe for the RESP charge scheme, and the RESP charge set generated at one conformation could be very inaccurate for another pose. The AM1-BCC model is somehow like a more flexible model that could describe different conformations in a relatively accurate way, thus serving as a good charge scheme for preliminary explorations of the conformational space.

The impact of refitting atomic charges on binding modes shows system-dependence. The refitted charge set produces significantly different binding thermodynamics and interaction patterns for Cocaine, while its influence on the binding patterns of protonated and deprotonated Ketamine is relatively small. For the protonated Cocaine, the old and new atomic charges have different preferences for the binding modes, and the binding thermodynamics is improved, especially for the RESP charge scheme. The binding affinity obtained with the new RESP charge set is in great agreement with the experimental value. For the protonated and deprotonated Ketamine, the refitted charge sets do not change the general interaction pattern within the CB8-Ketamine complex, and the conformational fluctuations in the bound state are still

observable. However, the binding thermodynamics is altered to some extent. It is worth noting that the presence and absence of the intra-molecular N-H \cdots Cl interaction in protonated Ketamine could lead to significant changes of inter-molecular interactions, and the AM1-BCC and RESP charge schemes behave differently. The binding affinities obtained with the AM1-BCC charge sets fitted in the presence and absence of the intra-molecular N-H \cdots Cl interaction differ significantly, while the RESP results seem less conformation-dependent. This behavior arises from the differences between the charge-derivation procedures of the two charge schemes. Specifically, as the BCC term is calculated from atom type and connectivity information, this correction term added to AM1 Mulliken charges is exactly the same for different conformations and thus cannot properly reflect the conformation-dependence. Overall, the impact of using different configurations in charge generation is more significant for the RESP charge scheme, while the binding information obtained with the AM1-BCC charges fitted from different configurations show smaller differences.

With the initially parameterized fixed-charge models, the predicted binding affinities show ~ 5 kcal/mol deviations from the experimental reference for the AM1-BCC charge scheme, and the rank of binding affinities cannot be correctly calculated. The AM1-BCC charge scheme has a tendency to overestimate the host-guest binding affinity. The reason for this systematic deviation should be highly related to the large ESP error of the CB8 host. Significant improvement is observed when the charge scheme is changed to RESP. The RMSE is lowered to 3.6 kcal/mol, and the ranking coefficients becomes 0.2. No systematic overestimation or underestimation is observed for this charge scheme. Although the prediction still deviates obviously from the experimental reference, this improvement suggests that more accurate electrostatics could be helpful. On this aspect, more accurate force fields such as polarizable ones and even QM calculations could be useful. As mentioned previously, we also refit the atomic charges of several guest molecules due to the conformational-dependence of the molecular ESP. The refitted charges improve the binding thermodynamics significantly. Specifically, the RMSEs of the computed binding affinities are lowered to 5.4 kcal/mol and 2.6 kcal/mol for the AM1-BCC and RESP charge schemes, respectively. The ranking coefficients are also improved, especially for the RESP charge scheme. Therefore, the iterative refitting procedure of using the initial charge set to explore the configurational space and refitting the atomic charges with newly explored conformation(s) is useful to improve the binding thermodynamics.

Finally, we discussed about the limitations and the directions of further investigations of the CB8-guest binding. For instance, the fixed-charge model could be possibly replaced with more complicated models, the bonded terms in force fields could also be improved, and alternative sampling strategies could be applied.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 21633001). Part of the simulation was performed on the high-performance computing platform of the Center for Life Science (Peking University). Dr. Zhaoxi Sun is supported by the PKU-Boya Postdoctoral Fellowship. 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 spherical CV, the comparison between the AM1 and B3LYP/6-31G* optimized structures used in ESP fitting, the time series of the host-guest contacts and the corresponding decomposition into atomic contributions during metadynamics simulations under the two charge schemes, the comparison between the AM1 and B3LYP/6-31G* optimized structures in the new round of charge fitting for the deprotonated and protonated forms of Ketamine (G5) and Cocaine (G7), and the 2D radius-contact free energy surfaces obtained with the new AM1 and RESP charge sets 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. Kim, K.; Selvapalam, N.; Oh, D. H., Cucurbiturils—a new family of host molecules. *Journal of inclusion phenomena and macrocyclic chemistry* **2004**, 50, 31-36.
2. Corma, A.; García, H.; Montes-Navajas, P.; Primo, A.; Calvino, J. J.; Trasobares, S., Gold nanoparticles in organic capsules: a supramolecular assembly of gold nanoparticles and cucurbituril. *Chem. Eur. J.* **2007**, 13, 6359-6364.
3. Wu, X.-L.; Luo, L.; Lei, L.; Liao, G.-H.; Wu, L.-Z.; Tung, C.-H., Highly efficient cucurbit [8] uril-templated intramolecular photocycloaddition of 2-naphthalene-labeled poly (ethylene glycol) in aqueous solution. *The Journal of organic chemistry* **2008**, 73, 491-494.
4. Masson, E.; Ling, X.; Joseph, R.; Kyeremeh-Mensah, L.; Lu, X., Cucurbituril chemistry: a tale of supramolecular success. *Rsc Advances* **2012**, 2, 1213-1247.
5. Saluja, V.; Sekhon, B. S., Calixarenes and cucurbiturils: Pharmaceutial and biomedical applications. *Journal of Pharmaceutical Education and Research* **2013**, 4, 16.
6. Xia, D.; Wang, P.; Ji, X.; Khashab, N. M.; Sessler, J. L.; Huang, F., Functional supramolecular polymeric networks: the marriage of covalent polymers and macrocycle-based host-guest interactions. *Chemical Reviews* **2020**, 120, 6070-6123.
7. Yahiaoui, K.; Seridi, L.; Mansouri, K., Temozolomide binding to Cucurbit[7]uril: QTAIM, NCI-RDG and NBO analyses. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2021**, 99, 61-77.
8. Sun, Z.; Wang, X.; Zhang, J. Z., Determination of Binding Affinities of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Inhibitors from Free Energy calculation. *Chemical Physics Letters* **2019**, 723, 1-10.
9. Nerattini, F.; Chelli, R.; Procacci, P., II. Dissociation free energies in drug-receptor systems via nonequilibrium alchemical simulations: application to the FK506-related immunophilin ligands. *Phys. Chem. Chem. Phys.* **2016**, 18, 15005-15018.
10. Procacci, P., Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics. *Curr. Opin. Struct. Biol.* **2021**, 67, 127-134.
11. Procacci, P.; Guarnieri, G., SAMPL7 blind predictions using nonequilibrium alchemical approaches. *Journal of Computer-Aided Molecular Design* **2021**, 35, 37-47.
12. Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chemical Physics Letters* **2017**, 684, 239-244.
13. 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.
14. Brotzakis, Z. F.; Bolhuis, P. G., Approximating free energy and committor landscapes in standard transition path sampling using virtual interface exchange. *J. Chem. Phys.* **2019**, 151, 174111.
15. Jurasz, J.; Czub, J.; Baginski, M.; Wieczor, M., Molecular mechanism of proton-coupled ligand translocation by the bacterial efflux pump EmrE. *bioRxiv* **2021**.
16. 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.
17. Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019**, 40, 1270-1289.
18. Wang, X.; Sun, Z., Determination of Base Flipping Free Energy Landscapes from Nonequilibrium Stratification. *J. Chem. Inf. Model.* **2019**, 59, 2980-2994.
19. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, 314, 141-151.
20. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, 113, 6042-6051.
21. Marinari, E.; Parisi, G., Simulated tempering: a new Monte Carlo scheme. *EPL (Europhysics Letters)* **1992**, 19, 451.
22. Mori, Y.; Okumura, H., Simulated tempering based on global balance or detailed balance conditions: S uwa-T odo, heat bath, and M etropolis algorithms. *J. Comput. Chem.* **2015**, 36, 2344-2349.
23. Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P., Generalization of the Jarzynski and Crooks nonequilibrium work

- theorems in molecular dynamics simulations. *Phys. Rev. E* **2007**, 75, 050101.
24. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **1985**, 107, 3902-3909.
25. Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P., RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, 27, 1101-1111.
26. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, 16, 1357-1377.
27. Sun, Z.; Wang, X., Thermodynamics of Helix formation in small peptides of varying length in vacuo, implicit solvent and explicit solvent: Comparison between AMBER force fields. *Journal of Theoretical and Computational Chemistry* **2019**, 1950015.
28. Desgranges, C.; Delhommelle, J., Towards a machine learned thermodynamics: exploration of free energy landscapes in molecular fluids, biological systems and for gas storage and separation in metal-organic frameworks. *Molecular Systems Design & Engineering* **2020**.
29. Sun, Z.; Zhu, T.; Wang, X.; Mei, Y.; Zhang, J. Z., Optimization of convergence criteria for fragmentation methods. *Chemical Physics Letters* **2017**, 687, 163-170.
30. Sahu, N.; Yeole, S. D.; Gadre, S. R., Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, 138, 104101.
31. Dahlke, E. E.; Truhlar, D. G., Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, 3, 46-53.
32. Li, W.; Li, S.; Jiang, Y., Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, 111, 2193-2199.
33. Findik, B. K.; Haslak, Z. P.; Arslan, E.; Aviyente, V., SAMPL7 blind challenge: quantum-mechanical prediction of partition coefficients and acid dissociation constants for small drug-like molecules. *Journal of Computer-Aided Molecular Design* **2021**.
34. Sun, Z.; Gong, Z.; Xia, F.; He, X., Ion Dynamics and Selectivity of Nav channels from Molecular Dynamics Simulation. *Chemical Physics* **2021**, 111245.
35. 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.
36. Bouchal, T.; Durník, I.; Illík, V.; Réblová, K.; Kulhánek, P., Importance of base-pair opening for mismatch recognition. *Nucleic Acids Research* **2020**, 48, 11322-11334.
37. Gervasio, F. L.; Chelli, R.; Massimo Marchi; Piero Procacci, A.; Schettino, V., Determination of the Potential of Mean Force of Aromatic Amino Acid Complexes in Various Solvents Using Molecular Dynamics Simulations: The Case of the Tryptophan-Histidine Pair. *J. Phys. Chem. B* **2001**, 105, 7835-7846.
38. Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V., Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *Journal of the American Chemical Society* **2002**, 124, 6133-6143.
39. Marsili, S.; Chelli, R.; Schettino, V.; Procacci, P., Thermodynamics of stacking interactions in proteins. *Phys. Chem. Chem. Phys.* **2008**, 10, 2673-2685.
40. Schepers, R. J.; Oyler, J. M.; Joseph Jr, R. E.; Cone, E. J.; Moolchan, E. T.; Huestis, M. A., Methamphetamine and amphetamine pharmacokinetics in oral fluid and plasma after controlled oral methamphetamine administration to human volunteers. *Clinical Chemistry* **2003**, 49, 121-132.
41. Hasan, A. A.; Ciancio, S., Relationship between amphetamine ingestion and gingival enlargement. *Pediatric dentistry* **2004**, 26, 396-400.
42. Subu, R.; Jayanthi, S.; Cadet, J. L., Compulsive methamphetamine taking induces autophagic and apoptotic markers in

the rat dorsal striatum. *Archives of Toxicology* **2020**, 94, 3515-3526.

43. Nave, R.; Schmitt, H.; Popper, L., Faster absorption and higher systemic bioavailability of intranasal fentanyl spray compared to oral transmucosal fentanyl citrate in healthy subjects. *Drug Deliv.* **2013**, 20, 216-223.
44. Taylor, D. R., Fentanyl buccal tablet: rapid relief from breakthrough pain. *Expert opinion on pharmacotherapy* **2007**, 8, 3043-3051.
45. DePriest, A. Z.; Puet, B. L.; Holt, A. C.; Roberts, A.; Cone, E. J., Metabolism and disposition of prescription opioids: a review. *Forensic Sci. Rev.* **2015**, 27, 115-145.
46. Sun, Z.; He, Q.; Li, X.; Zhu, Z., SAMPL6 host-guest binding affinities and binding poses from spherical-coordinates-biased simulations. *Journal of Computer-Aided Molecular Design* **2020**, 34, 589-600.
47. Kim, J.; Ham, S.; Hong, H.; Moon, C.; Im, H.-I., Brain reward circuits in morphine addiction. *Molecules and cells* **2016**, 39, 645.
48. Beltran-Campos, V.; Silva-Vera, M.; Garcia-Campos, M.; Diaz-Cintra, S., Effects of morphine on brain plasticity. *Neurología (English Edition)* **2015**, 30, 176-180.
49. Stuart-Harris, R.; Joel, S.; McDonald, P.; Currow, D.; Slevin, M., The pharmacokinetics of morphine and morphine glucuronide metabolites after subcutaneous bolus injection and subcutaneous infusion of morphine. *Br. J. Clin. Pharmacol.* **2000**, 49, 207-214.
50. Shigeev, S., Severity of opiate intoxication to gender and age. *Soudni lekarstvi* **2007**, 52, 21-24.
51. Drewes, A. M.; Jensen, R. D.; Nielsen, L. M.; Dronev, J.; Christrup, L. L.; Arendt-Nielsen, L.; Riley, J.; Dahan, A., Differences between opioids: pharmacological, experimental, clinical and economical perspectives. *Br. J. Clin. Pharmacol.* **2013**, 75, 60-78.
52. Felden, L.; Walter, C.; Harder, S.; Treede, R.-D.; Kayser, H.; Drover, D.; Geisslinger, G.; Lötsch, J., Comparative clinical effects of hydromorphone and morphine: a meta-analysis. *British journal of anaesthesia* **2011**, 107, 319-328.
53. Paramanandam, G.; Prommer, E.; Schwenke, D. C., Adverse effects in hospice patients with chronic kidney disease receiving hydromorphone. *Journal of palliative medicine* **2011**, 14, 1029-1033.
54. Hughes, M. M.; Atayee, R. S.; Best, B. M.; Pesce, A. J., Observations on the metabolism of morphine to hydromorphone in pain patients. *Journal of analytical toxicology* **2012**, 36, 250-256.
55. Kumar, M. G.; Lin, S., Hydromorphone in the management of cancer-related pain: an update on routes of administration and dosage forms. *Journal of Pharmacy & Pharmaceutical Sciences* **2007**, 10, 504-518.
56. Yang, Y.; Cui, Y.; Sang, K.; Dong, Y.; Ni, Z.; Ma, S.; Hu, H., Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. *Nature* **2018**, 554, 317-322.
57. Kirby, T., Ketamine for depression: the highs and lows. *The Lancet Psychiatry* **2015**, 2, 783-784.
58. Xu, J.; Lei, H., Ketamine-An Update on Its Clinical Uses and Abuses. *CNS neuroscience & therapeutics* **2014**, 20, 1015-1020.
59. Ilvento, L.; Rosati, A.; Marini, C.; L'Erario, M.; Mirabile, L.; Guerrini, R., Ketamine in refractory convulsive status epilepticus in children avoids endotracheal intubation. *Epilepsy & Behavior* **2015**, 49, 343-346.
60. Shikanai, H.; Hiraide, S.; Kamiyama, H.; Kiya, T.; Oda, K.; Goto, Y.; Yanagawa, Y.; Shimamura, K.-i.; Goda, Y.; Togashi, H., Subanalgesic ketamine enhances morphine-induced antinociceptive activity without cortical dysfunction in rats. *Journal of anesthesia* **2014**, 28, 390-398.
61. Mortenson, J. A.; Moriarty, K. M., Ketamine and midazolam anesthesia in Pacific martens (*Martes caurina*). *J. Wildl. Dis.* **2015**, 51, 250-254.
62. Baker, S. C.; Shabir, S.; Georgopoulos, N. T.; Southgate, J., Ketamine-Induced Apoptosis in Normal Human Urothelial Cells: A Direct, N-Methyl-d-Aspartate Receptor-Independent Pathway Characterized by Mitochondrial Stress. *The American journal of pathology* **2016**, 186, 1267-1277.
63. Balster, R. L.; Woolverton, W. L., Continuous-access phencyclidine self-administration by rhesus monkeys leading to physical dependence. *Psychopharmacology* **1980**, 70, 5-10.
64. Yamamoto, H.; Kamegaya, E.; Sawada, W.; Hasegawa, R.; Yamamoto, T.; Hagino, Y.; Takamatsu, Y.; Imai, K.; Koga, H.; Mishina, M., Involvement of the N-methyl-D-aspartate receptor GluN2D subunit in phencyclidine-induced motor

- impairment, gene expression, and increased Fos immunoreactivity. *Molecular brain* **2013**, 6, 1-16.
65. Domino, E. F., Neurobiology of phencyclidine (Sernyl), a drug with an unusual spectrum of pharmacological activity. *Int. Rev. Neurobiol.* **1964**, 6, 303-347.
66. McClatchy, D. B.; Savas, J. N.; Martínez-Bartolomé, S.; Park, S. K.; Maher, P.; Powell, S. B.; Yates, J., Global quantitative analysis of phosphorylation underlying phencyclidine signaling and sensorimotor gating in the prefrontal cortex. *Mol. Psychiatry* **2016**, 21, 205-215.
67. Jodo, E., The role of the hippocampo-prefrontal cortex system in phencyclidine-induced psychosis: a model for schizophrenia. *Journal of Physiology-Paris* **2013**, 107, 434-440.
68. Kuczkowski, K. M., The cocaine abusing parturient: a review of anesthetic considerations. *Canadian Journal of Anesthesia* **2004**, 51, 145-154.
69. Perreault, C. L.; Hague, N. L.; Morgan, K. G.; Allen, P. D.; Morgan, J. P., Negative inotropic and relaxant effects of cocaine on myopathic human ventricular myocardium and epicardial coronary arteries in vitro. *Cardiovascular Research* **1993**, 27, 262-268.
70. Freye, E.; Levy, J. V., Pharmacology and abuse of cocaine, amphetamines, ecstasy and related designer drugs. *Pharmacology of cocaine*, Springer **2010**, 49-60.
71. Phillips, K.; Luk, A.; Soor, G. S.; Abraham, J. R.; Leong, S.; Butany, J., Cocaine Cardiotoxicity. *American Journal of Cardiovascular Drugs* **2009**, 9, 177-196.
72. Volkow, N. D.; Wang, G.-J.; Fischman, M. W.; Foltin, R.; Fowler, J. S.; Franceschi, D.; Franceschi, M.; Logan, J.; Gatley, S. J.; Wong, C., Effects of route of administration on cocaine induced dopamine transporter blockade in the human brain. *Life sciences* **2000**, 67, 1507-1515.
73. Chen, J.; Kawamura, T.; Sethi, M. K.; Zaia, J.; Repunte-Canonigo, V.; Sanna, P. P., Heparan sulfate: Resilience factor and therapeutic target for cocaine abuse. *Scientific Reports* **2017**, 7, 13931.
74. O'Leary, M. E.; Chahine, M., Cocaine binds to a common site on open and inactivated human heart (Nav1.5) sodium channels. *The Journal of Physiology* **2002**, 541, 701-716.
75. McClung, C. A.; Sidiropoulou, K.; Vitaterna, M.; Takahashi, J. S.; White, F. J.; Cooper, D. C.; Nestler, E. J., Regulation of dopaminergic transmission and cocaine reward by the Clock gene. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 9377-9381.
76. Cunha-Oliveira, T.; Rego, A. C.; Garrido, J.; Borges, F.; Macedo, T.; Oliveira, C. R., Neurotoxicity of heroin-cocaine combinations in rat cortical neurons. *Toxicology* **2010**, 276, 11-17.
77. Minor, R. L.; Scott, B. D.; Brown, D. D.; Winniford, M. D., Cocaine-induced Myocardial Infarction in Patients with Normal Coronary Arteries. *Annals of Internal Medicine* **1991**, 115, 797-806.
78. Eken, Y.; Patel, P.; Díaz, T.; Jones, M. R.; Wilson, A. K., SAMPL6 host-guest challenge: binding free energies via a multistep approach. *Journal of computer-aided molecular design* **2018**, 32, 1097-1115.
79. Peerannawar, S. R.; Rao, S. S.; Gejji, S. P., Density functional investigations on 2-naphthalenecarbonitrile dimerization within cucurbit [8] uril cavitand. *Journal of molecular modeling* **2014**, 20, 1-9.
80. Chakraborty, D.; Chattaraj, P. K., Confinement induced thermodynamic and kinetic facilitation of some Diels-Alder reactions inside a CB[7] cavitand. *J. Comput. Chem.* **2018**, 39, 151-160.
81. 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.
82. Sun, Z., SAMPL7 Trimer/Trip Host-Guest Binding Poses and Binding Affinities from Spherical-Coordinates-Biased Simulations. *Journal of Computer-Aided Molecular Design* **2021**, 35, 105-115.
83. <https://github.com/samplchallenges/SAMPL8>.
84. Murkli, S.; Klemm, J.; Brockett, A. T.; Shuster, M.; Briken, V.; Roesch, M. R.; Isaacs, L., In Vitro and In Vivo Sequestration of Phencyclidine by Me4Cucurbit [8] uril. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2020**.
85. 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.
86. 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. *Journal of Physical Chemistry* **1992**, 97, 10269-10280.
87. 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.
88. 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.
89. 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.
90. 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.
91. Huai, Z.; Yang, H.; Li, X.; Sun, Z., SAMPL7 TrimerTrip host-guest binding affinities from extensive alchemical and end-point free energy calculations. *Journal of Computer-Aided Molecular Design* **2021**, 35, 117-129.
92. Huai, Z.; Yang, H.; Sun, Z., Binding thermodynamics and interaction patterns of human purine nucleoside phosphorylase-inhibitor complexes from extensive free energy calculations. *Journal of Computer-Aided Molecular Design* **2021**.
93. 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.
94. 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**.
95. 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.
96. 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.
97. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, 121, 10096-10103.
98. 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.
99. 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.
100. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Wiley Interdisip. Rev. Comput. Mol. Sci.* **2011**, 1, 826-843.
101. 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.
102. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters* **2008**, 100, 020603.
103. Tiwary, P.; Parrinello, M., A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015**, 119, 736-42.
104. 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.
105. 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.
106. Giovanni, B.; Davide, D.; Michele, P., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, 126, 2384.
107. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983**, 50, 1055-1076.
108. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, 52, 7182-7190.
109. 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.
110. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models, A Different Approach. *Journal of Physical*

Chemistry Letters **2014**, 5, 3863-3871.

