

Molecular Environment Specific Atomic Charges Improve Binding Affinity Predictions of SAMPL5 Host-Guest Systems

Duván González, Luis Macaya, and Esteban Vöhringer-Martinez*

*Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de
Concepción, 4070386 Concepción, Chile*

E-mail: evohringer@udec.cl

Phone: +56 41 2204986

Abstract

Host-guest systems are widely used in benchmarks as model systems to improve computational methods for absolute binding free energy predictions. Recent advances in sampling algorithms for alchemical free energy calculations and the increase in computational power have made their binding affinity prediction primarily dependent on the quality of the force field. Here, we propose a new methodology to derive the atomic charges of host-guest systems based on QM/MM calculations and the MBIS partitioning of the polarized electron density. A newly developed interface between the OpenMM and ORCA software package provides D-MBIS charges that represent the guest's average electrostatic interactions in the hosts or the solvent. The simulation workflow also calculates the average energy required to polarize the guest in the bound and unbound state. Alchemical free energy calculations using the GAFF force field parameters with D-MBIS charges improve the binding affinity prediction of six guests bound to two octa acid hosts compared to the AM1-BCC charge set after correction

with the average energetic polarization cost. This correction originates from the difference in potential energy that is required to polarize the guest in the bound and unbound state and contributes significantly to the binding affinity of anionic guests.

Introduction

Molecular dynamics simulations have been increasingly incorporated to predict binding free energies in early stage drug-discovery.¹ The accuracy of these simulations depends on the complete sampling of the bound and unbound configurations of the ligand, a correct description of the molecular system (e.g., ligand’s protonation or tautomeric state and experimental factors as buffer conditions), and the employed force field describing the interactions.² The final objective of these computer simulations is an accurate prediction of ligand-receptor binding affinity at low computational cost. However, the inherent flexibility and complexity of ligand-receptor systems makes the quantification and analysis of prediction errors from a correct description of the molecular system and the interactions difficult. The main problem is the unknown conformations of the ligand-receptor complex and how the ligand conformations may differ to the ones in the aqueous solution.³ This difference in the conformational space, which also involves the solvent degrees of freedom, contributes to the entropic contribution to the binding free energy. To circumvent the sampling problem of the receptor conformational space relatively rigid molecules have been proposed as hosts. These hosts have a few hundred atoms and tend to be more rigid and less complex than proteins, facilitating sampling all conformations. Host-guest systems, therefore, serve as model for absolute binding affinity prediction because they share molecular recognition events governed by the same basic forces as protein-ligand systems and allow a more complete sampling of the configurational space.⁴

Host-guest systems have been part of the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenges serving as an opportunity for the computer simulations community to test and improve computational methods, algorithms, and force fields.⁴⁻⁶

Supramolecular host families such as the cucurbiturils, gibb deep cavity cavitands, and cyclodextrins have diverse binding affinities and the ability to bind small drug-like compounds with protein-ligands affinities.⁷ The most representative hosts of cavitands are octa acid (OA) and tetra-endomethyl octa acid (TEMOA), a derivative with four added methyl groups at equivalent locations in the entry way.² These hosts possess a deep, basket-shaped, hydrophobic binding site with eight carboxylic acid groups responsible for their increased water solubility²(see Figure 1). OA was introduced as a host in challenge SAMPL4^{6,8} and has been considered in subsequent challenges, while TEMOA has been considered in challenges SAMPL5-6. The guests G1-G6 (see Figure 1) were proposed for OA and TEMOA in SAMPL5 and constitute a reference for the family of cavitands proposed by Mobley and Gilson.² The resulting host-guest system present chemical diversity, water solubility and exhibit a significant range of binding free energies from -4.25 to -9.37 kcal/mol and -3.90 to -5.94 kcal/mol for OA and TEMOA respectively. The host-guest combinations in the SAMPL5 benchmark set additionally exhibit stability at extreme pH values (pH=11.5) which allows for strict control of protonation states and eliminates one source of uncertainty present in the more challenging modeling of protein-ligand systems.⁹ These host-guest systems represent a standard benchmark set used by the computer simulation community to assess varying methodologies.¹⁰

The reduced flexibility and small size of host-guest systems has also stimulated new efficient algorithms and simulation protocols reducing the error from sampling and the differences between varying software packages.^{2,4,6,7,11-19} Yet, the computationally-efficient correct description of the molecular interactions by force fields presents an open challenge for binding host-guest predictions.¹¹ In the recently SAMPL7 challenge, Amezcua *et al.* raised concerns about the correct description of the polarization of the charged guest and the negatively charged OA hosts.²⁰ Polarization is a phenomenon where atoms and molecules induce changes in the electron distributions of the other molecules as they interact with each other²¹ and is predominant when charged groups are present. Having charged groups participating

in the binding process induces also a strong dependence on the salt concentration because ions reduce charge-charge interactions in condensed phases. This effect has only partly been addressed in computer simulations. The ion concentration and parameters affect the water dynamics and may lead to incorrect ion behavior (i.e. excessive ion pairing strength) due to unbalanced force field parameter combinations.^{22,23} Polarizable force fields have intensively been developed to address these issues in recent years, but the parameter derivation for new guests is often difficult and they lack compatibility with established non-polarizable protein force field used in drug discovery projects.

As an alternative to polarizable force fields, Cerutti and coauthors included the effect of the polarization by the aqueous solvent in the ff14ipq and ff15ipq^{24,25} protein force fields with the implicitly polarized charged method (IPolQ). This method draws on approximations of dipole interactions in external electrostatic fields and arrives at the solute’s atomic charges, which are halfway between the charges that would produce the electrostatic potential of the solute in the gas phase and those that would reproduce the solute’s electrostatic potential after solvent-induced polarization. Kelly and Smith^{26,27} proposed an alternative method to account for polarization without relying on the dipole approximation of the solute’s charge density and applied it to the calculation of hydration free energies of small organic molecules. Partitioning the solvent polarized electron density with the MBIS method²⁸ they derive atomic charges for different values of the coupling free energy parameter λ . This approach followed our recently proposed method using an implicit representation of the electrostatic field of the solvent together with the MBIS partitioning method²⁹ but neglects the potential energy required to polarize the solute’s electron density in the transfer from the gas to the condensed phase.

Here, we present a method which accounts for the potential energy required to polarize the ligand using the QM/MM methodology improving the description of electrostatic interactions. It provides molecular environment-specific atomic charges for the non-polarizable General Amber Force Fields, which are compatible with established protein force fields. Our

method consists of a computational simulation protocol that derives environment-specific atomic charges for the guest or the host using QM/MM calculations and the MBIS²⁸ partitioning method of the electron density (see Figure 1). These new D-MBIS atomic charges describe the average interaction of the polarized guest with its surrounding in the bound or unbound state. Additionally, we also include the average potential energy required to polarize the guest in the alchemical free energy calculations of annihilating the guest interactions in the host or in the aqueous solution. The self-adjusted mixture sampling method in YANK together with our validated simulation parameters result in a computational protocol that provides absolute binding affinities in 2-4 days on a standard graphical processing unit. The resulting absolute binding affinities account for the guest’s polarization in a specific molecular environment and possible variations of this polarization between the host and the solvent as well as its energetic contributions to the binding affinity. The new protocol improves over our recent methodology applied to hydration-free energies²⁹ and the work by Kelly and Smith^{26,27} and represents an alternative to polarizable force fields as AMOEBA,³⁰ which had considerable success in the recent SAMPL7 challenge.³¹

Methods

Systems preparation

The initial mol2 files were downloaded from the GitHub repository MobleyLab/benchmarksets (<https://github.com/MobleyLab/benchmarksets>). Protonation states of host and guest molecules were suggested by the SAMPL5 blind challenge organizers according to the expected pKas and the experimental pH values.⁹ To create the host-guest complex input files, the utilities parmed and tleap were used. The combined host-guest complex mol2 and guest mol2 file was loaded in tleap along with the provided host force field parameters (RESP atomic charges)³² for the complex phase including the GAFF1.8^{33,34} and AM1-BCC parameters^{35,36} for the guest as generated by antechamber in the AMBER18 package.³⁷ The

