# 1 Methods

## 1.1 Dataset and molecules parametrization

**1.Add image of the CBC Datasetand CBC (SAMPL5 paper HG)**

For the 2015 Statistical Assessment of Modelling Proteins and Ligands (SAMPL5) blind free energy predictions for a dataset of host-guest systems were requested from all the participants. Figure XX shows one of the three SAMPL5 datasets, which consists of 10 charged guests that binds to a cucurbituril clip (CBC) at a physiological pH (7.4). The CBC is flexible host, which has shown a high binding affinity to ferrocene, adamantane and bicyclooctane guests [1, 2, 3]. Experimental data for CBC host-guest complexes were obtained using UV, visible and fluorescent spectroscopic measurements. All data were measured in the laboratory of Bruce Gibb (Tulane University) [4].

Both host and guest were parametrized with AM1-BCC charges and general Amber force field (GAFF). Initially, guest free phase topology and coordinates files were created by using the protocol described in [5]. Guest force field parameters were extracted from bound phase files. Thus, guests were solvated in a rectangular box of TIP3P water molecules [6] with a minimum distance between solute and box surface of 12 , using *tleap*. Ions were added to neutralize the overall charge of the box. The system was energy minimized with 100 steps of steepest descent algorithm. Then, solute's atoms were position restrained with a force constant of $10 \, \text{kcal} \, \text{mol}^{-1} \text{Å}^{-2}$, to allow water to equilibrate in an NVT ensemble for 200 ps at 298 K, followed by an NPT equilibration for further 200 ps at 1 atm pressure, with Amber module Sander [7]. Lastly, 2 ns of NPT simulation were run with Sire/SOMD (rev 15.1) [8, 9] and a final coordinate file was extracted from the last trajectory snapshot with CCPTRAJ [10]. This protocol will be called "SAMPL5 box" through the paper. Recently, a second protocol, "iso box" protocol, was adopted to create guest files in solvated phased. In particular, host heavy atoms were substituted with water molecules. Thus, a minimization on the system was performed along with the equilibration steps described above. In this way, the final solvated phase box dimensions were equal to the bound phase ones, avoiding further box size effects problems.

## 1.2 Alchemical binding free energy

**2.add picture for the TD cycle double annihilation** Binding free energies $\Delta G_{\text{bind}}$ were estimated from MD simulations by means of a double annihilation technique proposed originally by Jorgensen et al.[11] and developed by Gilson et al.[12]. As depicted in fig XX $\Delta G_{\text{bind}}$ can be estimated as:

$$\Delta G_{\text{bind}} = -k_{\text{B}} T \ln \frac{Z_{\text{HG,solv}} Z_{\text{solv}}}{Z_{\text{G,solv}} Z_{\text{H,solv}}} \tag{1}$$

where $k_B$ is the Boltzmann constant, $T$ the temperature, $Z_{HG,solv}$, $Z_{G,solv}$, $Z_{H,solv}$, $Z_{solv}$ are the configuration integrals for host-guest complex, the guest, the host and the solvent molecules respectively. Figure XX depicts how the double annihilation approach may be used to evaluate $\Delta G_{bind}$ employing a thermodynamic cycle. First the guest's partial charges are turned off both in water and in the host-guest-complex phase (discharging step), giving the discharging free energy change $\Delta G_{elec}^{solv}$ and $\Delta G_{elec}^{host}$ respectively. Secondly, the guest is fully decoupled from the solvent or host, switching off the van der Waals terms (vanishing step), $\Delta G_{vdW}^{solv}$ and $\Delta G_{vdW}^{host}$. The discharging and vanishing steps are usually performed with a series of intermediate simulations that depend on a coupling parameter $\lambda \in [0,1]$. Closure of the thermodynamic cycle gives the free energy of binding $\Delta G_{bind}$ as[13]:

$$\Delta G_{bind} = (\Delta G_{elec}^{solv} + \Delta G_{vdW}^{solv}) - (\Delta G_{elec}^{host} + \Delta G_{vdW}^{host}) \tag{2}$$

In the actual MD simulation an empirical distance-restraint term is added to the potential energy function. This is done to prevent the non-interacting guest from drifting out of the host cavity. A flat-bottom restraining potential is used between one atom of the guest, chosen to be the one closest to the centre of mass, and four equivalent carbon atoms of the host. The restraint potential for atom $j$ of the guest is based on work by Michel et al.[14]:

$$U^{restr}(d_{j1}, ..., d_{jN_host}) = \sum_{i=1}^{N_{host}} \begin{cases} 0 & if |d_{ji} - Rji| \leq D_{ji} \\ \kappa_{ji}(|d_{ji} - R_{ji}| - D_{ji})^2 & if |d_{ji} - R_{ji}| > D_{ji} \end{cases} \tag{3}$$

where $U^{restr}(d_{j1}, ..., d_{jN_{host}})$ is the potential energy of the restraint as a function of the distance between a guest atom $j$ and a set of host atoms $i$, $d_{ji} = ||\mathbf{r}_i - \mathbf{r}_j||$, $D_{ji}$ is the restraint deviation tolerance, $R_{ji}$ the reference distance between host and guest atom, $\kappa$ the restraint force constant and $N_{host}$ the number of host atoms that contribute to the restraint. Introduction of the flat-bottom restraint brings to the calculation of a restraint free energy change $\Delta G_{restr}^{\circ}$, as fig XX shows, computed as:

$$\Delta G_{restr}^{\circ} = -k_B T \ln \left( \frac{Z_{H \bullet\bullet G^{ideal},solv}}{Z_{H,solv} Z_{G,gas}} \right) \tag{4}$$

where $Z_{H \bullet\bullet G^{ideal},solv}$ is the configuration integral for the restrained decoupled guest bound to the host, $Z_{H,solv}$ is the configuration integral for the solvated host and $Z_{G,gas}$ for the guest in an ideal thermodynamic state. Assuming that the restraint potential is decoupled form the solvent and host degrees of freedom, eq 4 simplifies to:

$$\Delta G_{restr}^{\circ} = -k_B T \ln \left( \frac{Z_{\bullet\bullet G^{ideal},solv}}{Z_{G,gas}} \right) \tag{5}$$

