# Ligand Affinities within the Open-Boundary Molecular Mechanics/Coarse-Grained Framework (I): Alchemical Transformations within the Hamiltonian Adaptive Resolution Scheme

Ksenia Korshunova[†,‡] and Paolo Carloni[*,†,‡]

†*Department of Physics, RWTH Aachen University, 52074 Aachen, Germany*

‡*Computational Biomedicine, Institute of Advanced Simulations IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, 52428 Jülich, Germany*

E-mail: p.carloni@fz-juelich.de

**Abstract**

Our recently developed Open-Boundary Molecular Mechanics/Coarse Grained (OB-MM/CG) framework predicts ligand poses in important pharmaceutical targets, such as G-protein Coupled Receptors, even when experimental structural information is lacking. The approach, which is based on GROMOS and AMBER force fields, allows for grand-canonical simulations of protein-ligand complexes by using the Hamiltonian Adaptive Resolution Scheme (H-AdResS) for the solvent. Here, we present a key step toward the estimation of ligand binding affinities for their targets within this approach. This is the implementation of the H-AdResS in the GROMACS code. The accuracy of our implementation is established by calculating hydration free energies of several molecules in water by means of alchemical transformations. The deviations of the GROMOS- and AMBER-based H-AdResS results from the reference fully-atomistic simulations are smaller than the accuracy of the force field and/or they are in the range of the published results. Importantly, our predictions are in good agreement with experimental data. The current implementation paves the way to the use of the OB-MM/CG framework for the study of large biological systems.

# Introduction

Human G-Protein Coupled Receptors (hGPCRs) are a superfamily of transmembrane proteins fundamental for cell signaling.[1] They are immensely important from a pharmaceutical perspective: as many as 34% of Food & Drug Administration (FDA) drugs target hGPCRs.[2] The structure-based drug design for this superfamily is unfortunately hampered by the lack of experimental structures (available for only 8% of 800 hGPCRs).[3] Thus, for most hGPCRs, molecular modeling is currently the only method to investigate ligand binding. However, the quality of structural models largely depends on the degree of sequence identity between the template and the target protein. Atomistic molecular dynamics (MD) simulations may improve the predictions.[4–9] Thanks to highly detailed representation of the protein-ligand com-

plex as well as its environment (lipid molecules comprising the membrane, water molecules, ions), MD simulations serve as a sophisticated and capable tool to examine protein-ligand interactions and to compute ligand-binding affinity. However, in the fairly common cases where the sequence identity between the target protein and the template is low, caution is advised: incorrectly predicted side chain orientations might cause long structure relaxation time, prevent proper equilibration of the system, and in some cases, lead to finding a wrong free energy minimum structure.

To face these challenges, our lab developed a GROMACS-based[10–13] multiresolution hybrid scheme, the so-called Open-Boundary Molecular Mechanics/Coarse-Grained framework (OB-MM/CG).[14,15] The core idea is to represent the region of interest (in this case, the ligand-binding site including the solvent molecules) on the highly detailed, atomistic level of resolution, while the description of the rest of the system is either coarse-grained (protein outside the binding site, solvent bulk) or implicit (membrane). Partial protein coarse-graining reduces the source of error caused by wrong side chain orientations often introduced in low-resolution models. A simplified implicit membrane reduces the computational cost of the simulation while preserving the most essential membrane functions: it prevents water from diffusing the hydrophobic region of the protein and keeps the protein fully folded.[16] In the latest version of our code, the solvent is treated using the Hamiltonian Adaptive Resolution Scheme (H-AdResS).[17] This approach, developed by Potestio et al.,[17] reduces the number of degrees of freedom of the water solvent by coupling a small atomistic region to a large coarse-grained reservoir. This results in a simulation of the grand-canonical ensemble in the atomistic region and allows, in principle, for rigorous calculations of ligand-binding free energies.[18] The current OB-MM/CG code was validated on a well-studied GPCR-ligand complex. The structural and dynamical properties, including the ligand poses, were shown to be in good agreement with the reference fully-atomistic simulations.[15]

Keeping the pharmacological relevance of the hGPCRs in mind, the next step of the OB-MM/CG framework's development consists in implementing predictions of ligand-binding

affinities, the key parameters for drug design. We address this issue using a multistep strategy: (i) Selecting a specific free energy method. Here, we chose alchemical transformations for a variety of reasons. First, it enables ranking and filtering the most promising drug molecule candidates in order to narrow down the molecule set for later stages of drug design.[19] Second, calculating relative — as opposed to absolute — binding free energies among a set of drug candidates cancels out the errors arising due to a chosen force field. These errors can be especially prominent in the case of drug-like molecules due to the difficulty of their parametrization.[20] Finally, these calculations are very much apt for drug screening as they are easy to implement and less CPU-demanding than other methods.[1] In combination with the staging approach chosen for sampling, the alchemical transformations can be parallelized in a straightforward fashion suitable for high-performance computing. This enables application for hGPCRs in the future. (ii) Implementing and validating alchemical transformations in the GROMACS-based H-AdResS code. To the best of our knowledge, this is the first time that such an implementation has been attempted.[2] (iii) Implementing the modified H-AdResS algorithm with enabled alchemical transformation into our OB-MM/CG code. This last step, currently in progress, will be presented in a subsequent paper.

The accuracy of our implementation is established by comparing the H-AdResS-based predictions of hydration free energies (HFEs) for a selected set of molecules against values obtained via standard fully atomistic MD calculations and from experiment. Our calculations are based on the force fields currently implemented in the OB-MM/CG framework,[26] namely, the GROMOS force field[27] and the General AMBER Force Field (GAFF).[28] We focus on neutral analogues of amino acid side chains (except glycine and proline[29–31]), commonly used in force-field parametrization.[32] They exhibit different size, hydrophobicity, and charge distribution and include common functional groups (alkanes, alcohols, amides, etc.). This

[1]However, alchemical transformation provides neither mechanisms nor kinetic constants of ligand binding/unbinding, which can be addressed by approaches such as well-tempered metadynamics[21,22] or Markov State Models[23]

[2]So far, hydration free energies have been calculated only using the AdResS approach.[24,25] Although both schemes dynamically couple regions of different resolution via a hybrid region, AdResS uses force interpolation while H-AdResS employs interpolation between potential energy functions.

allows us to assess the accuracy of our predictions for cases of varying difficulty.