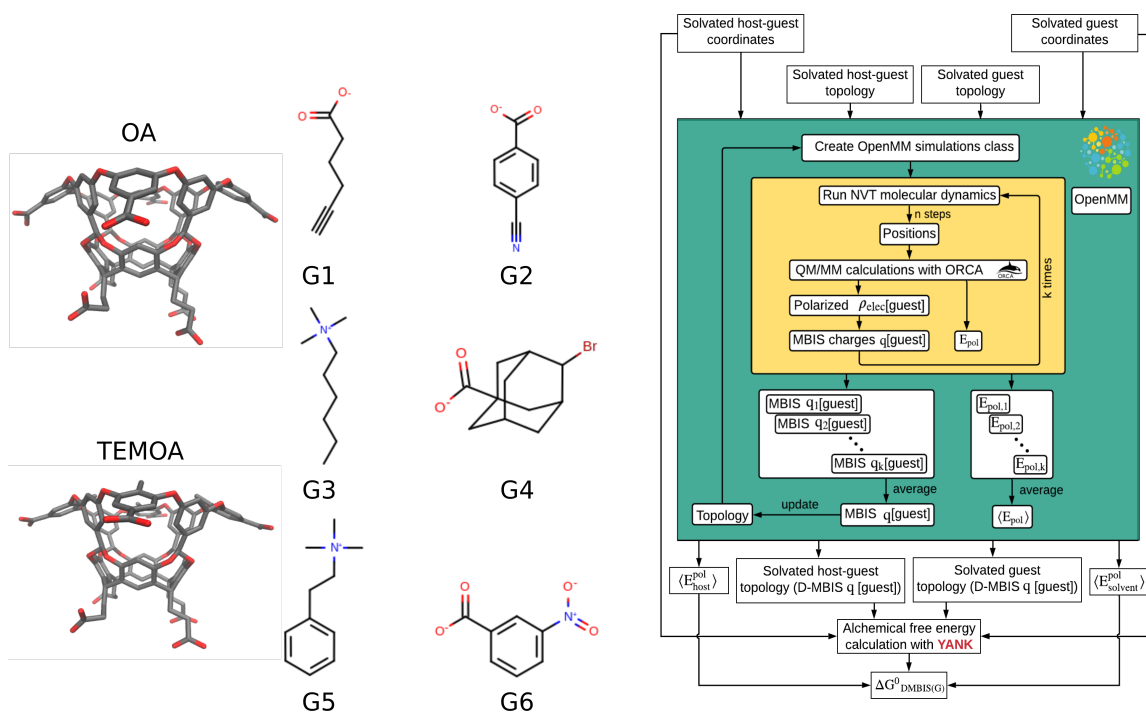


Figure 1: Left: Structures of host OA, TEMOA and their shared guests G1-G6. Protonation states of host and guest molecules were derived according to their pKas and the experimental pH values. Right: Computational flowchart used to calculate the D-MBIS atomic charges for the guest from different configurations together with the average energetic polarization cost of the guest in the bound state and aqueous solution. The average polarization energies correct the alchemical annihilation free energies of the guest in the two molecular environments obtained with the Yank software and provides the binding free energy.

systems were solvated in a cubic box with TIP3P water molecules, with a minimum distance between the solute and the box of 12 Å. A mass of 3 amu were used for the hydrogen atoms. The system’s net charge was neutralized with Na⁺ and Cl⁻ ions using Joung & Cheatham parameters.³⁸ To mimic experimental ionic strength produced by a 10 mM sodium phosphate buffer additional Na⁺ and Cl⁻ ions were added to the OA and TEMOA host-guest systems (NaCl concentration 60 mM). The resulting structures were minimized in 1000 steps using the L-BFGS algorithm and equilibrated running 1 ns of Langevin splitting dynamics (BAOAB splitting, 1 fs time step) at 298 K (except for the TEMOA-G3 system where 278 K was used in order to simulate the experimental conditions) with a Monte Carlo barostat set at 1 atm using OpenMM³⁹ and OpenMMTools.⁴⁰ Particle Mesh Ewald (PME) was used for long-range-electrostatic interactions with a cutoff of 12 Å. Lennard-Jones interactions used the same 12 Å cutoff radius combined with a switching function starting at the distance of 11 Å.

To prepare the simulations using the RESP atomic charges for the host and D-MBIS atomic charges for the guest, the guest’s AM1-BCC atomic charges were replaced in the topology files. The same protocol was followed for simulations using D-MBIS atomic charges for the OA host and the guests.

D-MBIS atomic charges

We derived atomic charges from the guest’s electron density polarized by the instantaneous configuration of the surrounding atoms in the bound state to the host and in the aqueous solution. For the atomic charges of the OA host, the host was empty and surrounded by water molecules and ions required to neutralize its net charge and emulate ionic strength. Electron densities were obtained with the QM/MM method at the B3LYP⁴¹ level of theory with the def2-TZVP basis set^{42,43} using ORCA 4.1.0.⁴⁴

We used the Minimal Basis Iterative Stockholder (MBIS)²⁸ method to calculate atomic charges from the polarized electron densities of the QM/MM calculations. As described in

the results section below, we programmed a simulation protocol in python(see Figure 1), which combines molecular dynamics simulation with OpenMM with QM/MM calculations for several configurations of the trajectory. The calculations are set up with a modified version of the QMhub interface⁴⁵ and run with ORCA 4.1.0.⁴⁴ The Horton2 software package⁴⁶ derives MBIS atomic charges from the polarized electron density from the QM/MM calculation of each configuration. The average MBIS atomic charges from all configurations are replaced in the topology, initiating a new cycle of MD simulations and QM/MM calculation until convergence. The simulation protocol uses up to 25 representative configurations to perform QM/MM calculations selecting the atoms of the molecular environment in a radius of 12 Å (a switching function starting at 10 Å scales the atomic charges to zero up to the cutoff radius) before each update of the topology. The OpenMM package runs the molecular dynamics simulations at 298 K and 1 bar with the following parameters: Langevin integrator with 2 fs time step, collision frequency of 1 ps, particle mesh Ewald electrostatics with 1.0 nm cutoff, constraints on bonds involving hydrogen atoms, restraint harmonic with spring constant of 0.2 kcal/mol · Å² in the bound state to ensure that the guest remains in the binding site⁴⁷ and Monte Carlo barostat. The total simulation time was five ns, after which we update the guest’s charges in the topology. Since this change in the topology alters the guest’s interaction with the molecular environment, the simulation protocol starts new MD simulations and calculates a new set of MBIS charges. After five topology updates, the difference between atomic charges from the previous update was below 0.02e.

We employ the same protocol for the D-MBIS atomic charges of the OA host, starting with a solvated OA host at the experimental ionic strength. We used the same five ns sampling time and reduced the QM/MM calculations to five per update due to the host’s number of electrons and performed only three charge updates in the host topology.

Chemically equivalent atoms, which result from molecular symmetry or free rotations, e.g., hydrogen atoms in methyl group, were identified in each solute with the OpenEye Python Toolkit (version 2017.2.1), averaging their converged D-MBIS charges. Because we

can assume that D-MBIS charges are not geometry-dependent, we treat them as constants during the MD simulations in this work. In other words, we do not include charge polarizability during the MD simulations.

Free Energy Calculations

The free energy calculations and analysis were performed with YANK 0.24.1, varying the atomic charges for guests between AM1-BCC and D-MBIS, and the OA host between resp and D-MBIS. The protocol followed the double decoupling methodology for guests in complex and solvent phases. Since the evaluated guest had a total net charge, a counter-ion of the opposite charge was selected to maintain a neutral simulation system in the alchemical transformation. Due to the observed free energy dependence on the alchemical ion position, we introduced a variation in the YANK package to automatically select the farthest ion to the guest’s initial position as a counter-ion. We applied a simple harmonic constraint on the heavy atoms’ center of mass of the host and the guest using a spring constant of $0.2 \text{ kcal/mol} \cdot \text{\AA}^2$ to keep the guest close to the binding site. The number and spacing of intermediate states λ in the double decoupling method were determined automatically for all systems by the trailblaze algorithm introduced in YANK.⁴⁸ The number of intermediate states for all systems in the bound and unbound state (solvent phase) respectively were OA-G1 (73-71), OA-G2 (72-71), OA-G3 (71-68), OA-G4 (77-75), OA-G5 (70-68), OA-G6 (74-72); TEMOA-G1 (71-71), TEMOA-G2 (67-71), TEMOA-G3 (71-68), TEMOA-G4 (81-78), TEMOA-G5 (69-69) and TEMOA-G6 (69-71).