where $Z_{\bullet\bullet G^{ideal},solv}$ is the configuration integral for the decoupled guest. Because the guest has no intermolecular interactions in both thermodynamic

states defined in eq. 5, and because the restraint does not hinder rotational motions, internal and rotational contributions cancels cancels out and the only term left is the translational contribution to the configuration integral. For $Z_{\mathrm{G,gas}}$ a standard volume of measurement $V^\circ$ is used, with the 1 M dilute solution convention $V^\circ=1660\,\text{Å}^3\,\text{mol}^{-1}$. Therefore eq. 5 simplifies further to:

$$\Delta G^\circ_{\mathrm{restr}} = -k_{\mathrm{B}}T\ln\left(\frac{V^{\mathrm{restr}}}{V^\circ}\right) \qquad (6)$$

where the restraint volume $V^{\mathrm{restr}}$ is :

$$V^{\mathrm{restr}} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx_j dy_j d_z j\, exp(-\beta U^{\mathrm{restr}}(d_{\mathrm{j1}},...,d_{\mathrm{jN_{host}}})) \qquad (7)$$

$V^{\mathrm{restr}}$ can be calculated numerically: first, the coordinates of the host-guest complex in the generate trajectory at $\lambda=1$ of the vanishing step were aligned to the first frame of the trajectory. Then, the average coordinates of each of the four host atoms used for the restraint were computed. Next, a grid of 0.1 Å and a rectangular cuboid given by the minimum/maximum coordinates of the four host atoms were defined. Thus, numerical integration was performed via multidimensional trapezoidal rule. By subtracting $\Delta G^\circ_{\mathrm{rest}}$ to eq. 1 the standard free energy of binding is computed.

## 1.3  Poisson-Boltzmann equation for charging free energy corrections

Corrections for discharging free energy were taken into consideration to estimate the final standard free energy of binding $\Delta G^\circ_{\mathrm{bind}}$. These corrections are based on the seminal works by Reif and Oostenbrink [15], Rocklin et al. [16] and earlier papers from Kastenholz and Hünenberger [17, 18]. Here corrections for free energy estimation were tuned for a Barker Watts Reaction field [19] with atom based cutoff. First, periodic boundary conditions (PBC), box size effects and polarization issues were treated. PBC allows to approximate an infinite real system by using a small part of it, the unit cell. In this way simulations can be computationally efficient, but a deviation from real system polarization may be introduced. Box size effects have been documented by Kastenholz and Hünenberger[17, 18] and Rocklin[16] and can introduce a bias up to $20\,\text{kcal}\,\text{mol}^{-1}$ for charged systems. Polarization issues arises mainly for the effective Coulombic potential employed. Furthermore, solvent model may have a dielectric constant different from the real solvent, introducing bias in polarization around the charged solute. To correct these three sources of errors a correction term $\Delta G_{\mathrm{POL}}$ was computed as:

$$\Delta G_{\mathrm{POL}} = \Delta G^{\mathrm{Coul}}_{\mathrm{NP}} - \Delta G^{\mathrm{Coul}}_{\mathrm{RF}} \qquad (8)$$

where $\Delta G^{\mathrm{Coul}}_{\mathrm{NP}}$ is the electrostatic charging free energy due to Coulombic interactions under non-periodic conditions. $\Delta G^{\mathrm{Coul}}_{\mathrm{RF}}$ is the electrostatic charging free energy obtained solving the Poisson equation under BWRF and

periodic boundary conditions, using a custom code kindly given us by Hü-nenberger [20]. In particular, $\Delta G_{\mathrm{NP}}^{\mathrm{Coul}}$ is computed by using *APBS* software [21]. First, the trajectory file is loaded and all the snapshots are aligned onto the first one. Secondly, 193 grid points per processor are selected for grid creation with a mesh-grid spacing of $0.50\,\text{Å}$. Thus, *APBS* performs a non periodic electrostatic potential calculation on the host-guest system retrieving the potential $\phi^{\mathrm{HG}}(\mathbf{r})$ as a function of coordinates $\mathbf{r}$. From this potential an electrostatic free energy change $\Delta G_{\mathrm{NP}}^{\mathrm{HG}}$ can be evaluted as:

$$\Delta G_{\mathrm{NP}}^{\mathrm{HG}} = \sum_{i=1}^{N_{\mathrm{solute}}} q_{\mathrm{i}}\phi^{\mathrm{HG}}(\mathbf{r}_{\mathrm{i}}) \qquad (9)$$

where the sum is extended to all the solute's atoms $N_{\mathrm{solute}}$, $q_{\mathrm{i}}$ is the $i$-th solute atom charge and $\phi^{\mathrm{HG}}(\mathbf{r}_{\mathrm{i}})$ is the electrostatic potential $\phi^{\mathrm{HG}}(\mathbf{r})$ at the $i$-th site. Then, the same grid is used to evaluate an electrostatic potential $\phi^{\mathrm{H}}(\mathbf{r})$ setting the guest's charges to zero, obtaining a $\Delta G_{\mathrm{NP}}^{\mathrm{H}}$ using eq 9 with $\phi^{\mathrm{H}}(\mathbf{r})$. As described in [22, 23, 24] for each electrostatic potential calculation a vacuum reference potential is evaluated, in order to cancel out the self the Coulombic terms. The final electrostatic charging free energy $\Delta G_{\mathrm{NP}}^{\mathrm{Coul}}$ is computed as a difference:

$$\Delta G_{\mathrm{NP}}^{\mathrm{Coul}} = \Delta G_{\mathrm{NP}}^{\mathrm{HG}} - \Delta G_{\mathrm{NP}}^{\mathrm{H}} \qquad (10)$$