111. Wang, X.; Yan, J.; Zhang, H.; Xu, Z.; Zhang, J. Z., An electrostatic energy-based charge model for molecular dynamics simulation. *J. Chem. Phys.* **2021**, 154, 134107.
112. Song, J.; Ji, C.; Zhang, J. Z., The critical effect of polarization on the dynamical structure of guanine quadruplex DNA. *Phys. Chem. Chem. Phys.* **2013**, 15, 3846-3854.
113. 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.
114. Shulga, D. A.; Titov, O. I.; Pisarev, S. A.; Palyulin, V. A., Multipole models of sulphur for accurate anisotropic electrostatic interactions within force fields. *Sar & Qsar in Environmental Research* **2017**, 29, 1-22.
115. Brown, I. D., Recent developments in the methods and applications of the bond valence model. *Chemical Reviews* **2009**, 109, 6858-6919.
116. Mohamed, N. A.; Bradshaw, R. T.; Essex, J. W., Evaluation of solvation free energies for small molecules with the AMOEBA polarizable force field. *J. Comput. Chem.* **2016**, 37, 2749-2758.
117. Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A., Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *J. Phys. Chem. B* **1998**, 102, 2293-2301.
118. Caldararu, O.; Olsson, M. A.; Riplinger, C.; Neese, F.; Ryde, U., Binding free energies in the SAMPL5 octa-acid host-guest challenge calculated with DFT-D3 and CCSD(T). *Journal of computer-aided molecular design* **2016**, 1-20.
119. Olsson, M. A.; Söderhjelm, P.; Ryde, U., Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J. Comput. Chem.* **2016**, 37, 1589-1600.
120. Jackson, N. E.; Webb, M. A.; de Pablo, J. J., Layered nested Markov chain Monte Carlo. *J. Chem. Phys.* **2018**, 149, 072326.
121. Wang, X.; Liu, J.; Zhang, J. Z.; He, X., Electrostatically embedded generalized molecular fractionation with conjugate caps method for full quantum mechanical calculation of protein energy. *J. Phys. Chem. A* **2013**, 117, 7149-7161.
122. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M., Fragment molecular orbital method: an approximate computational method for large molecules. *Chemical Physics Letters* **1999**, 313, 701-706.
123. Zhaoxi, S.; Zhirong, L., *BAR-based Multi-dimensional Nonequilibrium Pulling for Indirect Construction of QM/MM Free Energy Landscapes: Varying the QM Region*. 2021.
124. Zhaoxi, S.; Qiaole, H., *Seeding the Multi-dimensional Nonequilibrium Pulling for Hamiltonian Variation: Indirect QM/MM Free Energy Simulations*. 2021.
125. Heimdal, J.; Rydberg, P.; Ryde, U., Protonation of the proximal histidine ligand in heme peroxidases. *J. Phys. Chem. B* **2008**, 112, 2501-10.
126. Mikulskis, P.; Cioloboc, D.; Andrejić, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U., Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *Journal of computer-aided molecular design* **2014**, 28, 375-400.
127. Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O.; Essex, J. W.; Skylaris, C. K., Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies. *J. Phys. Chem. B* **2013**, 117, 9478-85.
128. Genheden, S.; Ryde, U.; Söderhjelm, P., Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J. Comput. Chem.* **2015**, 2114-2124.
129. Genheden, S.; Martinez, A. I. C.; Criddle, M. P.; Essex, J. W., Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. *Journal of Computer-Aided Molecular Design* **2014**, 28, 187-200.
130. Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C. K., Free Energies of Binding from Large-Scale First-Principles Quantum Mechanical Calculations: Application to Ligand Hydration Energies. *J. Phys. Chem. B* **2013**, 117, 9478-85.
131. Woods, C. J.; Manby, F. R.; Mulholland, A. J., An efficient method for the calculation of quantum mechanics/molecular

- mechanics free energies. *J. Chem. Phys.* **2008**, 128, 152-159.
132. Caveayland, C.; Skylaris, C. K.; Essex, J. W., Direct Validation of the Single Step Classical to Quantum Free Energy Perturbation. *J. Phys. Chem. B* **2014**, 119, 1017-25.
133. Lameira, J. S.; Kupchencko, I.; Warshel, A., Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J. Phys. Chem. B* **2016**, 120, 2155.
134. Sun, Z., BAR-based multi-dimensional nonequilibrium pulling for indirect construction of QM/MM free energy landscapes: from semi-empirical to ab initio. *Phys. Chem. Chem. Phys.* **2019**, 21, 21942-21959
135. Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019**, 21, 6672-6688
136. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Criteria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.

Table 1. Details of the host CB8 and 7 abused drugs investigated in the current work. The drug/host name, the IUPAC name, chemical formula, SMILES identifier, molecular weight in g/mol, experimental binding affinities in kcal/mol, experimental pKa, and net charge considered in the current work are given. Note that the net charge of each molecule considered in modelling is determined by the experimental pKa and the experimental condition for the measurement of the host-guest binding affinity, while the other data such as the molecular weight are valid for the net-neutral molecules. The experimental pKa of the guest G5 Ketamine is very close to the experimental condition for the binding affinity (pH 7.4). Thus, we consider the protonated and deprotonated forms of G5 in our calculation. The protonated G5 has the charge of +1, while the deprotonated form is net-neutral. The other guests have pKa values obviously larger than the experimental condition pH 7.4. Thus, all of them are protonated at the nitrogen atom. We also point out that the monoisotopic molecular weight is presented rather than the average one.

Molecule	Name	IUPAC Name	Chemical Formula	SMILES	Molecular Weight (g/mol)	ΔG_{exp}^*	$pK_{a,exp}$	Net charge
CB8	Cucurbit[8]uril	3,5,8,10,13,15,18,20,23,25,28,30,33,35,38,40,41,43,45,47,51,53,55,57,59,61,63,65,67,69,71,73-dotriacontazapentacosacyclo[3.5.3.3.3 ^{6,7} .3 ^{11,12} .3 ^{16,17} .3 ^{21,22} .3 ^{26,27} .3 ^{31,32} .2 ^{3,41} .2 ^{36,43} .1 ^{3,40} .1 ^{5,8} .1 ^{10,13} .1 ^{15,18} .1 ^{20,23} .1 ^{25,28} .1 ^{30,33} .1 ^{35,38} .1 ^{45,51} .1 ^{47,73} .1 ^{53,55} .1 ^{57,59} .1 ^{61,63} .1 ^{65,67} .1 ^{69,71}]octadecane-44,46,48,49,50,52,54,56,58,60,62,64,66,68,70,72-hexadecone	C ₄₈ H ₄₈ N ₁₂ O ₁₆	C1N2C3C4N(C2=O)CN5C6C7N(C5=O)CN8C9C2N(C8=O)CN5C8C%10N(C5=O)CN5C%11C%12N(C5=O)CN5C%13C%14N(C5=O)CN5C%15C%16N(C5=O)N7CN9C(=O)N2CN8C(=O)N%10CN%11C(=O)N%12CN%13C(-O)N%14CN%15C(-O)N%16CN%17C1=O	1329	-	-	0
G1	Metamfetamine	methyl[(2S)-1-phenylpropan-2-yl]amine	C ₁₀ H ₁₅ N	CN[C@@H](C)CC1=CC=CC=C1	149	-7.05	9.87	1
G2	Fentanyl	N-phenyl-N-[1-(2-phenylethyl)piperidin-4-yl]propanamide	C ₂₂ H ₂₈ N ₂ O	CCC(=O)N(C1CCN(CCC2=CC=CC=C2)CC1)C1=CC=CC=C1	336	-9.93	8.99	1
G3	Morphine	(1S,5R,13R,14S,17R)-4-methyl-12-oxa-4-azapentacyclo[9.6.1.0 ^{1,13} .0 ^{5,17} .0 ^{7,18}]octadeca-7(18),8,10,15-tetraene-10,14-diol	C ₁₇ H ₁₉ NO ₃	[H][C@@]12OC3=C(O)C=CC4=C3[C@@]11CCN(C)[C@]([H])(C4)[C@]1([H])C=C[C@@H]2O	285	-11.63	8.21	1
G4	Hydromorphone	(1S,5R,13R,17R)-10-hydroxy-4-methyl-12-oxa-4-azapentacyclo[9.6.1.0 ^{1,13} .0 ^{5,17} .0 ^{7,18}]octadeca-7(18),8,10-trien-14-one	C ₁₇ H ₁₉ NO ₃	[H][C@@]12OC3=C(O)C=CC4=C3[C@@]11CCN(C)[C@]([H])(C4)[C@]1([H])CCC2=O	285	-11.22	8.2	1
G5	Ketamine	2-(2-chlorophenyl)-2-(methylamino)cyclohexan-1-one	C ₁₃ H ₁₆ ClNO	CNC1(CCCCC1=O)C1=CC=CC=C1Cl	237	-12.32	7.5	0 or 1
G6	Phencyclidine	1-(1-phenylcyclohexyl)piperidine	C ₁₇ H ₂₅ N	C1CCN(CC1)C1(CCCCC1)C1=CC=CC=C1	243	-14.07	8.29	1
G7	Cocaine	methyl (1R,2R,3S,5S)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate	C ₁₇ H ₂₁ NO ₄	[H][C@]112CC[C@]([H])([C@H]1[C@H](C1)OC(=O)C1=CC=CC=C1)C(=O)OC)N2C	303	-7.92	8.61	1

Table 2. The ESP RRMSE and the dipole moments of the host CB8 and 7 guests under investigation. The number of ESP points in the scan is also given. We can see that the AM1-BCC charge model is comparable in reproducing the HF/6-31G* ESP and dipole for most systems, but for the large heterocyclic CB8 host, its performance is significantly worse than the RESP scheme. Note that neither of the two charge schemes performs well for the deprotonated G5, which should be triggered by the lone pair of nitrogen. The fitting accuracies of both charge schemes are significantly improved by adding an additional fitting point between N and Cl upon the protonation of G5, namely transforming the neutral -NH- to the cationic -NH₂⁺- and forming the intramolecular N-H···Cl hydrogen bond. For the deprotonated G5, we have tested that using a larger basis set (e.g., 6-311G**) and adding empirical dispersion corrections (e.g., D3) do not really change the geometry or help in improving the ESP fitting.

Molecule	ESP RRMSE		Dipole (Debye)		ESP points
	AM1-BCC	RESP	AM1-BCC	RESP	
CB8	0.33	0.05	0.01	0.01	25855
G1	0.03	0.01	11.03	10.96	6796
G2	0.03	0.02	8.10	8.45	11785
G3	0.04	0.02	12.76	12.86	8737
G4	0.04	0.02	14.75	14.81	8688
G5 prot	0.03	0.02	5.35	5.51	7982
G5 deprot	0.21	0.16	4.91	5.07	7918
G6	0.02	0.01	4.14	4.02	8910
G7	0.04	0.01	9.98	9.94	9988

Table 3. The CB8-guest binding affinities in kcal/mol. ΔG_{exp} is the experimental value. $\Delta G_{\text{AMI-BCC}}$ and ΔG_{RESP} denote the binding free energy obtained with AM1-BCC and RESP charges, respectively. SD represents the standard error of the free energy estimate, which is obtained from block analysis. MSE, MAE, RMSE, τ , and PI serve as quality measurements. For the guest G5, the ChemAxon predicted (pH 7.45) and the experimental (pH 7.5) pKa values are both very close to the experimental condition (pH 7.4) that the binding affinity is estimated. Thus, we consider the protonated and deprotonated forms to be approximately equally populated in the CB8-G5 complex, which results in the final calculated binding affinity being the average of the protonated and deprotonated G5 results. This averaged value is used in the calculation of error metrics and ranking coefficients, while the protonation-specified results are presented to enable a detailed comparison of the simulated results.

Host	Guest	ΔG_{exp}	$\Delta G_{\text{AMI-BCC}}$	SD	ΔG_{RESP}	SD
CB8	G1	-7.05	-10.1	0.4	-5.9	0.4
	G2	-9.93	-13.5	0.5	-8.2	0.5
	G3	-11.63	-17.6	0.5	-14.9	0.4
	G4	-11.22	-21.4	0.4	-9.9	0.5
	G5 prot	-12.32	-12.0	0.4	-4.1	0.5
	G5 deprot	-12.32	-10.4	0.5	-10.3	0.5
	G6	-14.07	-10.9	0.5	-12.4	0.5
	G7	-7.92	-13.8	0.4	-14.8	0.5
RMSE			5.4		3.6	
MSE			3.5		-0.1	
MAE			4.7		3.0	
τ			0.0		0.2	
PI			0.0		0.2	

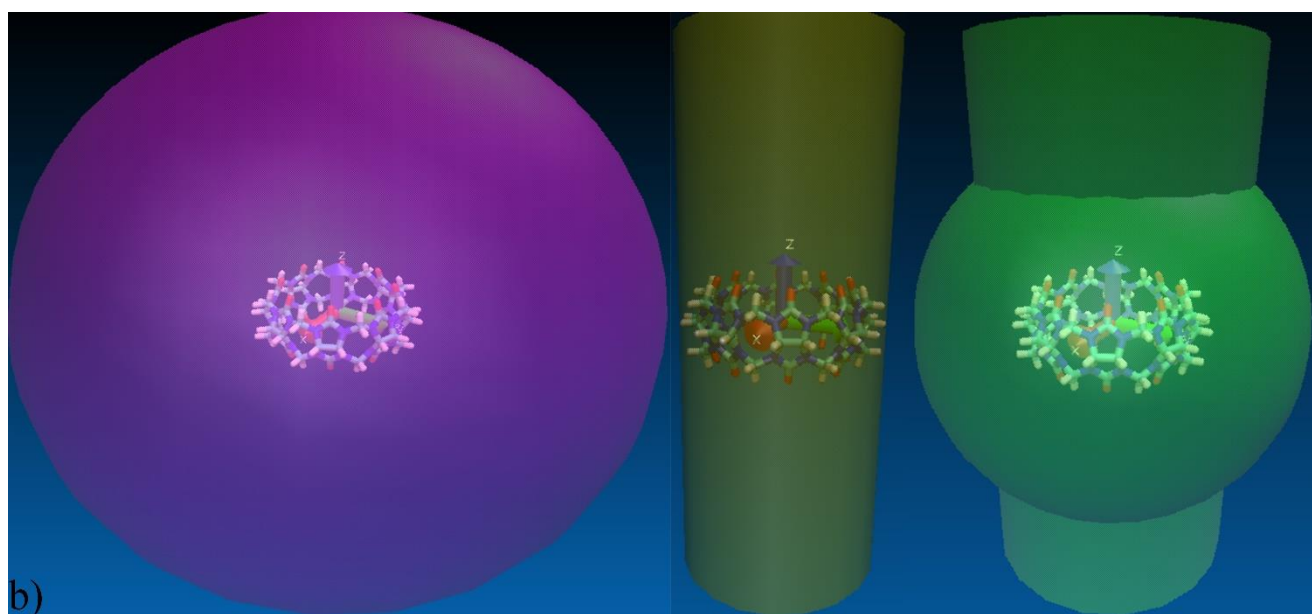
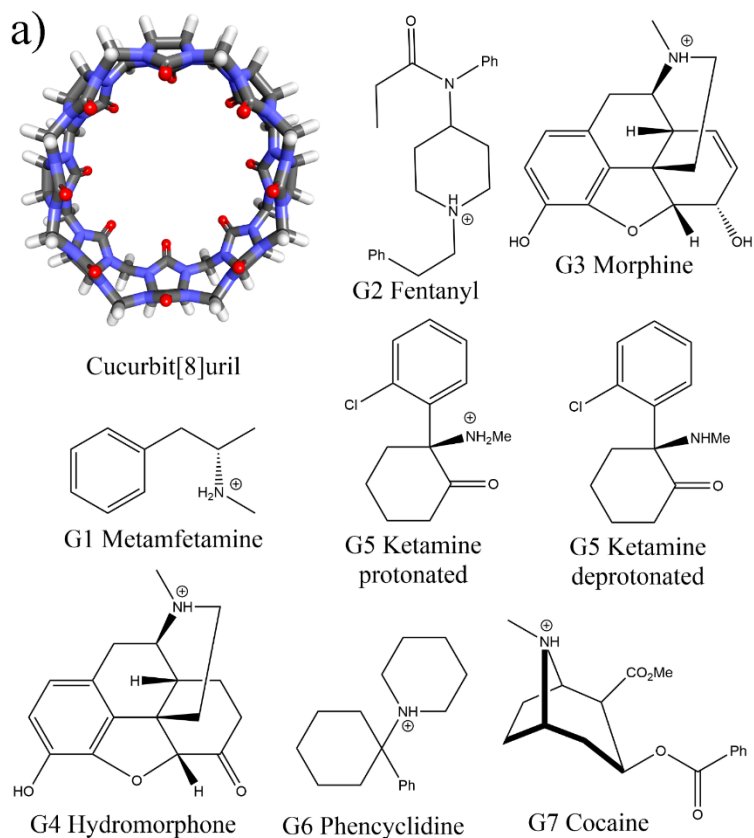


Fig. 1. a) The 3D structure of the supra-molecular container CB8 and the 2D chemical structures of the 7 abused drugs. The protonated and deprotonated forms of the guest G5 (Ketamine) are both considered. b) The spherical region (the magenta sphere) sampled in the current work. Note that alternative restraints such as the cylindrical one (the yellow one) and the combination of the cylindrical and spherical schemes (the green one) shown next to the spherical scheme could also be used in the host-guest simulations.