The self-adjusted mixture sampling (SAMS)⁴⁹ method sampled the whole range of thermodynamic states in one simulation, allowing the rapid computation of the free energy calculations. This method uses one replica to sample multiple intermediate states in an expanded ensemble adapted on the fly to achieve the desired target probabilities for each state. The global-jump state update scheme allows the sampler to jump to any thermodynamic state. We set the parameter gamma0 to 10 to control the initial rate of weight

adaptation and flatness threshold of 1.0 to control the number of (fractional) visits to each thermodynamic state.⁵⁰

To generate samples for each thermodynamic state, YANK uses compositions of Markov Chain Monte Carlo (MCMC)⁵¹ moves in the `openmmtools.mcmc` module. This method mixes metropolised rigid rotation and translation of the guest with one ps of Langevin splitting dynamics (BAOAB splitting⁵² using a four fs timestep, ten ps collision rate, and 2500 steps/iteration). The Monte Carlo Barostat attempts to scale the box size every 25 integration steps to maintain a pressure of 1 atm. In all calculations, we used 40000 iterations of the algorithm for both the complex and solvent phase resulting in converged free energy estimates (see results section) in combination with the automated equilibration detection to determine the optimal time range of the production trajectory used to calculate the expectations and other statistical properties.^{53,54}

The Multistate Bennet Acceptance Ratio (MBAR) implemented in PyMBAR package estimates the alchemical free energy difference and its error from the uncorrelated data with default values for the initial guest. MBAR requires the potential energy of each sampled configuration evaluated at every intermediate λ state, the number of samples drawn from each state, and any standard state corrections stemming from applied restraints to the system. This standard state correction is associated with the free energy of releasing the harmonic restriction on the guest in the bound state employing standard state volume V_0 and one mol/L concentration.^{47,50,53,54}

Polarization correction to the binding free energy

For the adequate description of the guest’s interactions with the molecular environment, we take the polarized guest’s electron density into account when calculating the D-MBIS atomic charges (see above). The polarization of the guest in the host in comparison to the non-interacting guest is associated with a potential energy cost that is always positive, counteracting the negative interaction energy with the molecular environment. Our binding free

energy calculations employ a double decoupling scheme. In principle, if this polarization energy would be the same in the bound state to the host and the solvent, it would cancel out. But, as we show in the results section, the polarization energy cost differs between the two molecular environments. This difference contributes to the binding free energy and is usually not considered in absolute binding free energy calculations with non-polarizable force fields.

The energetic cost associated with the guest’s polarization in the bound state and the solvent phase $E_{\text{host}}^{\text{pol}}$, $E_{\text{solvent}}^{\text{pol}}$ was calculated with an additional electronic structure calculation of each configuration in the simulation protocol described above. In this calculation, we use the standard electronic structure Hamiltonian without including the surrounding atomic charges (\hat{H}_{vac}) and calculate its expectation value using the polarized wave function of the guest from the prior QM/MM calculations in the host (Ψ_{complex}) or in solvent (Ψ_{solvent}). The energetic polarization cost is the difference between this expectation value and the converged self-consistent-field energy obtained in the single point calculation employing the same Hamiltonian:

$$E_{\text{host}}^{\text{pol}} = \langle \Psi_{\text{complex}} | \hat{H}_{vac} | \Psi_{\text{complex}} \rangle - \langle \Psi_{vac} | \hat{H}_{vac} | \Psi_{vac} \rangle \quad (1)$$

$$E_{\text{solvent}}^{\text{pol}} = \langle \Psi_{\text{solvent}} | \hat{H}_{vac} | \Psi_{\text{solvent}} \rangle - \langle \Psi_{vac} | \hat{H}_{vac} | \Psi_{vac} \rangle \quad (2)$$

The average polarization energy $\langle E_{\text{host}}^{\text{pol}} \rangle$ or $\langle E_{\text{solvent}}^{\text{pol}} \rangle$ of each guest in the two molecular environments is obtained averaging the 25 configurations from the new simulation protocol (see Figure 1) and represents the ensemble-averaged value. We added this contribution to the alchemical free energy from the bound state or the solvent phase. Consequently, two contributions related to the polarization energy cost are considered for calculation the binding free energy $\Delta G_{\text{bind}}^{\circ}$.

Results

Host-guest systems serve as benchmarks for predicting the absolute binding free energies of drugs to receptors with computational methods. The two main challenges are the accurate, computationally efficient description of the interactions in the molecular system and the sampling of the entire configurational space. To address these two challenges, we first establish a simulation protocol that derives environment-specific atomic charges for the bound and solvated state of six guests combined with two hosts improving the description of the guest’s electrostatic interactions. After validating a computationally efficient simulation protocol based on the Self Adjusted Mixture Sampling (SAMS) method in YANK, we obtain reproducible and converged absolute free energies with the new atomic charges. Our methodology includes correcting the binding affinity by the average potential energy required to polarize each guest in the unbound and the bound state from QM/MM calculations. The new derived protocol provides absolute binding free energies within 51-98 hours on one standard GPU (RTX2080), depending on the size of the guest.

In the first part, we present the new python interface that allows the derivation of environment-specific atomic D-MBIS charges combining classical molecular dynamics with the OpenMM package and QM/MM calculations with the ORCA software package. In the second section, we validate a computationally efficient simulation protocol to obtain the binding free energy with YANK using the SAMS free energy estimation method. Finally, the last section confers the absolute binding affinities of the six guest to the OA and TEMOA hosts with different combinations of D-MBIS atomic charges and their correction with the average polarization energy in the host and aqueous solution.

New interface between OpenMM and ORCA provides environment specific D-MBIS atomic charges and average polarization energies

In a molecule’s transfer from the vacuum to condensed phases, its electron density is distorted, resulting in its polarization. The energy required to polarize the electron density for a fixed position of the nuclei is the polarization energy and is always positive. This energy usually is not considered in the calculation of binding free energies or hydration free energies with non-polarizable force fields, as already pointed out by Swope *et al.*^{55,56} We demonstrated in our earlier work that accounting for this polarization energy in hydration free energy calculations of the FreeSolv database results in good agreement with the experimental values when combined with MBIS atomic charges.²⁹ However, alchemical absolute binding free energy simulations have mostly neglected a possible contribution of polarization to the binding affinity. These computer simulations calculate the free energy cost of decoupling the guest’s interaction from its environment in two different molecular systems: the bound and the unbound state (surrounded by solvent molecules). Neglecting the contribution of polarization assumes that both states share the same value resulting in its cancellation in the thermodynamic free energy cycle. This approximation may not always be accurate, especially when the guest possesses a large polarizability and the binding site is hydrophobic. Here, we tested if this approximation holds for the host-guest systems in SAMPL5 using six different guests bound to the OA or the TEMOA host.

After the guest becomes polarized either by the solvent or the host, the electrostatic interactions with its molecular environment will differ from its unpolarized state, e.g., a vacuum. To maintain a description compatible with a non-polarizable force field, we assign atomic charges for the guest, which best describes the new distribution of the polarized electron density. We have chosen the MBIS partitioning method of the electron density to calculate atomic charges due to its characteristics (see ref⁵⁷ for a detailed discussion) and our previous results for the FreeSolv hydration free energies.²⁹ Due to the heterogeneous distribution of atoms in host-guest systems, we use the QM/MM methodology to obtain

MBIS atomic charges from the guest’s polarized electron density. These electronic structure calculations include atomic charges from the surrounding molecular environment represented by the force field in the electronic Hamiltonian, resulting in a polarized electron density.

The polarization energy and atomic charges will depend on the atoms’ instantaneous position and present fluctuations due to their dynamics in a canonical ensemble. We are interested in the MBIS atomic charges at the force field level, which best describe the average interactions and the ensemble-averaged polarization energy of the guest. Therefore, we created a new simulation protocol that generates new configurations of the system with classical molecular dynamics simulations at the force field level using the OpenMM software and performs a QM/MM calculation with ORCA 4.1 using the new positions of the system. MBIS atomic charges are then obtained from the guest’s polarized electron density, and the polarization energy calculated as the difference of the energy expectation value of the vacuum Hamiltonian applied to the polarized wave function and the wave function in a vacuum (see Methods section). After extensive sampling of the configuration space, we average the obtained polarization energies and the various MBIS atomic charges. The latter replace the guest’s old charges in the force field, starting a new cycle of molecular dynamics simulations and QM/MM calculations. Finally, we repeat the charge update in the force field until convergence is reached (see Figure 1 with a graphical representation of the simulation workflow).