This paper is organized as follows: the brief overview of the H-AdResS theory and the computational details are reported in the Theory and Methods Section. The Results and Discussion Section presents the HFE results obtained using the H-AdResS setup and compares them with the previously published calculated and experimental HFE values. The Conclusions Section provides final remarks on the present work and the outlook for the future development of the setup.

# Theory and Methods

## Hamiltonian Adaptive Resolution Scheme (H-AdResS)

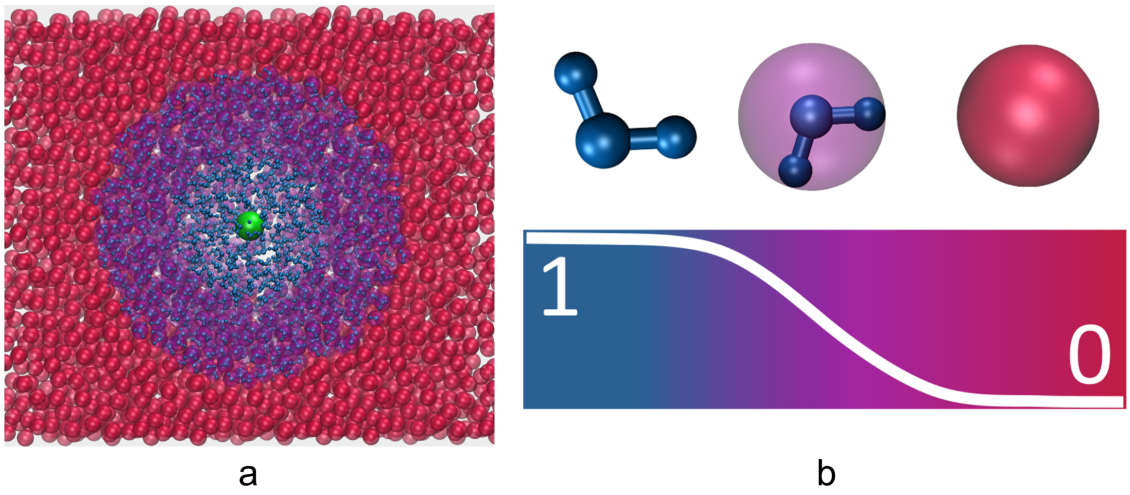
**Short review of the method.** The H-AdResS[17] represents the water solvent at different levels of resolution simultaneously in a single simulation. The levels of resolution — in our case, atomistic and coarse-grained — are assigned to subregions in space and smoothly coupled via a transitional hybrid region. This allows the water molecules to change their resolution on the fly as they freely move through the hybrid region. In our setup, the solute molecule is placed in the center of a spherical atomistic ($AT$) region, surrounded by a hybrid ($Hy$) shell, which is coupled to a coarse-grained ($CG$) region (Figure 1a).

The total Hamiltonian for the system of a solute molecule and N water molecules reads:

$$\mathcal{H} = K^{tot} + V^{sol/sol-w} + \sum_{\alpha}^{N} \left[\lambda_\alpha V_\alpha^{AT} + (1-\lambda_\alpha) V_\alpha^{CG}\right] - \sum_{\alpha}^{N} \Delta\mathcal{H}(\lambda_\alpha) \tag{1}$$

Here, $K^{tot}$ is the total kinetic energy of the entire system (solute plus solvent), and $V^{sol/sol-w}$ is the potential energy of the intra- and intermolecular interactions of the solute. The third term describes the potential energy of intermolecular solvent-solvent interactions of N water molecules (labeled $\alpha$) on the atomistic ($V^{AT}$) and the coarse-grained ($V^{CG}$) levels of resolution. The resulting potential energy of a molecule $\alpha$ is a sum weighted by

Figure 1: **a.** H-AdResS setup: a solute molecule (green) is located in the center of the spherical $AT$ region (blue), surrounded by a spherical $Hy$ shell (purple), with the rest of the solvent in the $CG$ region represented by CG beads (red). **b.** The switching function $\lambda_\alpha$ responsible for the transition between the $AT$ and $CG$ regions smoothly varies between 1 and 0 in the $Hy$ region. The atomistic water molecule is mapped to a single CG virtual site positioned at the molecule's CoM.




the switching function $\lambda_\alpha = \lambda_\alpha(\mathbf{R}_\alpha)$. The latter depends on the position of the molecule's center of mass (CoM) and is defined to be 1 in the $AT$ region, 0 in the $CG$ region, and $1{>}\lambda_\alpha{>}0$ in the $Hy$ region, where it smoothly changes its value. This results in a linearly interpolated representation of the water molecule transitioning between the $AT$ and $CG$ regions (Figure 1b). In the present setup, $V^{AT}$ is the SPC/E model,[33] which describes water molecules as rigid; therefore, there is no intramolecular term for the solvent in Eq. 1. $V^{CG}$ is a Lennard-Jones potential derived to match the density of the SPC/E model and applied to a single CG virtual site positioned at the water molecule's CoM.

The last term in the Hamiltonian (Eq. 1) is the so-called compensation term $\Delta\mathcal{H}(\lambda_\alpha)$, which is a time-independent function of the molecule's position. The pairwise CG potential used in practice is an approximation of a many-body CG potential that would be able to reproduce the properties of an atomistic system exactly. The latter, however, is very difficult and time-consuming to obtain.[34] Although the pairwise CG potential does reproduce the chosen properties of the $AT$ potential (here: density of water) well, the chemical potentials of the $AT$ and $CG$ regions differ, which leads to the net flux of water molecules in the $Hy$

region in the absence of the compensation term. This term can be constructed by imposing either the same pressure or density in the $AT$ and $CG$ regions.[17,18] Since the $V^{CG}$ potential was derived based on the density of the atomistic water, the latter method is used. In this case, $\Delta\mathcal{H}(\lambda_\alpha)$ reads:[17]

$$\Delta\mathcal{H}(\lambda_\alpha) = \frac{\Delta F(\lambda_\alpha)}{N} + \frac{\Delta p(\lambda_\alpha)}{\rho} \equiv \Delta\mu(\lambda_\alpha) \tag{2}$$

Here, the Helmholz free energy difference is $\Delta F(\lambda_\alpha) = F(\lambda_\alpha) - F(\lambda = 0)$, the pressure difference is $\Delta p(\lambda_\alpha) = p(\lambda_\alpha) - p(\lambda = 0)$, and $\rho$ is the reference water density. The resulting compensation is equivalent to the change in chemical potential $\Delta\mu(\lambda_\alpha)$ across the $Hy$ region.[17] We also note that this setup results in a grand-canonical description of the $AT$ region, since $\mu^{AT}$ and $\mu^{CG}$ are preserved and the water molecules are freely exchanged across the $Hy$ region.[35] (more details in the SI).