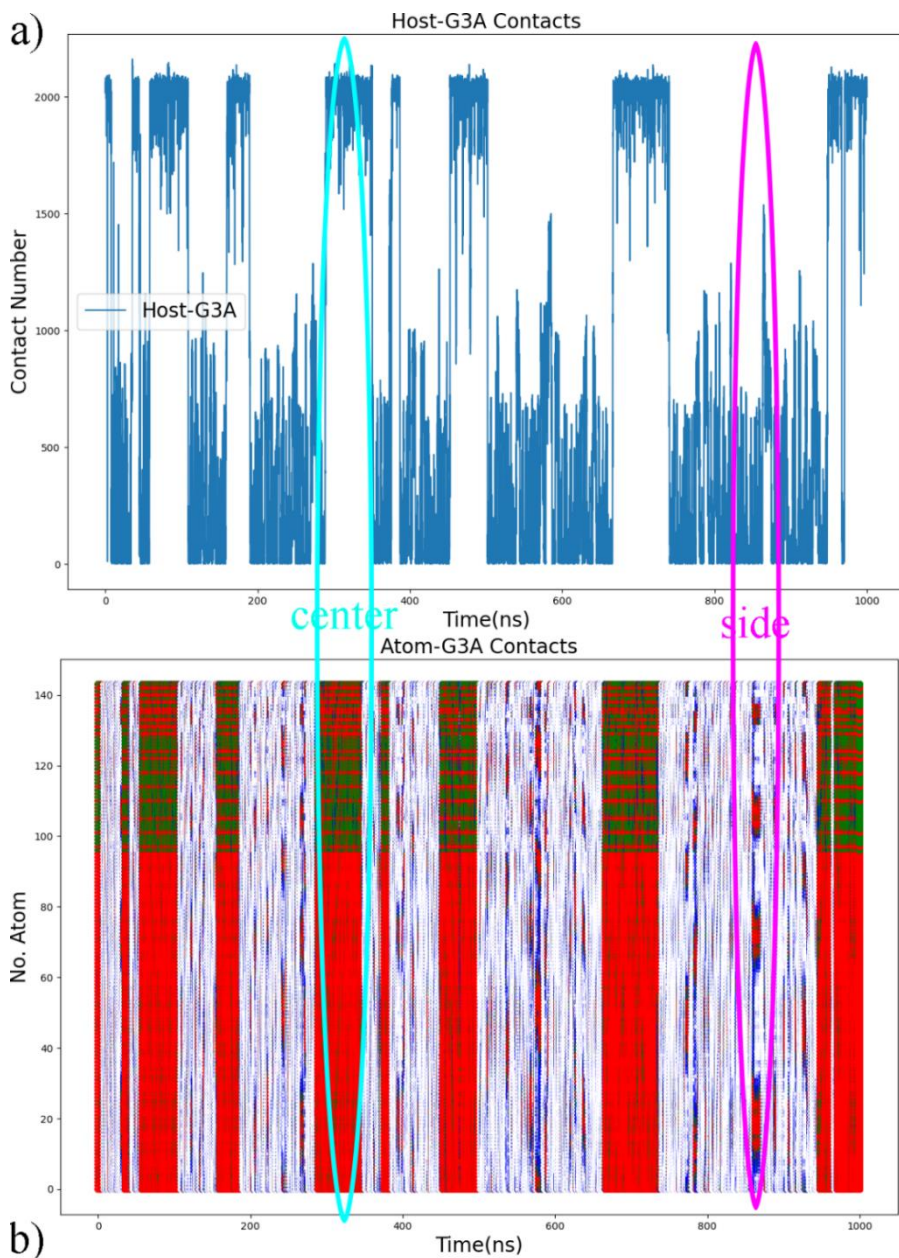
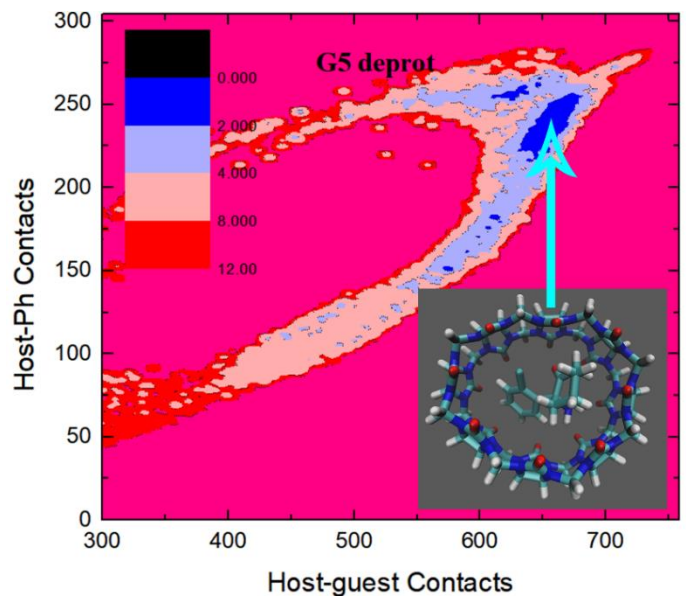
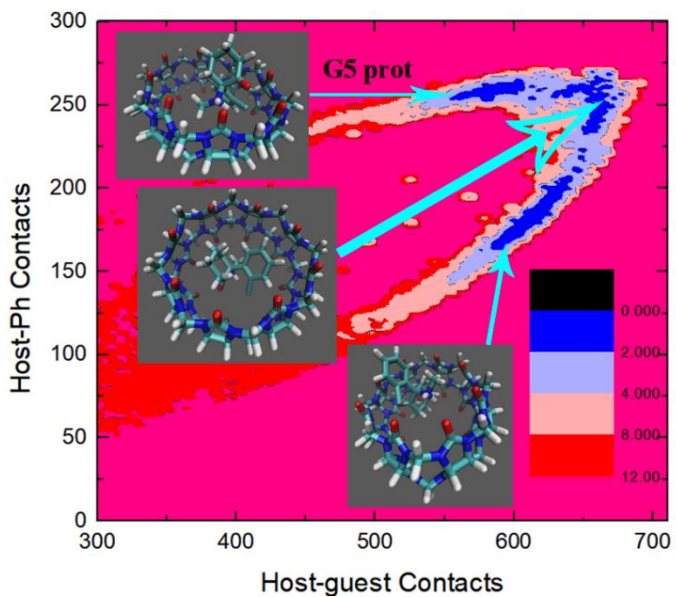
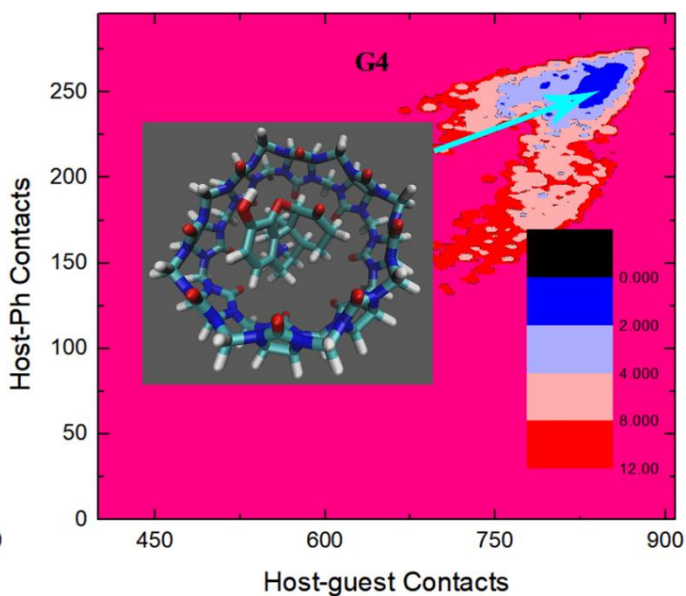
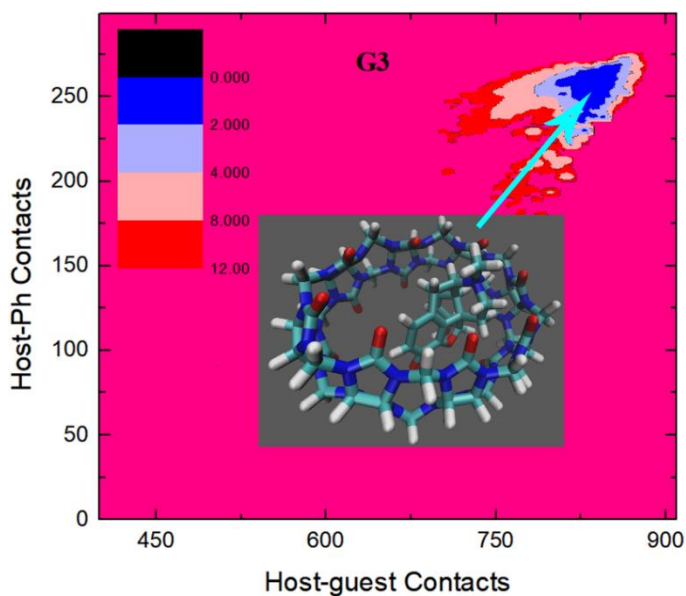
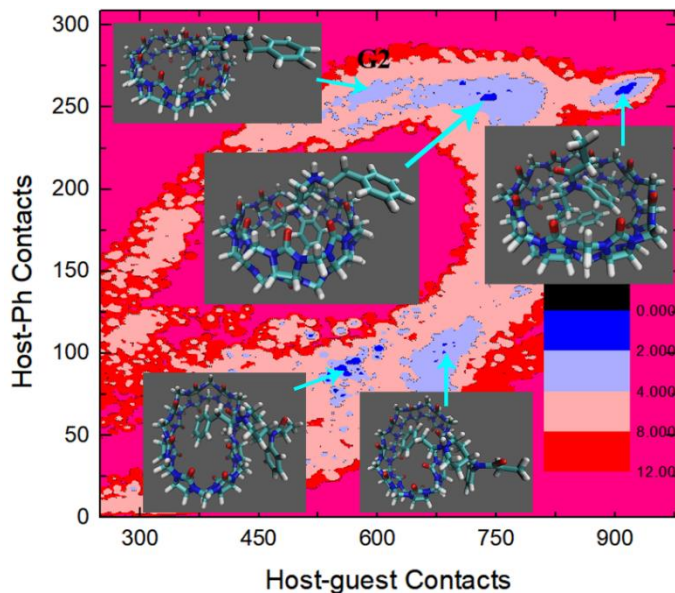
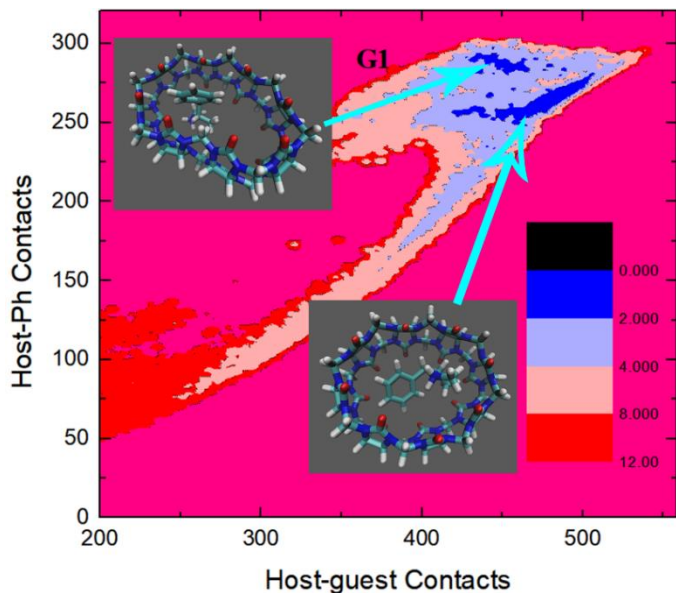


Fig. 2. a) The number of contacts between all atoms of the host CB8 and the guest G3 and b) its decomposition by each atom of the host during metadynamics simulations with the AM1-BCC charge scheme. The y-axis represents the serial number of host atom. Among the 144 atoms of the host CB8, the first 96 atoms are heavy atoms, and the other 48 atoms are hydrogen atoms. As all atoms of the host and the guest are included in the calculation, the number of contacts differs from the later projection of the free energy landscapes. 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. The cyan and magenta ovals provide examples of center-binding and side-binding poses explored during enhanced sampling simulations. The results of all host-guest complexes under two charge schemes are provided in Fig. S3 and S4.



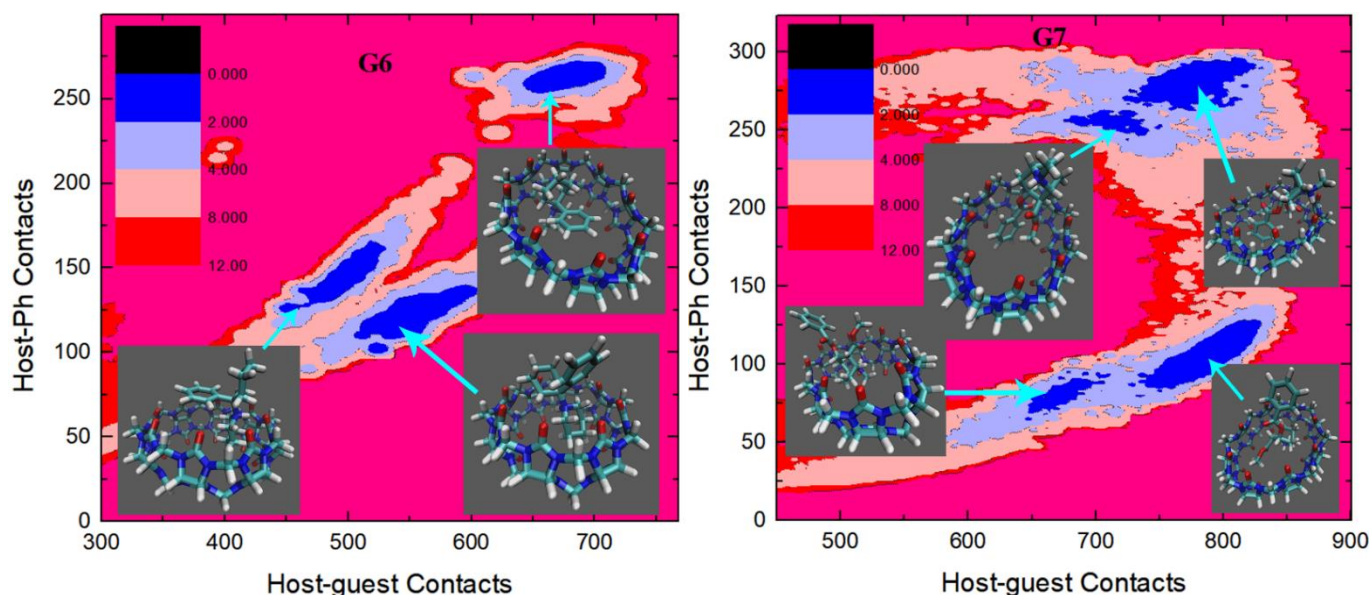
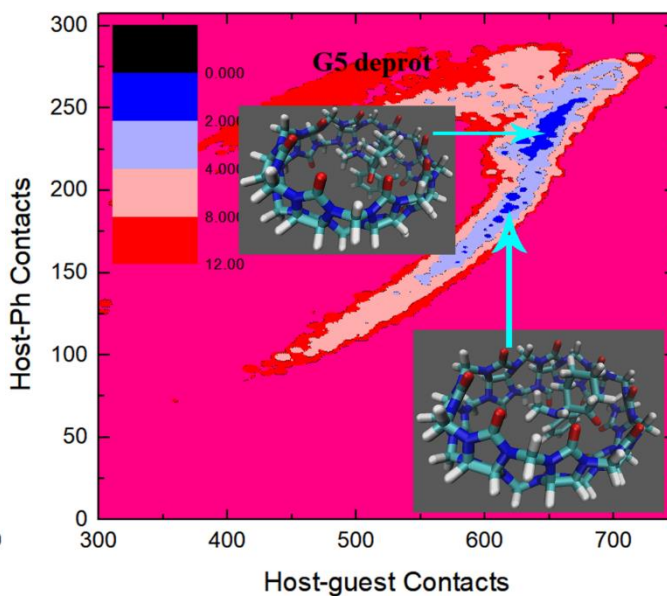
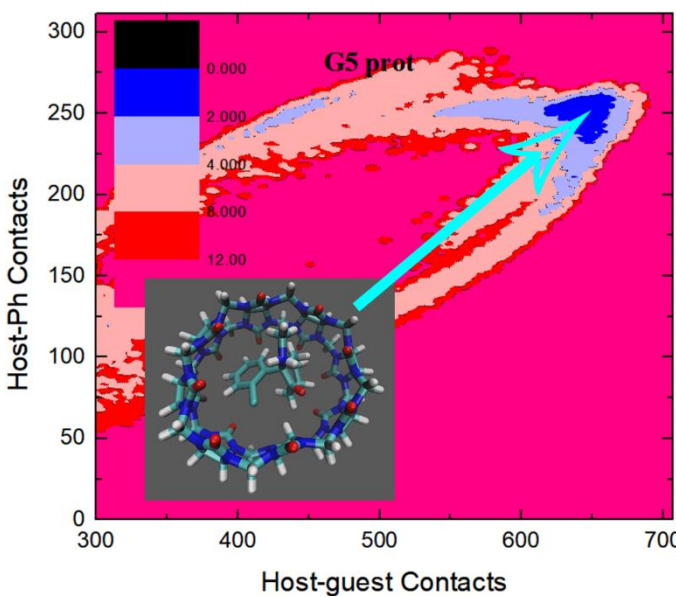
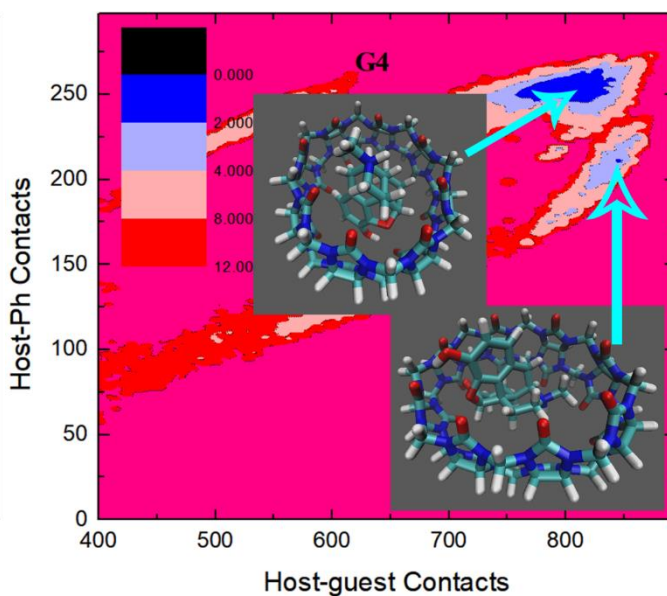
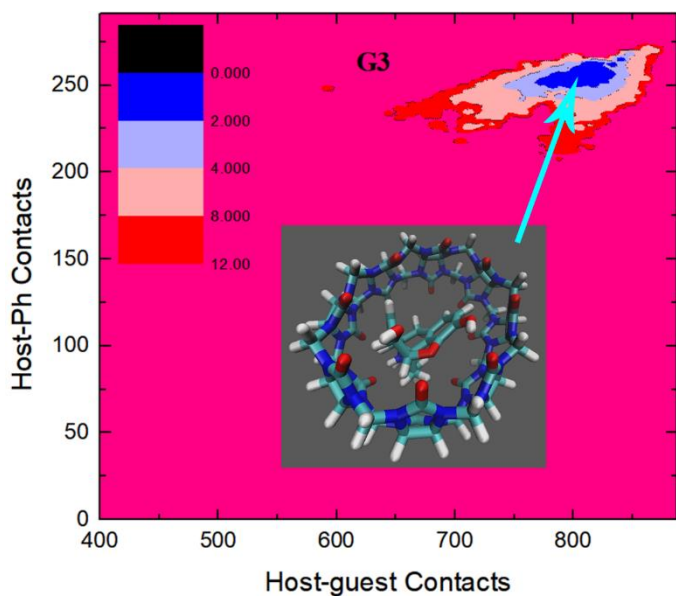
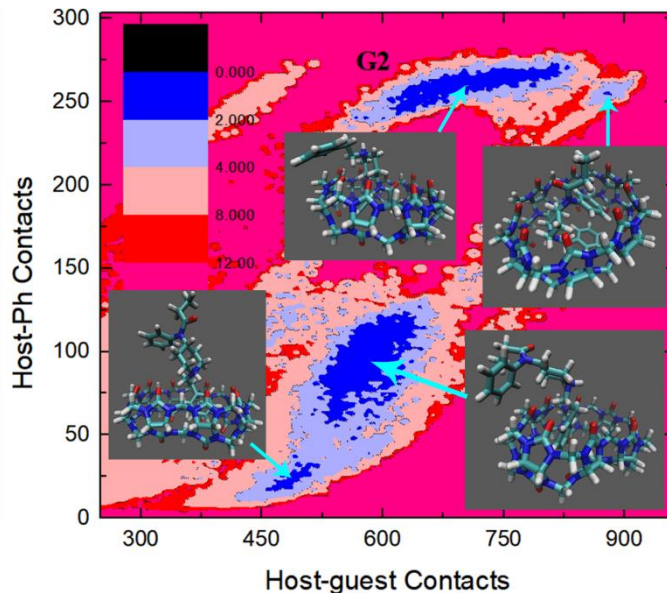
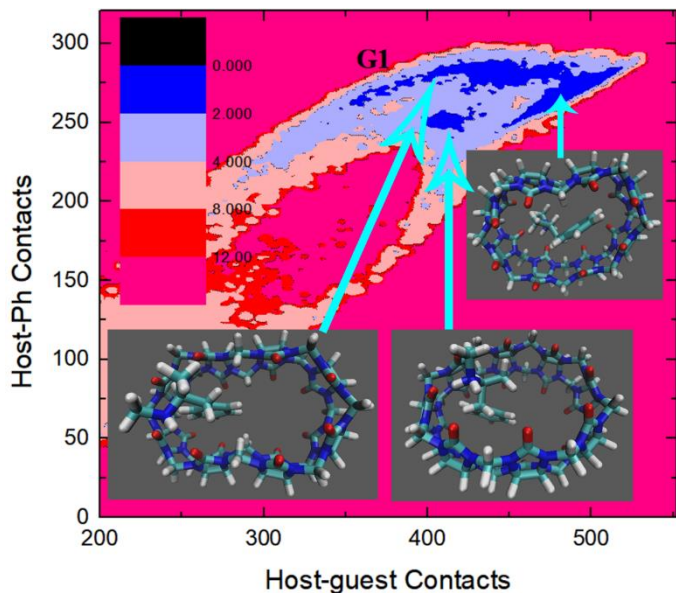