Secondly, the presence of an atom-based cutoff introduces errors in computing solute-solvent interactions. This summation scheme causes an apparent solvation of negatively charged species and a de-solvation of positively charged molecules. For atom-based BWRF a $\Delta G_{\mathrm{PSUM}}$ correction term was evaluated as:

$$\Delta G_{\mathrm{PSUM}} = -\frac{N_A}{6\epsilon_0}Q_{\mathrm{mol}}\gamma_s\left[\left(\frac{2(\epsilon_{\mathrm{BW}}+1)}{2\epsilon_{\mathrm{BW}}+1}\times\frac{\langle N_{\mathrm{s}}(r_{\mathrm{c}})\rangle}{4\pi\frac{r_{\mathrm{c}}^3}{3}}\right)+\frac{3}{2\epsilon_{\mathrm{BW}}+1}\right] \quad (11)$$

where $N_A$ is Avogadro number, $\epsilon_0$ the experimental permittivity for the solvent, $\epsilon_{\mathrm{BW}}$ the dielectric constant of the water model used, $\gamma_s = \sum_{\mathrm{i}=1}^{\mathrm{N}} q_{\mathrm{i}}\mathbf{r}_{\mathrm{i}}$ is the trace of the quadrupole-moment tensor of the solvent model, where the sum is over all $N$ atoms in a solvent molecule, $q_{\mathrm{i}}$ is the charge of the $i$-th atom in a solvent molecule, $\mathbf{r}_{\mathrm{i}}$ is the coordinate vector, $Q_{\mathrm{mol}}$ is the net charge of the solute,$r_{\mathrm{c}}$ is the reaction field cutoff length and $\langle N_{\mathrm{s}}(r_{\mathrm{c}})\rangle$ the average number of solvent molecules within $r_{\mathrm{c}}$. $\langle N_{\mathrm{s}}(r_{\mathrm{c}})\rangle$ was evaluated considering the distance between solutes'atoms and the centre of mass of each water molecule.

Finally, employing an effective Coulombic potential gives a bias in the guest-host interactions. Thus a $\Delta G_{\mathrm{dir}}$ term was computed as:

$$\Delta G_{\mathrm{dir}} = U_{\mathrm{NP}}^{\mathrm{dir}} - U_{\mathrm{RF}}^{\mathrm{dir}} \qquad (12)$$

where $U_{\text{NP}}^{\text{dir}}$ is the sum of the intermolecular and intramolecular solute energies under Coulomb's law and $U_{\text{RF}}^{\text{dir}}$ is the sum of intermolecular and intramolecular solute energies with BWRF Coulombic potential.

The final charging correction term is:

$$\Delta G_{\text{COR}} = (\Delta G_{\text{NP}}^{\text{HG}} + \Delta G_{\text{NP}}^{\text{H}} + U_{\text{NP}}^{\text{dir}}) - (\Delta G_{\text{RF}}^{\text{HG}} + \Delta G_{\text{RF}}^{\text{H}} + U_{\text{RF}}^{\text{dir}}) \qquad (13)$$

giving a final standard free energy of binding $\Delta G_{\text{bind}}^{\circ}$ :

$$\Delta G_{\text{cor}}^{\circ} = \Delta G_{\text{bind}} + \Delta G_{\text{restr}}^{\circ} + \Delta G_{\text{COR}}^{\text{PBcode}} \qquad (14)$$

## 1.4 A termodynamic cycle for charging free energy corrections

To have a comparison term and assess the magnitude of the charging free energy corrections computed as described in 1.3, a thermodynamic cycle approach was tested, too. As depicted in fig XX, in order to maintain the electroneutrality at the beginning and at the end of each leg of the thermodynamic cycle as many dummy counterions as the charge of the solute can be added to the box. Thus, once the solute is fully decoupled, van der Waals parameters and charge of each ion are sequentially switched on both in solvated and bound phases. In this way, considering $n$ counterions, a correction term $\Delta G_{\text{COR}}^{\text{ion}}$ can be computed as:

$$\Delta G_{\text{COR}}^{\text{TDion}} = \sum_{i=1}^{n} \left\{ \Delta G_{\text{ion}_i}^{\text{solv}} - \Delta G_{\text{ion}_i}^{\text{host}} \right\} \qquad (15)$$

where the sum is extended up to all the $n$ counterions, $\Delta G_{\text{ion}_i}^{\text{solv}}$ is the free energy change for turning on van der Waals and charges of the $i$-th ion in the solvated phase, $\Delta G_{\text{ion}_i}^{\text{host}}$ is the free energy change for the $i$-th ions in the bound phase.

## 1.5 Alchemical free energy production simulations

For the discharging step nine equidistant $\lambda$ windows were selected for the host-guest complex and the guest in water phase. Differently, for the vanishing step eighteen equidistant windows were employed, in order to guarantee good overlap of the potential energy distributions of neighbouring $\lambda$ windows. Initially, for the SAMPL5 challange all the host-guest systems were run for 16 ns with a velocity-Verlet integrator timestep of 4 fs using a hydrogen mass repartitioning (HMR) scheme [25]. For more recent simulations, HMR was not adopted, once convergence tests were made to assure the same results were attained. All simulations were performed in an NPT ensemble and temperature control was achieved with an Andersen Thermostat with a coupling constant of $10 \, \text{ps}^{-1}$ [26]. Pressure was manintained by a Monte

Carlo barostat, attempting isotropic box edge scaling every $100\,\mathrm{fs}$. Periodic boundary conditions were imposed with a $12\,\text{Å}$ atom-based cutoff distance for the non-bonded interactions, using a Barker Watts reaction field with a dielectric constant of 82.0 [19]. In the host-guest complex the molecules were restrained according to eq. 3. The parameters were $R_{\mathrm{ji}}{=}5\,\text{Å}$, $D_{\mathrm{ji}}{=}2\,\text{Å}$ and $\kappa_{\mathrm{ji}}{=}10\,\mathrm{kcal\ mol}^{-1}\ \text{Å}^{-2}$. Individual free energy contributions from the discharging and vanishing steps were estimated by using MBAR [27]. To estimate the accuracy and consistency of the computed binding free energy, each simulation was repeated twice using different initial assignments of velocities drawn from the Maxwell-Boltzmann distribution. Final binding free energies are reported as average of both runs and statistical uncertainties were calculated according to:

$$err(\Delta G) = \frac{\sigma}{\sqrt{n}} \tag{16}$$

where $\sigma$ is the standard deviation of both runs and $n{=}2$.