**Practical implementation.** In the present work, the previously existing H-AdResS implementation was modified in order to enable the alchemical free energy calculations by integrating the H-AdResS-related functions into the free energy sampling routine of GROMACS 4.5.[10–13]

## Computational Details

The solute molecules reported in Table 1 are all the neutral amino acid side chain analogues including the $N\delta$- and $N\epsilon$-protonated tautomers of histidine, namely, Hid and Hie (glycine and proline were not considered, as mentioned in the Introduction). The resulting 15 molecules were parametrized in two ways. In the first case, the topology and coordinate files were taken directly from the FreeSolv database,[36] where they were parametrized using the GAFF[28] force field with AM1-BCC[37,38] charges.[3] In the second case, the topology and coordinate files were generated using the PRODRG server,[42] with the RESP[39–41] charges

[3]With the exception of histidine analogues, for which B3LYP/6-31G** RESP[39–41] charges were calculated.

calculated at the B3LYP/6-31G** level using the Gaussian09 package.[43]

**Table 1: Molecules undergoing HFE calculations in this study. The test set consists of nine nonpolar and four polar amino acid side chain analogues, along with the two neutral tautomers of histidine. The side chain analogue of cysteine (methanethiol) was replaced with ethanethiol, since the server for one of the force fields (PRODRG[42] for GROMOS[27]) does not allow input topologies with only two heavy atoms such as methanethiol.**

| Molecule | Residue |
|---|---|
| methane | Ala (alanine) |
| propane | Val (valine) |
| isobutane | Leu (leucine) |
| n-butane | Ile (isoleucine) |
| toluene | Phe (phenylalanine) |
| 3-methyl-1H-indole | Trp (tryptophan) |
| 5-methylimidazole | Hid (histidine) |
| 4-methylimidazole | Hie (histidine) |
| methylsulfanylethane | Met (methionine) |
| methanol | Ser (serine) |
| ethanol | Thr (threonine) |
| acetamide | Asn (asparagine) |
| propionamide | Gln (glutamine) |
| p-cresol | Tyr (tyrosine) |
| ethanethiol | "Cys" (cysteine) |

For the solvent, either the fully atomistic (using $V^{AT}$ in Eq. 1) or the hybrid (using $V^{AT}$ and $V^{CG}$) descriptions were employed, along with an effectively fully atomistic H-AdResS setup (i.e. $\lambda_\alpha = 1$ for $\alpha = 1, .., N$). The latter was used to check for possible errors associated with the implementation of the modified H-AdResS code (see the SI). Overall, this led to six different setups reported in Table 2.

The SPC/E model[33] was used for $V^{AT}$. Although TIP3P[44] and SPC[45] are the "native" water models for AMBER/GAFF and GROMOS, respectively, SPC/E was tested for HFE calculations and is compatible with both force fields.[46,47] The $V^{CG}$ Lennard-Jones potential was derived in a separate simulation of a water box using the iterative Boltzmann inversion to match the density of the SPC/E model.[48]

Each system consisted of one of the solute molecules in Table 1, placed in the center of the cubic box and solvated with water molecules using the `genbox` tool in GROMACS 4.5.[49] Peri-

**Table 2: Approaches to calculate HFE used in this study. Notice that, in all cases, SPC/E[33] model was used for the atomistic description of water.**

| Abbreviation | Solute force field | Water description | GROMACS code version |
|---|---|---|---|
| AMB-ref[a] | GAFF[28] with AM1-BCC[37,38] | atomistic | 4.5[10–13] |
| GRO-ref | GROMOS[27] with RESP[39–41] | atomistic | 4.5[10–13] |
| AMB-HAd[a] | GAFF[28] with AM1-BCC[37,38] | hybrid, $\lambda_\alpha = \lambda_\alpha(\mathbf{R}_\alpha)$ | H-AdResS[c] |
| GRO-HAd | GROMOS[27] with RESP[39–41] | hybrid, $\lambda_\alpha = \lambda_\alpha(\mathbf{R}_\alpha)$ | H-AdResS[c] |
| AMB-HAd-AT[a,b] | GAFF[28] with AM1-BCC[37,38] | hybrid, $\lambda_\alpha = 1$ | H-AdResS[c] |
| GRO-HAd-AT[b] | GROMOS[27] with RESP[39–41] | hybrid, $\lambda_\alpha = 1$ | H-AdResS[c] |

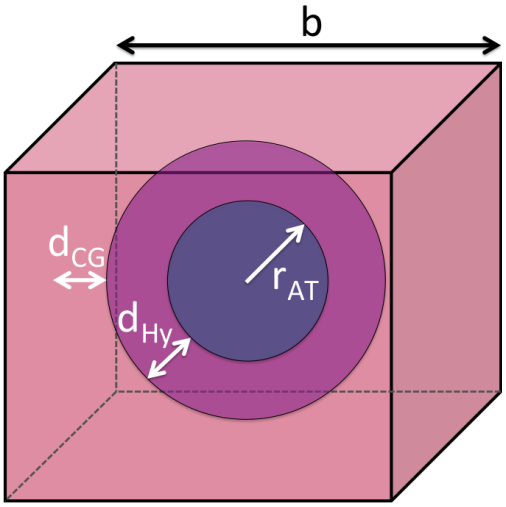
[a] For the sake of clarity, the names of the approaches are derived from the names of their general force fields (AMBER or GROMOS) instead of the names of small-molecule parameter sets;
[b] The results of these setups are reported in Table 1 of the SI;
[c] In-house version of GROMACS 4.5 with modified H-AdResS.