To obtain representative ensemble-averaged polarization energies and D-MBIS atomic charges, we validated the parameters of the simulation workflow. The polarization energy turned out to be most sensitive in the workflow’s parameter validation. As a test system, we selected the guest G1 in the TEMOA host and addressed first the sampling time and then the total number of charge updates. Increasing the sampling time from 5 to 25 ns with the initial AM1-BCC atomic charges, we observed fluctuations in the polarization energy between different trajectories for longer sampling times (see Figure 2). To reduce these variations and maintain a reduced computational cost, we decided to set the sampling time

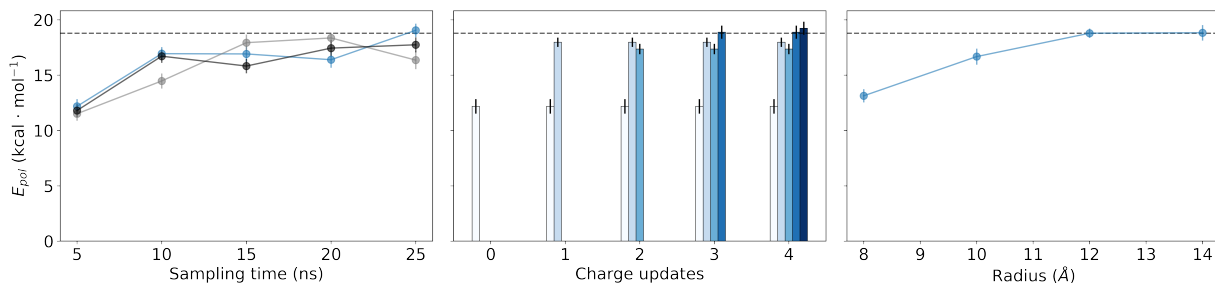


Figure 2: Validation of the simulation workflow’s parameter shown in Figure 1. The average polarization energy of guest G1 in host TEMOA is shown for three independent trajectories as function of the total simulation time using AM1-BCC atomic charges (no topology update) (left), for varying number of charge updates of the guest’s topology with 5 ns sampling for each update (middle), and using five updates and 25 ns of sampling and varying the cutoff radius defining the atomic charges for the QM/MM calculations surrounding the guest (right). For all three parameters convergence of the polarization energy is achieved at the values used in the simulation workflow.

to 5 ns and analyze the effect of updating the guest’s atomic charges in the topology. After the guest’s first charge update in the topology, the average polarization energy undergoes a considerable change (see Figure 2). In this first update, the guest’s charges also vary significantly, altering the configuration of the molecular environment and the polarization energy. After five updates, however, the average polarization energy converges to a constant value. This is accompanied by marginal changes in each topology update resulting in the converged ensemble-averaged polarization energy and D-MBIS atomic charges. As pointed out by Pan *et al.*,⁴⁵ the polarization energy might also depend on the radius used to select the atomic charges of the molecular environment for the QM/MM calculations. The selection uses the distance to the guest as criteria and combines it with a switching function that scales the charges to zero starting 2 Å before the cutoff radius.⁴⁵ Using five updates and 25 ns of sampling per update, we tested various selection radii from 8 to 14 Å. A radius of 12 Å leads to converged average polarization energy for the guest G1 in the TEMOA host (see Figure 2).

We did not carry out an extensive benchmark of the electronic structure method used in the QM/MM calculations. The triple zeta basis set (def2-TZVP) combined with the B3LYP functional provides a robust and computationally efficient approach to obtain the molecular

electron density from the QM/MM calculations. Although local functionals (GGA) are computationally less expensive, they suffer from the self-interaction error leading to small dipole moments and an inaccurate description of the electron distribution in the molecule.^{58–60} Hybrid functionals describe the distribution better, and a triple zeta basis set represents an excellent computational compromise between accuracy and computational cost.

With five topology charge updates using 25 configurations for the QM/MM calculations taken from five nanoseconds trajectories (12 Å distance cutoff), we derived environment-specific D-MBIS atomic charges together with the average polarization energy for each guest in the OA/TEMOA host and in aqueous solution using our validated simulation workflow (see Figure 1). These D-MBIS atomic charges, together with the other force field parameters and the initial structures, serve as input for the alchemical free energy calculation in the bound and unbound state in YANK (see below). After correcting the alchemical free energies by the two average polarization energies in each molecular environment, we obtain the binding free energy $\Delta G_{\text{DMBIS(G)}}^{\circ}$.

Host-guest binding affinities at low computational cost with YANK

To assess a force field’s accuracy, binding free energies derived from molecular simulation should be initial configuration and simulation time independent and account for most experimental conditions. Simultaneously, a compromise for computational efficiency has to be made by selecting a method and parameters that provide results in a reasonable time frame compared with the time for the experiment. As a model system to establish such a method and optimal parameters in Yank, we chose the TEMOA host with the guest G1. We tested two ways to obtain the binding free energies: Hamiltonian Replica Exchange (HREX) and Self Adjusted Mixture Sampling (SAMS). We validated the simulation parameters, number of iterations, number of thermodynamic states, and the initial conformation of the guest to obtain initial configuration and simulation time independent results for each of the methods. Finally, we also addressed the ionic strength used in the experiment and the water model

dependence on the calculated binding free energy.

All simulations were performed initially with YANK 0.23.7 using 1000 steps of minimization, hydrogen mass of 3.0 amu, a time step of 4 fs, and a harmonic restraint to keep the guest in the host’s cavity and automated detection of equilibration before the calculation of the free energy.⁵³ The initial configurations and force field parameters from the benchmark sets include the GAFF parameters combined with RESP atomic charges for the host and AM1-BCC for the guest and the TIP4P-Ew water model.

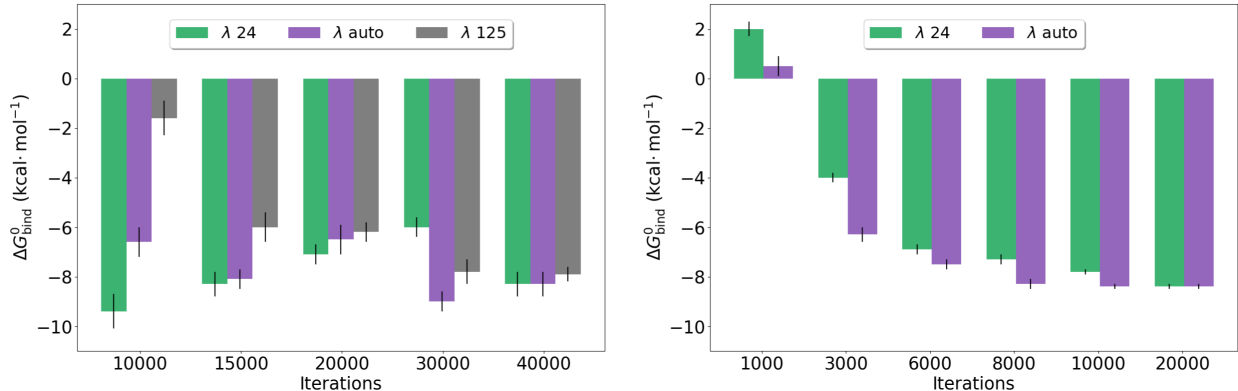


Figure 3: Binding affinities for the TEMOA-G1 system obtained with different number of iterations for sampling the configurational space using Self Adjusted Mixture Sampling (SAMS, left) and Hamiltonian Replica Exchange method (HREX, right) in Yank. Different number of intermediate λ states are compared to the automatically determined ones by the trailblaze algorithm (60 and 56 intermediate states for the complex and solvent phase respectively)

First, we seek to find the sampling method, combined with an optimal number of alchemical intermediate states (λ values), which provides converged host-guest binding free energies in the shortest time with limited computational resources. We based our analysis on the computational time it takes to obtain the results on one GPU (RTX2080). Optimizing the number of intermediate λ states for optimal overlap (thermodynamic trailblazing), we first varied the number of iterations. For the SAMS free energy method, 40000 iterations presented converged results. In contrast, in the HREX sampling method, binding free energies become independent of the number of iterations at a value of 8000 (see Figure 3). If we reduce the number of intermediate λ states ($\lambda = 24$) we observe large fluctuations of the calculated binding free energies in the SAMS method. In contrast, in HREX, the affinity

converges to the same value but at a more significant number of iterations. Although the HREX method seems more robust, its computational time is approximately more than ten times larger than the SAMS method, considering 8000 and 40000 iterations together with the optimal number of intermediate λ states (SAMS 41 hours; HREX 520 hours).

We chose SAMS with 125 λ values for further validations because the computational cost of SAMS lies in the range of 1-2 days on one GPU. We optimized the number of thermodynamic states for each system with the "thermodynamic trailblazing" method to further reduce the computational cost for the production runs. The binding free energy obtained with this method was -8.3 ± 0.5 kcal/mol in perfect agreement with the converged value of the validation study of -8.0 ± 0.3 kcal/mol.

During the validation of the simulation parameters, we observed in some simulations that the obtained binding free energies differ significantly depending on which ion was selected as the alchemical counterion (Na^+). The interactions of the alchemical counterion are switched off together with the negatively or positively charged guest to maintain a neutral system. Specifically, the distance to the negatively charged guest G1 was the primary determinant leading to significant changes in the calculated values (see Table 1), even when the guest's and solvent's initial positions were kept constant. We swapped the coordinates of the alchemical Na^+ ion with another positively charged ion in the simulation box at different distances to the carbon atom of the negatively charged carboxylic group of the guest, maintaining all initial atom positions (see Table 1). The binding free energies evidence a pronounced dependence on the counterion's distance to the guest, which reaches the reference value of -8.0 ± 0.3 kcal/mol at distances exceeding 25 Å. Additionally, we also considered a system for which new initial positions for the solvent and ions were generated and selected the alchemically transformed ion's coordinate as the farthest one. The match of the obtained binding free energy with the reference value confirms that the results are independent of the initial configuration of the ions and the water molecules in the system.