The computed binding free energies are then compared to experimental values considering the determination coefficient $R^2$ and the mean unsigned error MUE. To gain insight into the distribution of the two differnet measures a bootstrapping scheme is used in which each computed free energy point is considered to parametrize a normal distribution with its mean given by the computed free energy and $\sigma$ the associated computed error. One thousand samples are then drawn from the artificial normal distributions for each data point and correlated with the experimental values, giving rise to a distribution of $R^2$ and MUE. The resulting distributions are typically not symmetric around the mean and uncertainties in the dataset metrics are reported with a 95% confidence interval.

## 2 Results

### 2.1 Test on sodium ion

Before running the charging free energy correction with Poisson-Boltzmann code (PB code) on host–guest molecules, a test on the hydration free energy of a sodium ion was run. In particular, a sodium ion was solvated in two boxes: one called "big" whose edge was $60\,\text{Å}$ and one "small" with $30\,\text{Å}$ edges. Furthermore, two simulations for each box were run: one with a non-bonded interactions cutoff of $10\,\text{Å}$ and one with $12\,\text{Å}$ cutoff. Each simulation was $2\,\mathrm{ns}$ long and 11 equidistant $\lambda$ windows were used both for discharging and vanishing step, and two independent runs were made. In this way, the cutoff and box size dependence of the PB code was verified. Table 1 show the free energy estimation as an average from two independent runs along with the standard error (err). The final corrected hydration free energy $\langle \Delta G_{\mathrm{solv-cor}} \rangle$

Table 1: Sodium hydration free energy corrected with PB code approach. $\Delta G_{\text{solv}}$ is the solvation free energy obtained from normal MD simulations, $\Delta G_{\text{COR}}$ is the average correction term from PB code, as explain in 1.3 and $\Delta G_{\text{solv}-\text{cor}}$ is the final correct hydration free energy estimation. Each free energy is expressed in kcal mol$^{-1}$ along with the standard error (err)

| Na ion | $\Delta G_{\text{solv}}$ | err | $\Delta G_{\text{COR}}^{\text{PBcode}}$ | err | $\Delta G_{\text{solv}-\text{cor}}$ | err |
|---|---|---|---|---|---|---|
| 10 Åbig | -63.72 | 0.01 | -42.18 | 0.06 | -105.96 | 0.07 |
| 10 Åsmall | -62.75 | 0.02 | -41.17 | 0.08 | -103.92 | 0.10 |
| 12 Åbig | -67.85 | 0.03 | -38.37 | 0.03 | -106.22 | 0.01 |
| 12 Åsmall | -66.31 | 0.03 | -37.70 | 0.03 | -104.02 | 0.07 |

Table 2: Hydration free energy estimations for acetate and methylammonium with PB code charging corrections. $\Delta G_{\text{solv}}$ is the solvation free energy obtained from normal MD simulations, $\Delta G_{\text{COR}}$ is the average correction term from PB code and $\Delta G_{\text{solv}-\text{cor}}$ is the final corrected hydration free energy. Each free energy is expressed in kcal mol$^{-1}$ along with the standard error (err)

| molecule | $\Delta G_{\text{solv}}$ | err | $\Delta G_{\text{COR}}^{\text{PBcode}}$ | err | $\Delta G_{\text{solv}-\text{cor}}$ | err |
|---|---|---|---|---|---|---|
| acetate | -72.68 | 0.15 | -9.10 | 0.04 | -81.78 | 0.19 |
| methylammonium | -37.55 | 0.04 | -39.46 | 0.14 | -61.73 | 0.10 |

shows a dependence on the box size with discrepancies between results, differently from [17, 18]. A further test was done, to predict hydration free energy for acetate and methlyammonium molecules. Here, the two solutes were solvated in TIP3P box of 30 Å edge and parametrized with GAFF. 2 ns simulations with 11 equidistant $\lambda$ windows were run twice, employing a cutoff length of 10 Å. Table 2 show results for initial hydration free energy estimation $\Delta G_{\text{solv}}$ and final corrected free energy $\Delta G_{\text{solv}-\text{cor}}$. Results are in agreement with experimental $\Delta G_{\text{solv}}$ [28] of $-77.6$ kcal mol$^{-1}$ and $-61.4$ kcal mol$^{-1}$ for acetate and methylammonium respectively.

## 2.2 Binding free energy with charging correction

Fig. 1 and table 3 report a comparison between the standard free energy of binding $\Delta G_{\text{bind}}^{\circ}$, the standard free energy of binding after charging correction $\Delta G_{\text{bind}-\text{cor}}^{\circ}$ and experimental values $\Delta G_{\text{EXP}}$ for the host-guest system parametrized with the "SAMPL5 box" protocol 1. $\Delta G_{\text{bind}}^{\circ}$ are in line with experimental results with a $R^2$: $0.69 < 0.70 < 0.72$ and MUE: $5.33 < 5.40 < 5.47$ kcal mol$^{-1}$. Large errors are present for guests G2 and G3, which are suppose to bind substantially worse than observed in experiments. Furthermore, these two molecules are made up of linear flexible alkyl chains

Figure 1: Comparison between standard free energy of binding $\Delta G^\circ_\text{bind}$, standard free energy of binding after charging correction $\Delta G^\circ_\text{bind−cor}$ and experimental binding free energy $\Delta G_\text{EXP}$ adopting "SAMPL5 box" protocol
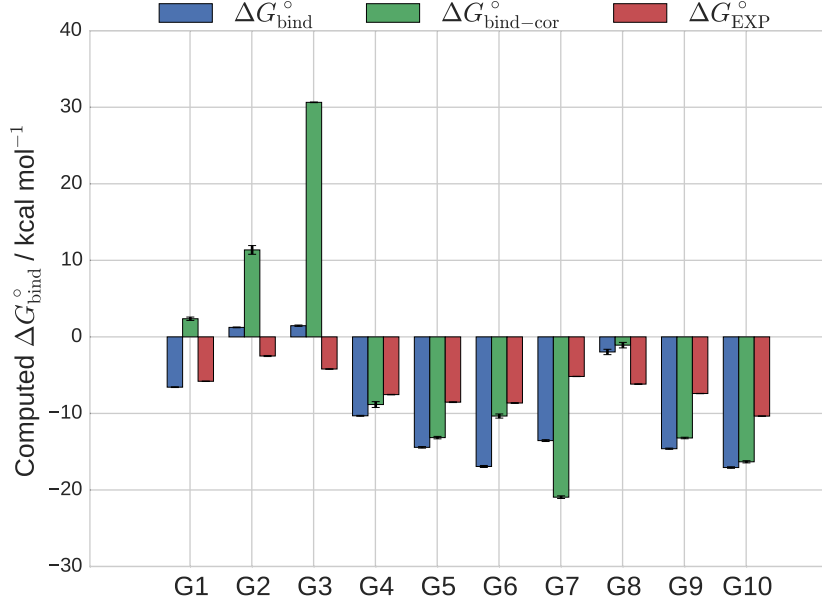
Table 3: Details of charging free energy correction for "SAMPL5 box" protocol: $Q_\text{L}$ is the solute charge in $e$, $\Delta G^\circ_\text{bind}$ the standard free energy of binding, $\Delta G_\text{COR}$ the charging free energy correction term evaluated with PB code, $\Delta G^\circ_\text{bind−−cor}$ the standard free energy of binding with charging correction, $\Delta G_\text{EXP}$ the experimental binding free energy. All the values are expressed in kcal mol$^{-1}$ along with standard error of the mean (err)