odic boundary conditions in all three dimensions were applied. Harmonic position restraints were applied on the solute molecule to prevent it from drifting from the center of the $AT$ region. The side length of the cubic box $b$ was calculated as follows: $b = 2*(r_{AT}+d_{Hy}+d_{CG})$, where $r_{AT}$ is the radius of the atomistic region, and $d_{Hy}$ and $d_{CG}$ are the widths of the $Hy$ and $CG$ regions, respectively (Figure 2). The radius of the $AT$ region was calculated based on the molecule size with the margin of 1.2 nm. The latter corresponds to the cutoff radii for the nonbonded interactions (see the next paragraph) to ensure that the sampling of the microstates occurs strictly within the $AT$ region. The widths of the $Hy$ and $CG$ regions are system-independent: $d_{Hy} = 1.2$ nm,[14] $d_{CG} = 0.8$ nm.

A cutoff radius of 1.2 nm was used for both electrostatic and van der Waals interactions. Long range electrostatic interactions were treated using the Reaction Field[50–52] method. As for van der Waals interactions, tabulated user-defined potential functions for repulsion and dispersion were used group-wise, with separate tables for the atomistic and coarse-grained representations, respectively. In the case of the fully atomistic reference simulations, the tabulated potential for atomistic interactions was used for the sake of consistency.

Free energy differences were estimated using the staging method with multiple indepen-

Figure 2: **Geometry of the simulation box.** $b$ is the side length of the box, $r_{AT}$ is the radius of the $AT$ region, $d_{Hy} = 1.2$ nm is the width of the $Hy$ shell, and $d_{CG} = 0.8$ nm is the distance from the outer radius of the $Hy$ shell to the side of the cubic box.




dent intermediate states simulated in parallel.[53] The two end states were defined via the intermolecular solute-solvent interactions: solute in water versus solute in vacuum. Decoupling the Coulomb interactions was performed linearly with 11 staging windows. van der Waals interactions were decoupled using soft-core potentials[54–57] with 21 staging windows. The analysis of the results was performed using the `g_bar` (Bennett Acceptance Ratio)[58] module of the GROMACS 4.5 package.[10–13]

Each of the intermediate window stages for each solute (Table 1) in water underwent the following sequence of calculations:

1. Minimization of the fully atomistic systems (solute in explicit water) was performed using the steepest descent algorithm with the maximum step size of 0.01 nm and the force tolerance of 100 kJ/(mol*nm) was used.

2. Equilibration of the fully atomistic systems. The 100 ps-long NVT equilibration at $T = 300$ K was followed by the NpT equilibration of the same length. The reference pressure of 1 bar was maintained by coupling the system to a Parrinello-Rahman barostat.[59]

3. Free energy production simulations were performed in the NVT ensemble ($T = 300$ K) with the trajectory length of 2 ns for all six setups (Table 2):

   (a) For the AMB/GRO-ref setups, the output of the NpT equilibration was used as a starting point for the production simulation.

   (b) In the case of the AMB/GRO-HAd and AMB/GRO-HAd-AT setups, coordinate and topology files from step 2 were first modified by adding a virtual site at the CoM of each water molecule in order to apply the H-AdResS algorithm. Then, an additional 100 ps-long NVT equilibration was performed before starting the production simulations.

In steps 2 and 3, the accurate leapfrog stochastic dynamics integrator[60] was chosen with the time step set to 2 fs and the inverse friction constant set to 0.4 ps. Hydrogen bonds of the solute molecule were constrained using the LINCS algorithm.[61]

# Results and Discussion

After integrating H-AdResS[17] into the alchemical free energy sampling routines of GROMACS 4.5,[10–13] we calculated the free energy changes associated with the transfer of the molecules in Table 1 from vacuum to water (hydration free energies, HFE hereafter). For each solute, the calculations were performed using either the GAFF[28] or the GROMOS[27] force fields, with either a hybrid atomistic/coarse-grained description or a fully atomistic one (Table 2). The HFE values with the latter description have been previously reported for various generations of AMBER and GROMOS force fields.[29–31,36,46,47,62,63] The solute parametrization was done either by truncating the corresponding amino acid from the protein force field[29–31,46,47,62] or by using the small-molecule parametrization.[36,63] In this study, we adopted the latter approach.[4]

[4]Because the PRODRG server[42] does not allow input structures with only two heavy atoms, the analogue of cysteine (methanethiol) was replaced by ethanethiol.

We compare our H-AdResS results (AMB-HAd and GRO-HAd in Table 2) with (i) reference fully atomistic simulations from this work (AMB/GRO-ref in Table 2), (ii) fully atomistic simulations from the available literature, and (iii) published experimental values.

**(i) Comparison with all-atom MD simulations from this work.** The H-AdResS (AMB/GRO-HAd) and the reference fully atomistic (AMB/GRO-ref) HFE calculations are performed using the same simulation run parameters and differ only in the implementation of the water solvent. The AMB/GRO-HAd HFE values are expected to deviate slightly from the AMB/GRO-ref values for a few reasons. First, the calculation of the pairwise solvent-solvent interactions via the so-called *kernel* functions[49] is implemented slightly differently in the in-house H-AdResS package[5]. This purely computational error $\delta_{kernel}$ is calculated as the difference between the HFE values of effectively fully atomistic H-AdResS simulations AMB/GRO-HAd-AT (i.e. $\lambda_\alpha = 1$ for $\alpha = 1, .., N$) and the reference fully-atomistic results of AMB/GRO-ref (see SI for details): $\delta_{kernel} = \Delta F_{ref} - \Delta F_{HAd-AT}$. The slightly different equilibration procedures used for reference fully atomistic and H-AdResS setups (see Computational Details) might also contribute to $\delta_{kernel}$[6]. Second, the correction forces acting on the solvent in the hybrid region may contribute to discrepancies between the two approaches. However, their influence is indirect, since the sampling of the solute-solvent system states occurs only in the *AT* region. The estimation of the *correction force* influence $\delta_{corr}$ was done by comparing the H-AdResS simulations with hybrid (AMB/GRO-HAd) and fully atomistic (AMB/GRO-HAd-AT) solvent among themselves: $\delta_{corr} = \Delta F_{HAd-AT} - \Delta F_{HAd}$. The deviation $\delta$ of the H-AdResS results from the reference fully atomistic ones is the sum of both contributions. The corrections were also averaged either over the full or partial set of molecules listed in Table 1: $\overline{\delta} = \sum_i^M = \delta_i/M$, where $M$ is the number of molecules.