Fig. 3. 2D $C-C_{ph}$ free energy surfaces in kcal/mol obtained under the AM1-BCC charge scheme. Multiple binding poses are observed for structurally complex G2, G5, G6 and G7 and the structurally simple G1. Despite the simplicity of the G1 structure, two binding poses that are of similar thermodynamic stabilities are observed. One binding mode features the whole guest lying at the center of the host, while only a part of the host is directly holding the guest in the other binding pose. The structurally complex G2 has multiple binding modes that have similar thermodynamic stabilities. The two phenyl rings could stay at the center of the CB8 host simultaneously, or only one aromatic ring penetrates the host and the other parts of the guest form a half-closed pocket and nestle against the host. The guests G3 and G4 are structurally similar, and their 2D free energy surfaces also show similar features. The cyclic structures of G3 and G4 limit their conformational flexibility and lead to their rigid-body dynamical behaviors, which results in only one binding mode available. The protonated and deprotonated forms of G5 have similar 2D free energy surfaces. The crescent on the G5 surface is triggered by the fluctuations of the guest in its bound state. The center of the crescent state has the whole guest inside the cavity, while either of its 6-membered rings could fluctuate and become more solvent-exposed. The 3 binding modes of the guest G6 could be classified according to the degrees of penetrations of the 3 rings of this guest and the orientation of the $-NH^+$ group. There is an obvious free energy barrier between the two small- C/C_{ph} states and the large- C/C_{ph} one, which is caused by the need of distorting the CB8 ring in order to fit the penetrating groups of the large guest. The two small- C_{ph} binding patterns of G7 feature the phenyl ring outside the CB8 cavity. The other groups could penetrate the host or stay outside the cavity. Both of the two large- C_{ph} poses have the phenyl ring and the $-COOMe$ group inside the cavity and the other outside the cavity, and their difference lies in the orientation of the hydrogen

bond donor -NH^+ . When the inter-molecular $\text{-N-H}\cdots\text{O=C-}$ hydrogen bond is formed in the large-C conformation, the host-guest complex becomes more compact and more stable than the small-C conformation.



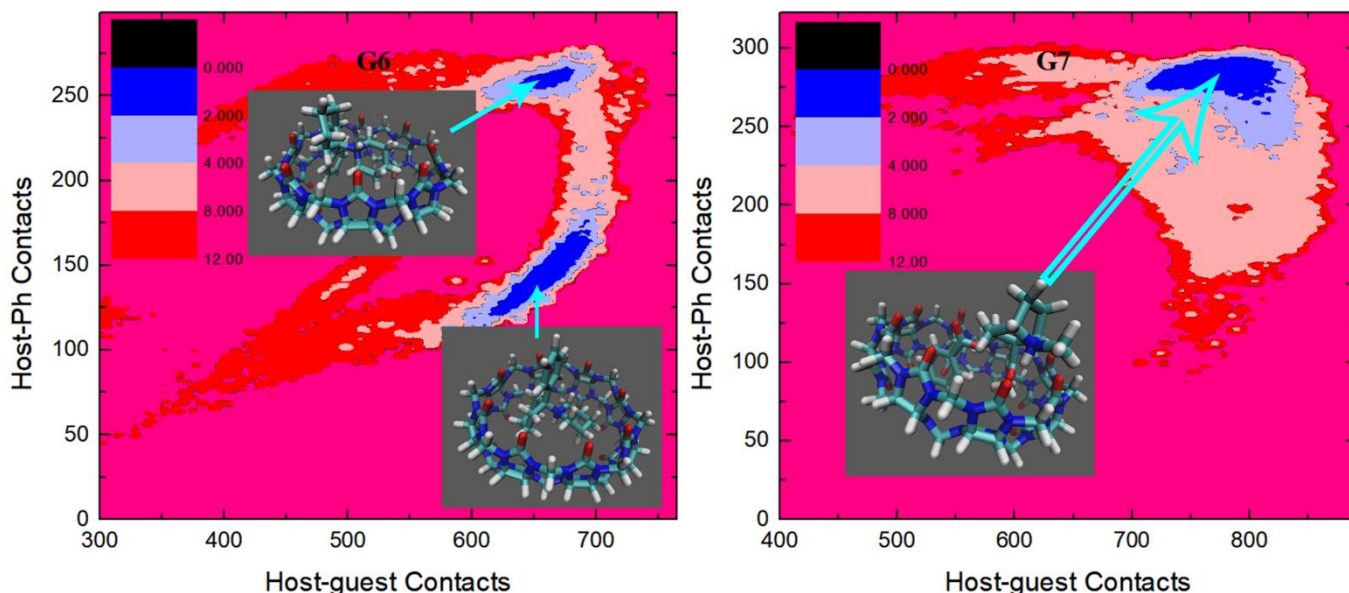


Fig. 4. 2D $C-C_{ph}$ free energy surfaces in kcal/mol obtained under the RESP charge scheme. The free energy landscapes and the binding poses are similar to the previous AM1-BCC cases, but obvious differences still exist. For the structurally simple G1, the bound state with large host-guest and host-phenyl contacts has the whole guest inside the CB8 cavity, while fluctuations of the host-guest system lead to other possible binding modes that have the phenyl ring inside the cavity and the other parts of the group outside and their differences mainly lie in the distortion of the host backbone. The free energy landscape for the structurally complex G2 Fentanyl shows obvious differences with the AM1-BCC results. 4 conformational states could be identified. The large-C small- C_{ph} state is the global minimum and features the guest nestling against the host with the phenyl ring connected to the amide bond binding to the outer surface of the host and the other phenyl ring penetrating the host, which is similar to the AM1-BCC result. In the small-C large- C_{ph} state, the whole guest clings to the host and phenyl at the amide bond side penetrates the host. The typical center-binding pose is in the large-C large- C_{ph} state, where the whole guest is in the CB8 cavity, the two phenyl rings form intra-molecular stacking interactions, and $-NH^+$ forms hydrogen bond with one $-C=O$ portal. In the small-C small- C_{ph} state, phenyl at the $-NH^+$ side penetrates the cavity, $-NH^+$ forms inter-molecular hydrogen bond with $-C=O$ portal, the other parts of the guest stay away from the host. The guest G3 only has one minimum, which is similar to the AM1-BCC case. For the guest G4, the global minimum is similar to the AM1-BCC result. The $-OH$ group forms inter-molecular hydrogen bond with $-C=O$ portal. $-NH^+$ is close to the $-C=O$ portal, but the donor-acceptor distance is larger than the normal 3.5 Å criterion for hydrogen bonds. The other less stable pose is similar to the most stable binding pose, and the difference is

that the -NH^+ group is perpendicular to the CB8 ring and far from the host. For protonated G5, the conformational fluctuation of the host-guest complex is smaller than the AM1-BCC case, and the two 6-membered rings are stably coordinated by the host. The deprotonated G5 shows similar behaviors, but its conformational fluctuation seems more significant. The surface of deprotonated G5 is similar to the AM1-BCC one. The free energy landscape for the structurally complex G6 also shows obvious differences compared with the AM1-BCC result. The two binding poses have similar host-guest contacts, and the difference lies in the degree of penetration of the phenyl ring. In the small- C_{ph} binding pose, the piperidinium and cyclohexyl moieties penetrate the cavity and the phenyl ring is solvent-exposed, and the orientation of -NH^+ makes it impossible for the formation of intermolecular hydrogen bonds. Note that this pose is unstable under the AM1-BCC charge scheme. The other large- C_{ph} minimum has the phenyl and piperidinium rings inside the host cavity, which is also observed under the AM1-BCC charge scheme. For the last guest G7, only the pose with the phenyl and -COOMe groups in the cavity is observed to be stable, and this guest conformation is extremely similar to the configuration used to generate the RESP charges.

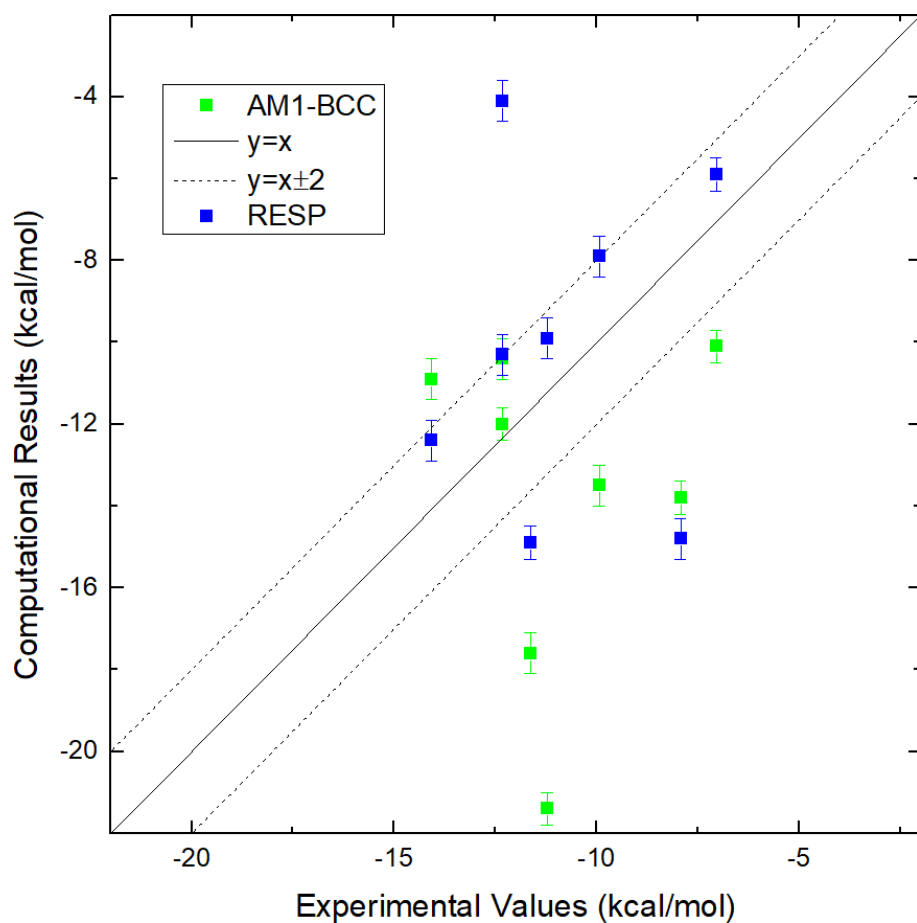


Fig. 5. Correlation between the binding affinities obtained from our computational modeling and the experimental reference for CB8-guest systems. The exact values of the binding affinities are presented in Table 3.

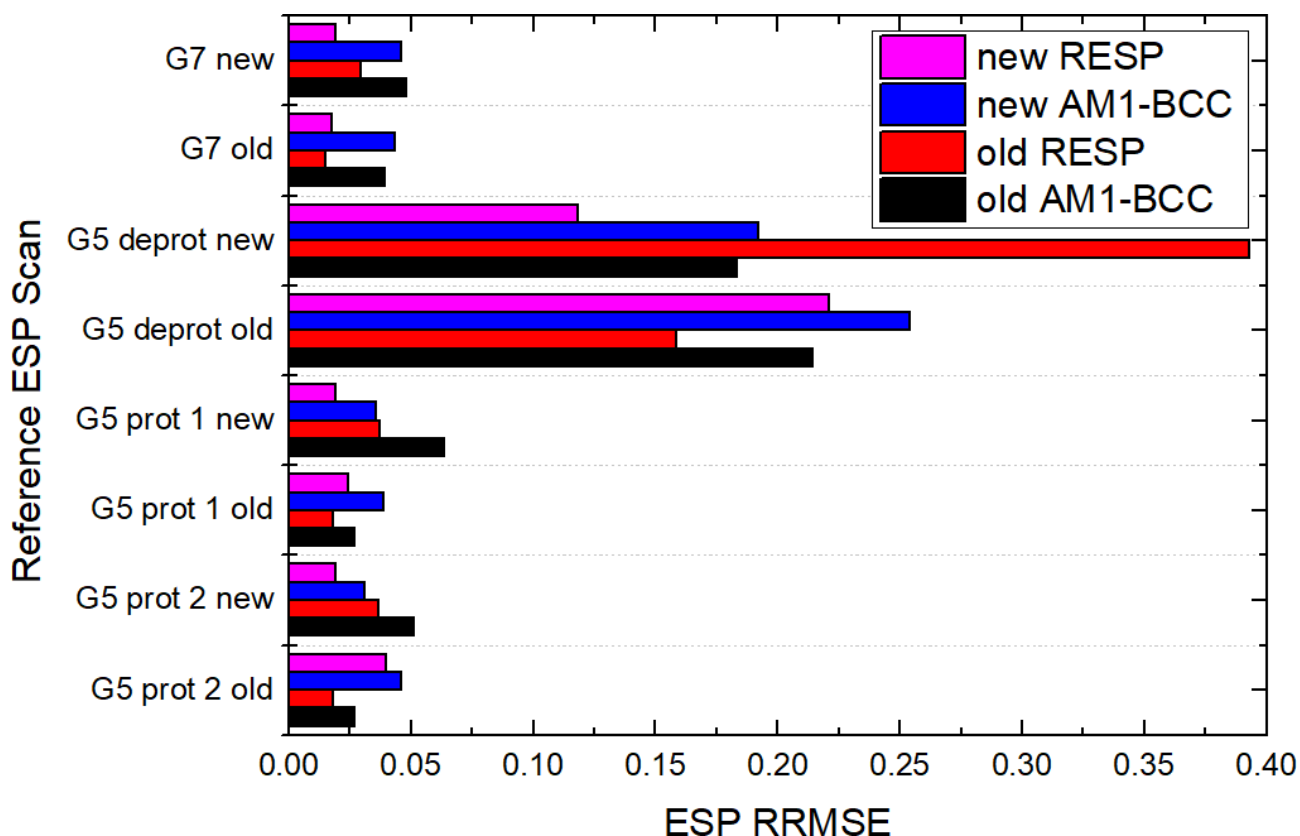
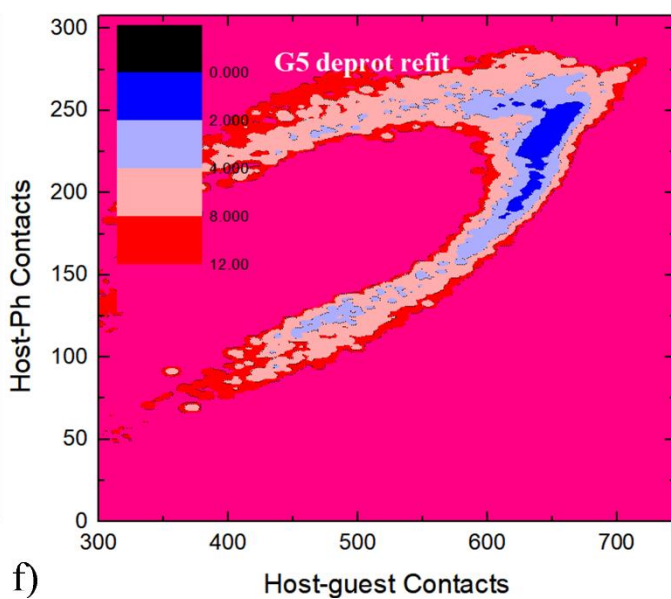
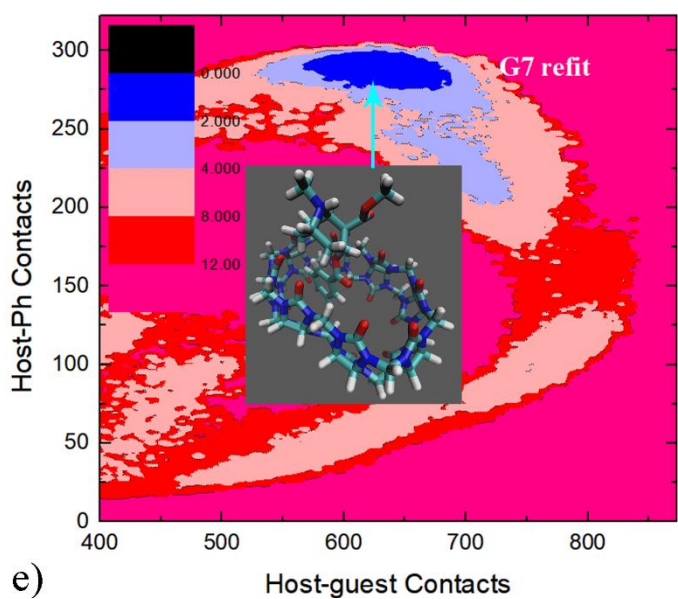
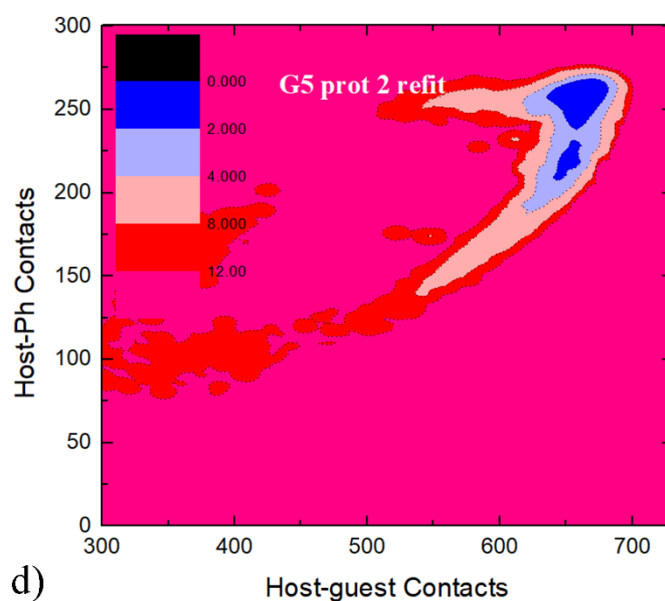
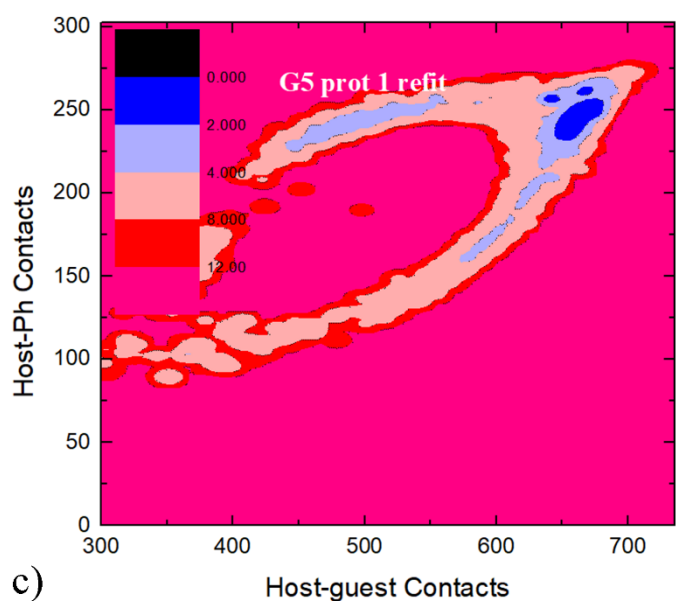
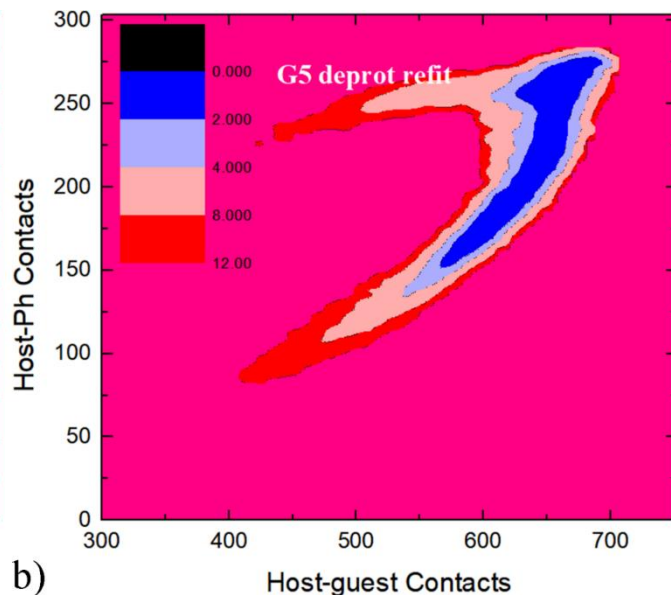
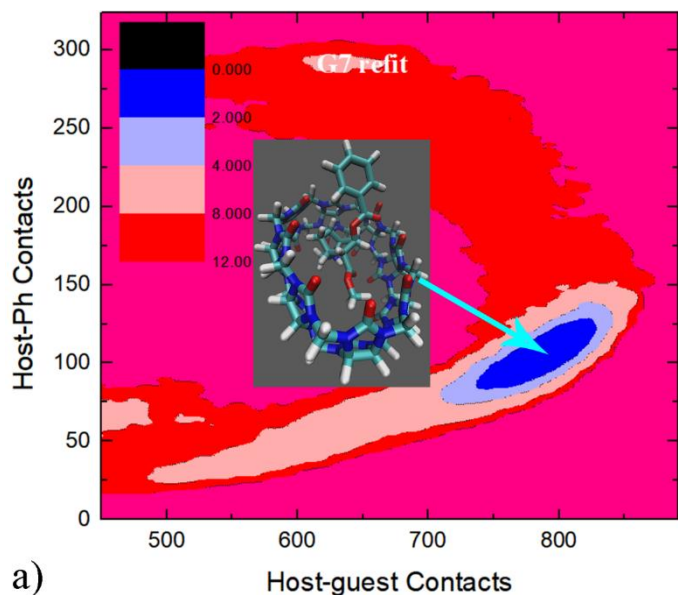