Based on these results, we added a new feature to YANK, which automatically selects

Table 1: Binding affinities calculated for various initial distances of the alchemical Na^+ counter ion to the guest and different initial binding pose of the guest (distance is measured between Na^+ counter ion and the carbon atom of the negatively charged carboxylic group of the guest in Å). Values of binding free energies are in kcal/mol

Simulation	$d(\text{Na}^+ - \text{guest})$	ΔG_{bind}^0
Same configuration different counter ion	11.3	-3.5 ± 0.5
	22.7	-5.3 ± 0.5
	27.9	-8.0 ± 0.4
Different initial positions of ions	34.4	-7.9 ± 0.4
Different binding pose and ions positions	34.0	-8.2 ± 0.5

the ion with the longest distance as the counter ion and makes the binding free energies independent of the counter ion chosen and the initial configuration of solvent and ions. This version can be downloaded from https://github.com/QCMM/yank_autoselect_ions and will be added in the future to the YANK software package.

Release YANK 0.24.1 introduced a better treatment of the long-range interactions (PME),⁶¹ which motivated us to recalculate the binding affinity with the automatic selection of counterion and the same simulation parameters. The obtained value of -8.3 ± 0.3 kcal/mol resembles our previous result within the error margins suggesting that the effect of long-range electrostatic interactions is minor. Interestingly, simulations with the new version of YANK, which presented a distance of Na^+ counter ion to the guest of 11.3 Å result in binding free energy of -6.4 ± 0.3 kcal/mol closer to the reference value. Thus, improving long-range electrostatic interaction in version 0.24.1 reduces the effect of the alchemical ion’s distance to the guest on the binding affinity. However, it still deviates significantly from the converged value of -8.3 kcal/mol, which enforces our feature in Yank, selecting the farthest counterion from the guest as the alchemical one. Yet, we switched to the new 0.24.1 version for further validation and production runs due to better treatment of long-range electrostatic interactions.

Binding free energies should also be independent of the initial position of the guest in the host. We addressed the effect of the guest’s initial position in the host generating an

alternative conformation from their SMILES string representation through the OMEGA⁶² library in the OpenEye Toolkit (version 2018.10.1) (see Figure 4). The differences in binding affinity were only 0.05 kcal/mol, which confirms that binding free energies obtained with Yank for this system do not depend on the initial guest’s conformation. The main reason might be related to the algorithm in YANK, which attempts rotation and translations of the guest in the host in each iteration. The Monte Carlo rotational and translational moves sample more relative orientations in the binding pocket than standard molecular dynamics simulations and reduce the dependence on the initial configuration.

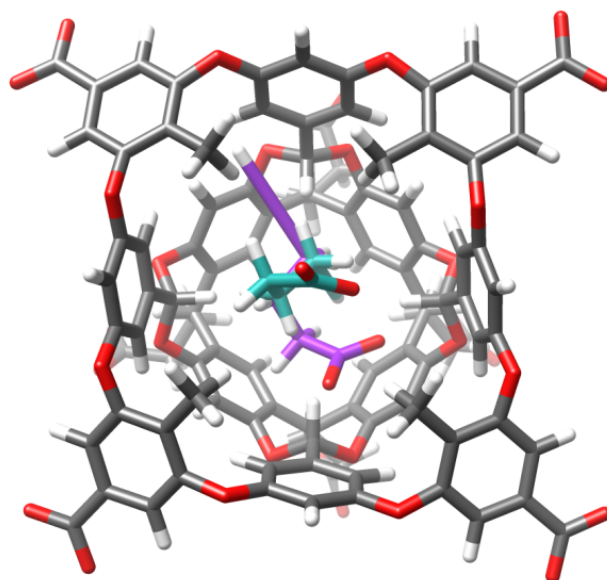


Figure 4: Three dimensional structure of TEMOA and G1 in different initial conformations (carbon atom color cyan and purple respectively) used in the binding free energies reported in Table 1

We also evaluated changing the atomic mass of the hydrogen atoms to 1 amu with a 2 fs time step maintaining all other parameters. The binding affinity was -7.9 ± 0.7 kcal/mol without a statistically significant difference to the one reported previously. However, the computational time duplicates so that we maintained the larger time step for the production runs.

Based on the analysis discussed above, we chose the SAMS free energy method with 40000 iterations and automatic detection of intermediate λ states and the feature of selecting the

farthest alchemical counter-ion from the guest together with a 4 fs time step and the hydrogen atomic mass of 3 amu for the production runs. This combination of parameters and method provides a host-guest binding free energy approximately in 41 hours on a RTX2080 GPU.

Finally, we also studied the effect of including ions in the system to mimic the experimental ionic strength and the change of the water model (TIP3P and TIP4P-Ew) on the calculated host-guest binding free energies. We added three and two extra sodium and chloride ions in the simulations with the host and the aqueous solution, respectively, to mimic the 10 mM experimental buffer concentration. One should notice that due to the net charge of the TEMOA host, the number of Na^+ ions in the simulation box is eight to maintain a neutral system. The calculated binding free energies were -3.7 ± 0.4 and -3.5 ± 0.3 kcal/mol using the TIP3P water model with and without buffer, respectively. Interestingly, the results with TIP3P water model show only a slight effect on the buffer. For the TIP4P-Ew water model, the buffer’s impact is more significant than the TIP3P water model, resulting in -4.0 ± 0.4 and -8.3 ± 0.3 kcal/mol with and without buffer, respectively. A possible explanation for the difference between the two water models is the dielectric constant, which might be determinant at this high ion concentration.¹⁷

Prior studies of TEMOA-G1 in the SAMPL5 challenge using similar methods and the TIP3P model report identical values with and without buffer (-4.2 ± 0.4 and -4.2 ± 0.0 kcal/mol).⁶³ The same authors report that accounting for the experimental ionic strength in the simulation with similar methods and the same force field for all TEMOA guests reduced the RMSE from 3.5 to 1.9 kcal/mol. Pal *et al.* report binding free energy change from -7.5 ± 0.1 to -4.8 ± 0.1 kcal/mol in TEMOA-G1 after accounting for the ionic strength using the OPLSA2005 force field for the host and the guest combined with the AGBNP2 water model and an RMSE reduction from 6.7 to 2.1 kcal/mol.⁶⁴

Based on our results and previous observation for the whole guests in the OA and TEMOA hosts, we decided to use the TIP3P water model due to its compatibility with the GAFF force field and account for the ionic strength resembling the experimental condition.⁶⁵

In summary, the SAMS free energy method with the simulation parameters described above provides reproducible, converged binding free energies with YANK in less than two days on a standard GPU independent of initial conformation. Furthermore, we simulated all OA/TEMOA systems at the experimental ionic strength and maintained the TIP3P water for compatibility with the GAFF force field used for the host and the guest.

Polarization Corrected Binding Free Energies with D-MBIS atomic charges

With the validated simulation protocol for the host-guest binding affinities described above, we studied the Gibbs binding free energies variation when AM1-BCC atomic charges in the General Amber Force Field (GAFF) are replaced with the new D-MBIS charge set. After including the polarization correction, we compare the results to the experimental reference values. First, we maintained the RESP charges on the host and change only the guest’s charges ($\Delta G_{\text{bind}}^{\text{D-MBIS(G)}}$), and then we also calculate D-MBIS atomic charges for the OA host followed by a new charge derivation for the guest ($\Delta G_{\text{bind}}^{\text{D-MBIS(H-G)}}$). All calculated binding free energies with D-MBIS atomic charges account for the difference in the guest’s polarization energy between the host and the aqueous solution ranging from 1 to 5 kcal/mol.

Table 2 summarizes the experimental binding free energies of the six guests in the OA and TEMOA hosts and the computational results obtained using either the AM1-BCC or the D-MBIS guest atomic charges combined with the RESP charges for the two hosts. The last two columns display the mean potential energy required to polarize the guest when either bound to the host or in the aqueous solution. The Root Mean Square Errors (RMSE) were 2.6 ± 0.3 and 2.1 ± 0.3 kcal/mol using AM1BCC or D-MBIS atomic charges for the OA host. The other statistical descriptors as the correlation coefficient R or the Kendall τ do not show a significant difference between the two guest’s charge sets. For the TEMOA host, the RMSE drops from 1.9 ± 0.3 kcal/mol obtained with the AM1-BCC atomic charges to 1.4 ± 0.2 kcal/mol when using the D-MBIS charge model. The other mentioned statistical

descriptors also evidence better agreement with the experimental reference values. The Kendall τ value of one in the TEMOA host evidences that D-MBIS atomic charges with polarization correction order the six guests correctly from weak to strong binders. In some drug discovery applications, this correct classification of strong to weak binders is more valuable than the absolute binding free energies.

