JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

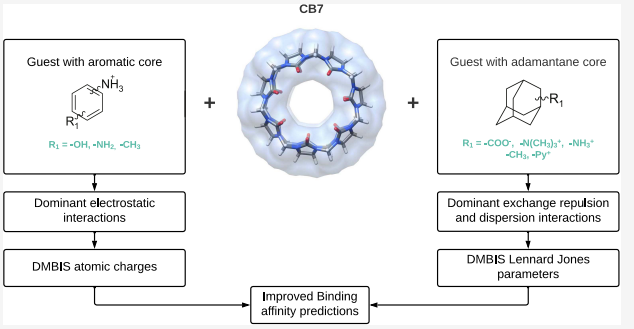
Nonbonded Force Field Parameters from Minimal Basis Iterative Stockholder Partitioning of the Molecular Electron Density Improve CB7 Host−Guest Affinity Predictions

Duván González, Luis Macaya, Carlos Castillo-Orellana, Toon Verstraelen, Stefan Vogt-Geisse, and Esteban Vöhringer-Martinez*

Cite This: J. Chem. Inf. Model. 2022, 62, 4162−4174 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Binding affinity prediction by means of computer simulation has been increasingly incorporated in drug discovery projects. Its wide application, however, is limited by the prediction accuracy of the free energy calculations. The main error sources are force fields used to describe molecular interactions and incomplete sampling of the configurational space. Organic host−guest systems have been used to address force field quality because they share similar interactions found in ligands and receptors, and their rigidity facilitates configurational sampling. Here, we test the binding free energy prediction accuracy for 14 guests with an aromatic or adamantane core and the CB7 host using molecular electron density derived nonbonded force field parameters. We developed a computational workflow written in Python to derive atomic charges and Lennard-Jones parameters with the Minimal Basis Iterative Stockholder method using the polarized electron density of several configurations of each guest in the bound and unbound states. The resulting nonbonded force field parameters improve binding affinity prediction, especially for guests with an adamantane core in which repulsive exchange and dispersion interactions to the host dominate.



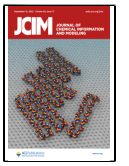
INTRODUCTION

Computer simulations are increasingly used in drug discovery projects for ligand receptor binding free energy prediction because of increasing computational resources and efficient algorithms.[1] The prediction accuracy of these simulations is crucial for their applications and depends on sampling all conformations of the ligand−receptor system in the bound and unbound states and the accurate description of the interactions in the molecular system.[2] To fulfill both requirements, polarizable and nonpolarizable force fields are used to describe the interactions between atoms, and computational efficient algorithms have been established to predict the binding free energy.[3] Rigid molecules acting as hosts have been mostly used to validate the prediction accuracy of force fields and free energy methods. These hosts consist of a few hundred atoms, but they share the same intermolecular forces as proteins and allow complete sampling of the configurational space.[4]

Cucurbiturils, Gibb's deep-cavity cavitands, and cyclodextrins represent examples of these hosts used in recent Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenges because of their diverse binding affinities and the ability to bind small drug-like compounds.[4−7] The family of cucurbit[n]urils forms a highly symmetric structure of covalently bound glycoluril monomers (CB[n], n = 5, 6, 7, 8, 10) with an ureidyl carbonyl portal and a relatively nonpolar inner cavity.[8] The inner cavity, however, displays a very high quadrupole moment that allows higher-order electrostatic interactions.[9] CB[n] hosts bind guests with a hydrophobic core and at least one cationic moiety interacting with the oxygen atoms of the carbonyl groups surrounding both portals of the host.[10] CB6 is the oldest of the family and has been extensively studied by the pioneering work of Mock[11,12] and Kim.[13] In 2005, the Isaacs group made a comparative binding affinity study of the same guests to the CB6, CB7, and CB8 hosts and discovered that CB7 and CB8 retain much of the remarkable binding characteristics of CB6 complexes.[10] In contrast to CB8, the CB7 host exhibits very high selectivity for adamantane derivatives as guests.[10]

In this study, we focus on CB7 shown in Figure 1 to assess the quality of nonbonded force field parameters. CB7 was used in many SAMPL initiatives[4,7] as the host to benchmark the affinity prediction of guests with an adamantane or aromatic

Received: March 17, 2022
Published: August 12, 2022



https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

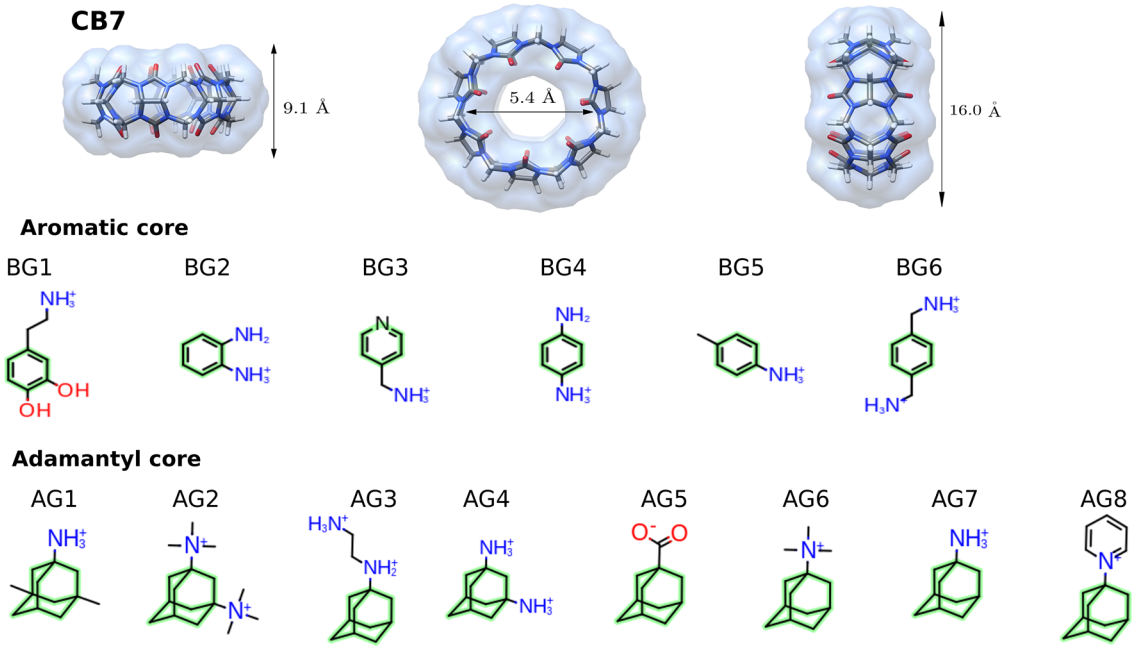


Figure 1. Host−guest systems. On top, the curcubituril CB7 host is shown in different orientations, and at the bottom, the guest is grouped by the aromatic or adamantane core with the labels used throughout the text.

core, disclosing binding affinities that exceed even the tightest binding protein−ligand systems.[8,10,14,15] The measured binding affinities of guests with aromatic and adamantane cores, shown in Figure 1, cover a range of −6.3 to −12.6 kcal/mol and −6.0 to −18.2 kcal/mol, respectively.[10]

This wide range of binding free energies was partially attributed to the nature of the interactions, described by Nau et al.[16] They conclude that these hosts bind guests by a combination of electrostatic interactions with the carbonyl rims and hydrophobic interactions in a very apolar inner cavity. From a computational and theoretical perspective, the interactions have been addressed qualitatively using the noncovalent index by Contreras-García et al.[17] For one guest with hydroxyl groups, they conclude that van der Waals interactions dominate inside the cavity and that electrostatic interactions dominate at the top of the host. Parish et al. applied Symmetry Adapted Perturbation Theory with functional group partitioning (F-SAPT) to quantify different types of interaction and the contribution of specific functional groups.[18] Their analysis assigns electrostatic interactions an important role, accompanied by a significant contribution of dispersion interactions and exchange repulsion for guests with an adamantane core. Here, we extend their F-SAPT analysis in a comparative study of CB7 guests with an adamantane or aromatic core to quantify the different types of interactions for these two groups of guests and rationalize the contribution of varying functional groups to the total interaction energy.