Fig. 6. ESP RRMSE calculated with the original and newly generated atomic charges with the two charge schemes. The old charges are obtained from the old conformation used in the original configurational space exploration, while the new charges are refitted from the bound conformation extracted from the global free energy minimum explored with the old charges. For the guest G7, the conformation has minor impacts on the AM1-BCC ESP reproduction, while its influence on the RESP ESP RRMSE is relatively large. The ESP RRMSE of RESP charges is always smaller than AM1-BCC irrespective of conformations. For deprotonated G5, due to the significant difference between the guest conformation used in initial charge generation and the bound state obtained from enhanced sampling simulations, the conformational dependence of charge fitting is much more significant. The ESP RRMSEs of RESP charges obtained from the old and new configurations show significant differences, while those for AM1-BCC charges are relatively insensitive to the conformation, suggesting its mean-field-like behavior compared with RESP and its suitability for preliminary conformational exploration. As for the protonated form of G5, the ESP reproductions are much better for the two charge schemes, and the conformation-dependent behavior is moderate. Two models are generated for the protonated G5. The first one (G5 prot 1) is based on the structure that has the intra-molecular hydrogen bond formed by Cl and -NH_2^+ , while the other (G5 prot 2) is obtained by adding one hydrogen atom to the deprotonated G5 and thus does not have intra-molecular hydrogen bond. Note that the two old reference ESP scans for protonated G5 are the same.



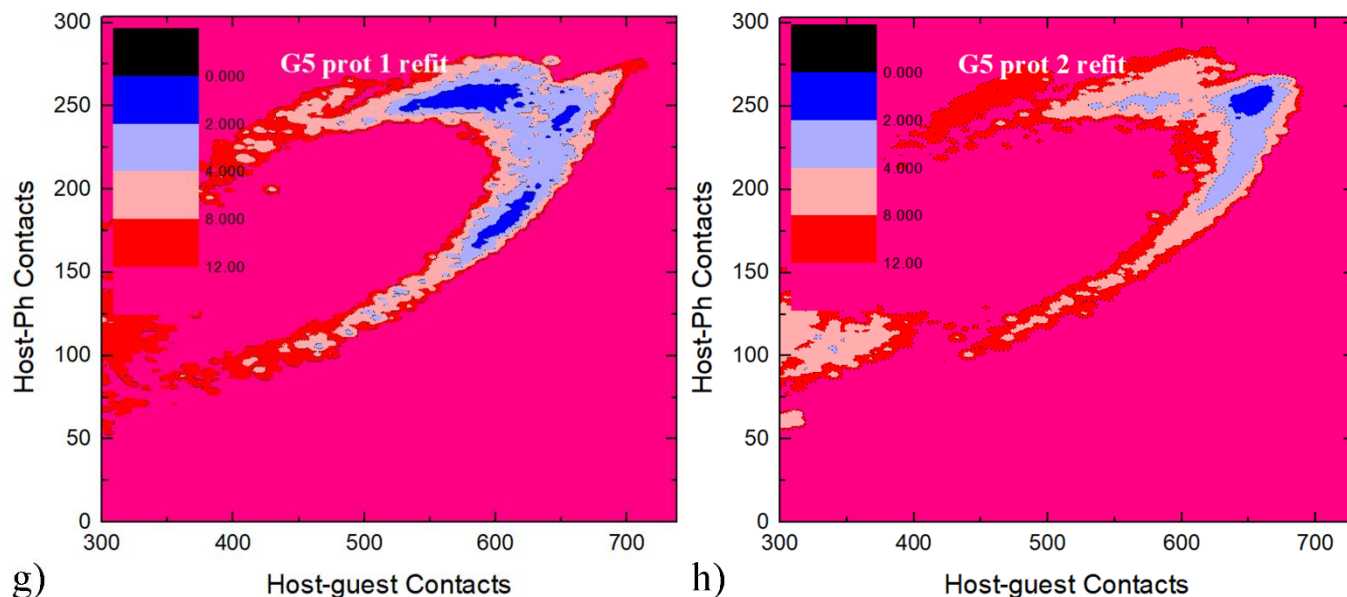


Fig. 7. 2D $C - C_{ph}$ free energy surfaces in kcal/mol with the newly fitted atomic charges: a-d) AM1-BCC and e-h) RESP. Two models are refitted for the protonated G5. The configuration used to generate the first model (G5 prot 1) is similar to the that used in the initial charge generation and has the intra-molecular hydrogen bond formed by Cl and $-NH_2^+$, while the other (G5 prot 2) is obtained by adding one hydrogen atom to the refitted deprotonated G5 and regenerating the charges and thus does not have intra-molecular hydrogen bond.

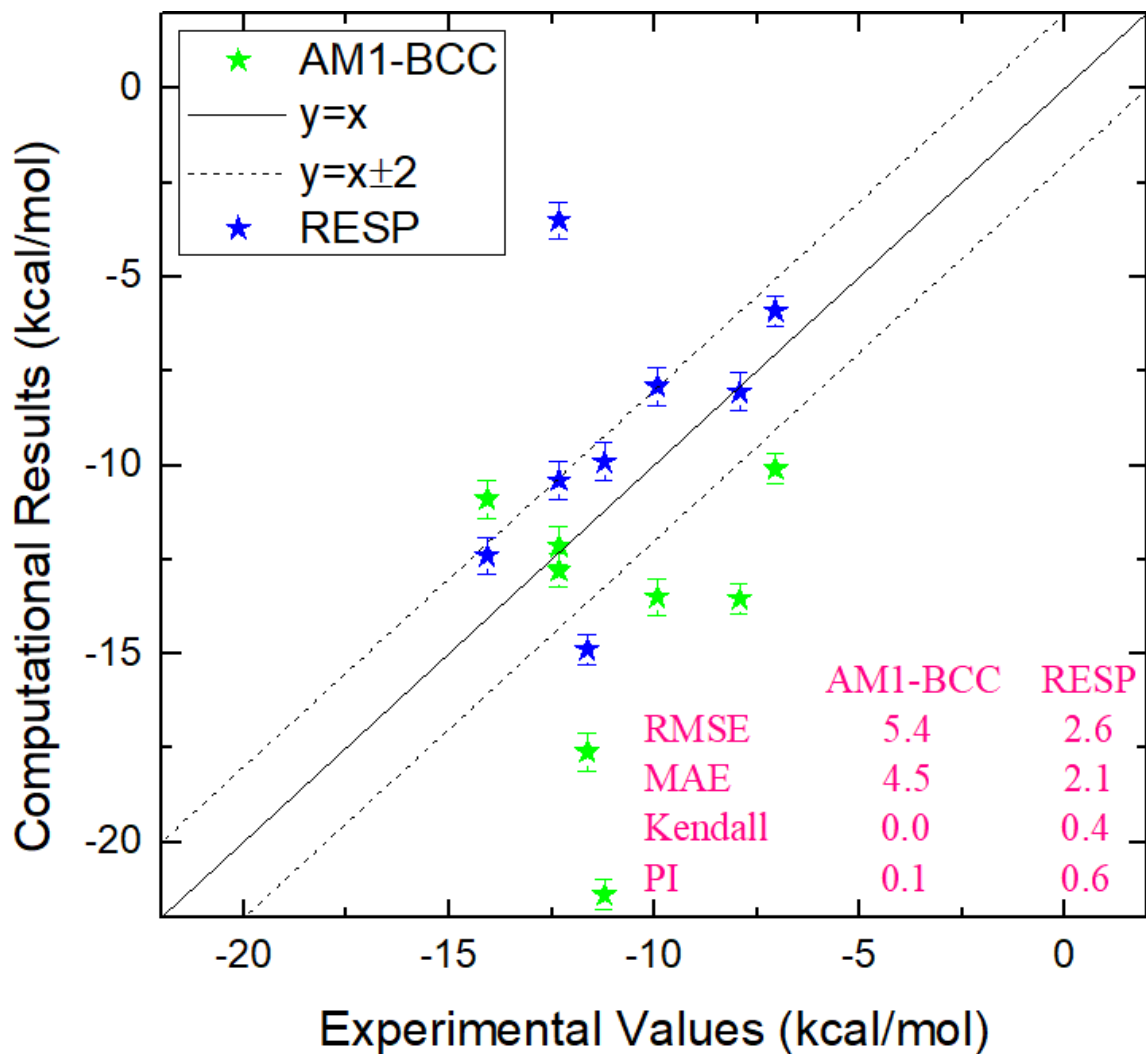


Fig. 8. Correlation between the computed binding affinities with the newly fitted charges and the experimental reference for CB8-guest systems. The error metrics and ranking coefficients are also provided in the plot. The quality metrics calculated with the two models of the protonated G5 are almost identical.

Supporting Information: A General Picture of Cucurbit[8]uril Host-Guest Binding

Zhaoxi Sun^{1*}, Zhe Huai², Qiaole He³, Zhirong Liu¹

¹*Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China*

²*XtalPi - AI Research Center (XARC), 9F, Tower A, Dongsheng Building, No.8, Zhongguancun East Road, Haidian District, Beijing 100083, P.R. China*

³*Enzymaster (Ningbo) Bio-Engineering Co., Ltd., North Century Avenue 333, 315100 Ningbo, China*

*To whom correspondence should be addressed: z.sun@pku.edu.cn

Fig. S1. An illustration of the 3D spherical coordinates CV.

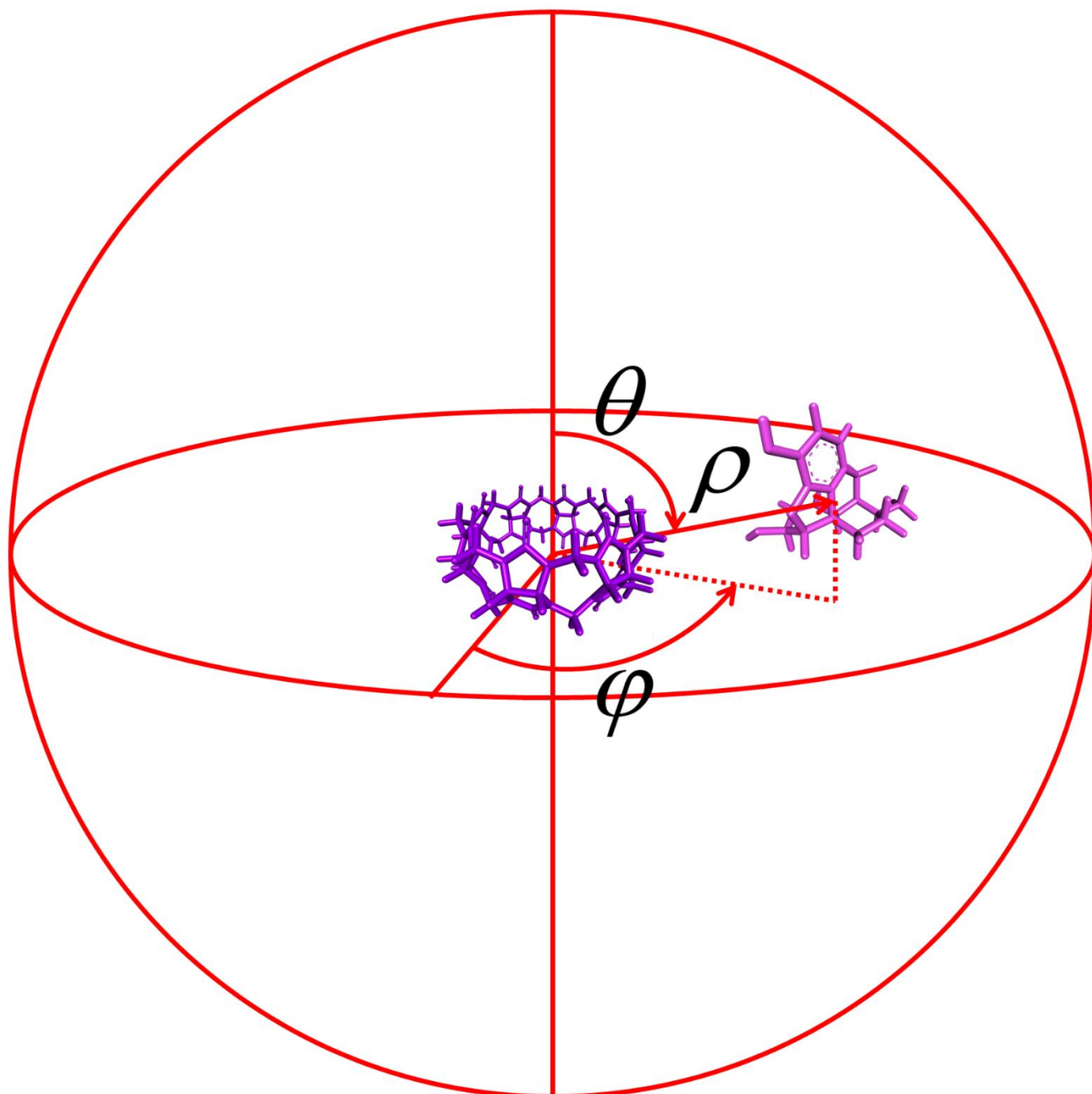


Fig. S2. Comparison between the AM1 and B3LYP/6-31G* optimized structures used in ESP fitting. The optimized structures at the two levels are virtually identical for G4 and protonated G5, have negligible difference for CB8, G1 and G6, and minor differences exist for the other molecules. However, as shown in Table 1, for the macrocyclic host CB8 and the deprotonated G5, the two charge schemes show obvious differences, and neither of the two charge schemes performs well for the deprotonated G5. The failure in the former CB8 case suggests that the AM1-BCC model, which is fitted using small drug-like molecules, could fail to reproduce the HF/6-31G* ESP for large macrocycles, emphasizing the use of more accurate charge models. The failure in the latter deprotonated G5 case arises from the inability of the atomic-centered fixed-charge model to reproduce the ESP around the molecule. By adding one more atomic-centered fitting point upon the protonation at the N atom and producing the protonated form of G5, both charge models could reproduce the ESP with small errors. This phenomenon indicates that the addition of some dummy points to describe the lone pairs could be necessary.