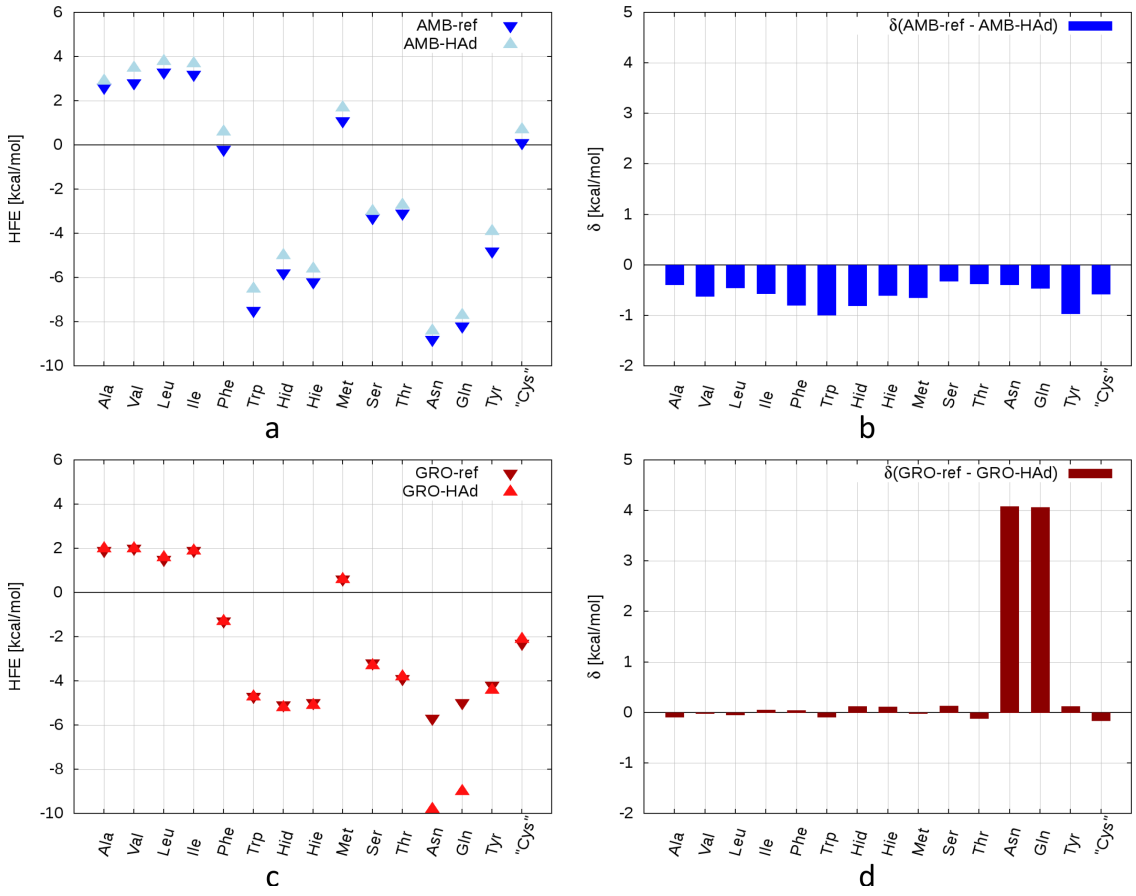
The HFE values from the AMB-HAd and AMB-ref calculations, along with the corresponding $\delta$ values, are plotted in Figure 3a,b. For all molecules in the set, AMB-HAd slightly

[5]This is done to handle both atomistic and coarse-grained types of interactions.

[6]However, we expect this contribution to be very small compared to the kernel one.

overestimates AMB-ref with an average deviation $\overline{\delta}$ = **-0.60 kcal/mol** ($\overline{\delta}_{kernel}$ = +0.69 kcal/mol, $\overline{\delta}_{corr}$ = -1.29 kcal/mol). For fixed-charge force fields such as AMBER (GAFF) and GROMOS, Mobley et al.[63] proposed an accuracy limit of **1 kcal/mol**. By using this as an upper-bound accuracy estimation for this study, we conclude that the overall average deviation in the AMB-HAd case is smaller than the force-field accuracy.

Figure 3: **Calculated HFE values. a.** AMB-ref and AMB-HAd HFE values. **b.** Deviations $\delta$ of AMB-HAd from AMB-ref. **c.** GRO-ref and GRO-HAd values. **d.** Deviations $\delta$ of GRO-HAd from GRO-ref.