The quantification and characterization of the types of molecular interaction and the contribution of functional groups to the total interaction energy are crucial to model atomic interactions in force fields. However, accurate force fields alone do not result directly in better binding affinity prediction because of a significant entropic contribution to the free energy. Binding of the guest releases water molecules from the cavity, resulting in a change of the solvent entropy.[16] Alchemical free energy calculations combined with molecular dynamics in explicit solvents account for the entropic contribution to the binding affinity and should be combined with a computationally inexpensive but accurate description of the four classes of interactions commonly defined in SAPT: electrostatic, induction, dispersion, and exchange repulsion. Existing models account for these effects in different ways, with most of the new developments focusing on the induction (or polarization) contribution. Our recent approach to model the induction relies on MD simulations with static polarized atomic charges, which are refined iteratively in a QM/MM setup, and a polarization cost associated with the polarized charges, which can be computed separately. More specifically, we have shown that Minimal Basis Iterative Stockholder atomic charges derived from the polarized electron density of various conformations (D-MBIS) combined with the General Amber Force Field (GAFF) reduce binding affinity prediction errors in octa acid hosts and order the guests correctly from weak to strong binders.[19] We were able to show that accounting for the difference in the potential energy required to polarize the electron density in the bound and unbound states improves affinity predictions of anionic guests. Many other techniques have been proposed. For example, polarization effects of solutes have been included also in a protein force field by Cerutti and coauthors.[20,21] Kelly and Smith[22,23] proposed an alternative method to account for solute polarization and applied it to the calculation of hydration free energies of small organic molecules. Partitioning the polarized electron density of solutes with the MBIS method,[24] they derive atomic charges for different values of the coupling free energy parameter lambda. This approach is similar to our previously proposed method using an implicit representation of the electrostatic reaction field of the solvent together with the MBIS partitioning method[25] but neglects the potential energy required to polarize the solute's electron density in the transfer from the gas to the condensed phase. We have shown recently that this contribution becomes important for anionic guest in the octa acid hosts.[19]

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

Although we achieved better binding free energy predictions in octa acid hosts improving electrostatic interactions with the D-MBIS atomic charges, in CB7, the interactions present a considerable contribution of dispersion and exchange repulsion.[18] Nonpolarizable force fields describe these interactions by atom specific Lennard-Jones (LJ) parameters. A self-consistent derivation of atomic charges and LJ parameters is crucial to improve force fields. The recently proposed RESP2 approach is a good example where a more sophisticated atomic charge method does not always correspond to improved accuracy.[26] Mohebifar et al. have shown recently that atomic C_6 LJ coefficients in various force fields are overestimated compared to the ones derived from quantum chemical calculations using the exchange-dipole model.[27] The discrepancy between the LJ parameters found in common force fields and the ones from quantum chemical calculations has motivated different approaches to derive them. Cole et al. had previously proposed to use Atoms In Molecules (AIM) partitioning of the electron density to derive dispersion coefficients optimizing the van der Waals radii to reproduce liquid densities.[28] Kantonen et al. optimized element specific mapping parameters between a property of the atom in a molecule from the MBIS method and the LJ parameters to reproduce thermodynamic properties of molecular liquids.[29] Visher et al. proposed to derive the dispersion coefficients from an AIM partitioning of the electron density using an empirical relation between the van der Waals radius and the atom polarizability.[30] Deriving all force field parameters from first-principles accounting for polarization with the Drude oscillators, they were able to reproduce experimental estimates for thermodynamic pure liquid properties of alkanes. Most of the mentioned approaches used electronic structure calculations of one representative conformation of the molecule in the vacuum, without accounting for conformational variability and the polarizing effects of the solvent molecules on the electron density. To the best of our knowledge, no method has derived parameters for the Lennard-Jones potential from the polarized electron density using different configurations of the solute and the solvent.

Here, we present a computational framework that derives nonbonded force field parameters from the polarized electron density of molecules in condensed phases. Lennard-Jones parameters are calculated for each guest using the method by Federov et al.[31] to obtain van der Waals radii and dispersion coefficients from MBIS[24] partitioning of the polarized electron density. We use our previously introduced environment specific D-MBIS atomic charges and account for the potential energy required to polarize the guest in different environments using the QM/MM methodology.[19] Our computational workflow written in Python provides nonbonded force field parameters, combining molecular dynamics simulation in OpenMM[32] and QM/MM calculations with the ORCA software package.[33] The polarized electron density from the QM/MM calculations is used to derive the parameters using IOData[34] and other Python libraries from Horton3. The D-MBIS atomic charge set together with the van der Waals radii and dispersion coefficients averaged over various configurations of the molecular system replaces the GAFF nonbonded force field parameters.

The affinity prediction accuracy of the new set of parameters for various guests of the CB7 host is assessed with the self-adjusted mixture sampling method in YANK[35] together with previously validated simulation parameters.[19] 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 extends our recent methodology applied to the octa acid host[19] providing a complete set of nonbonded force field parameters derived from the polarized electron density of the guest in the bound and unbound states.

METHODS

System Preparation. The bound conformation of each host−guest system was downloaded from the GitHub repository MobleyLab/benchmark sets,[36] and for some systems, the most likely binding mode was obtained with the OpenEye toolkit.[37−39] The protonation states of the guests were taken as suggested by Mobley[36] and correspond to the most probable states based on the experimental pH value of 4.74 and the pK_a values of the titratable functional groups shown in Table S1 of the Supporting Information. For 1,2-diaminobenzene (BG2) and 1-adamantanecarboxylic acid (AG5), which pK_a values are close to 4.74, we calculated the binding free energy of both possible protonation states and report the one closest to the experimental value (see Table S13 in the Supporting Information). The input files were created with YANK 0.24.1, using RESP[40] atomic charges for the host and AM1-BCC[41,42] for guests, respectively. These charges were combined with the parameters from the GAFF1.8[43,44] force field generated by antechamber with the Ambertools17 package.[45] The host and guests were solvated in a cubic box with explicit water molecules described by the TIP3P water model,[46] with a clearance distance away from any host/guest atom to the edge of the solvation box of 12 Å. A mass of 3 amu was used for the hydrogen atoms combined with constraints on hydrogen bonds to allow larger time steps in the production run. To emulate experimental ionic strength produced by a 50 mM sodium acetate buffer, additional Na^+ and Cl^- ions were added to the CB7-guests systems using Joung and Cheatham parameters.[47] The resulting structures were minimized in 1000 steps using the Limited-memory-Broyden-Fletcher-Goldfarb-Shanno[48] (L-BFGS) algorithm and equilibrated running 1 ns of Langevin splitting dynamics (BAOAB splitting, where the letters A and B stand for the linear drift and kick and O stands for the Ornstein−Uhlenbeck equation as parts of the integration algorithm,[49] 1 fs time step) at 298 K with a Monte Carlo barostat set at 1 atm using OpenMM[32] and OpenMMTools.[50] 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 create input files for simulations using the D-MBIS atomic charges, the initial atomic charge set RESP/AM1-BCC for the host and guest respectively were replaced by derived D-MBIS atomic charges preserving the same Lennard-Jones parameters from GAFF. The same protocol was performed for simulations using D-MBIS atomic charges combined with LJ parameters derived with MBIS partitioning of the polarized electron density for the guest. In this approach, in addition to the replacement by the D-MBIS atomic charges, the GAFF Lennard-Jones parameters of the guest were changed by the derived MBIS LJ parameters.

Nonbonded Force Field Parameters from the Polarized Electron Density. For the CB7 host, we combined

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

our recently introduced D-MBIS atomic charges[19] with Lennard-Jones parameters and bonded force field parameters from GAFF1.8 as implemented in the AMBER18 package.[45] To calculate the D-MBIS atomic charges, the host was surrounded by water molecules and ions required to emulate the ionic strength. Our simulation protocol in Python[19] combines molecular dynamics simulation carried out with OpenMM[32] with the QM/MM methodology applied to several configurations of the trajectory. Electron densities were obtained with the QM/MM method at the B3LYP[51] level of theory with the def2-TZVP basis set,[52,53] using the new support to import force fields combined with PDB files in ORCA 4.2.1[33] in which all MM atoms are included as point charges. We used the Minimal Basis Iterative Stockholder (MBIS)[24] method to calculate atomic charges from the polarized electron densities of the QM/MM calculations. Combining IOData[34] and other libraries from the Horton3 package, we derive MBIS atomic charges from the polarized electron density of each configuration. We average MBIS atomic charges from all configurations and replace them in the topology before we initiate a new cycle of MD simulations and QM/MM calculations. The simulation protocol uses up to 25 representative configurations for the QM/MM calculations before each update of the topology. The OpenMM package[32] performs molecular dynamics simulations at 298 K and 1 bar using a Langevin integrator with a 2 fs time step, collision frequency of 1 ps, particle mesh Ewald electrostatics with a 1.0 nm cutoff, constraints on bonds involving hydrogen atoms, harmonic restraint with a spring constant of 0.2 kcal/(mol·Å)^2 in the bound state to ensure that the guest remains in the binding site,[54] and Monte Carlo barostat. The total simulation time was 5 ns between each update of the guest's charges in the topology. With five topology updates, the difference between atomic charges from two subsequent cycles was below 0.05e.[19]

Chemically equivalent atoms, such as hydrogen atoms in methyl groups, were identified for the host with the OpenEye Python Toolkit (version 2017.2.1), and their converged D-MBIS charges were averaged. We assume that D-MBIS charges are not geometry-dependent, and we do not include charge polarizability during the MD simulations.