Table 2: Absolute binding free energy results using RESP atomic charges for the OA and TEMOA host and AM1-BCC or D-MBIS atomic charges for the guest. Values in kcal/mol

Complex	$\Delta G_{\text{exp}}^{\circ}$	$\Delta G_{\text{AM1-BCC}}^{\circ}$	$\Delta G_{\text{DMBIS(G)}}^{\circ}$	$\langle E_{\text{host}}^{\text{pol}} \rangle$	$\langle E_{\text{solvent}}^{\text{pol}} \rangle$
OA					
OA-G1	-5.04 ± 0.01	-7.5 ± 0.2	-3.5 ± 0.5	13.4 ± 0.3	14.0 ± 0.2
OA-G2	-4.25 ± 0.01	-7.8 ± 0.3	-7.5 ± 0.4	18.5 ± 0.4	15.1 ± 0.2
OA-G3	-5.06 ± 0.01	-7.1 ± 0.2	-8.2 ± 0.3	3.4 ± 0.1	2.5 ± 0.1
OA-G4	-9.37 ± 0.00	-12.2 ± 0.3	-10.6 ± 0.8	18.9 ± 0.4	20.9 ± 0.5
OA-G5	-4.50 ± 0.00	-3.4 ± 0.2	-3.5 ± 0.2	4.8 ± 0.1	4.7 ± 0.2
OA-G6	-5.33 ± 0.00	-8.4 ± 0.2	-4.1 ± 0.5	13.6 ± 0.3	18.6 ± 0.3
RMSE		2.6 ± 0.3	2.1 ± 0.3		
MAE		2.5 ± 0.3	1.9 ± 0.3		
R		0.8 ± 0.4	0.7 ± 0.4		
τ		0.4 ± 0.4	0.4 ± 0.4		
TEMOA					
TEMOA-G1	-5.24 ± 0.05	-3.8 ± 0.3	-5.5 ± 0.4	18.9 ± 0.3	13.8 ± 0.1
TEMOA-G2	-5.04 ± 0.03	-7.8 ± 0.2	-5.2 ± 0.5	15.0 ± 0.3	15.8 ± 0.3
TEMOA-G3	-5.94 ± 0.12	-7.9 ± 0.2	-8.7 ± 0.2	3.6 ± 0.1	2.6 ± 0.1
TEMOA-G4	-2.38 ± 0.02	-2.5 ± 0.4	-1.8 ± 0.9	23.2 ± 0.5	19.6 ± 0.4
TEMOA-G5	-3.90 ± 0.02	-4.0 ± 0.2	-2.4 ± 0.3	4.0 ± 0.1	4.5 ± 0.1
TEMOA-G6	-4.52 ± 0.02	-7.5 ± 0.2	-3.3 ± 0.5	11.5 ± 0.3	18.3 ± 0.3
RMSE		1.9 ± 0.3	1.4 ± 0.2		
MAE		1.6 ± 0.2	1.1 ± 0.2		
R		0.7 ± 0.3	0.9 ± 0.1		
τ		0.6 ± 0.3	1.0 ± 0.0		

To rationalize the different outcomes of the D-MBIS atomic charges for the TEMOA and the OA host, we analyzed the binding conformation of the same guest to the two hosts. Binding free energy variations of the same guest bound to two hosts might result only from different interactions with the molecular environment or also contain a contribution from the guest's different orientation in the two hosts.^{66,67} We addressed different possible orientations in binding modes of guests G1 and G4 to the two hosts. These guests present the most

considerable binding free energy difference between the two hosts using the AM1-BCC atomic charges. Both guests in TEMOA align with the symmetry axis of the host. In contrast, in the OA host, they are slightly tilted in the host cavity allowing additional water molecules in the upper opening of the host (see Figure S1-2 in Supporting Information). Methyl groups constricting the upper opening in TEMOA align the guest with its symmetry axis restricting the guest’s conformations compatible with solvent-exposed carboxyl groups of both guests and the access of water molecules in the cavity. We conclude that the guests present different orientations in the two hosts and are exposed to different molecular environments in OA and TEMOA. Interestingly, the average polarization energy of G1 reflects the different molecular environments in both hosts.

Variation of the guest’s conformations in the respective host leads to significant differences in their binding affinity. Thus, a change in the charge sets describing the electrostatic interactions from AM1-BCC to D-MBIS, in principle, could alter the preferred conformation of the guest in the respective host. However, for the guests G1 and G4 discussed above, we observed no significant change in their conformation bound to the OA and TEMOA host replacing the AM1-BCC charge set with the D-MBIS one (see Figure S3 in Supporting Information). Thus, the atomic charges describing the interactions contribute significantly to the binding free energy differences. To analyze this contribution, we analyzed the host-guest systems OA-G1, OA-G6, and TEMOA-G6, which present the most notable difference in binding affinity of 4-5 kcal/mol between the two charge sets. For the G1 guest, the differences arise mainly from varying electrostatic interactions of the guest originating from $\sim 0.2 e$ larger D-MBIS atomic charges in the alkyne carbon atom. In G6, the most significant variations between the two charge sets are $\sim 0.4 e$ for the nitro’s group nitrogen atom and $\sim 0.2 e$ for its oxygen atoms and $0.2 e$ for the aromatic carbon atom bound to the nitro group. In all three cases (OA-G1, OA-G6, and TEMOA-G6), the results obtained with D-MBIS atomic charges, including the polarization energy correction, are closer to the experimental value.

Together with the different absolute values of the atomic charges, one crucial difference between our methodology and the AM1-BCC atomic charges is that D-MBIS atomic charges are environment-specific and may vary between the guest’s bound and the unbound state. To test the effect of using environment-specific atomic charges, we selected guest G2 and G3 in host OA. G2 presents the most prominent D-MBIS atomic charge difference between the bound and unbound state and G3 the smallest (see Table S1 in Supporting Information). For both guests, we repeated the binding free energy calculations using the same atomic charges derived in an aqueous solution for the bound state and the unbound state without correcting for the polarization energy. Guest G2 resulted in binding free energy of -3.6 kcal/mol, which is 3.9 kcal/mol more positive than the value in Table 2 and G3 presented a deviation of 0.4 kcal/mol. Using the atomic charges derived when the guest is bound to the host also leads to more positive binding free energies compared to the ones reported in Table 2 (see Table S2 in Supporting Information). From this analysis, it becomes evident that environment-specific atomic charges and the average polarization energy correction of the alchemical free energies improve the calculated binding free energies.

To analyze how the difference of average polarization energy between host and solvent contributes to the binding free energies, we explored its dependence on the guest’s chemical nature and the molecular environment: aqueous solution, OA, or TEMOA host. Based on the chemical nature of the guest, we expect that the cations bearing the ammonium group (G3, G5) would present smaller polarization energies than the guest with anionic character with an extra charge increasing its polarizability. The average polarization energies displayed in Figure 5 confirm this expectation for the three molecular environments. Only minor variations between the hosts’ values or aqueous solutions exist in the G3 and G5 guests. The differences between the molecular environments are more significant for the anionic systems G1, G2, G4, and G6. The average polarization energy increases with the guest’s size, as is expected due to its larger polarizability. The adamantyl group containing bromine presents the most prominent values, and the alkyne or cyano group show similar values. Interestingly,

the values differ considerably between both hosts and the aqueous solution. The difference between OA and TEMOA host for G1 is mainly related to the carboxyl group’s orientation inside the host discussed above. In OA, G1 is fully solvated in a slightly tilted conformation, resulting in similar average polarization energy in the aqueous solution. The constricted opening of the TEMOA host by the methyl groups impeding water access results in more considerable polarization energy than in OA or the aqueous solution. The same holds for G4, where we have established a similar variation of the guest’s conformation in the two hosts and less solvent-exposed TEMOA cavity. We observe the most considerable difference in polarization energy between hosts and the aqueous solution for G6, possibly originated by the nitro group and the resonance between the carboxyl and the nitro group. It is most prominent in an aqueous solution and smaller in both hosts. We conclude that in anionic guests, the difference in the average polarization energy between the bound state and aqueous solution contributes considerably to the binding affinity and depends on the guest’s chemical nature, governed mainly by the polarizability. For cationic guests, the contribution to the binding affinity is in the 1-2 kcal/mol range.