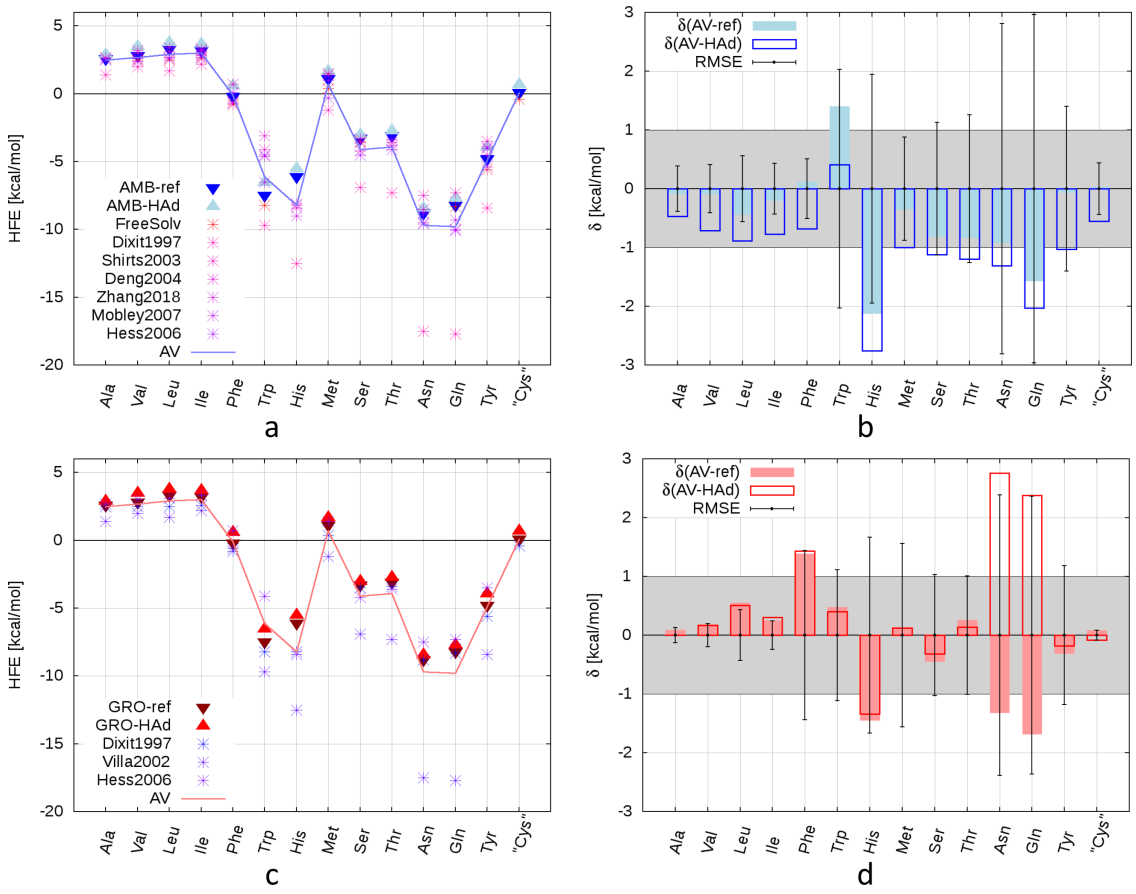

The overall agreement of GRO-HAd with GRO-ref turns out to be excellent (see Figure 3c,d), with the exception of Asn and Gln analogues, acetamide and propionamide. Indeed, without these two solutes, $\overline{\delta}$ < **-0.1 kcal/mol** ($\overline{\delta}_{kernel}$ = -0.40 kcal/mol and $\overline{\delta}_{corr}$ = +0.39 kcal/mol), where 0.1 kcal/mol is the precision limit of the calculated uncertainties due to finite sampling.[30] In the case of the amides, GRO-HAd underestimates GRO-ref by ≈ 4 kcal/mol. The

source of this large deviation can be narrowed down to pairwise force calculations, since $\overline{\delta}_{corr}$ does not change its value when calculated both with and without the amides. This point is discussed further in the next section in the context of the previously published calculated results.

**(ii) Comparison with previously published all-atom MD simulations.** We now compare our results with the published calculated HFE values for AMBER[30,31,36,46,47,62,63] and GROMOS.[29,47,62] To avoid introducing a bias toward a specific model used in one of the published calculations, we first calculate the average over all calculated HFE values (ours and published) for each molecule. In the case of the histidine analogues, a weighted HFE value is computed as in Zhang et al.[46] (0.2*Hid+0.8*Hie) for the sake of consistency with the available results. The averages are then used as reference values to calculate the deviations of the H-AdResS and reference fully atomistic values (denoted as $\delta_{AV-HAd}$ and $\delta_{AV-ref}$, respectively). We use root-mean-square error (RMSE) to characterize the dispersion of the HFE values for each molecule: $RMSE = \sqrt{\sum_i^L \delta_i^2/L}$, where $L$ is the number of HFE values available for a molecule.

For the AMBER force field (Figure 4a,b), the $\delta_{AV-HAd}$ and $\delta_{AV-ref}$ values for most molecules lie within or close to the force-field accuracy limit of 1 kcal/mol, with the exception of Trp, His, Asn, and Gln analogues. In the case of Trp, Asn, and Gln, however, the RMSE values are fairly large (2-3 kcal/mol), so that AV-HAd and AV-ref lie within their respective dispersion ranges, thus, only the His analogue exceeds its dispersion range. The deviations $\overline{\delta}_{AV-HAd}$ and $\overline{\delta}_{AV-ref}$ averaged over the full set of molecules are **-1.01 kcal/mol** and **-0.42 kcal/mol**, respectively.

Figure 4: **Comparison with the published calculated HFE values. a.** AMB-ref, AMB-HAd, and previously published AMBER HFE values. **b.** Deviations of AMB-ref and AMB-HAd from the average values. Gray-shaded area represents the accuracy limit adopted here (1 kcal/mol).[63] The error bars are the RMSE values calculated for each molecule over the available results. **c.,d.**: Same as a. and b. but for the GROMOS force field.