| molecule | $Q_\text{L}$ | $\Delta G^\circ_\text{bind}$ | err | $\Delta G^\text{PBcode}_\text{COR}$ | err | $\Delta G^\circ_\text{bind−cor}$ | err | $\Delta G_\text{EXP}$ | err |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 2 | -6.57 | 0.05 | 8.94 | 0.22 | 2.37 | 0.22 | -5.80 | 0.03 |
| G2 | 2 | 1.24 | 0.05 | 10.12 | 0.57 | 11.36 | 0.57 | -2.50 | 0.07 |
| G3 | 4 | 1.46 | 0.09 | 29.18 | 0.04 | 30.64 | 0.10 | -4.02 | 0.03 |
| G4 | 1 | -10.31 | 0.06 | 1.47 | 0.38 | -8.84 | 0.38 | -7.24 | 0.03 |
| G5 | 1 | -14.43 | 0.10 | 1.27 | 0.12 | -13.16 | 0.16 | -8.53 | 0.07 |
| G6 | 2 | -16.93 | 0.13 | 6.58 | 0.24 | -10.34 | 0.27 | -8.64 | 0.05 |
| G7 | -2 | -13.55 | 0.12 | -7.39 | 0.13 | -20.94 | 0.17 | -5.17 | 0.02 |
| G8 | 1 | -1.97 | 0.35 | 0.88 | 0.09 | -1.08 | 0.36 | -6.17 | 0.04 |
| G9 | 1 | -14.91 | 0.08 | 1.70 | 0.05 | -13.21 | 0.10 | -7.40 | 0.02 |
| G10 | 1 | -17.07 | 0.10 | 0.75 | 0.09 | -16.32 | 0.14 | -10.35 | 0.02 |

and contain several positively charged ammonium groups.On the contrary, charging free corrections perform much worse with $R^2$: $0.07< 0.08 < 0.09$ and MUE: $11.65< 11.79 < 11.94$kcal mol$^{-1}$. In particular, G2 and G3 show high positive free energy correction term $\Delta G_{\mathrm{COR}}= (10.12 \pm 0.57)$ kcal mol$^{-1}$ and $(29.18 \pm 0.04)$ kcal mol$^{-1}$ respectively. Additionally, negative charged molecules, as G7, has a more negative free energy correction term. $\Delta G_{\mathrm{COR}}^{\mathrm{PBcode}}=(-7.39 \pm 0.13)$ kcal mol$^-$ Therefore, PBcode correction seems to be proportional to the solute charge: the higher charge the solute has the higher will be the $\Delta G_{\mathrm{COR}}$.

Three main reasons can be drawn, to explain such a bias in the PB corrections:

- First, the host is too flexible and theory 1.3 cannot work properly. Indeed, both [15] and [16] have shown very good results by measuring binding free energy only for large and rigid host molecules.

- Secondly, the force field parametrization may play a role, as found in [29] with cyclohexane solvation free energies.

- Finally, the PB code may have some errors, especially in the Poisson-Boltzmann reaction field calculations.

In order to avoid possible box size effects which may influence the PB code itself, "iso box" protocol 1 was applied to all the host-guest systems and charging free energy correction were re computed and fig. 2 and table 4 show results. In this case, $\Delta G_{\mathrm{bind}}^{\circ}$ predictions show the same trend as before, with $R^2$=$0.57 < 0.64< 0.70$ and MUE $= 4.85 < 5.10< 5.36$ kcal mol$^{-1}$. On the other side, $\Delta G_{\mathrm{bind--cor}}^{\circ}$ has a better trend with respect to previous calculations, with an R$^2$= $0.39 < 0.41< 0.44$ and MUE $= 7.89< 8.14< 8.39$ kcal mol$^{-1}$, although G2, G3 and G7 corrections are of the order of ten kcal mol$^{-1}$. Higher standard error are present in these simulations, mainly due to the vanishing step in bound phase. Further tests will be done to assess the convergence of results between both kind of protocols.

<span style="color:red">3.This is very strange, SAMPL5 sire 15.1 HMR was more stable in boundphase vanishing step vs same box dim for hostguest sire 16.1 allbonds constraints</span>