With the CB7 host force field, we calculate D-MBIS atomic charges for each guest by applying the computational workflow described above: guest atoms are described quantum mechanically, and the rest is modeled with a force field.[19] With this protocol, D-MBIS atomic charges are obtained for each guest in the bound state and in the aqueous solution. The potential energy required to polarize the electron density in the different molecular environments is included in the free energy calculation as described below.

We additionally derive Lennard-Jones parameters for all guests in the aqueous solution to replace the GAFF parameters with a new computational workflow shown in Figure 2. We use a similar alternation of QM/MM calculation and molecular dynamics simulations described above to obtain the polarized electron density of the guest for each configuration (blue area in Figure 2). After partitioning the molecular electron density in atomic contributions with the MBIS method, we calculate the effective volume of the AIM density of each atom in the molecule $\langle r^3\rangle_{\mathrm{aim}} = \int r^3 \rho_{\mathrm{aim}} d\tau$. We average the values of $\langle r^3\rangle_{\mathrm{aim}}$ for each atom over all configurations sampled for each cycle. In parallel, we also derive the MBIS atomic charges for each configurations, and their average value of all configurations replaces their values in the original guest topology. With the new D-MBIS atomic charges, a new cyle of molecular dynamics simulations starts until the change in the D-MBIS atomic charges is below a threshold of Δe = 0.05, which

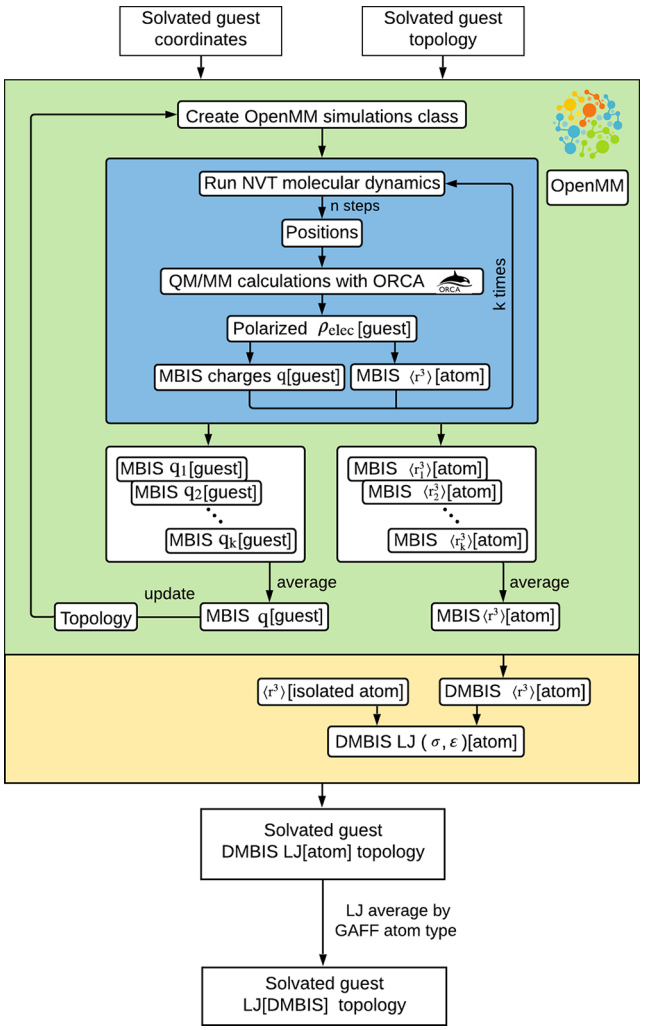


Figure 2. Computational flowchart used to derive the Lennard-Jones parameters for the guests based on MBIS partitioning. The blue area represents the inner cycle of molecular dynamics simulations in the NVT ensemble with OpenMM followed by QM/MM calculations with ORCA. The number of steps (n) defines the number of configurations used for the QM/MM calculations given a total simulation time. MBIS charges and the effective volume of each atom in the guest are obtained from the polarized electron densities before continuing the molecular dynamics simulations. This cycle is repeated k times, where k represent a value for the relation between the number of QM/MM calculations and total sampling time. For this case, with 25 total QM/MM calculations and 5 ns of sampling, the inner cycle is executed 5 times in the first update, 10 in the second, and so forth. The ensemble average value of all MBIS atomic charges replaces their original value in the topology (green area). With the new topology, a new cycle of molecular dynamics simulations and QM/MM calculations begins, and the cycle is repeated five times after which the change in atomic charges is below $\Delta e \leq 0.05$. σ and ϵ parameters of the Lennard-Jones potential for each atom in the guest are calculated from the averaged effective volume of the atom in the molecule obtained in the last cycle and the respective volume of the isolated atom (yellow area). The final LJ parameters for each atom type in the GAFF force field become the σ and ϵ parameter values of all atoms in the guest of the same type.

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

corresponds to 5 topology updates. From the last cycle, the values of $\langle r^3 \rangle_{aim}$ are used to derive Lennard-Jones parameters.

We apply the relation first proposed by Becke and Johnson[55] and later by from Tkatchenko[56] to derive the C_6 dispersion coefficients for each atom in the molecule according to

$$C_6 = \left(\frac{\langle r^3 \rangle_{aim}}{\langle r^3 \rangle_{free}} \right)^2 C_{6,free} \quad (1)$$

where $C_{6,free}$ corresponds to the dispersion coefficient of the isolated atom taken from Chu,[57] and $\langle r^3 \rangle_{free}$ corresponds to the expectation value for the isolated atom calculated at the same level of theory as the guest (B3LYP/def2-TZVP). The van der Waals radii of each atom in the molecule are calculated with the recently introduced relation from Fedorov and Tkatchenko[31,58]

$$R_{vdw} = 2.54\alpha^{1/7} \quad (2)$$

where α is the static polarizability of the atom in the molecule obtained by scaling the polarizability of the isolated atom recommended in Chu et al.[57] with the factor $\frac{\langle r^3 \rangle_{aim}}{\langle r^3 \rangle_{free}}$.

We maintain the GAFF atom type of the guest but assign new values of Lennard-Jones parameters averaging the values of all atoms in the guest of the same type. With the new set of LJ parameters for the guest atom types, we derive a new set of D-MBIS atomic charges for the guest in the bound and the unbound states together with their polarization energies (see Figure S1 in the Supporting Information).

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. The polarization of the guest in the host in comparison to the noninteracting isolated guest is associated with a potential energy cost that is always positive. Because the binding free energy calculations employ a double decoupling scheme, this polarization energy would cancel out only if it would be the same in the bound state and the solvent; but as we have shown previously, the polarization energy cost differs between the two molecular environments.[19] This difference between environments contributes to the binding free energy and is usually neglected in absolute binding free energy calculations with nonpolarizable force fields.

To estimate the energetic cost associated with the guest's polarization in the bound state and the solvent phase E^{pol}_{host} and $E^{pol}_{solvent}$, we used the same method as described previously. Briefly, we perform an additional electronic structure calculation of each configuration using 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 ($\Psi_{complex}$) or in the solvent ($\Psi_{solvent}$). The energetic polarization cost results as the difference between this expectation value and the converged self-consistent-field energy obtained in the single point calculation employing the same Hamiltonian:

$$E^{pol}_{host} = \langle \Psi_{complex} | \hat{H}_{vac} | \Psi_{complex} \rangle - \langle \Psi_{vac} | \hat{H}_{vac} | \Psi_{vac} \rangle \quad (3)$$

$$E^{pol}_{solvent} = \langle \Psi_{solvent} | \hat{H}_{vac} | \Psi_{solvent} \rangle - \langle \Psi_{vac} | \hat{H}_{vac} | \Psi_{vac} \rangle \quad (4)$$

The average polarization energy $\langle E^{pol}_{host} \rangle$ or $\langle E^{pol}_{solvent} \rangle$ of each guest in the two molecular environments is obtained averaging all configurations from the new simulation protocol representing an ensemble-averaged value. The two contributions related to the polarization energy cost are included in the calculation of the binding free energy ΔG°_{bind} (see eq 1 in the Supporting Information).

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 CB7 host between RESP and D-MBIS and keeping the Lennard-Jones parameters from the GAFF force field. Additionally, binding free energies were also calculated considering the D-MBIS atomic charges for the host and guests together with the Lennard-Jones parameters for the guests derived from MBIS partitioning of the molecular electron density as described above. YANK has implemented the double decoupling methodology which annihilates the guest's interaction in the bound and unbound states. Since all guests studied have an associated net charge, a counterion of the opposite charge must undergo an alchemical transformation in parallel with the guest to keep the neutrality of the system. Due to the observed free energy dependence on the alchemical ion position analyzed in our previous investigation on octa acids,[19] we select the ion farthest from the guest's initial position as an alchemical counterion. We applied a simple harmonic constraint on the heavy atoms center of mass of CB7 and the guest using a spring constant of 1.1 kcal/(mol·Å^2) to maintain 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.[35] This resulted in the following number of intermediate states for all systems in the bound and unbound states, respectively: CB7-BG1(77−72), CB7-BG2(72−68), CB7-BG3(73−67), CB7-BG4(71−68), CB7-BG5(72−67), CB7-BG6(96−92), CB7-AG1(77−72), CB7-AG2(102−93), CB7-AG3(99−97), CB7-AG4(97−95), CB7-AG5(73−72), CB7-AG6(72−68), CB7-AG7(69−68), and CB7-AG8(73−69).