Finally, we asked if deriving D-MBIS atomic charges for the host in an aqueous solution rather than using the provided RESP charges would improve the results. We chose the OA host where D-MBIS atomic charges for the guest did not improve the results obtained with the original AM1-BCC set. In addition, the ordering of the guest according to their affinity was unsatisfactory. With the new D-MBIS atomic charges for the OA host, we recalculated the guest’s average polarization energies in the host or the aqueous solution. The calculated Gibbs binding free energies, including the polarization energy correction $\Delta G_{\text{DMBIS(H/G)}}^{\circ}$ are summarized in Table 3. Most statistical descriptors reflect better agreement with the experimental reference reducing the RMSE to 1.8 ± 0.2 kcal/mol and increasing the Kendall τ to 0.6 ± 0.4 . This result evidences that D-MBIS atomic charges improve the results considerably over the AM1-BCC/RESP charges and that the host’s charge distribution also alters the binding free energy.

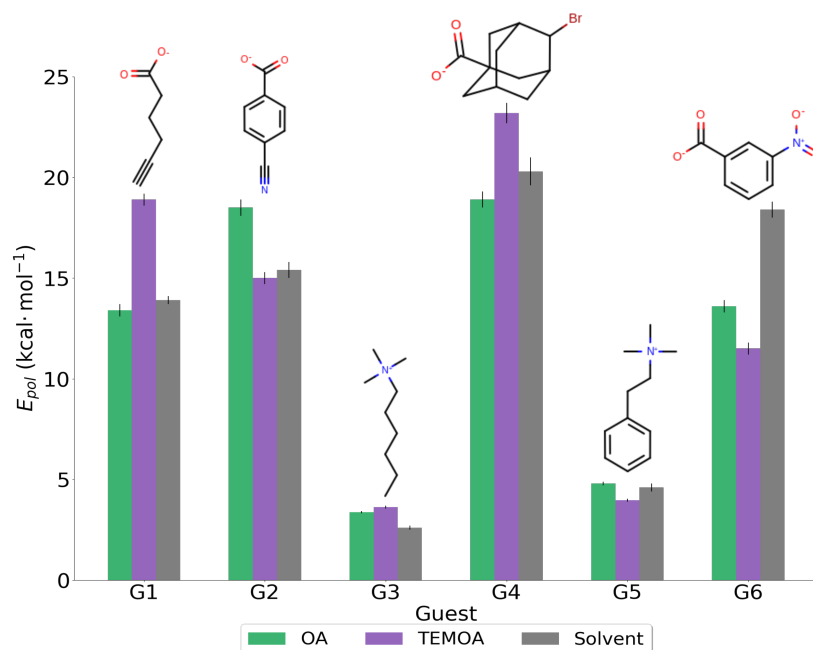


Figure 5: Energy required to polarize the guest in different molecular environments. The green and purple bars consider the OA and TEMOA host using RESP atomic charges for the host and D-MBIS ones for guest. The gray bars represent the unbound state in the aqueous solution (the value corresponds to the average of the average polarization energy in the solvent shown in Table 2).

Table 3: Absolute binding free energies using RESP atomic charges for the OA host and D-MBIS for the guests (D-MBIS(G)) or D-MBIS atomic charges for the OA host and the guests (D-MBIS(H-G)). The values of $\langle E_{\text{host}}^{\text{pol}} \rangle$ and $\langle E_{\text{solvent}}^{\text{pol}} \rangle$ were obtained with the D-MBIS atomic charges for host and guests (D-MBIS(H-G)). Values in kcal/mol

Complex	$\Delta G_{\text{exp}}^{\circ}$	$\Delta G_{\text{DMBIS(G)}}^{\circ}$	$\Delta G_{\text{DMBIS(H/G)}}^{\circ}$	$\langle E_{\text{host}}^{\text{pol}} \rangle$	$\langle E_{\text{solvent}}^{\text{pol}} \rangle$
OA					
OA-G1	-5.04 ± 0.01	-3.5 ± 0.5	-2.5 ± 0.5	12.1 ± 0.2	14.6 ± 0.3
OA-G2	-4.25 ± 0.01	-7.5 ± 0.4	-5.9 ± 0.5	14.3 ± 0.3	15.4 ± 0.3
OA-G3	-5.06 ± 0.01	-8.2 ± 0.3	-4.2 ± 0.3	4.2 ± 0.1	2.6 ± 0.1
OA-G4	-9.37 ± 0.00	-10.5 ± 0.8	-11.0 ± 0.7	18.8 ± 0.5	20.8 ± 0.3
OA-G5	-4.50 ± 0.00	-3.5 ± 0.2	-2.1 ± 0.2	6.1 ± 0.1	4.8 ± 0.1
OA-G6	-5.33 ± 0.00	-4.1 ± 0.5	-6.6 ± 0.5	17.6 ± 0.3	18.6 ± 0.3
RMSE		2.1 ± 0.3	1.8 ± 0.2		
MAE		1.9 ± 0.3	1.7 ± 0.2		
R		0.7 ± 0.4	0.8 ± 0.4		
τ		0.4 ± 0.4	0.6 ± 0.4		

Conclusions and outlook

We have developed a new simulation protocol written in Python that combines QM/MM calculations and molecular dynamics simulations to derive D-MBIS atomic charges that describe the average electrostatic interaction of guests in the host or aqueous solution. Additionally, it also provides the average energetic polarization cost of each guest in its specific molecular environment. The new simulation protocol combined with alchemical free energy calculations using the self-adjusted mixture sampling method in Yank enables absolute binding free energy calculations for host-guest systems in approximately 50-90 hours on a standard GPU. Our results confirm that D-MBIS atomic charges and inclusion of the energetic polarization cost improve the prediction of binding affinities of the six guests bound to the TEMOA host compared to the AM1-BCC charge set. Using D-MBIS atomic charges also for the OA host result in better agreement with the experimental reference values and ordering of the guest from weak to strong binders. The compatibility of the D-MBIS atomic charges with existing non-polarizable force fields and the automated computational framework present a computationally efficient alternative to include polarization effects in free energy predictions.

Our computational workflow for parameter derivation is not restricted to atomic charges and the improvement of electrostatic interactions. First attempts to derive Lennard-Jones parameters describing the repulsive and dispersion interactions have shown promising results.

Acknowledgement

The authors thank financial support from FONDECYT 1200369.

Supporting Information Available

Table S1 shows the average atomic charge difference between the bound and unbound state for each guest in the OA host and Table S2 the binding free energies obtained without inclusion of the polarization energy correction. Figure S1 the distribution of angles characterizing the representative conformation of guest G1 and G6 in the OA and the TEMOA and Figure S2 the same distribution for different atomic charge sets of the guest: AM1-BCC to D-MBIS. Initial coordinates and topologies used in the alchemical free energy calculations are provided at [DOI:10.5281/zenodo.4912084](https://doi.org/10.5281/zenodo.4912084).

Data and Software Availability

OpenMM and Yank are open source software and can be used free of charge. The simulation workflow written in Python is available under [DOI:10.5281/zenodo.4912084](https://doi.org/10.5281/zenodo.4912084). ORCA 4.1.0 is free for academic use open request and Horton 2.0 is free open source software. All figures were created with matplotlib package from python.

References

- (1) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153–4169.
- (2) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.
- (3) Amezcua, M.; El Khoury, L.; Mobley, D. L. SAMPL7 Host–Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 1–35.
- (4) Muddana, H. S.; Daniel Varnado, C.; Bielawski, C. W.; Urbach, A. R.; Isaacs, L.; Geballe, M. T.; Gilson, M. K. Blind prediction of host–guest binding affinities: a new SAMPL3 challenge. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 475–487.
- (5) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259–279.
- (6) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host–guest blind prediction challenge: an overview. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305–317.
- (7) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 host–guest binding affinity prediction challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 937–963.
- (8) Gibb, C. L. D.; Gibb, B. C. Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 319–325.

- (9) Sullivan, M. R.; Sokkalingam, P.; Nguyen, T.; Donahue, J. P.; Gibb, B. C. Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA deep-cavity cavitands. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 21–28.
- (10) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The role of water in host-guest interaction. *Nat. Commun.* **2021**, *12*, 93.
- (11) Yin, J.; Fenley, A. T.; Henriksen, N. M.; Gilson, M. K. Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. *J. Phys. Chem. B* **2015**, *119*, 10145–10155.
- (12) Gao, K.; Yin, J.; Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Binding Enthalpy Calculations for a Neutral Host–Guest Pair Yield Widely Divergent Salt Effects across Water Models. *J. Chem. Theory Comput.* **2015**, *11*, 4555–4564.
- (13) Yin, J.; Henriksen, N. M.; Gilson, M. K. The SAMPL5 host–guest challenge: computing binding free energies and enthalpies from explicit solvent simulations by the attach-pull-release (APR) method. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 133–145.
- (14) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1–19.
- (15) Henriksen, N. M.; Gilson, M. K. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host–Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *J. Chem. Theory Comput.* **2017**, *13*, 4253–4269.
- (16) Assaf, K. I.; Florea, M.; Antony, J.; Henriksen, N. M.; Yin, J.; Hansen, A.; Qu, Z.-w.; Sure, R.; Klapstein, D.; Gilson, M. K.; Grimme, S.; Nau, W. M. HYDROPHOBE Challenge: A Joint Experimental and Computational Study on the Host–Guest Binding of Hydrocarbons to Cucurbiturils, Allowing Explicit Evaluation of Guest Hydration Free-Energy Contributions. *J. Phys. Chem. B* **2017**, *121*, 11144–11162.