Figure 2: Comparison between standard free energy of binding $\Delta G_{\text{bind}}^{\circ}$, standard free energy of binding after charging correction $\Delta G_{\text{bind-cor}}^{\circ}$ and experimental binding free energy $\Delta G_{\text{EXP}}$ using "iso box" protocol to prepare the free phase system
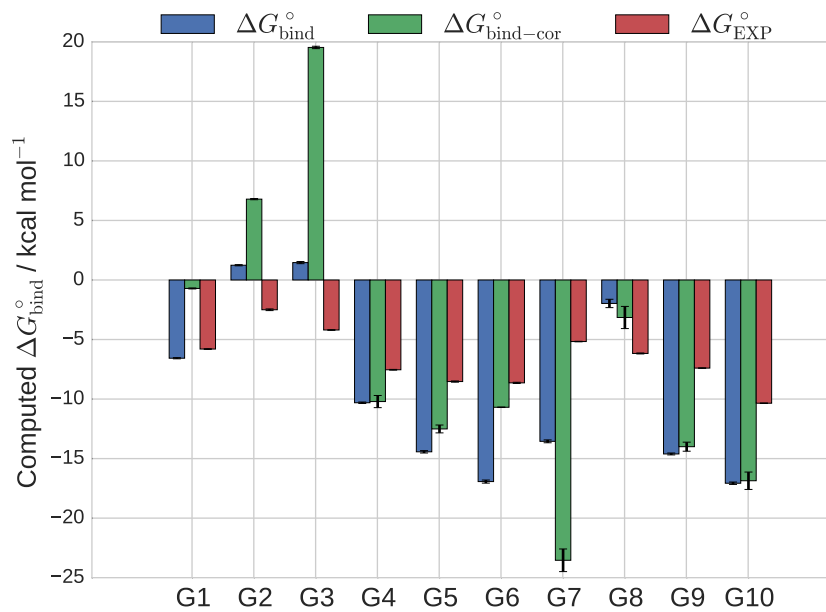
Table 4: Details of charging free energy correction with "iso box" protocol: $Q_L$ is the solute charge in $e$, $\Delta G_{\text{bind}}^{\circ}$ the standard free energy of binding, $\Delta G_{\text{COR}}$ the charging free energy correction term evaluated with PB code, $\Delta G_{\text{bind--cor}}^{\circ}$ the standard free energy of binding with charging correction, $\Delta G_{\text{EXP}}$ the experimental binding free energy. All the values are expressed in kcal mol$^{-1}$ along with standard error of the mean (err)

| molecule | $Q_L$ | $\Delta G_{\text{bind}}^{\circ}$ | err | $\Delta G_{\text{COR}}^{\text{PBcode}}$ | err | $\Delta G_{\text{bind-cor}}^{\circ}$ | err | $\Delta G_{\text{EXP}}$ | err |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 2 | -8.52 | 0.02 | -0.71 | 0.04 | -5.80 | 0.03 | | |
| G2 | 2 | -1.93 | 0.12 | 6.79 | 0.05 | -2.50 | 0.07 | | |
| G3 | 4 | -6.00 | 0.35 | 19.53 | 0.10 | -4.02 | 0.03 | | |
| G4 | 1 | -11.28 | 0.59 | -10.21 | 0.51 | -7.24 | 0.03 | | |
| G5 | 1 | -13.77 | 0.13 | -12.51 | 0.33 | -8.53 | 0.07 | | |
| G6 | 2 | -17.37 | 0.07 | -10.69 | 0.02 | -8.64 | 0.05 | | |
| G7 | -2 | -15.45 | 0.92 | -23.54 | 0.95 | -5.17 | 0.02 | | |
| G8 | 1 | -4.31 | 0.81 | -3.15 | 0.93 | -6.17 | 0.04 | | |
| G9 | 1 | -15.45 | 0.52 | -14.00 | 0.38 | -7.40 | 0.02 | | |
| G10 | 1 | -17.93 | 0.04 | -16.86 | 0.07 | -10.35 | 0.02 | | |

## 2.3 Thermodynamic cycle approach for charging free energy corrections

**4.Fix results of SAMPL5 box with SAMPL5 real results. Some of them are different since I did not add LJLRC, but why G8 is so different?**

Table 5: Standard free energy of binding $\Delta G_{\text{bind}}^{\circ}$, charging free energy correction $\Delta G_{\text{COR}}^{\text{TDion}}$ evaluated with thermodynamic cycle 1.4 and standard free energy of binding with charging correction $\Delta G_{\text{bind-cor}}^{\circ}$ along with standard error (err) for "SAMPL5 box" protocol

| molecule | $Q_L$ | $\Delta G_{\text{bind}}^{\circ}$ | err | $\Delta G_{\text{COR}}$ | err | $\Delta G_{\text{bind-cor}}^{\circ}$ | err | $\Delta G_{\text{EXP}}$ | err |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 2 | -8.32 | 0.05 | 0.18 | 0.22 | -8.14 | 0.14 | -5.80 | 0.03 |
| G2 | 2 | -1.67 | 0.05 | 0.10 | 0.15 | -1.57 | 0.12 | -2.50 | 0.07 |
| G3 | 4 | 1.34 | 0.09 | -5.31 | 0.01 | 30.64 | 0.10 | -4.02 | 0.03 |
| G4 | 1 | -10.21 | 0.06 | -0.47 | 0.06 | -10.69 | 0.22 | -7.24 | 0.03 |
| G5 | 1 | -14.16 | 0.32 | -0.21 | 0.02 | -14.37 | 0.32 | -8.53 | 0.07 |
| G6 | 2 | -17.35 | 0.26 | -0.39 | 0.23 | -17.74 | 0.49 | -8.64 | 0.05 |
| G7 | -2 | -12.46 | 0.12 | -0.43 | 0.05 | -12.89 | 1.27 | -5.17 | 0.02 |
| G8 | 1 | -3.10 | 1.12 | -2.19 | 0.01 | -5.29 | 1.13 | -6.17 | 0.04 |
| G9 | 1 | -15.03 | 0.08 | -2.34 | 0.08 | -17.37 | 0.16 | -7.40 | 0.02 |
| G10 | 1 | -17.35 | 0.21 | -0.01 | 0.11 | -17.36 | 0.33 | -10.35 | 0.02 |

Table 5 and fig 3 show the standard free energy of binding $\Delta G_{\text{bind-cor}}^{\circ}$