The self-adjusted mixture sampling (SAMS),[59] also known as optimally adjusted mixture sampling, was used to sample the whole range of thermodynamic states, allowing the rapid computation of the free energy calculations. SAMS uses one replica in an expanded ensemble[60] to sample multiple thermodynamic states, adapted on the fly[59] 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 visits to each thermodynamic state.[61] Each iteration consists of 1 ps Langevin splitting dynamics (BAOAB splitting[49] using a 4 fs time step, 10 ps collision rate, and 2500 steps/iteration) followed by Monte Carlo (MC) displacement and rotation moves of the ligand without leaving the same thermodynamic state. These MC moves are carried out with the openmmtools.mcmc module drawing samples from the equilibrium distribution at each iteration according to the Markov Chain Monte Carlo (MCMC) method.[62] To maintain a pressure of 1 atm, a Monte Carlo barostat attempts to scale the box size for every 25 integration steps that were performed. In all free energy calculations, we used 40000 iterations for bonded and nonbonded states to ensure converged free energy estimates,

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

based on the validation of the binding free energy with respect to the number of iterations carried out previously.[19] All simulations in YANK start with the automated equilibration detection to maximize the number of effectively uncorrelated samples in the production simulation used to compute equilibrium averages.[61,63,64]

The analysis of the samples was performed with the Multistate Bennet Acceptance Ratio (MBAR)[61] estimator implemented in the PyMBAR 3.0.3 package.[65] YANK computed an estimate of the statistical inefficiency of the sampling process in order to decorrelate the SAMS samples. The statistical inefficiency was then used to discard the equilibrium phase by maximizing the number of effective samples and to subsample the data before the MBAR analysis. Due to the harmonic type restraint applied to the heavy atoms' center of mass of the CB7 and guest (complex phase), an analytical correction was added to bring the affinity in units of standard concentration and correct for the restraint volume in the decoupled state. 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 $V_0 = \frac{1L}{N_A}$ where N_A is Avogadro's constant. This calculation results in V_0 = 1660.53928 Å^3 where V_0 is treated as a per mole quantity.[54,61,63,64]

The obtained free energies were named according to the atomic charges that were used for the guest and the host: $\Delta G^{\circ}_{q[AM1\text{-}BCC]}$ and $\Delta G^{\circ}_{q[D\text{-}MBIS]}$ were obtained with the AM1-BCC and the D-MBIS atomic charges, respectively. For the latter, we applied the polarization correction in eq 1 of the Supporting Information to calculate the binding free energy according to the thermodynamic cycle of the double-decoupling scheme (Figure S2 in the Supporting Information). Free energies labeled $\Delta G^{\circ}_{LJ\text{-}q[D\text{-}MBIS]}$ were obtained with new Lennard-Jones parameters of the guest in aqueous solution and D-MBIS atomic charges for the bound and unbound states including the polarization correction. $\Delta G^{\circ}_{LJ\text{-}q[D\text{-}MBIS(solv)]}$ used D-MBIS atomic charges from the aqueous solution in the unbound and bound states without considering the polarization correction.

Symmetry Adapted Perturbation Theory Calculations. We apply zeroth-order symmetry adapted perturbation theory (SAPT0) to analyze the noncovalent interactions of the host−guest system, in terms of electrostatic, exchange, induction, and dispersion energy.[66,67] The monomers in the CB7-AG1 and CB7-AG7 systems for the SAPT0 computation correspond to CB7 and the respective guest (AG1 and AG7). We also analyzed the effect of including a solvation layer surrounding the ammonium group of the guests on the SAPT0 interaction energy and its components. The selected water molecules correspond to the first solvation sphere of the ammonium group defined by molecules with a maximum distance of 3.0 Å of the ammonium nitrogen atom. For the solvated host−guest systems, we computed the interaction of the CB7 host including the water molecules with two different guests: AG7 and BG5.

To gauge the pairwise contributions of functional groups with the interaction energy, we use functional-group symmetry adapted perturbation (F-SAPT0).[18] In this method, a single interaction energy is computed between guest and host monomers, and the interaction energy components are divided between the functional groups. The fragmentation of the monomers into chemically relevant functional groups is based on assigning local orbitals to functional groups and linking bonds, while leaving the functional groups neutral. These σ-bond linkers are then evenly assigned between the connecting fragments and can contribute to the electrostatic interaction between functional groups. For the F-SAPT0 analysis, the host monomer (CB7) was taken as a single fragment, while for the guests, each was separated into three and two fragments, respectively. For AG1, the three fragments correspond to the adamantyl core, the NH_3^+ amino group, and the two methyl substituents, while for AG7, the two fragments correspond to the adamantane core and the NH_3^+ amino group.

For the FSAPT0 analysis, the host system was separated into two fragments corresponding to CB7 and water molecules. The AG7 guest was fragmented into three fragments as in the system without solvation, while the BG5 guest was fragmented into three parts: the NH_3^+ amino group, the phenyl group, and the methyl substituent. For all studied systems, the most representative conformation of the guest was selected for SAPT analysis. We also analyzed the geometry dependence of the SAPT0 results, by optimizing for CB7-AG7 the two most probable conformations (conformations A-B in Figure S9 of the Supporting Information) from the trajectory using the HF3c/MINIX method.[68] The different conformations did not present significant changes in the relative contributions of the different types of interactions, and therefore for CB7-BG5 and AG8, the most relevant conformations of the dynamics were considered (conformations D and B, respectively, in Figures S7 and S9 in the Supporting Information). We used Psi4 for all SAPT0 and F-SAPT0 computations with the jun-cc-pvdz basis set.[69]

RESULTS AND DISCUSSION

Future application of free energy calculations in drug discovery projects is conditioned by the accuracy of the predicted binding free energies of ligands to receptors. One limitation to reach better prediction accuracy is force fields and their ability to correctly describe interactions between the ligand, the receptor, and the solvent. To quantify force field related binding affinity errors from free energy calculations, host−guest systems have been used because they share the same type of interactions as in proteins but allow complete sampling of the configurational space. Here, we analyze the binding affinity prediction accuracy of various sets of atomic charges and Lennard-Jones parameters for guests with an aromatic or adamantane core bound to the CB7 host.

First, we apply our recently introduced D-MBIS atomic charges for guests BG1-6 sharing an aromatic core shown in Figure 1 and evaluate the absolute binding free energy with the YANK package using the self-adjusted mixture sampling method. When the same D-MBIS atomic charges were used for the guests with an adamantane core AG1-8, we observed a poor prediction accuracy compared to the AM1-BCC/GAFF force field. Electronic structure calculations combined with symmetry adapted perturbation theory evidence dominant repulsive exchange and dispersion interactions in these guests, compared to the ones with an aromatic core. To improve the description of these interactions in force fields, we derived Lennard-Jones parameters from the partitioned molecular electron density of the guest. Combining these parameters with D-MBIS atomic charges and the inclusion of the polarization energy reduces the binding free energy errors considerably. Our proposed computational workflow written in Python allows one to derive nonbonded force field parameters from

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

Table 1. Absolute Binding Free Energies for Guests with an Aromatic Core (BG1-6) to the CB7 Host Using RESP Atomic Charges for the CB7 Host and AM1-BCC for the Guest or D-MBIS Atomic Charges for the Host and Guest and GAFF Atom Types and Bonded Parameters[a]

system	ΔG°_{exp}	$\Delta G^{\circ}_{q[AM1\text{-}BCC]}$	$\Delta G^{\circ}_{q[D\text{-}MBIS]}$	$\langle E^{pol}_{host} \rangle$	$\langle E^{pol}_{solvent} \rangle$
CB7-BG1	−6.3 ± 0.1	−10.1 ± 0.4	−5.1 ± 0.4	14.7 ± 0.2	10.7 ± 0.2
CB7-BG2	−6.7 ± 0.1	−5.5 ± 0.2	−6.8 ± 0.3	7.3 ± 0.2	8.4 ± 0.2
CB7-BG3	−7.6 ± 0.1	−5.4 ± 0.4	−8.3 ± 0.4	9.1 ± 0.2	10.1 ± 0.2
CB7-BG4	−8.6 ± 0.1	−8.1 ± 0.3	−4.9 ± 0.5	6.0 ± 0.2	7.1 ± 0.1
CB7-BG5	−9.4 ± 0.1	−8.9 ± 0.4	−8.1 ± 0.3	6.0 ± 0.1	6.3 ± 0.1
CB7-BG6	−12.6 ± 0.1	−7.3 ± 0.5	−10.0 ± 0.4	6.2 ± 0.1	5.4 ± 0.1
RMSE		2.9 ± 0.6	2.0 ± 0.5		
MAE		2.3 ± 0.6	1.6 ± 0.4		
R		0.0 ± 0.5	0.7 ± 0.3		
τ		−0.1 ± 0.5	0.5 ± 0.4		

[a]The values of $\langle E^{pol}_{host} \rangle$ and $\langle E^{pol}_{solvent} \rangle$ were obtained with the D-MBIS atomic charges for host and guests. Values are in kcal/mol.