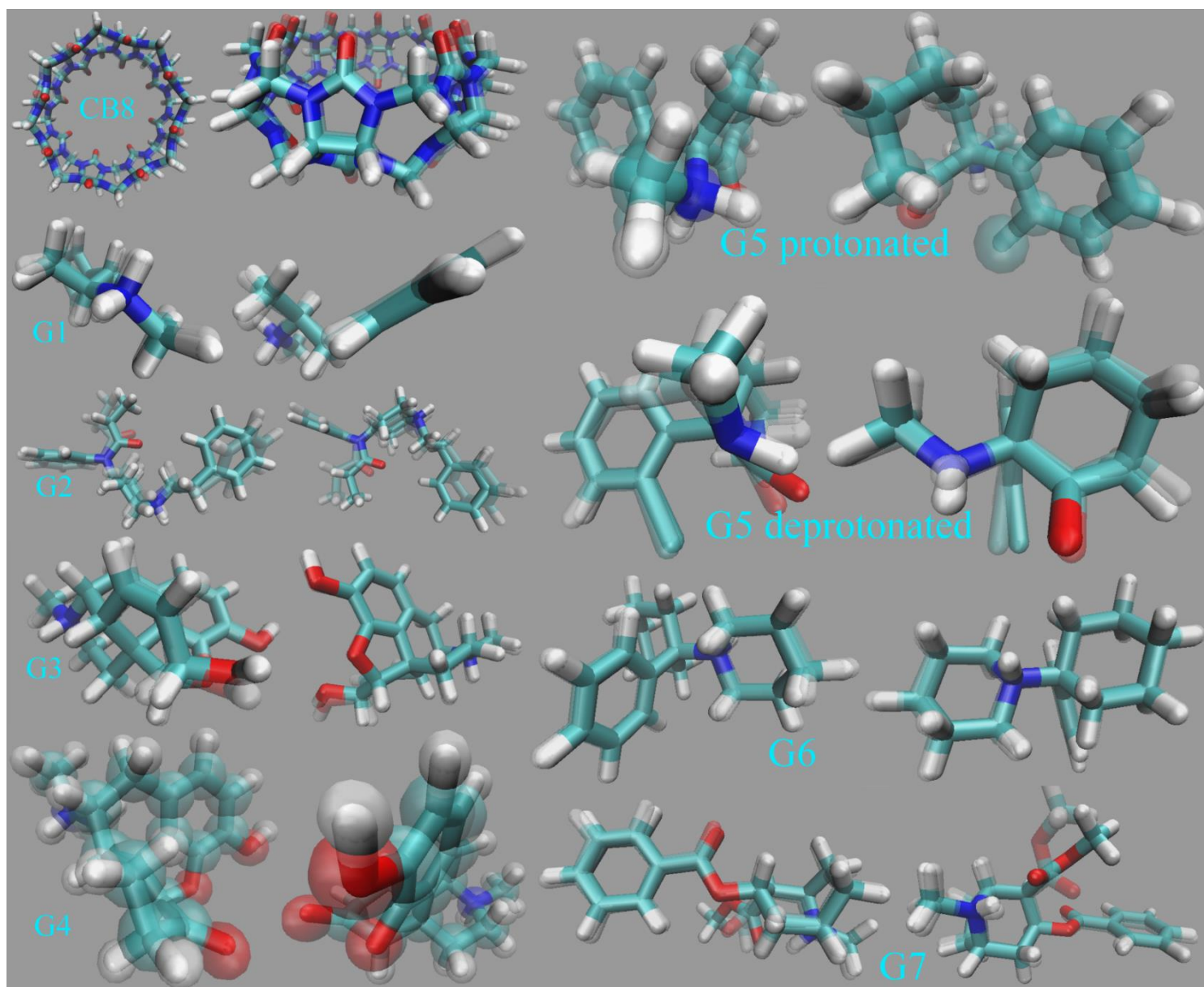
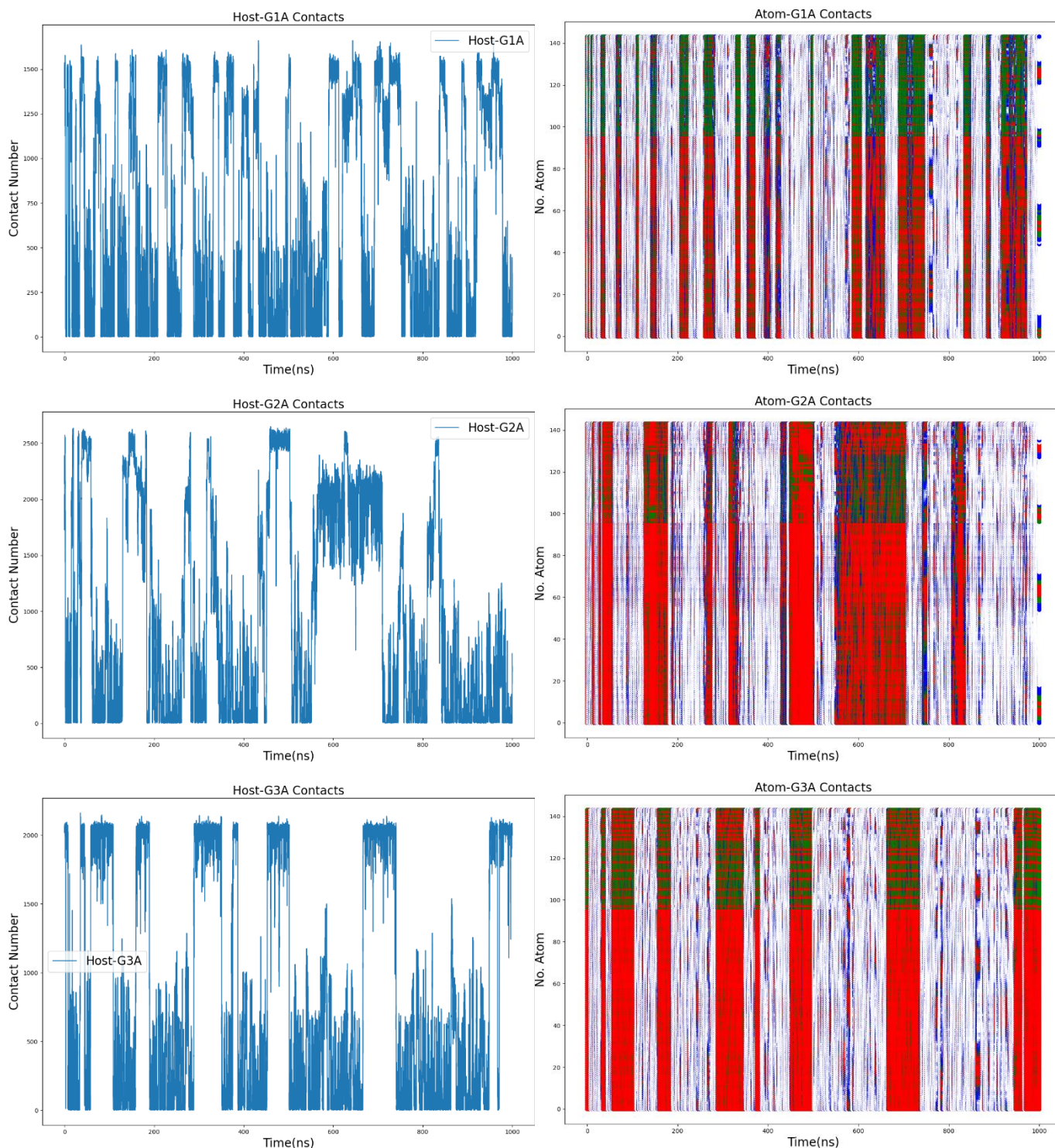
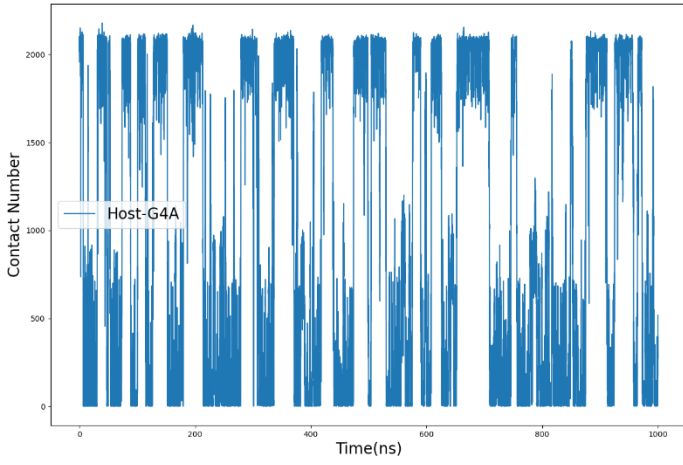


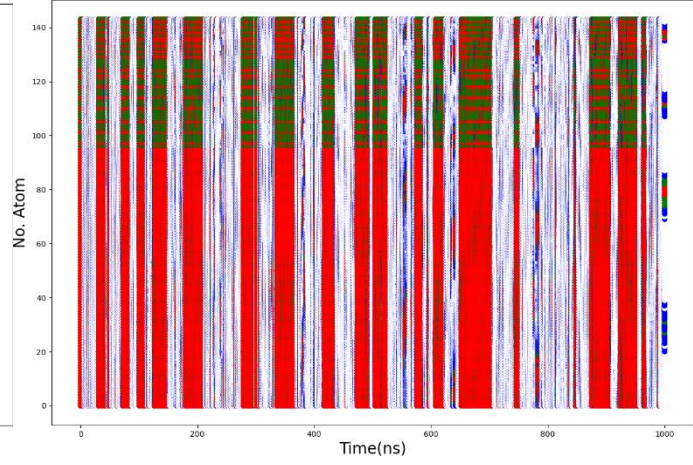
Fig. S3. The number of contacts between all atoms of the host CB8 and the guests and the by-host-atom decomposition during metadynamics simulations under the AM1-BCC charge scheme. The y-axis represents the serial number of host atom. Among the 144 atoms of the host CB8, the first 96 atoms are heavy atoms, and the other 48 atoms are hydrogen atoms. All atoms of the host and the guest are included in the calculation. 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.



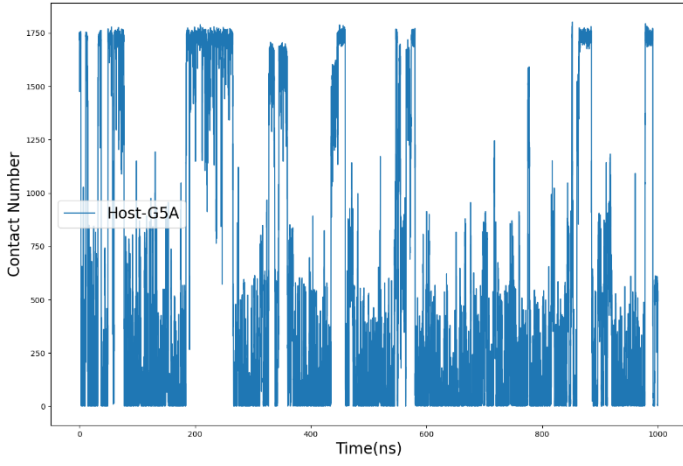
Host-G4A Contacts



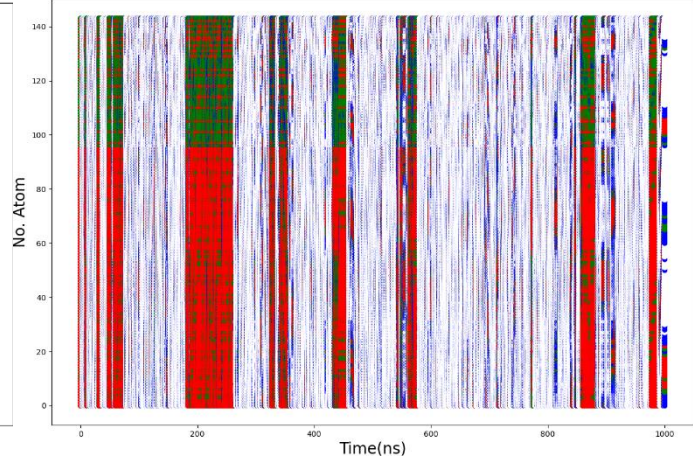
Atom-G4A Contacts



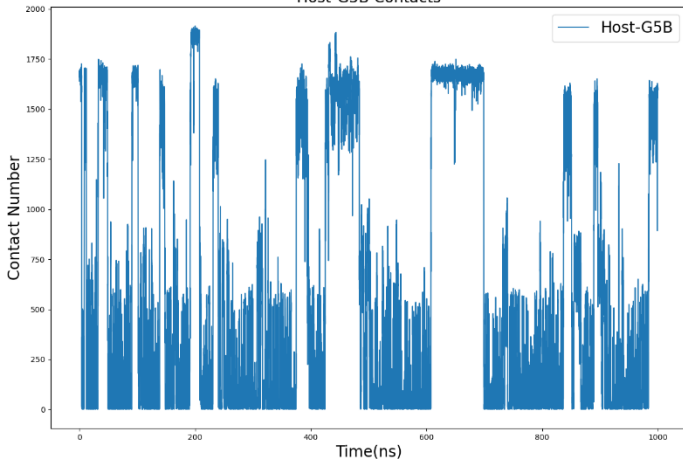
Host-G5A Contacts



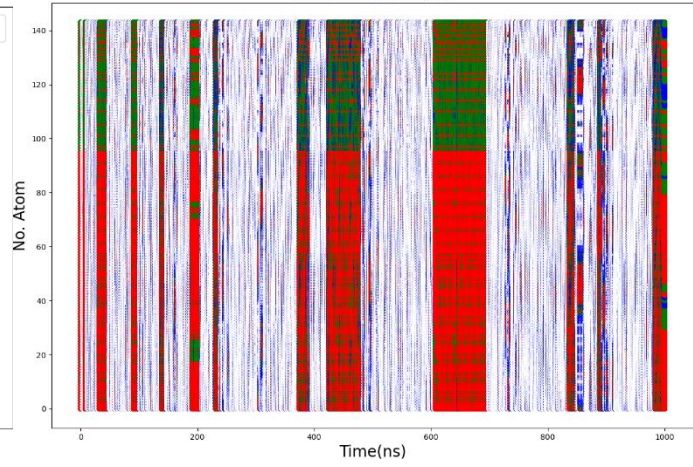
Atom-G5A Contacts



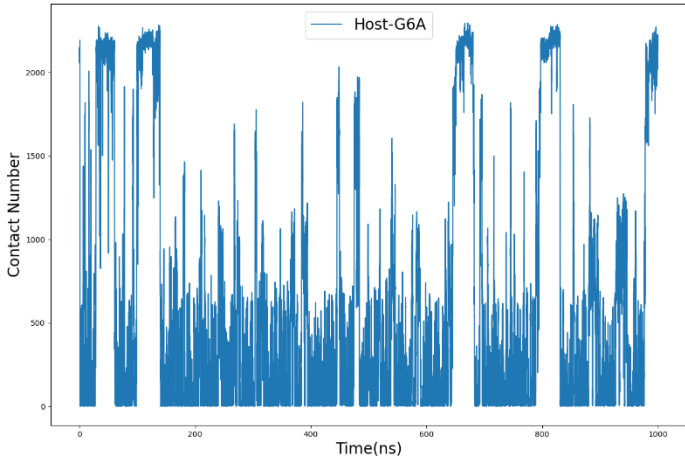
Host-G5B Contacts



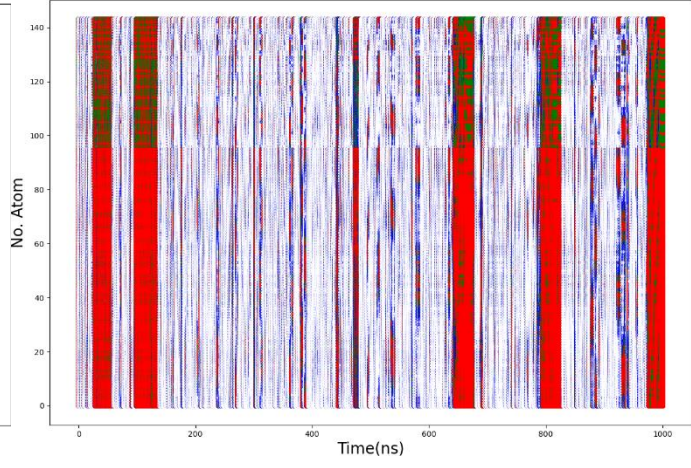
Atom-G5B Contacts



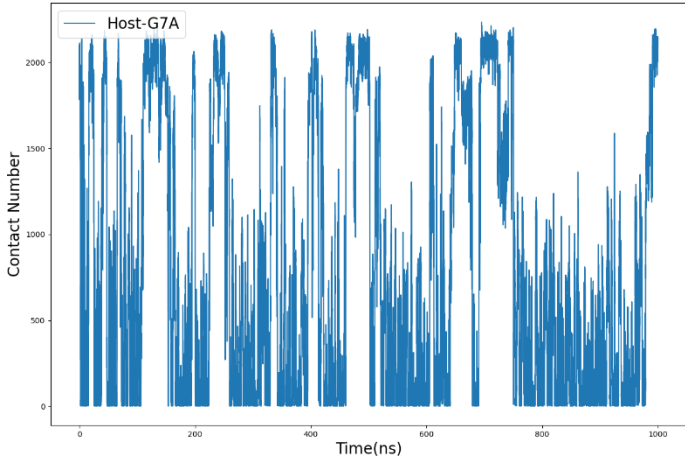
Host-G6A Contacts



Atom-G6A Contacts



Host-G7A Contacts



Atom-G7A Contacts

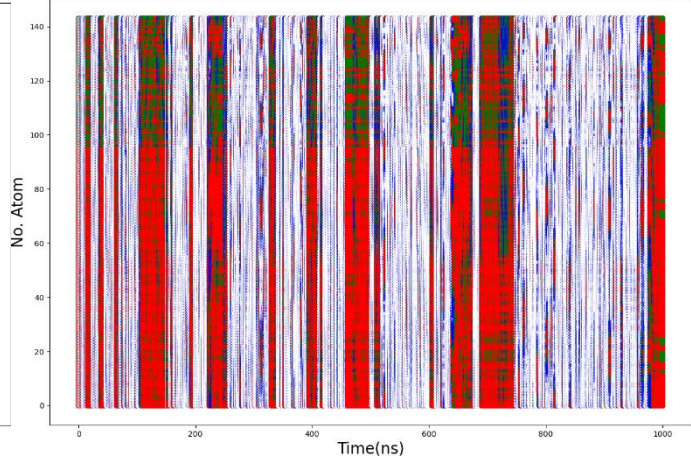
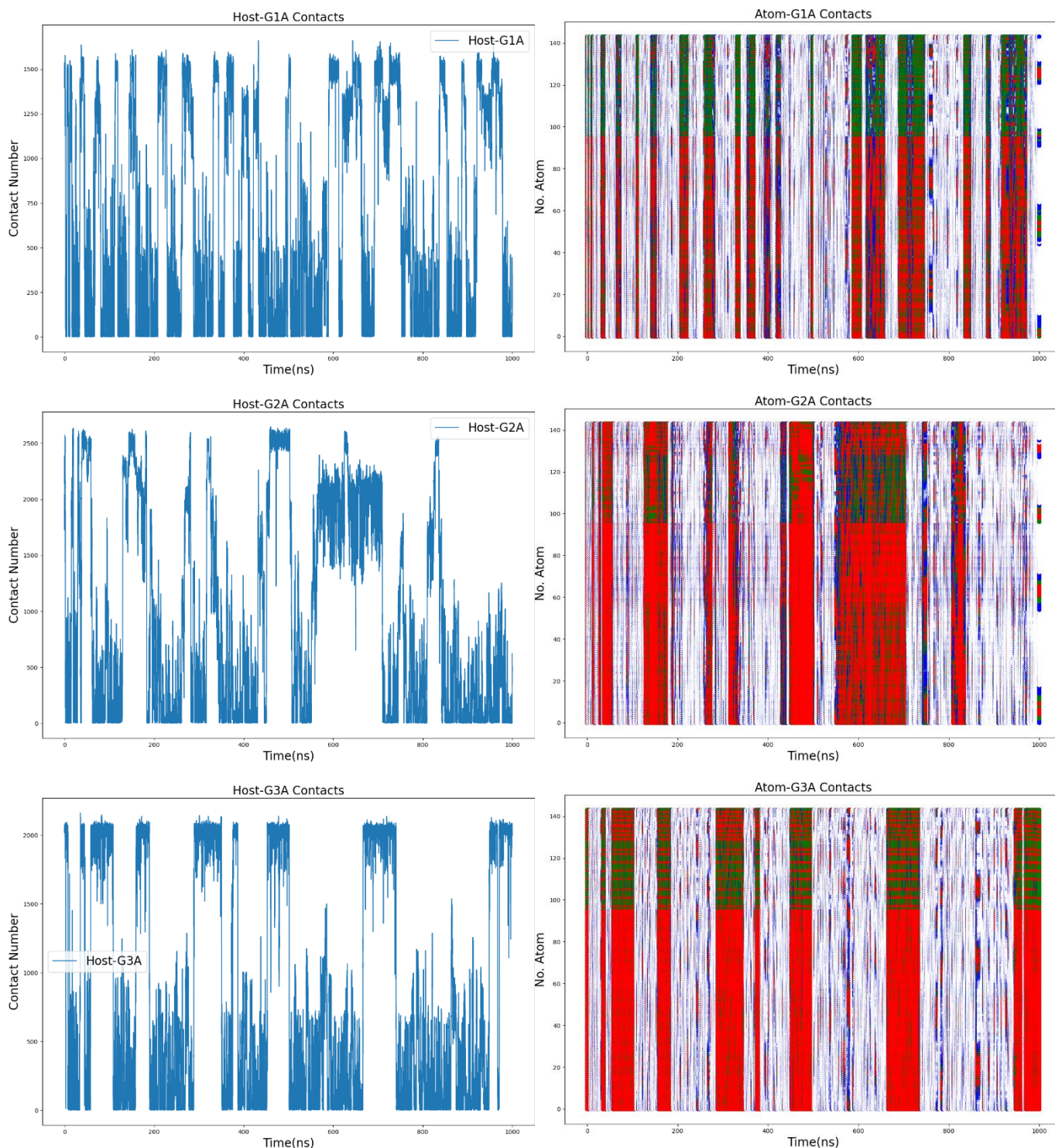
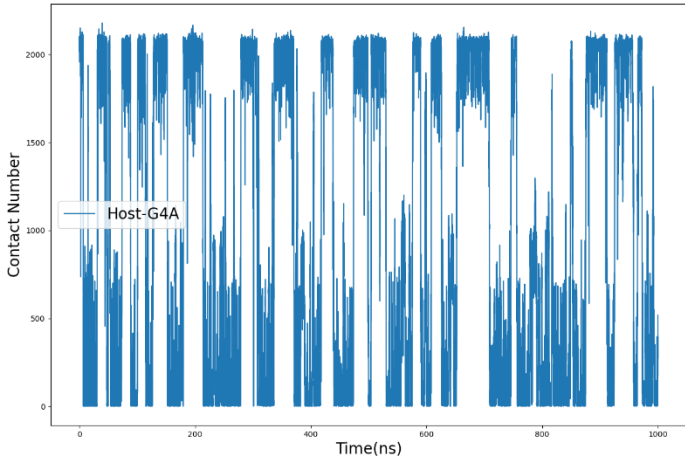


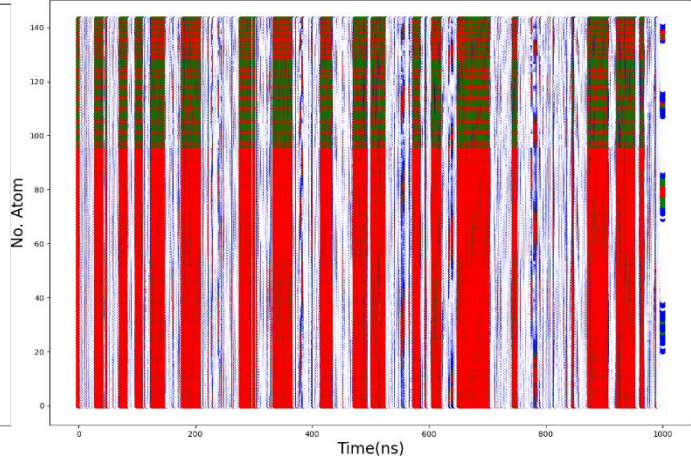
Fig. S4. The number of contacts between all atoms of the host CB8 and the guests and the by-host-atom decomposition during metadynamics simulations under the RESP charge scheme. The y-axis represents the serial number of host atom. Among the 144 atoms of the host CB8, the first 96 atoms are heavy atoms, and the other 48 atoms are hydrogen atoms. All atoms of the host and the guest are included in the calculation. 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.