Figure 3: Comparison between standard free energy of binding $\Delta G_{\mathrm{bind}}^{\circ}$, standard free energy of binding after charging correction with thermodynamic cycle $\Delta G_{\mathrm{bind-cor}}^{\circ}$ 1.4 and experimental binding free energy $\Delta G_{\mathrm{EXP}}$ with "SAMPL5 box" parametrization
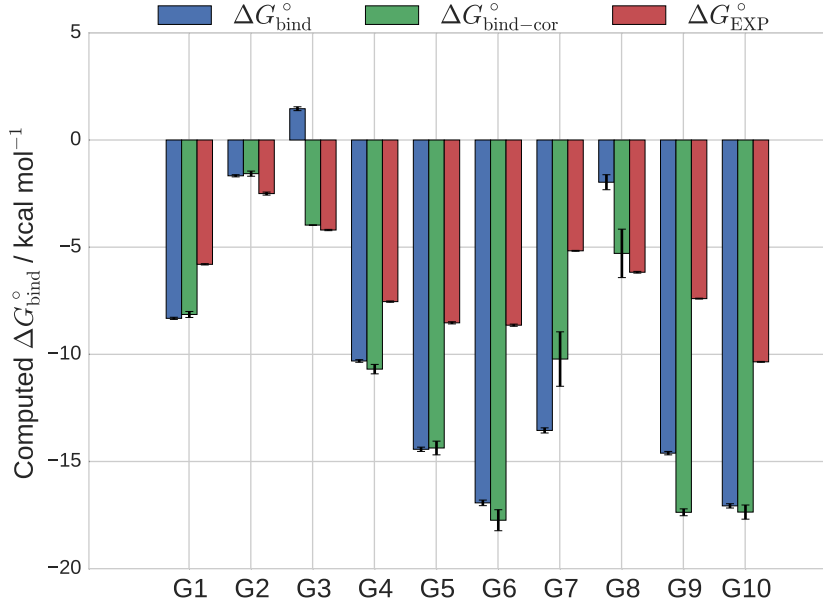


corrected with the thermodynamic approach depicted in fig XX, with host-guest systems parametrized with "SAMPL5 box" protocol.

In this case the final standard free energy of binding $\Delta G_{\mathrm{bind-cor}}^{\circ}$ presents a better trend with respect to experimental data, with an $R^2$: 0.65< 0.73< 0.79 and MUE of 4.47< 4.74< 5.04 kcal mol$^{-1}$.

Thus, to test the consistency of the thermodynamic cycle approach, all the CBC guests were solvated according to the "iso box" protocol and tab. 6 and fig. 4 show the final estimations. $\Delta G_{\mathrm{bind-cor}}^{\circ}$ is in line with results in tab. 5 with an R$^2$=0.59< 0.66< 0.74 and MUE=4.43< 4.69< 4.96 kcal mol$^{-1}$. High standard errors are present, which are mainly due to fluctuations in the bound phase vanishing step, for molecule G4, G6, G7, G8 and G9. A further run will be made for these molecules to enhance convergence of results. Overall, thermodynamic cycle approach seems to give a more reasonable charging correction, but a strict convergence of free energy changes is required to be more accurate and consistent.

Table 6: Details of charging free energy correction: $Q_{\mathrm{L}}$ is the solute charge in $e$, $\Delta G_{\mathrm{bind}}^{\circ}$ the standard free energy of binding, $\Delta G_{\mathrm{COR}}$ the charging free energy correction term, $\Delta G_{\mathrm{bind-cor}}^{\circ}$ the standard free energy of binding with charging correction, $\Delta G_{\mathrm{EXP}}$ the experimental binding free energy. All the values are expressed in kcal mol$^{-1}$ along with standard error (err)

| molecule | $Q_{\mathrm{L}}$ | $\Delta G_{\mathrm{bind}}^{\circ}$ | err | $\Delta G_{\mathrm{COR}}^{\mathrm{TDion}}$ | err | $\Delta G_{\mathrm{bind-cor}}^{\circ}$ | err | $\Delta G_{\mathrm{EXP}}$ | err |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 2 | -8.52 | 0.02 | 0.53 | 0.06 | -7.98 | 0.03 | -5.80 | 0.03 |
| G2 | 2 | -1.93 | 0.12 | 0.09 | 0.11 | -1.83 | 0.01 | -2.50 | 0.07 |
| G3 | 4 | -6.00 | 0.35 | 1.48 | 0.10 | -4.52 | 0.24 | -4.02 | 0.03 |
| G4 | 1 | -11.28 | 0.59 | 0.03 | 0.02 | -11.24 | 0.61 | -7.24 | 0.03 |
| G5 | 1 | -13.77 | 0.13 | 0.20 | 0.01 | -13.57 | 0.14 | -8.53 | 0.07 |
| G6 | 2 | -17.37 | 0.07 | 0.70 | 0.33 | -16.67 | 0.40 | -8.64 | 0.05 |
| G7 | -2 | -15.45 | 0.92 | 0.66 | 0.17 | -14.79 | 1.09 | -5.17 | 0.02 |
| G8 | 1 | -4.31 | 0.81 | 0.22 | 0.07 | -4.09 | 0.73 | -6.17 | 0.04 |
| G9 | 1 | -15.45 | 0.52 | 0.29 | 0.06 | -15.16 | 0.46 | -7.40 | 0.02 |
| G10 | 1 | -17.93 | 0.04 | 0.35 | 0.09 | -17.57 | 0.05 | -10.35 | 0.02 |

Figure 4: Comparison between "iso box" standard free energy of binding $\Delta G_{\mathrm{bind}}^{\circ}$, standard free energy of binding after charging correction with thermodynamic cycle $\Delta G_{\mathrm{bind-cor}}^{\circ}$ 1.4 and experimental binding free energy $\Delta G_{\mathrm{EXP}}$
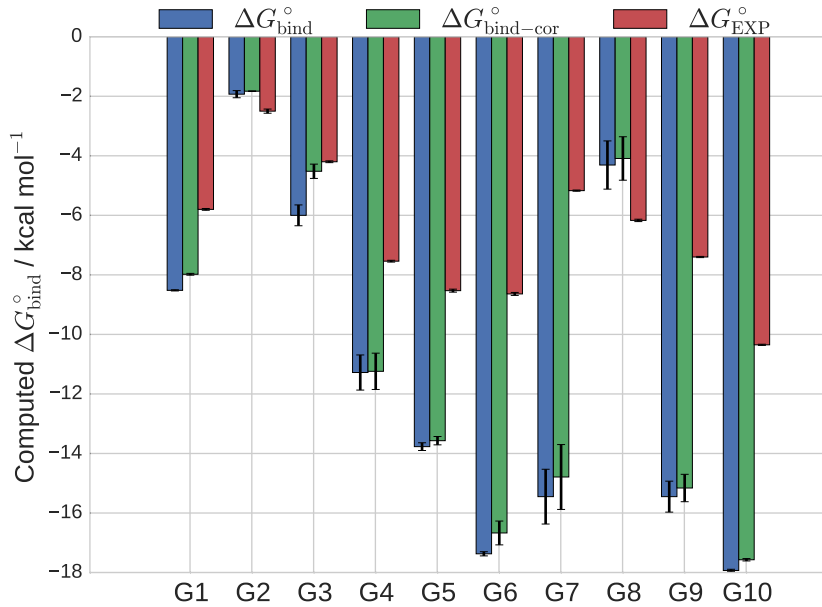
# References

[1] Laura Gilberg, Ben Zhang, Peter Y. Zavalij, Vladimir Sindelar, and Lyle Isaacs. *Org. Biomol. Chem.*, 13:4041–4050, 2015.