Table 2. Absolute Binding Free Energies for Guests with an Adamantane Core (AG1-8) Combining the GAFF Force Field Parameters with RESP Atomic Charges for the CB7 Host and AM1-BCC for the Guest or D-MBIS Atomic Charges for the Host and Guest[a]

system	ΔG°_{exp}	$\Delta G^{\circ}_{q[AM1\text{-}BCC]}$	$\Delta G^{\circ}_{q[D\text{-}MBIS]}$	$\langle E^{pol}_{host} \rangle$	$\langle E^{pol}_{solvent} \rangle$
CB7-AG1	−6.0 ± 0.1	−6.8 ± 0.3	−10.1 ± 0.7	7.9 ± 0.1	6.5 ± 0.1
CB7-AG2	−6.6 ± 0.1	−3.0 ± 0.2	−2.1 ± 0.5	3.6 ± 0.1	3.4 ± 0.1
CB7-AG3	−18.2 ± 0.1	−19.4 ± 0.7	−24.0 ± 0.6	15.5 ± 0.2	12.8 ± 0.1
CB7-AG4	−11.3 ± 0.1	−4.2 ± 0.5	−12.0 ± 0.4	7.9 ± 0.1	7.8 ± 0.1
CB7-AG5	−11.6 ± 0.1	−17.0 ± 0.5	−17.7 ± 0.6	19.6 ± 0.2	20.2 ± 0.4
CB7-AG6	−16.7 ± 0.1	−22.4 ± 0.2	−23.5 ± 0.3	3.4 ± 0.1	2.5 ± 0.1
CB7-AG7	−17.2 ± 0.1	−20.6 ± 0.4	−22.7 ± 0.3	6.4 ± 0.1	6.0 ± 0.1
CB7-AG8	−16.8 ± 0.8	−22.4 ± 0.4	−24.7 ± 0.2	2.7 ± 0.1	2.0 ± 0.1
RMSE		4.6 ± 0.6	5.6 ± 0.6		
MAE		4.1 ± 0.7	5.2 ± 0.7		
R		0.9 ± 0.1	0.9 ± 0.0		
τ		0.5 ± 0.3	0.7 ± 0.2		

[a]The values of $\langle E^{pol}_{host} \rangle$ and $\langle E^{pol}_{solvent} \rangle$ were obtained with the D-MBIS atomic charges for host and guests. Values are in kcal/mol.

the polarized electron density, providing an alternative method to obtain nonbonded force field parameters for ligands in drug discovery projects.

Force Fields with D-MBIS Atomic Charges Identify Correctly the Strongest and Weakest Binding Guests with an Aromatic Core. We first addressed the guests with an aromatic core BG1-6, which are of the same size as the octa acid host−guests in our previous study.[19] The main difference is that the CB7 host is neutral, lacking the negatively charged carboxyl groups of the octa acids and that all BG1-6 guests are cationic in contrast to mostly anionic guests in the octa acids. Because of their cationic character, the energy required to polarize the electron density in the transfer from the noninteracting gas phase to the condensed phase is small. Furthermore, the difference between these two polarization energies in the bound and unbound states contributes only ≃1 kcal/mol to the binding free energy, as shown in the last two columns of Table 1. Only dopamine (BG1) presents larger absolute values and a difference of 4.0 kcal/mol because the two hydroxyl groups are exposed to different molecular environments in the bound and in the unbound states. When bound to CB7, they interact with the carbonyl groups at its top or bottom, whereas in the unbound state, intramolecular hydrogen bonds between the two groups compete with intermolecular ones to water molecules.

In the next step, we analyzed if the polarization energies combined with the D-MBIS atomic charges improve the prediction of the binding affinity. D-MBIS atomic charges derived for the host and guest predict binding free energy better than the AM1-BCC charge set, evidenced by smaller root-mean-square error (RMSE) and mean absolute error (MAE) and increased Pearson (R) and Kendall (τ) correlation coefficients shown in the last four lines of Table 1. The improved performance of D-MBIS atomic charges confirms our previous results of the two octa acid hosts with 6 different guests. Interestingly, the strongest and weakest binders are classified correctly.

1,4-Diaminobenzene (BG4) presents the largest deviation from the measured value, which might stem from alternating protonation states of both ammonium groups caused by different levels of solvent exposure. Varying protonation states might also affect 1,2-diaminobenzene (BG2). Its pK_a value is close to the experimental pH of 4.74 (see Table S1 in the Supporting Information) which suggests that a neutral, deprotonated form of the guest can coexist in the aqueous solution. We carried out free energy calculations of this neutral form with the D-MBIS atomic charges, but they showed larger deviation from the experimental value (see Table S13 in the Supporting Information). When the pK_a value is close to the experimental pH, multiple protonation states of the guest coexist, and binding to the host might also induce protonation state changes. To model protonation changes of BG2 and BG4 with computer simulations is feasible in constant pH simulations but is out of the scope of this work.

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

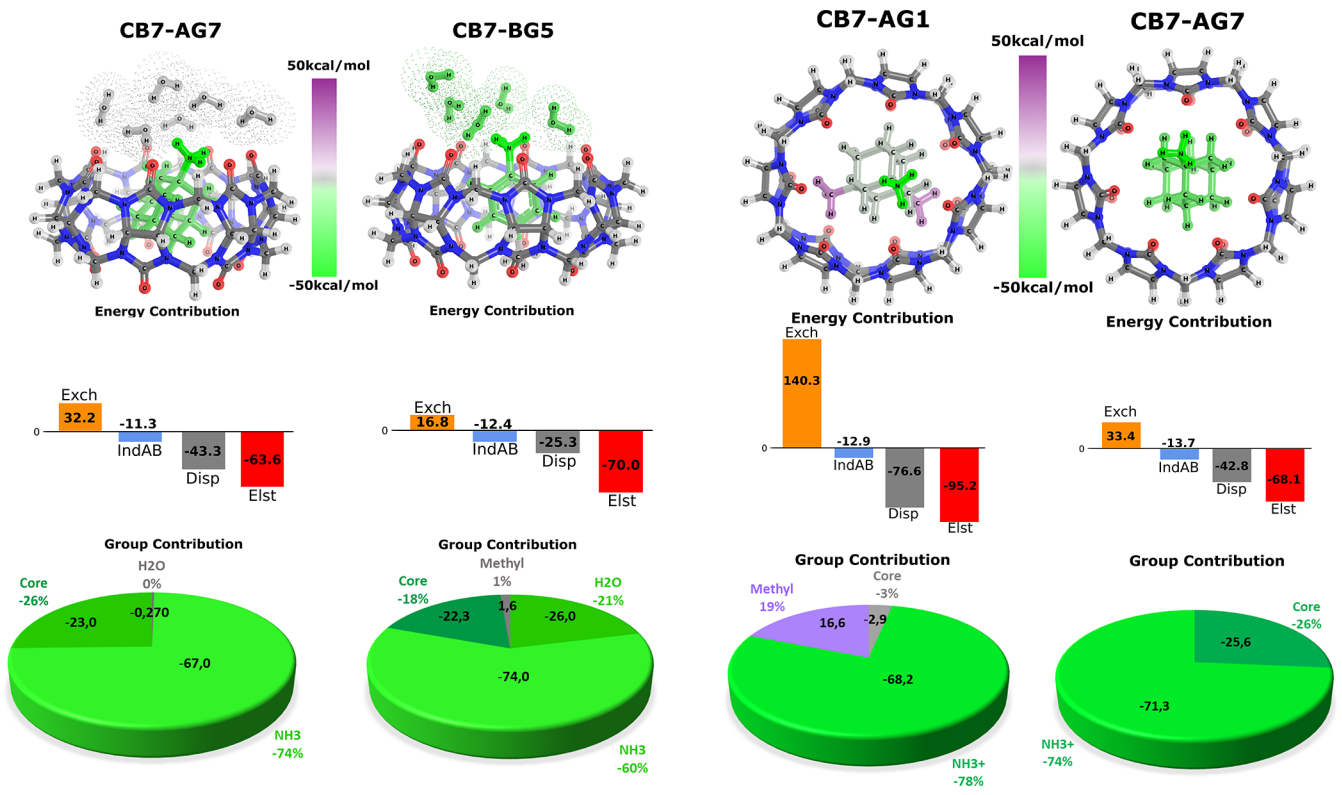


Figure 3. Symmetry Adapted Perturbation Analysis (SAPT) of the interactions in guests with aromatic and adamantane cores at the SAPT0/jun-ccVDZ level. Left: Comparison of the interactions between guests with aromatic and adamantane cores from a representative conformation in the aqueous solution, including water molecules solvating the charged ammonium group. Atoms of the guest and water molecules are colored according to their contribution to the interaction energy. The bars in the middle show electrostatic (Elst), dispersion (Disp), induction from host (A) to guest (B) (IndAB), and exchange (Exch) part of the total interaction energy expressed in kcal/mol. The pie charts at the bottom show the contribution from each functional group to the total interaction energy in percentage and the interaction energy in kcal/mol. Right: The same representations for two guests with an adamantane core highlight the large repulsive exchange contribution of memantine (AG1) and the repulsive contribution of its two methyl groups.