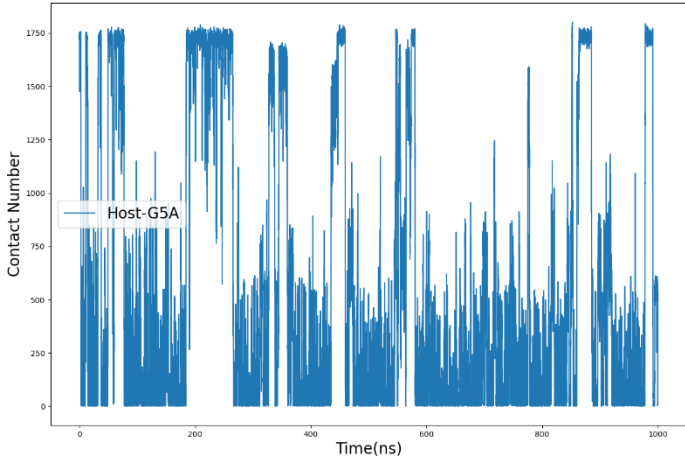
Host-G4A Contacts



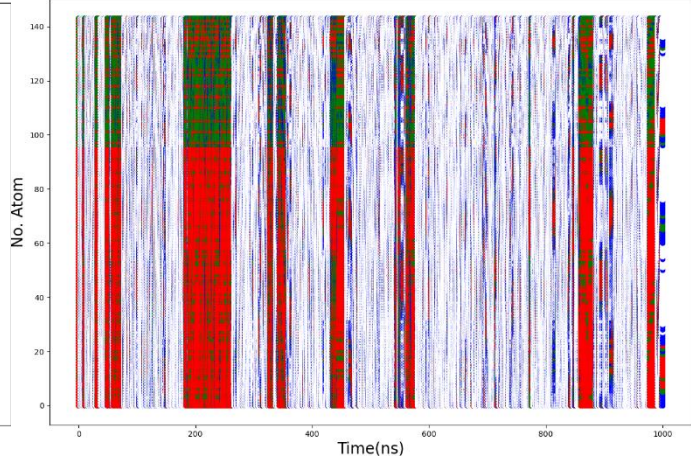
Atom-G4A Contacts



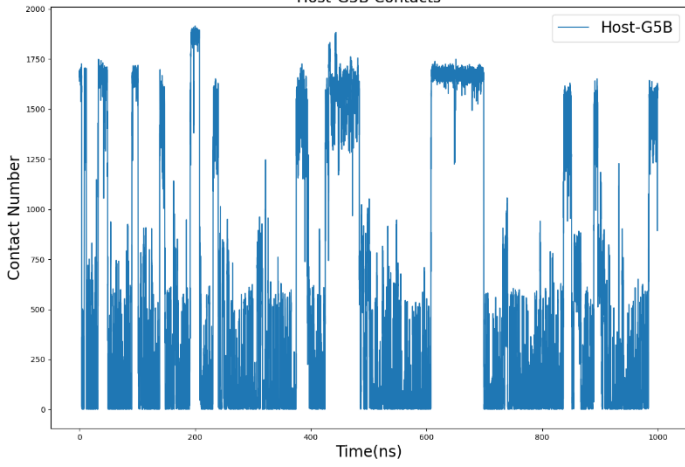
Host-G5A Contacts



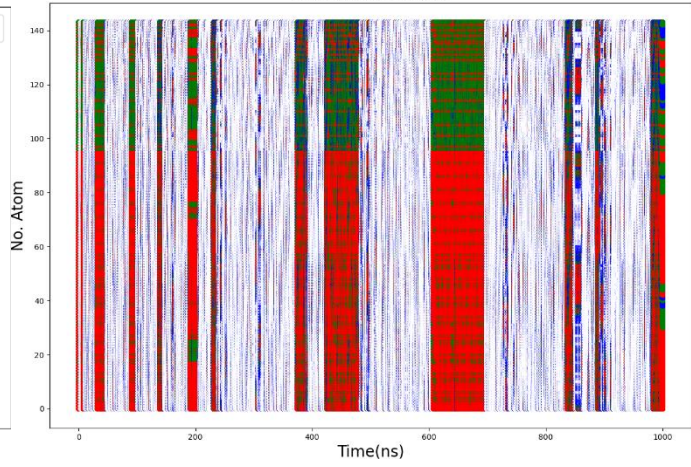
Atom-G5A Contacts



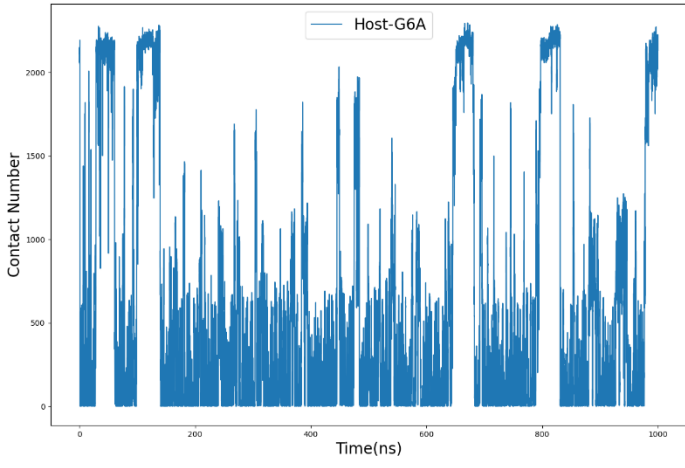
Host-G5B Contacts



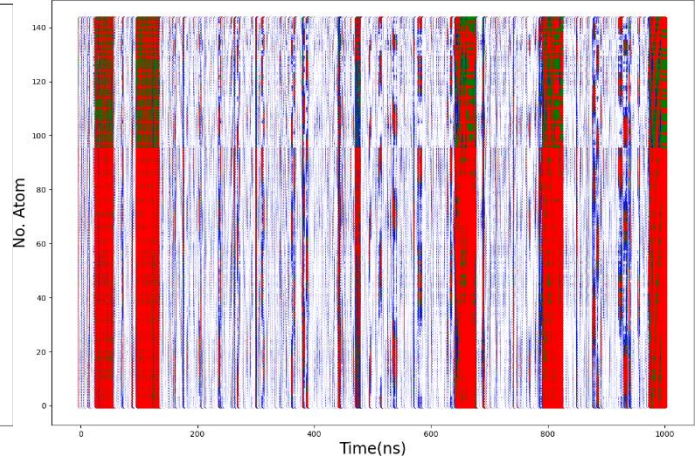
Atom-G5B Contacts



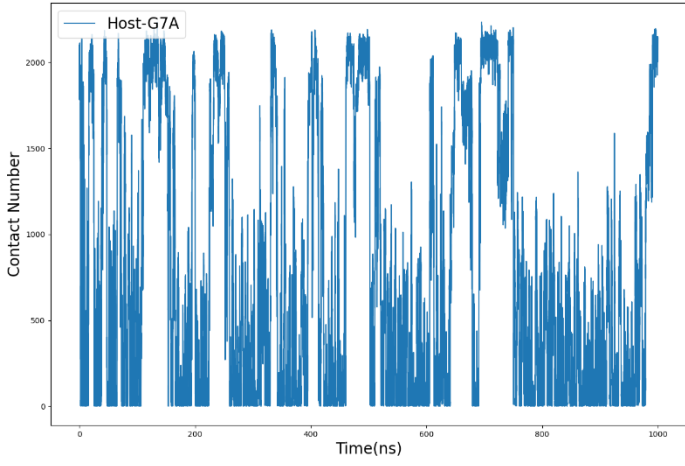
Host-G6A Contacts



Atom-G6A Contacts



Host-G7A Contacts



Atom-G7A Contacts

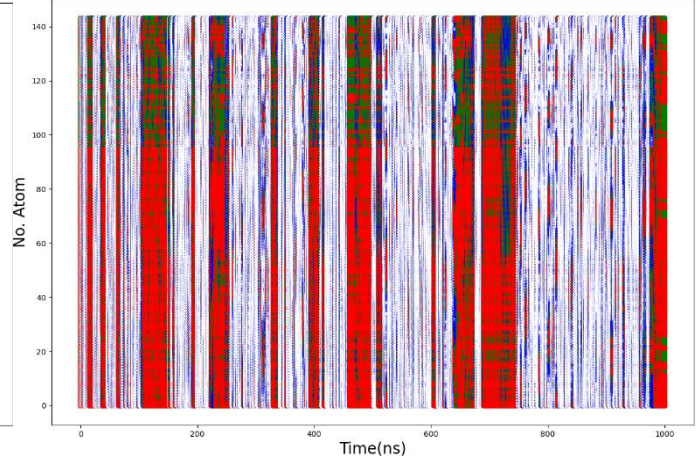
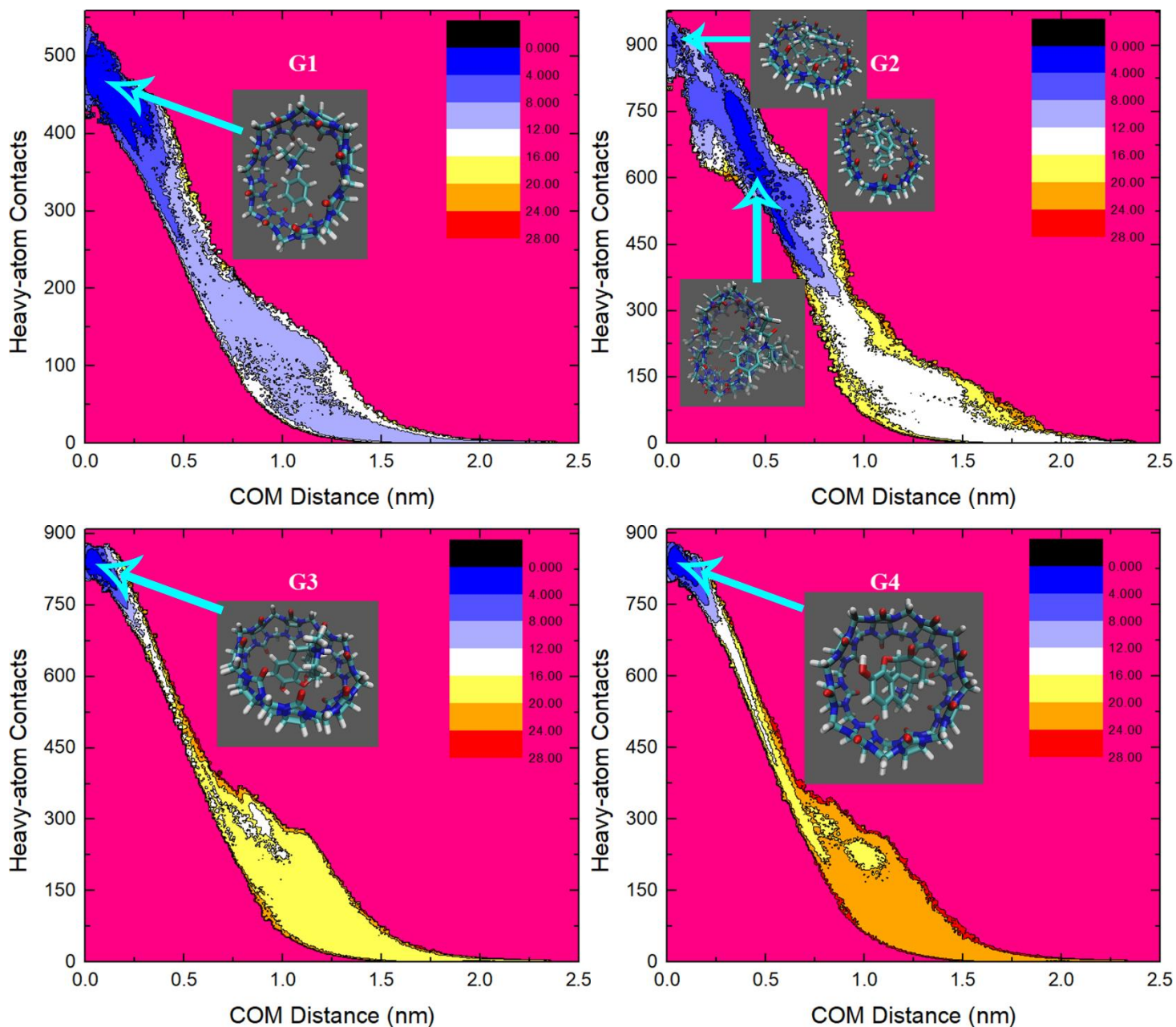


Fig. S5. 2D $\rho-C$ free energy surfaces in kcal/mol under the AM1-BCC charge scheme. Representative structures extracted from the global minimum and the other low-energy local minima are also shown.



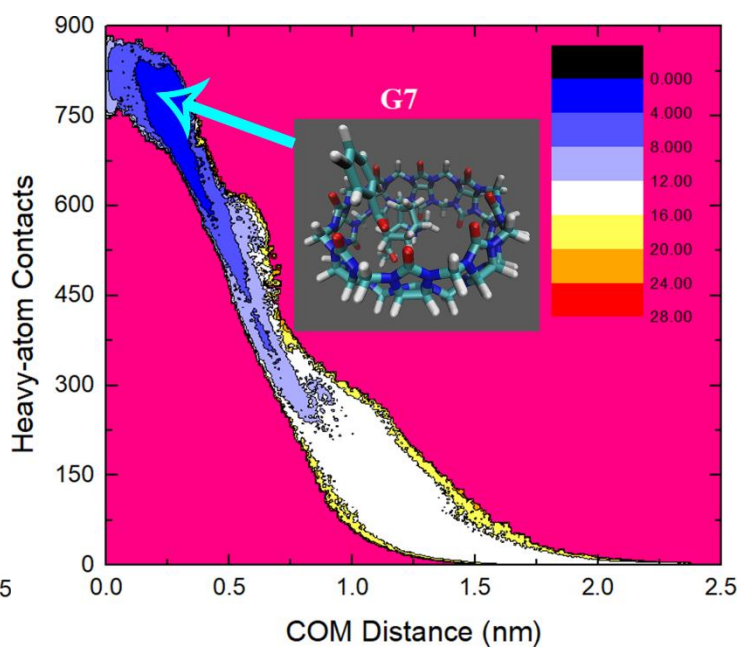
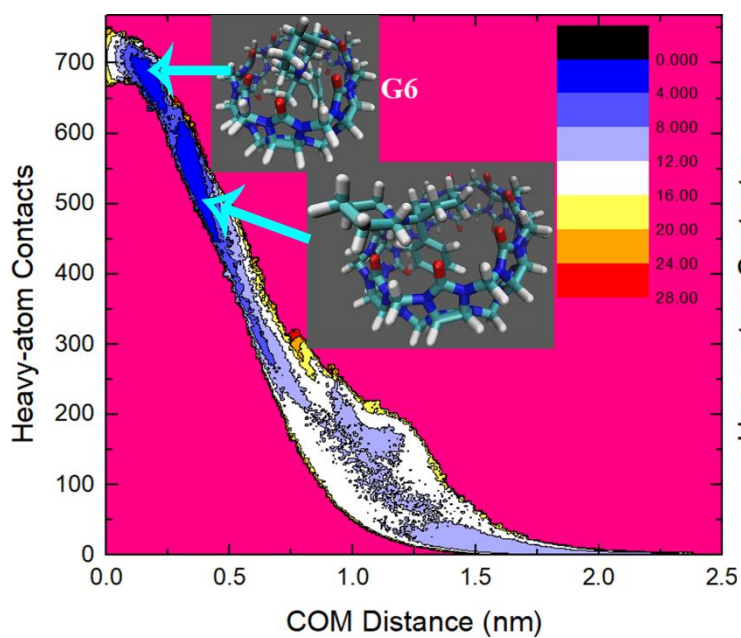
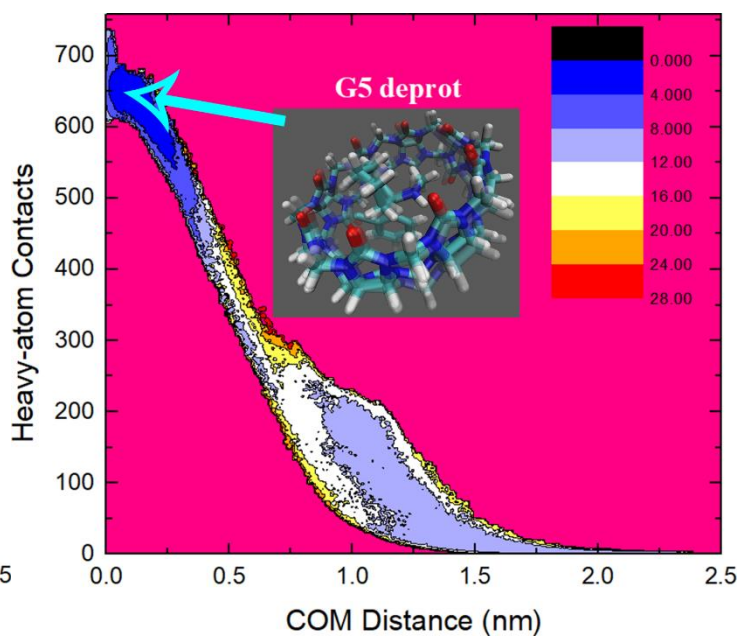
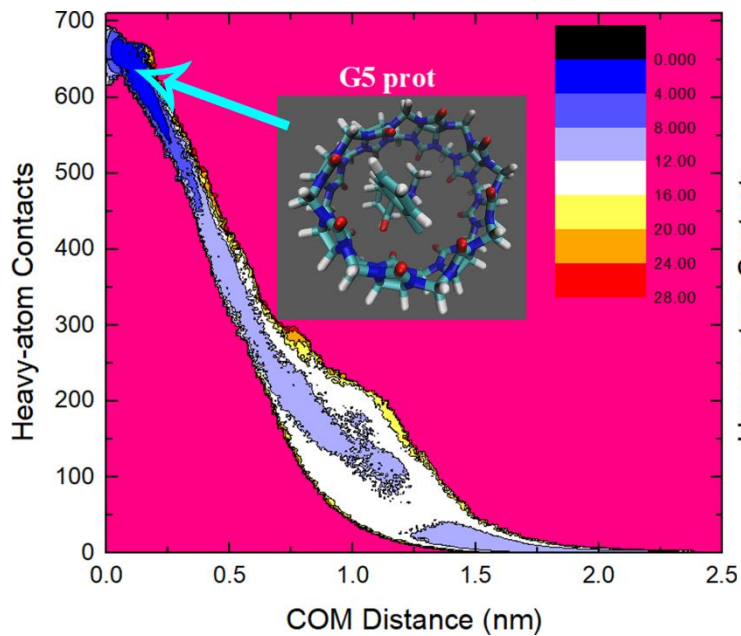
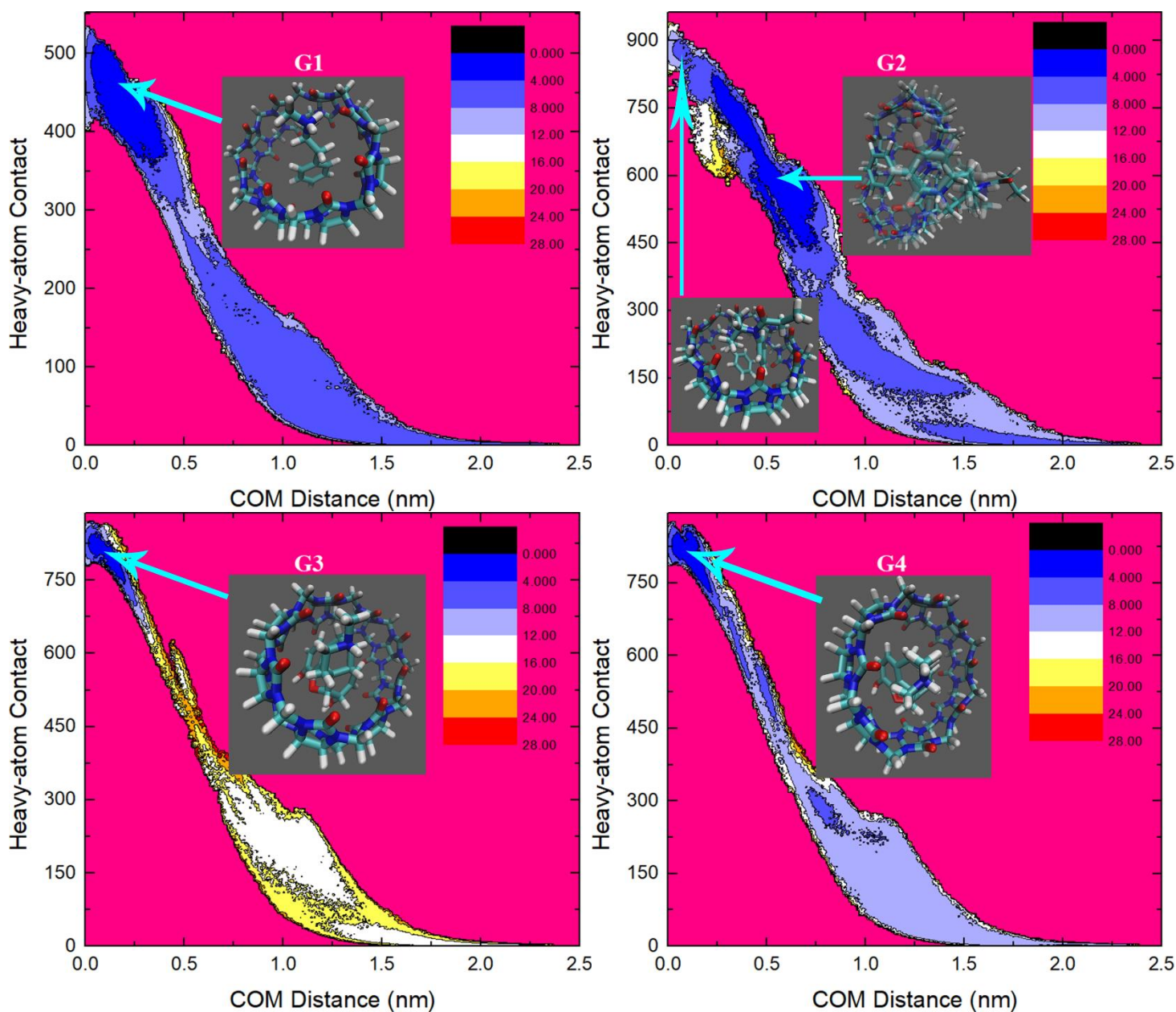