[2] Ben Zhang and Lyle Isaacs. *J. Med. Chem.*, 57(22):9554–9563, 2014. PMID: 25369565.

[3] Rebecca Sure, Jens Antony, and Stefan Grimme. *J. Phys. Chem. B*, 118(12):3431–3440, 2014.

[4] Jian Yin, Niel M Henriksen, David R Slochower, Michael R Shirts, Michael W Chiu, David L Mobley, and Michael K Gilson. Overview of the sampl5 host–guest challenge: Are we doing better? *Journal of computer-aided molecular design*, pages 1–19, 2016.

[5] Stefano Bosisio, Antonia SJS Mey, and Julien Michel. Blinded predictions of host-guest standard free energies of binding in the sampl5 challenge. *Journal of Computer-Aided Molecular Design*, pages 1–10, 2016.

[6] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. *J. Chem. Phys.*, 79(2):926–935, 1983.

[7] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. Amber 12, 2012.

[8] G. Calabro J. Michel C. Woods, A.S.J. Mey. Sire molecular simulations framework, (n.d.). (accessed May 31, 2016).

[9] Peter Eastman, Mark S. Friedrichs, John D. Chodera, Randall J. Radmer, Christopher M. Bruns, Joy P. Ku, Kyle A. Beauchamp, Thomas J. Lane, Lee-Ping Wang, Diwakar Shukla, Tony Tye, Mike Houston, Timo Stich, Christoph Klein, Michael R. Shirts, and Vijay S. Pande. *J. Chem. Theory Comput.*, 9(1):461–469, 2013. PMID: 23316124.

[10] Daniel R. Roe and III Thomas E. Cheatham. *J. Chem. Theory Comput.*, 9(7):3084–3095, 2013. PMID: 26583988.

[11] William L. Jorgensen, J. Kathleen Buckner, Stephane Boudon, and Julian TiradoâĂŘRives. *J. Chem. Phys.*, 89(6):3742–3746, 1988.

[12] Michael K Gilson, James A Given, Bruce L Bush, and J Andrew Mc-Cammon. *Biophys. J.*, 72(3):1047, 1997.

[13] Julien Michel and Jonathan W Essex. *J. Comput. Aided Mol. Des.*, 24(8):639–658, 2010.

[14] Julien Michel, Richard H Henchman, Georgios Gerogiokas, Michelle WY Southey, Michael P Mazanetz, and Richard J Law. *J. Chem. Theory Comput.*, 10(9):4055–4068, 2014.

[15] Maria M Reif and Chris Oostenbrink. *J. Comput. Chem.*, 35(3):227–243, 2014.

[16] Gabriel J Rocklin, Sarah E Boyce, Marcus Fischer, Inbar Fish, David L Mobley, Brian K Shoichet, and Ken A Dill. *J. Mol. Biol.*, 425(22):4569–4583, 2013.

[17] Mika A Kastenholz and Philippe H Hünenberger. *J. Chem. Phys.*, 124(12):124106, 2006.

[18] Mika A Kastenholz and Philippe H Hünenberger. *J. Chem. Phys.*, 124(22):224501, 2006.

[19] Ilario G. Tironi, RenÃĺ Sperb, Paul E. Smith, and Wilfred F. van Gunsteren. *J. Chem. Phys.*, 102(13):5451–5459, 1995.

[20] Christine Peter, Wilfred F van Gunsteren, and Philippe H Hünenberger. A fast-fourier transform method to solve continuum-electrostatics problems with truncated electrostatic interactions: Algorithm and application to ionic solvation and ion–ion interaction. *The Journal of chemical physics*, 119(23):12205–12223, 2003.

[21] N Baker, M Holst, and Feng Wang. *J. Comput. Chem.*, 21(15):1343–1352, 2000.

[22] Michael K Gilson, Malcolm E Davis, Brock A Luty, and J Andrew Mc-Cammon. Computation of electrostatic forces on solvated molecules using the poisson-boltzmann equation. *The Journal of Physical Chemistry*, 97(14):3591–3600, 1993.

[23] Jeffry D Madura, James M Briggs, Rebecca C Wade, Malcolm E Davis, Brock A Luty, Andrew Ilin, Jan Antosiewicz, Michael K Gilson, Babak Bagheri, L Ridgway Scott, et al. Electrostatics and diffusion of molecules in solution: simulations with the university of houston brownian dynamics program. *Computer Physics Communications*, 91(1-3):57–95, 1995.

[24] Federico Fogolari, Alessandro Brigo, and Henriette Molinari. The poisson–boltzmann equation for biomolecular electrostatics: a tool for

structural biology. *Journal of Molecular Recognition*, 15(6):377–392, 2002.

[25] Chad W Hopkins, Scott Le Grand, Ross C Walker, and Adrian E Roitberg. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of chemical theory and computation*, 11(4):1864–1874, 2015.

[26] H. C. Andersen. *J. Chem. Phys.*, 72:2384–2393, 1980.

[27] Michael R. Shirts and John D. Chodera. *J. Chem. Phys.*, 129(12), 2008.

[28] Aleksandr V Marenich, Casey P Kelly, Jason D Thompson, Gregory D Hawkins, Candee C Chambers, David J Giesen, Paul Winget, Christopher J Cramer, and Donald G Truhlar. Minnesota solvation database. *University of Minnesota, Minneapolis*, 2009.

[29] Michel J. Bosisio S., Mey A. S. J. S. *J. Comput. Aided Mol. Des.*