To confirm that the D-MBIS charge set improves the description of host−guest interactions, one has to ensure that the observed difference in binding free energy between the two charge sets does not originate from differing conformations of the guests. Our conformational analysis in Figures S3−S8 and Tables S2−S7 of the Supporting Information shows similar trends in the relative weight of the different conformations for most guests upon changes in the charge sets.

We conclude, therefore, that D-MBIS atomic charges for guests BG1-6 in the CB7 host combined with the inclusion of the polarization energy improve binding affinity predictions and correctly classify the strongest and weakest binder from the set. For the cationic guests BG2-6, differences in polarization energies between bound and unbound states are small, contributing only ≃1 kcal/mol to the binding free energy.

Host−Guest Interactions Are Dominated by Exchange Repulsion and Dispersion Interactions for Guests with an Adamantane Core. After the improvement in binding free energy prediction of the guest with an aromatic core using D-MBIS atomic charges, we focused on the ones with the adamantane core. In contrast to the guests with an aromatic core, however, D-MBIS atomic charges combined with the GAFF parameters deviate considerably from the measured binding free energies, reducing the prediction accuracy reflected by worse statistical descriptors as shown in the third column of Table 2.

To rationalize the poor performance of the D-MBIS atomic charges for guests with an adamantane core compared to the aromatic ones BG1-6, we analyzed which type of interaction dominated in both host−guest systems with Symmetry Adapted Perturbation Theory (SAPT0). We selected representative conformations of *p*-toluidine (BG5) and amantadine (AG7) representing both groups and included the coordinating water molecules. Figure 3 shows that exchange interactions in the bulkier amantadine guest with an adamantane core is larger compared to the guest with aromatic guest BG5. Amantadine also presents a larger contribution of attractive dispersion interaction compared to the favorable electrostatic interaction (dispersion(D)/electrostatic(E) ratio: 0.36(BG5) vs 0.68(AG7)). Water molecules have a significant participation in the interaction energy of BG5, possibly because its smaller size allows better access to the top part of CB7. The adamantane core in the bulky amantadine (AG7) contributes 26% to the total interaction energy; whereas the benzene ring contributes only 18%, and the methyl group is slightly repulsive, not contributing to the binding energy.

If the origin of the large deviation in the binding free energy of guests with an adamantane core corresponds to incorrectly described repulsive and dispersion interactions, these types of interactions should dominate in the guests with large deviations in binding affinity. The calculated binding free energy of memantine (AG1) possessing two methyl groups deviates by 4.1 kcal/mol from the measured value with the D-

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

Table 3. Measured Absolute Binding Free Energies for Guests with an Adamantane Core (AG1-8) (exp) and Calculated Values Using RESP Atomic Charges for the CB7 Host and AM1-BCC for the Guest with GAFF Lennard-Jones (LJ) Parameters (q[AM1-BCC]) or LJ Parameters Derived with MBIS Partitioning of the Polarized Electron Density of the Guest Combined with D-MBIS Atomic Charges Derived for the Guest in the Bound and Unbound States (LJ-q[D-MBIS]) or Using Only Atomic Charges from the Unbound State in the Aqueous Solution (LJ-q[D-MBIS(solv)])[a]

system	ΔG°_{exp}	$\Delta G^{\circ}_{q[AM1\text{-}BCC]}$	$\Delta G^{\circ}_{LJ\text{-}q[D\text{-}MBIS]}$	$\Delta G^{\circ}_{LJ\text{-}q[D\text{-}MBIS(solv)]}$
CB7-AG1	−6.0 ± 0.1	−6.8 ± 0.3	−6.3 ± 0.3	−4.0 ± 0.6
CB7-AG2	−6.6 ± 0.1	−3.0 ± 0.2	−2.5 ± 0.4	0.8 ± 0.4
CB7-AG3	−18.2 ± 0.1	−19.4 ± 0.7	−16.3 ± 0.4	−20.5 ± 0.3
CB7-AG4	−11.3 ± 0.1	−4.2 ± 0.5	−9.4 ± 0.5	−10.5 ± 0.5
CB7-AG5	−11.6 ± 0.1	−17.0 ± 0.5	−14.0 ± 0.5	−9.5 ± 0.5
CB7-AG6	−16.7 ± 0.1	−22.4 ± 0.2	−19.2 ± 0.3	−14.0 ± 0.2
CB7-AG7	−17.2 ± 0.1	−20.6 ± 0.4	−18.1 ± 0.3	−18.6 ± 0.3
CB7-AG8	−16.8 ± 0.1	−22.4 ± 0.4	−19.5 ± 0.3	−15.0 ± 0.3
RMSE		4.6 ± 0.6	2.4 ± 0.3	3.2 ± 0.9
MAE		4.1 ± 0.7	2.1 ± 0.4	2.5 ± 0.6
R		0.9 ± 0.1	0.9 ± 0.1	0.9 ± 0.0
τ		0.5 ± 0.3	0.6 ± 0.2	0.9 ± 0.1

[a]Values are in kcal/mol.

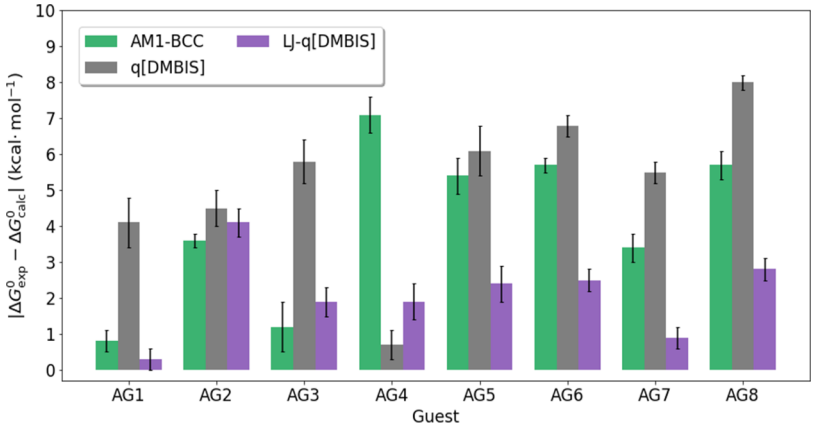


Figure 4. Absolute error in binding affinity for guests with an adamantane core, using different atomic charges and Lennard-Jones (LJ) parameters, AM1-BCC charges/GAFF LJ (bars in green), D-MBIS charges/GAFF LJ (bars in gray), and D-MBIS charges/D-MBIS LJ (bars in violet). Bonded force field parameters were taken for all guests and hosts from the GAFF force field.

MBIS charge set. The composition of the total interaction energy for this guest in Figure 3 shows a significant larger repulsive exchange contribution compared to AG7 combined with a larger contribution of dispersion compared to the electrostatic attractive interaction (dispersion(D)/electrostatic(E) ratio: 0.80(AG1) vs 0.63(AG7)). The large repulsive contribution originates from the overlap of the guest's and the host's electron density pointing to unfavorable steric interactions. The two methyl groups in menantine (AG1) contribute significantly to this repulsive interaction as evidenced by the F-SAPT partitioning of the total interaction by groups displayed at the bottom of Figure 3. *N*-(1-Admantyl)pyridinium (AG8) also presents a large deviation in the binding free energy of 7.9 kcal/mol. Compared to amantadine (AG7) this guest possesses a positively charged pyridine group instead of the ammonium group. Interestingly, it reveals a similar repulsive interaction as memantine (AG7) but a larger contribution of dispersion to the attractive interactions (dispersion(D)/electrostatic(E) ratio: 0.86(AG8) vs 0.80(AG1); Figure S10 in the Supporting Information). The increased dispersion interaction almost equals the electrostatic one for this guest and stems from the partial aromatic character of the pyridine group. One possible explanation of the large deviations in the obtained binding free energy might be an overestimation of this dispersion interaction by too large C_6 coefficients of the employed GAFF force field, as reported recently by Rowley.[70,71]

From the SAPT analysis, we conclude that guests with an adamantane core have a large contribution of the repulsive exchange interaction and a considerable attractive dispersion interaction. The guests AG1 and AG8 represent limiting cases where either the repulsion or dispersion interaction governs the interactions compared to AG7. Based on this analysis, improving parameters in nonpolarizable force fields that represent the repulsive exchange and attractive dispersion interaction could lead to an improvement in the binding affinity prediction for guests with an adamantane core.