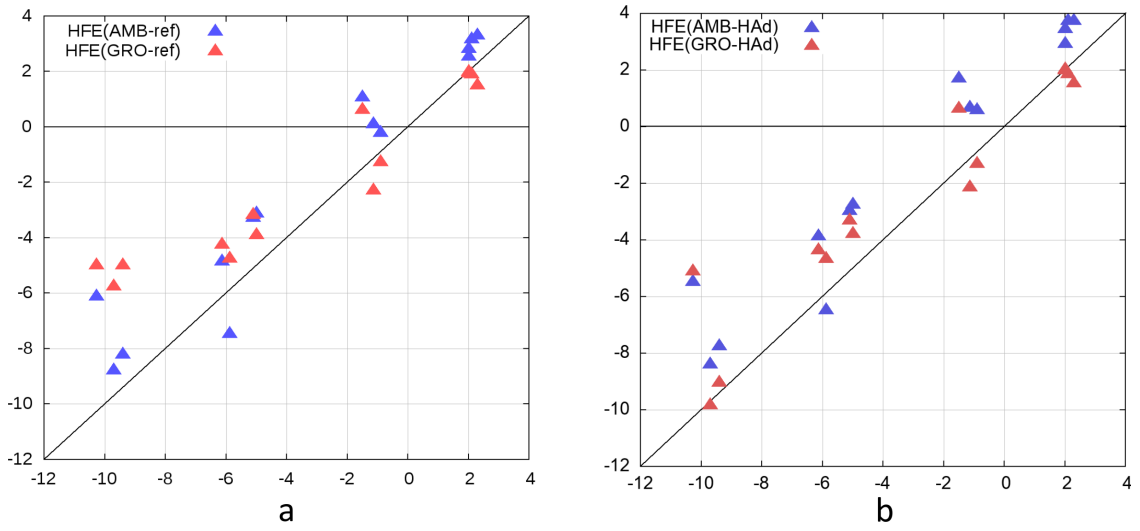

The corresponding deviations for the GROMOS-based calculations (Figure 4c,d) are also within the force-field accuracy, with the exception of the Phe, His, Asn, and Gln analogues. However, in these cases, the $\delta_{AV-HAd}$ and $\delta_{AV-ref}$ values still lie either within (Phe, His) or close (Asn, Gln) to their respective dispersion ranges (Figure 4d). Hence, the GRO-HAd and GRO-ref HFE deviations for Asn and Gln analogues, also calculated in **(i)**, actually turn out to be close to the expected dispersion for amides reported in literature. The averaged deviations are $\overline{\delta}_{AV-HAd} = +\mathbf{0.45\ kcal/mol}$ and $\overline{\delta}_{AV-ref} = \mathbf{-0.13\ kcal/mol}$.

Since, in both AMBER and GROMOS cases, the $\overline{\delta}_{AV-HAd}$ values turned out to be equal or less than 1 kcal/mol, the agreement of the H-AdResS results with the corresponding

averaged HFE values is rather good. As it may be expected, the values $\delta_{AV-ref}$ for the reference fully atomistic calculations are even smaller.

**(iii) Comparison with experiment.** Finally, we address the predictive power of the chosen solute parametrization by computing correlation coefficients $R^2$ between the calculated and published experimental values (Figure 5). Our fully atomistic reference HFE calculations correlate rather well with experiment: $R^2(AMB\text{-}ref)$=**0.925** and $R^2(GRO\text{-}ref)$=**0.909**.[7] This establishes the accuracy of the simulation setup used in all presented calculations. The correlations for the H-AdResS HFE calculations hold very well in comparison with the reference case: $R^2(AMB\text{-}HAd)$=**0.933** and $R^2(GRO\text{-}HAd)$=**0.888**.

Figure 5: **Comparison with the experimental results. a.** Comparison of the experimental results[36] with the fully-atomistic results (AMB-ref and GRO-ref) and **b.** with the H-AdResS results (AMB-HAd and GRO-HAd).




[7]The value of $R^2(GRO\text{-}ref)$ is lower than $R^2(AMB\text{-}ref)$ due to higher deviation of GRO-ref HFE values for the Asn and Gln analogues. In particular, GRO-ref in our case overestimates the hydrophobicity of both amide molecules stronger than AMB-ref. The observed discrepancy likely origins from the GROMOS force field's overestimation of polar side chains' HFE values.[29]

# Conclusions

We have presented the first step required to efficiently predict the ligand-binding affinities within our Open-Boundary Molecular Mechanics/Coarse-Grained framework (OB-MM/CG) for hGPCRs.[15] We showcased the ability of the Hamiltonian Adaptive Resolution Scheme (H-AdResS)[17] with enabled alchemical free energy calculations to reproduce fully atomistic reference simulations. We calculated hydration free energies (HFE) for a set of commonly investigated neutral amino-acid analogues using two different force fields for the solutes, namely, GAFF[28] and PRODRG/GROMOS.[27,42] The implementation turns out not only to reproduce the reference fully atomistic HFE calculations within acceptable accuracy for both AMBER and GROMOS parameterizations of the solute but also to agree well with the experimental data. As the next and final step, our code will be included in the OB-MM/CG framework. The implementation is currently underway.

# Acknowledgement

The authors thank Thomas Tarenzi, Emiliano Ippoliti, and Jan Meinke for fruitful discussions and helpful comments regarding the H-AdResS implementation. Simulations were performed with computing resources granted by RWTH Aachen University under project rwth0571.

# Supporting Information Available

Supporting Information Available: Further details about the theory of H-AdResS; additional information regarding the setup types used in the calculations of the hydration free energies (HFEs); all numerical results (HFE, $\delta$, and RMSE values) presented as tables.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

# References


(1) Gilman, A. G. G proteins: transducers of receptor-generated signals. *Annu. Rev. Biochem.* **1987**, *56*, 615–649.

(2) Hauser, A. S.; Chavali, S.; Masuho, I.; Jahn, L. J.; Martemyanov, K. A.; Gloriam, D. E.; Babu, M. M. Pharmacogenomics of GPCR drug targets. *Cell* **2018**, *172*, 41–54.

(3) GPCRdb. `https://gpcrdb.org/`, (accessed August 30, 2020).

(4) Kufareva, I.; Katritch, V.; Stevens, R. C.; Abagyan, R. Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* **2014**, *22*, 1120–1139.

(5) Cavasotto, C. N.; Palomba, D. Expanding the horizons of G protein-coupled receptor structure-based ligand discovery and optimization using homology models. *Chem. Commun.* **2015**, *51*, 13576–13594.

(6) Esguerra, M.; Siretskiy, A.; Bello, X.; Sallander, J.; Gutiérrez-de Terán, H. GPCR-ModSim: A comprehensive web based solution for modeling G-protein coupled receptors. *Nucleic Acids Res.* **2016**, *44*, W455–W462.

(7) Heifetz, A.; James, T.; Morao, I.; Bodkin, M. J.; Biggin, P. C. Guiding lead optimization with GPCR structure modeling and molecular dynamics. *Curr. Opin. Pharmacol.* **2016**, *30*, 14–21.