- (17) Yin, J.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bind3P: Optimization of a Water Model Based on Host–Guest Binding Data. *J. Chem. Theory Comput.* **2018**, *14*, 3621–3632.
- (18) Slochower, D. R.; Henriksen, N. M.; Wang, L.-P.; Chodera, J. D.; Mobley, D. L.; Gilson, M. K. Binding Thermodynamics of Host–Guest Systems with SMIRNOFF99Frosst 1.0.5 from the Open Force Field Initiative. *J. Chem. Theory Comput.* **2019**, *15*, 6225–6242.
- (19) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 601–633.
- (20) Mobley, D. L.; Amezcua, M.; Ponder, J.; Khalak, Y.; Eken, Y.; Almeida, N.; Isaacs, L.; Gibb, B.; Kellett, K.; Serrilon, D. The SAMPL7 Host-Guest Challenge Virtual Workshop. *Zenodo*. **2020**, DOI:10.5281/zenodo.3674155.
- (21) Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47*, 2795–2803.
- (22) Kognole, A. A.; Aytenfisu, A. H.; MacKerell, A. D. Balanced polarizable Drude force field parameters for molecular anions: phosphates, sulfates, sulfamates, and oxides. *J. Mol. Model.* **2020**, *26*, 152.
- (23) Saric, D.; Kohns, M.; Vrabec, J. Dielectric constant and density of aqueous alkali halide solutions by molecular dynamics: A force field assessment. *J. Chem. Phys.* **2020**, *152*, 164502.
- (24) Debiec, K. T.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further

- along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926–3947.
- (25) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515–4534.
- (26) Kelly, B. D.; Smith, W. R. A Simple Method for Including Polarization Effects in Solvation Free Energy Calculations When Using Fixed-Charge Force Fields: Alchemically Polarized Charges. *ACS Omega* **2020**, *5*, 17170–17181.
- (27) Kelly, B. D.; Smith, W. R. Alchemical Hydration Free-Energy Calculations Using Molecular Dynamics with Explicit Polarization and Induced Polarity Decoupling: An On-the-Fly Polarization Approach. *J. Chem. Theory Comput.* **2020**, *16*, 1146–1161.
- (28) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894–3912.
- (29) Riquelme, M.; Lara, A.; Mobley, D. L.; Matamala, A. R.; Vöhringer-Martinez, E. Hydration Free Energies in the FreeSolv Database Calculated with Polarized Iterative Hirshfeld Charges. *J. Chem. Inf. Model.* **2018**, *58*, 1779–1797.
- (30) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.
- (31) Shi, Y.; Laury, M. L.; Wang, Z.; Ponder, J. W. AMOEBA binding free energies for the SAMPL7 TrimerTrip host–guest challenge. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 79–93.

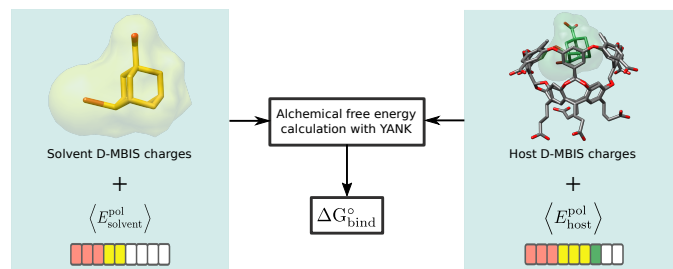
- (32) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (34) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (35) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (36) 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.
- (37) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs. Comput. Mol. Sci.* **2013**, *3*, 198–210.
- (38) Joung, I. S.; Cheatham, 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.
- (39) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

- (40) Chodera, J.; Rizzi, A.; Naden, L.; Beauchamp, K.; Grinaway, P.; Fass, J.; Rustenburg, B.; Ross, G. A.; Simmonett, A.; Swenson, D. W. choderalab/openmmtools: 0.14.0 - Exact treatment of alchemical PME electrostatics, water cluster test system, optimizations. *Zenodo*. **2018**, DOI:10.5281/zenodo.1161149.
- (41) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (42) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (43) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- (44) Neese, F. The ORCA program system. *WIREs. Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (45) Pan, X.; Rosta, E.; Shao, Y. Representation of the QM Subsystem for Long-Range Electrostatic Interaction in Non-Periodic Ab Initio QM/MM Calculations. *Molecules* **2018**, *23*, 2500.
- (46) Verstraelen, T.; Tecmer, P.; Heidar-Zadeh, F.; Boguslawski, K.; Chan, M.; Zhao, Y.; Kim, T. D.; Vandenbrande, S.; Yang, D.; González-Espinoza, C. E.; Fias, S.; Límacher, P. A.; Berrocal, D.; Malek, A.; Ayers, P. W. HORTON 2.0.0. 2015; <http://theochem.github.com/horton/>.
- (47) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.
- (48) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Grinaway, P.; Rustenburg, B.; Al-

- banese, S.; Saladi, S. choderalab/yank: 0.20.1 - Exact treatment of PME electrostatics and optimizations. *Zenodo*. **2018**, DOI:10.5281/zenodo.1161274.
- (49) Tan, Z. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. *J. Comput. Graph. Stat.* **2017**, *26*, 54–65.
- (50) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (51) Liu, J. S. *Monte Carlo Strategies in Scientific Computing*, 2nd ed.; Springer-Verlag: New York, 2002.
- (52) Leimkuhler, B.; Matthews, C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *App. Math. Res. Express* **2013**, *2013*, 34–56.
- (53) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799–1805.
- (54) Chodera, J. D.; Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.* **2011**, *135*, 194110.
- (55) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. II. Method and Application for Computing Effect of Polarization Cost on Free Energy of Hydration. *J. Phys. Chem. B* **2010**, *114*, 8631–8645.
- (56) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. I. Method for Computing Energy. *J. Phys. Chem. B* **2010**, *114*, 8621–8630.
- (57) Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partitioning Schemes. *J. Phys. Chem. A* **2018**, *122*, 4219–4245.

- (58) Hait, D.; Liang, Y. H.; Head-Gordon, M. Too big, too small, or just right? A benchmark assessment of density functional theory for predicting the spatial extent of the electron density of small chemical systems. *J. Chem. Phys.* **2021**, *154*, 074109.
- (59) Hait, D.; Head-Gordon, M. How accurate are static polarizability predictions from density functional theory? An assessment over 132 species at equilibrium geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800–19810.
- (60) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.
- (61) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Albanese, S.; Grinaway, P.; Prada-Gracia, D.; Rustenburg, B.; ajsilveira,.; Saladi, S.; Boehm, K.; Rodríguez-Guerra, J. choderalab/yank: 0.24.1 - Bugfix release. *Zenodo*. **2019**, DOI:10.5281/zenodo.3236834.
- (62) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (63) Tofoleanu, F.; Lee, J.; Pickard IV, F. C.; König, G.; Huang, J.; Baek, M.; Seok, C.; Brooks, B. R. Absolute binding free energies for octa-acids and guests in SAMPL5. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 108–118.
- (64) Pal, R. K.; Haider, K.; Kaur, D.; Flynn, W.; Xia, J.; Levy, R. M.; Taran, T.; Wickstrom, L.; Kurtzman, T.; Gallicchio, E. A combined treatment of hydration and dynamical effects for the modeling of host-guest binding thermodynamics: the SAMPL5 blinded challenge. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 29–44.

- (65) Papadourakis, M.; Bosisio, S.; Michel, J. Blinded predictions of standard binding free energies: lessons learned from the SAMPL6 challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1047–1058.
- (66) 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. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 375–400.
- (67) Caldararu, O.; Olsson, M. A.; Misini Ignjatović, M.; Wang, M.; Ryde, U. Binding free energies in the SAMPL6 octa-acid host–guest challenge calculated with MM and QM methods. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1027–1046.



Graphical TOC Entry

We present a computational framework to derive environment specific D-MBIS atomic charges together with the energy required to polarize the guest in the bound and unbound state absolute which combined with Yank result in accurate host-guest binding free energies.