Lennard-Jones Parameters Derived from MBIS Partitioning of the Electron Density Improve Binding Affinity Predictions in Guests with an Adamantane Core. Based on the large contribution of repulsive exchange and dispersion interactions of guests with an adamantane core described above, we derived Lennard-Jones parameters for these guests partitioning their electron density. When we partition the molecular electron density in atomic contributions, an effective volume of an atom in a molecule is defined.

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

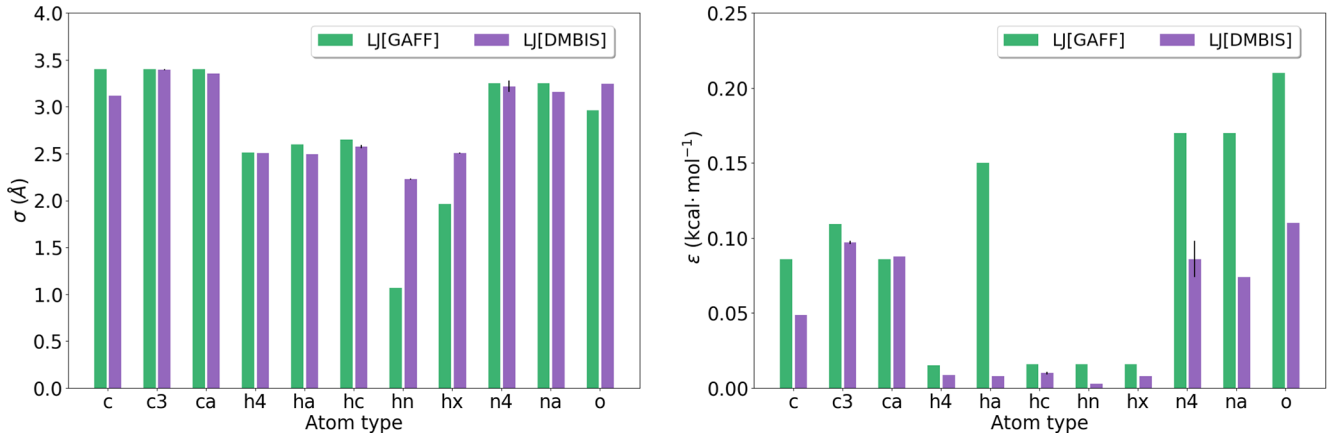


Figure 5. Lennard-Jones parameters derived from MBIS partitioning of the polarized electron density averaged for the atom types of the guests with an adamantane core. The MBIS derived radius at the minimum of the potential is compared to the values in the GAFF force field for different atom types on the left and the MBIS derived depth of the potential at the right.

With the effective volume of the atom in the molecule, C_6 dispersion coefficients are calculated by rescaling tabulated coefficients of free atoms[57] with the ratio of the volume of the atom in the molecule and the free atom as described in the Methods section. The van der Waals radius, where the Lennard-Jones potential becomes zero, is obtained from the recent relation of polarizability and volume proposed by Tkatchenko.[58] We derived the two parameters of the Lennard-Jones potential for all atoms in each guest with our computational workflow considering various configurations in the aqueous solution and averaged the coefficients for each atom type in the GAFF force field. With the obtained Lennard-Jones parameters, we calculated a new set of D-MBIS atomic charges for each guest and repeated the free energy calculations using the new set of nonbonded force field parameters. Interestingly, the fourth column of Table 3 shows that the D-MBIS binding affinities for the 8 adamantane core guests are in better agreement with the measured values, reducing the RMSE and MAE by a factor of 2, maintaining a Pearson correlation coefficient R of 0.9, and an almost equal value of Kendall τ.

For most guests, MBIS derived force field parameters reproduce binding affinities better than the ones using D-MBIS atomic charges with GAFF Lennard-Jones parameters, except for the adamantane-1,3-diamine guest (AG4) as shown in Figure 4. Adamantane-1,3-diamine possesses two charged ammonium groups at the adamantane core, where changes in the protonation state might occur depending on the two groups' solvent exposure; but as mentioned above, to include protonation state changes in the simulations is out of the scope of this work.

Before we can conclude that nonbonded force field parameters from MBIS partitioning improve the description of the interactions, we have to compare the conformations adopted by the guests for different force field parameters. Our conformational analysis in Figure S9 and Tables S8−S10 in the Supporting Information shows that for most guests the relative weight of the different conformations is maintained for different charge sets and Lennard-Jones parameters. The parameters do, however, alter the radial distribution function of water molecules as shown in Figure S11 of the Supporting Information. Introducing the Lennard-Jones parameters reduces the number of molecules in the first solvation shell of the guest because of the increased repulsive interaction and the presence of the host atoms at this distance. At the same time, it increases the probability of finding water molecules at distances larger than 2.5 Å.

To rationalize these observations, we compared the Lennard-Jones parameters of the GAFF force field with the ones obtained with MBIS partitioning. In Figure 5, the van der Waals radii σ of the hydrogen atom types *hn* and *hx* bound to the nitrogen atom of the ammonium group increase significantly with the MBIS partitioning method and explain the observed difference in the radial distribution function of the solvent. For the other atom types, changes in σ are minor, except for the oxygen atoms and carbon atom (*c* and *o*) in the carboxyl group.

As already reported by the Rowley group,[70] Figure 5 shows that the depth of the potential, ϵ, obtained with our method is smaller than the one from GAFF. Differences are more pronounced for electronegative atom types *n4*, *na*, and *o*. In general, hydrogen atoms present smaller MBIS-derived ϵ values, especially for *ha*. One possible explanation is that these parameters were increased in the GAFF force field to reproduce thermodynamic properties in the empirical fitting of parameters.

In general, we conclude that MBIS partitioning of the polarized electron density provides atomic charges and Lennard-Jones parameters which improve binding affinity predictions for this type of host−guest systems.

Finally, to reduce the computational cost associated with the parameter derivation, we employ the same D-MBIS atomic charge set and Lennard-Jones parameters of the aqueous solution for the bound and unbound states and neglect the contribution of differences in the polarization energy. With our current workflow, the derivation of parameters only in the aqueous solution reduces the total computational time by one-third (see Table S12 in the Supporting Information). The obtained binding free energies present a larger deviation from the measured values as shown in the last column of Table 3, but they still show good correlation and correct ordering of weak to strong binders. Only AG5 presents a large deviation because of its anionic character and the significant difference between the polarization energy in the two molecular environments, in agreement with our earlier results.[19] The pK_a of AG5 is very close to the experimental pH value, and deviations in the calculated binding free energy from the experimental value could also arise because of the contribution

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

of different protonation states of the guest and a possible protonation change upon binding.

This reduced computational cost methodology for nonbonded force field parameter derivation with our workflow written in Python would only require the electronic structure calculation in the aqueous solution for cationic or neutral ligands irrespective of the host. In the future, we will study if it also leads to better binding affinity prediction to other hosts or proteins. This will open the possibility to apply our methodology in drug affinity prediction projects.

CONCLUSIONS AND OUTLOOK

A new methodology is proposed to obtain nonbonded force field parameters from partitioning the molecular electron density with the MBIS method from various configurations of the guest, which improves binding affinity predictions for CB7 host−guest systems. Guests with an adamantane core have a large contribution of repulsive exchange and attractive dispersion interactions compared to the smaller guests with an aromatic core as revealed by an F-SAPT0 analysis. These interactions are better described by our modified Lennard-Jones parameters. Applying our Lennard-Jones parameters to the guest with an aromatic core did not reduce the errors, but as for the guests with an adamantane core, they discriminate correctly between strong and weak binders (see Table S11 in the Supporting Information), which is crucial in drug discovery projects. Although significant improvement in the binding affinity prediction accuracy has been achieved, a root-mean-square error of approximately 2 kcal/mol remains. For guests BG2 and AG5 with pK_a values close to the experimental pH, different protonation states could be involved in the binding to CB7, which were not considered in this work. The coexistence of multiple protonation states and their changes upon binding might explain part of the observed deviations from the experimental binding affinities.

In the future, we plan to study binding to more complex protein systems with our established computational workflow. For cationic or neutral ligands, we will explore the computationally less expensive alternative to our method, which does not include the polarization energy correction and requires electronic structure calculations only once in the aqueous solution for each ligand.

DATA AND SOFTWARE AVAILABILITY

OpenMM and Yank are open source software and can be used free of charge. The simulation workflow written in Python can be obtained from the following Github repository:https://github.com/QCMM/ffparaim_amber. Initial coordinates and topologies are available under the following DOI: DOI: 10.5281/zenodo.6359479. ORCA 4.2.1 is free for academic use open request, and Horton3 is free open source package. All figures were created with matplotlib.