Fig. S6. 2D $\rho-C$ free energy surfaces in kcal/mol under the RESP charge scheme. Representative structures extracted from the global minimum and the other low-energy local minima are also shown. For all guests except G2, the global minimum features the center-binding mode and the distorted cavity to hold the guest. For G2, the most stable binding pose is similar to the AM1-BCC case, i.e., the extended guest clinging to the outer side of the squashed CB8 ring. For the guest G6, unlike the AM1-BCC results, only one center-binding pose is observed. For deprotonated G5, only one stable interaction pattern/orientation for the inter-molecular $-N-H\cdots O=C-$ hydrogen bond is observed.



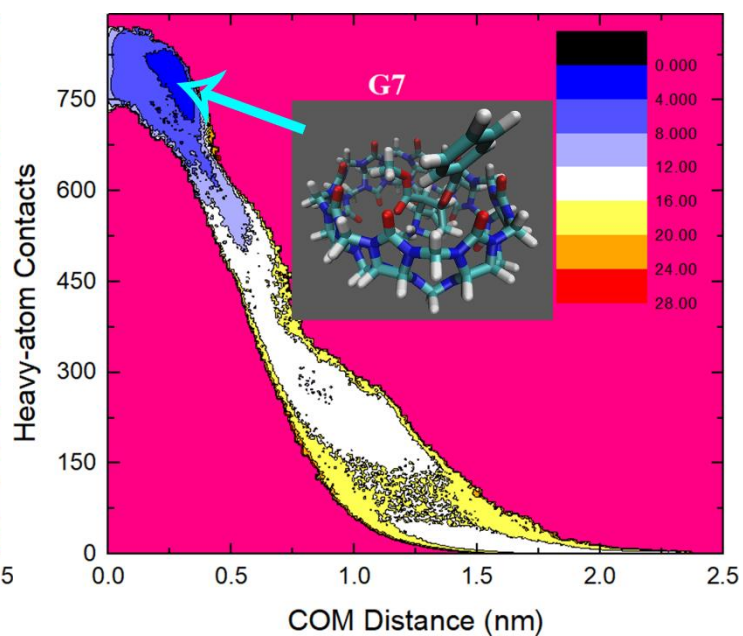
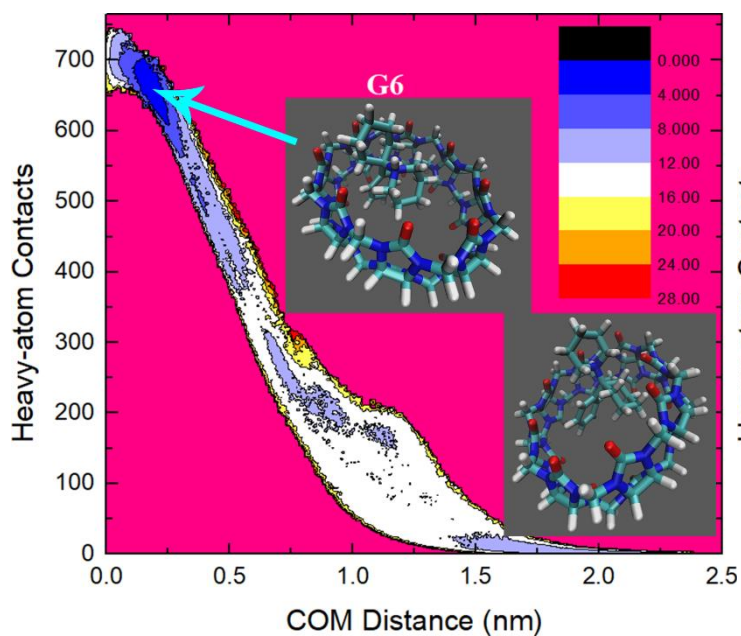
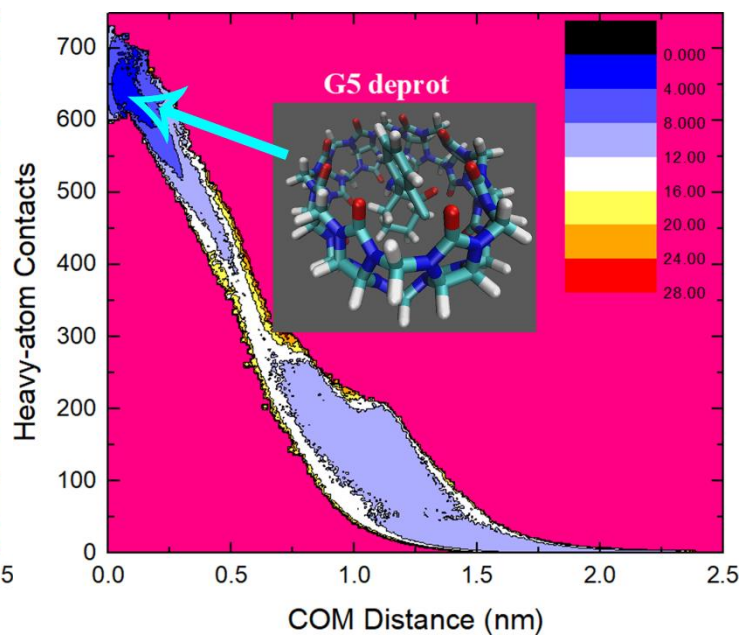
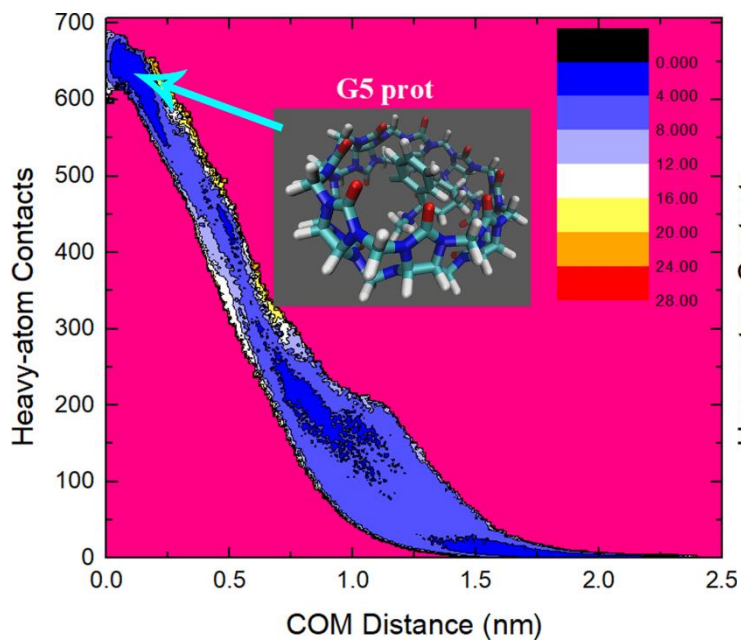
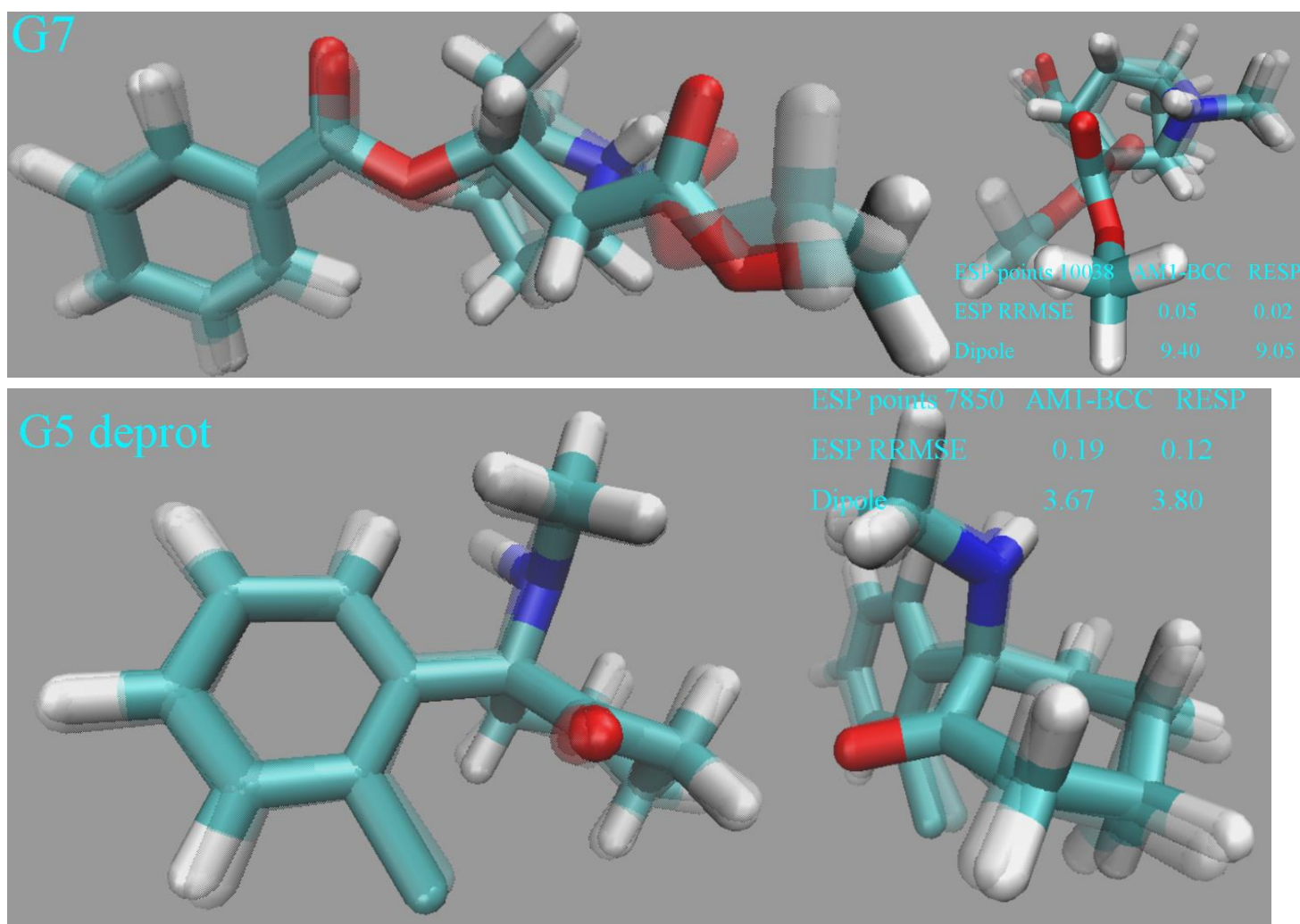
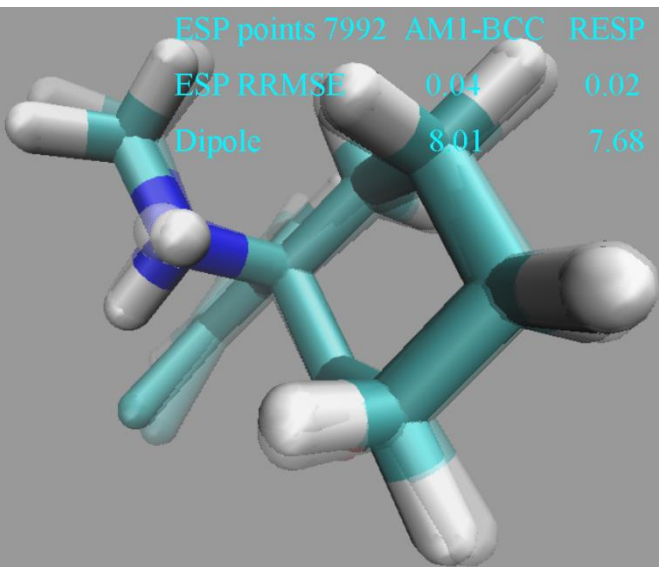
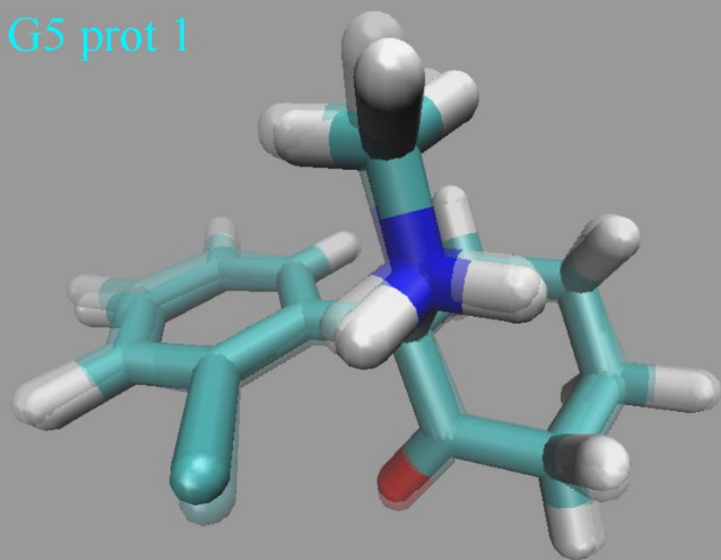


Fig. S7. Comparison between the AM1 and B3LYP/6-31G* optimized structures in the new round of charge fitting for the deprotonated and protonated forms of Ketamine (G5) and Cocaine (G7). The optimized structures under the two Hamiltonians are extremely similar for protonated and deprotonated G5, while differences exist in some regions (e.g., the orientation at -COOMe) for Cocaine. For the deprotonated Ketamine, due to the separation of N and Cl atoms, the reproduction of the ESP around the molecule is improved for both charge schemes. The ESP reproductions for other two molecules seem unchanged. The fitting qualities of protonated G5 and G7 are similar to the previous case. Two models are refitted for the protonated G5. One of them (G5 prot 1) has the intra-molecular hydrogen bond formed by Cl and $-NH_2^+$, while the structure used to fit the other model (G5 prot 2) is obtained by adding one hydrogen atom on the -NH group of the refitted deprotonated G5 structure and does not have intra-molecular hydrogen bond.



G5 prot 1



G5 prot 2

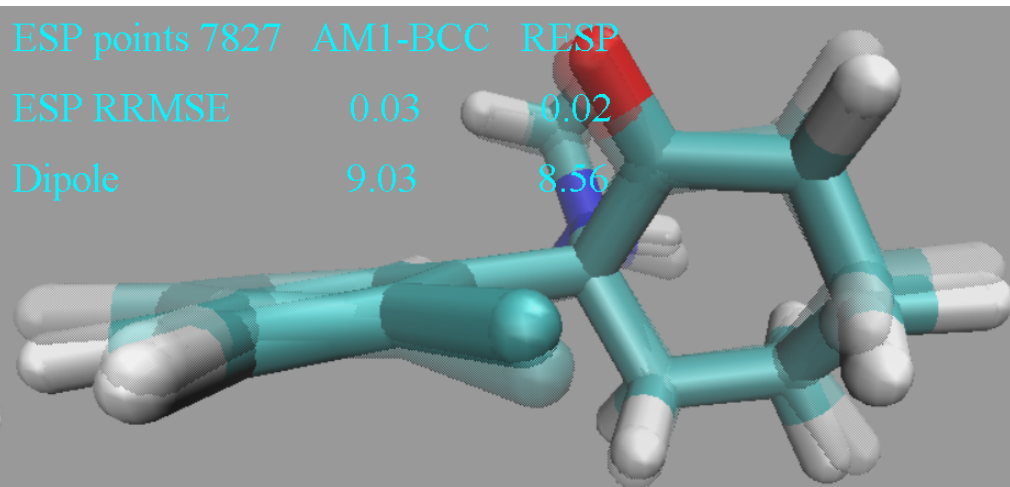
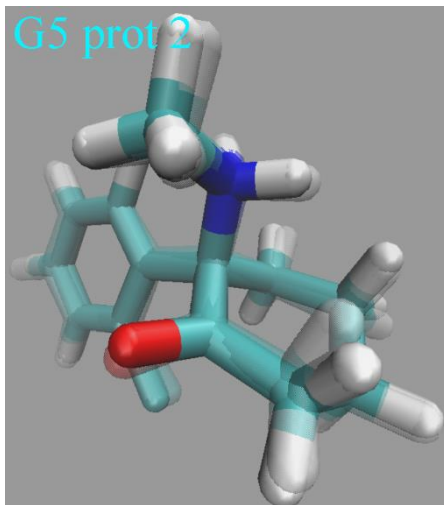


Fig. S8. 2D $\rho-C$ free energy surfaces in kcal/mol obtained with the new charge sets generated with a-d) the AM1-BCC scheme and e-h) the RESP scheme. Two models are refitted for the protonated G5. The configuration used to generate the first model (G5 prot 1) is similar to the that used in the initial charge generation and has the intra-molecular hydrogen bond formed by Cl and $-\text{NH}_2^+$, while the other (G5 prot 2) is obtained by adding one hydrogen atom to the refitted deprotonated G5 and regenerating the charges and thus does not have intra-molecular hydrogen bond.