(8) Fierro, F.; Suku, E.; Alfonso-Prieto, M.; Giorgetti, A.; Cichon, S.; Carloni, P. Agonist binding to chemosensory receptors: a systematic bioinformatics analysis. *Front. Mol. Biosci.* **2017**, *4*, 63.

(9) Lupala, C. S.; Rasaeifar, B.; Gomez-Gutierrez, P.; Perez, J. J. Using molecular dynamics for the refinement of atomistic models of GPCRs by homology modeling. *J. Biomol. Struct. Dyn.* **2018**, *36*, 2436–2448.

(10) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; Van Drunen, R.; Van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. Gromacs: A parallel computer for molecular dynamics simulations. Physics computing. 1993; pp 252–256.

(11) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(12) Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. annual* **2001**, *7*, 306–317.

(13) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(14) Tarenzi, T.; Calandrini, V.; Potestio, R.; Giorgetti, A.; Carloni, P. Open boundary simulations of proteins and their hydration shells by Hamiltonian adaptive resolution scheme. *J. Chem. Theory Comput.* **2017**, *13*, 5647–5657.

(15) Tarenzi, T.; Calandrini, V.; Potestio, R.; Carloni, P. Open-Boundary Molecular Mechanics/Coarse-Grained Framework for Simulations of Low-Resolution G-Protein-Coupled Receptor–Ligand Complexes. *J. Chem. Theory Comput.* **2019**, *15*, 2101–2109.

(16) Leguèbe, M.; Nguyen, C.; Capece, L.; Hoang, Z.; Giorgetti, A.; Carloni, P. Hybrid molecular mechanics/coarse-grained simulations for structural prediction of G-protein coupled receptor/ligand complexes. *PloS one* **2012**, *7*, e47332.

(17) Potestio, R.; Fritsch, S.; Espanol, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.* **2013**, *110*, 108301.

(18) Español, P.; Delgado-Buscalioni, R.; Everaers, R.; Potestio, R.; Donadio, D.; Kre-

mer, K. Statistical mechanics of Hamiltonian adaptive resolution simulations. *J. Chem. Phys.* **2015**, *142*, 02B606_1.

(19) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free energy methods in drug design: Prospects of âĂIJalchemical perturbationâĂİ in medicinal chemistry: miniperspective. *Eur. J. Med. Chem.* **2018**, *61*, 638–649.

(20) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770.

(21) Capelli, R.; Bochicchio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the full free energy landscape of neuroreceptor/ligand unbinding by metadynamics simulations. *J. Chem. Theory Comput.* **2019**, *15*, 3354–3361.

(22) Capelli, R.; Carloni, P.; Parrinello, M. Exhaustive search of ligand binding pathways via volume-based metadynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3495–3499.

(23) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144.

(24) Praprotnik, M.; Delle Site, L.; Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* **2005**, *123*, 224106.

(25) Fiorentini, R.; Kremer, K.; Potestio, R.; Fogarty, A. C. Using force-based adaptive resolution simulations to calculate solvation free energies of amino acid sidechain analogues. *J. Chem. Phys.* **2017**, *146*, 244113.

(26) Schneider, J.; Korshunova, K.; Si Chaib, Z.; Giorgetti, A.; Alfonso-Prieto, M.; Carloni, P. Ligand Pose Predictions for Human G Protein-Coupled Receptors: Insights from

the Amber-Based Hybrid Molecular Mechanics/Coarse-Grained Approach. *J. Chem. Inf. Model.* **2020**,

(27) Daura, X.; Mark, A. E.; Van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* **1998**, *19*, 535–547.

(28) 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.

(29) Villa, A.; Mark, A. E. Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.* **2002**, *23*, 548–553.

(30) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(31) Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.

(32) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.

(33) Berendsen, H.; Grigera, J.; Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(34) Wang, H.; Junghans, C.; Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E* **2009**, *28*, 221–229.

(35) Wang, H.; Hartmann, C.; Schütte, C.; Delle Site, L. Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique. *Phys. Rev. X* **2013**, *3*, 011018.

(36) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711–720.

(37) 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.

(38) 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.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(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. A* **1993**, *97*, 10269–10280.

(41) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(42) Schüttelkopf, A. W.; Van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Cryst. D* **2004**, *60*, 1355–1363.

(43) Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. et al. Gaussian 09, Revision A.02. *Wallingford CT: Gaussian, Inc.* **2009**,

(44) 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.

(45) Berendsen, H.; Postma, J.; Van Gunsteren, W.; Hermans, a. J. Intermolecular forces. 1981.

(46) Zhang, H.; Yin, C.; Jiang, Y.; van der Spoel, D. Force field benchmark of amino acids: I. hydration and diffusion in different water models. *J. Chem. Inf. Model.* **2018**, *58*, 1037–1052.

(47) Hess, B.; van der Vegt, N. F. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.

(48) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.

(49) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D.; Sijbers, A.; Feenstra, K.; van Drunen, R. et al. GROMACS User Manual version 4.0. 2005.

(50) Onsager, L. Electric moments of molecules in liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(51) Barker, J. A.; Watts, R. O. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789–792.

(52) van der Spoel, D.; Van Maaren, P. J.; Berendsen, H. J. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* **1998**, *108*, 10220–10230.

(53) Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* **2010**, *31*, 1569–1582.

(54) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics. *J. Phys. Chem. A* **1997**, *101*, 5926–5930.

(55) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(56) Pillardy, J.; Piela, L. Molecular dynamics on deformed potential energy hypersurfaces. *J. Phys. Chem. A* **1995**, *99*, 11805–11812.

(57) Shao, C.-S.; Byrd, R.; Eskow, E.; Schnabel, R. B. Global optimization for molecular clusters using a new smoothing approach. *J. Glob. Optim.* **2000**, *16*, 167–196.

(58) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(59) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(60) Van Gunsteren, W. F.; Berendsen, H. J. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1988**, *1*, 173–185.

(61) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(62) Dixit, S. B.; Bhasin, R.; Rajasekaran, E.; Jayaram, B. Solvation thermodynamics of amino acids assessment of the electrostatic contribution and force-field dependence. *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 1105–1113.

(63) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

## Graphical TOC Entry