ASSOCIATED CONTENT

Supporting Information

. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00316.

pK_a values of all titratable groups for all guests, thermodynamic cycle of double-decoupling scheme used to calculate absolute binding free energy, conformational analysis of guests described with different nonbonded force field parameters, and computational time required to calculate Lennard-Jones parameters and D-MBIS atomic charges (PDF)

AUTHOR INFORMATION

Corresponding Author

Esteban Vöhringer-Martinez − *Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, 4070386 Concepción, Chile;* orcid.org/0000-0003-1785-4558; Phone: +56 41 2204986; Email: evohringer@udec.cl

Authors

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

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

Carlos Castillo-Orellana − *Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, 4070386 Concepción, Chile;* orcid.org/0000-0003-2703-1678

Toon Verstraelen − *Center for Molecular Modeling (CMM), Ghent University, B-9052 Ghent, Belgium;* orcid.org/0000-0001-9288-5608

Stefan Vogt-Geisse − *Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, 4070386 Concepción, Chile*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00316

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are thankful for the financial support from FONDECYT 1200369. T.V. acknowledges the Research Board of Ghent University (BOF) for its financial support.

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) 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.

(6) 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.

(7) 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.

(8) Cao, L.; Šekutor, M.; Zavalij, P. Y.; Mlinarić-Majerski, K.; Glaser, R.; Isaacs, L. Cucurbit[7]uril Guest Pair with an Attomolar Dissociation Constant. *Angew. Chem., Int. Ed.* **2014**, *53*, 988−993.
(9) Assaf, K. I.; Nau, W. M. Cucurbiturils: from synthesis to high-affinity binding and catalysis. *Chem. Soc. Rev.* **2015**, *44*, 394−418.
(10) Liu, S.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems. *J. Am. Chem. Soc.* **2005**, *127*, 15959−15967.
(11) Mock, W. L.; Shih, N. Y. Organic ligand-receptor interactions between cucurbituril and alkylammonium ions. *J. Am. Chem. Soc.* **1988**, *110*, 4706−4710.
(12) Mock, W. L.; Shih, N. Y. Structure and selectivity in host-guest complexes of cucurbituril. *J. Org. Chem.* **1986**, *51*, 4440−4446.
(13) Whang, D.; Heo, J.; Park, J. H. A Molecular Bowl with Metal Ion as Bottom: Reversible Inclusion of Organic Molecules in Cesium Ion Complexed Cucurbituril. *Angew. Chem.* **1998**, *37*, 78−80.
(14) Moghaddam, S.; Inoue, Y.; Gilson, M. K. Host-guest complexes with protein-ligand-like affinities: computational analysis and design. *J. Am. Chem. Soc.* **2009**, *131*, 4012−4021.
(15) Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L.; Chen, W.; Moghaddam, S.; Gilson, M. K.; Kim, K.; Inoue, Y. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy−entropy compensation. *Proc. Natl. Acad. Sci. U. S. A* **2007**, *104*, 20737−20.
(16) Nau, W. M.; Florea, M.; Assaf, K. I. Deep Inside Cucurbiturils: Physical Properties and Volumes of their Inner Cavity Determine the Hydrophobic Driving Force for Host-Guest Complexation. *Isr. J. Chem.* **2011**, *51*, 559−577.
(17) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **2011**, *7*, 625−632.
(18) Parrish, R. M.; Parker, T. M.; Sherrill, C. D. Chemical Assignment of Symmetry-Adapted Perturbation Theory Interaction Energy Components: The Functional-Group SAPT Partition. *J. Chem. Theory Comput.* **2014**, *10*, 4417−4431.
(19) González, D.; Macaya, L.; Vöhringer-Martinez, E. Molecular Environment-Specific Atomic Charges Improve Binding Affinity Predictions of SAMPL5 Host−Guest Systems. *J. Chem. Inf. Model.* **2021**, *61*, 4462−4474.
(20) 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.
(21) 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.
(22) 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.
(23) 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.
(24) 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.
(25) 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.
(26) Schauperl, M.; Nerenberg, P. S.; Jang, H.; Wang, L.-P.; Bayly, C. I.; Mobley, D. L.; Gilson, M. K. Non-bonded force field model with advanced restrained electrostatic potential charges (RESP2). *Commun. Chem.* **2020**, *3*, 44.
(27) Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model. *J. Chem. Theory Comput.* **2017**, *13*, 6146−6157.
(28) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312−2323.
(29) Kantonen, S. M.; Muddana, H. S.; Schauperl, M.; Henriksen, N. M.; Wang, L.-P.; Gilson, M. K. Data-Driven Mapping of Gas-Phase Quantum Calculations to General Force Field Lennard-Jones Parameters. *J. Chem. Theory Comput.* **2020**, *16*, 1115−1127.
(30) Visscher, K. M.; Geerke, D. P. Deriving Force-Field Parameters from First Principles Using a Polarizable and Higher Order Dispersion Model. *J. Chem. Theory Comput.* **2019**, *15*, 1875−1883.
(31) Fedorov, D. V.; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. Quantum-Mechanical Relation between Atomic Dipole Polarizability and the van der Waals Radius. *Phys. Rev. Lett.* **2018**, *121*, 183401.
(32) 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.
(33) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs. Comput. Mol. Sc* **2018**, *8*, e1327.
(34) Verstraelen, T.; Adams, W.; Pujal, L.; Tehrani, A.; Kelly, B. D.; Macaya, L.; Meng, F.; Richer, M.; Hernández-Esparza, R.; Yang, X. D.; Chan, M.; Kim, T. D.; Cools-Ceuppens, M.; Chuiko, V.; Vöhringer-Martinez, E.; Ayers, P. W.; Heidar-Zadeh, F. IOData: A python library for reading, writing, and converting computational chemistry file formats and generating input files. *J. Comput. Chem.* **2021**, *42*, 458−464.
(35) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Grinaway, P.; Rustenburg, B.; Albanese, S.; Saladi, S. choderalab/yank: 0.20.1 - Exact treatment of PME electrostatics and optimizations. *Zenodo*; 2018. https://zenodo.org/record/1161274#.YvBrd3bMLIU (accessed 2022-08-07).
(36) Mobley, D. L.; Slochower, D. MobleyLab/benchmarksets: Version 1.2. *Zenodo*; 2017. https://zenodo.org/record/839047/export/hx#.YvBrtXbMLIU (accessed 2022-08-07).
(37) McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578−596.
(38) McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 897−906.
(39) Kelley, B. P.; Brown, S. P.; Warren, G. L.; Muchmore, S. W. POSIT: Flexible Shape-Guided Docking For Pose Prediction. *J. Chem. Inf. Model.* **2015**, *55*, 1771−1780.
(40) 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.
(41) 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.
(42) 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.
(43) 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.
(44) 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.
(45) 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.

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174

(46) 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.
(47) 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.
(48) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **1989**, *45*, 503−528.
(49) Leimkuhler, B.; Matthews, C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Appl. Math. Res. Express* **2013**, *2013*, 34−56.
(50) 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. https://zenodo.org/record/1161149#.YvBq9XbMLIU (accessed 2022-08-07).
(51) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(52) 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.
(53) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.
(54) 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.
(55) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction: high-order dispersion coefficients. *J. Chem. Phys.* **2006**, *124*, 014104.
(56) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(57) Chu, X.; Dalgarno, A. Linear response time-dependent density functional theory for van der Waals coefficients. *J. Chem. Phys.* **2004**, *121*, 4083−4088.
(58) Tkatchenko, A.; Fedorov, D. V.; Gori, M. Fine-Structure Constant Connects Electronic Polarizability and Geometric van-der-Waals Radius of Atoms. *J. Phys. Chem. Lett.* **2021**, *12*, 9488−9492.
(59) Tan, Z. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. *J. Comput. Graph. Stat.* **2017**, *26*, 54−65.
(60) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* **1992**, *96*, 1776−1783.
(61) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(62) Liu, J. S. *Monte Carlo Strategies in Scientific Computing*, 2nd ed.; Springer-Verlag: New York, 2002.
(63) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799−1805.
(64) 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.
(65) Beauchamp, K.; Chodera, J.; Naden, L.; Shirts, M.; Martiniani, S.; Stern, C.; McGibbon, R. T.; Gowers, R.; Barnett, J. W.; Dotson, D. choderalab/pymbar: Fix PyPi Release. *Zenodo*; 2017. https://zenodo.org/record/886537#.YvBt2HbMLIU (accessed 2022-08-07).
(66) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *WIREs. Comput. Mol. Sci.* **2012**, *2*, 254−272.
(67) Kodrycka, M.; Holzer, C.; Klopper, W.; Patkowski, K. Explicitly Correlated Dispersion and Exchange Dispersion Energies in Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2019**, *15*, 5965−5986.
(68) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(69) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.
(70) Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model. *J. Chem. Theory Comput.* **2017**, *13*, 6146−6157.
(71) Walters, E. T.; Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-Hole Dipole Moment Model. *J. Phys. Chem. B* **2018**, *122*, 6690−6701.

https://doi.org/10.1021/acs.jcim.2c00316
J. Chem. Inf. Model. 2022, 62, 4162−4174