Assessment of Halogen Off-Center Point-Charge Models Using Explicit Solvent Simulations

Andreia Fortuna[†,‡,¶] and Paulo J. Costa[*,†,‡]

†*BioISI – Instituto de Biosistemas e Ciências Integrativas, Faculdade de Ciências, Universidade de Lisboa, 1749-016, Lisboa, Portugal*

‡*Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal*

¶*Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, University of Lisbon, Av. Professor Gama Pinto, 1649-003 Lisbon, Portugal*

E-mail: pjcosta@ciencias.ulisboa.pt,pjcosta@fc.ul.pt

Phone: +351-21-7500196

Abstract

Compounds containing halogens can form halogen bonds (XBs) with biological targets such as proteins and membranes due to their anisotropic electrostatic potential. To accurately describe this anisotropy, off-center point-charge (EP) models are commonly used in force field methods allowing the description of XBs at the molecular mechanics and molecular dynamics level. Various EP implementations have been documented in the literature and despite being efficient in reproducing protein-ligand geometries and sampling of XBs, it is unclear how well these EP models predict experimental properties such as hydration free energies (ΔG_{hyd}), which are often used to validate force field performance. In this work, we report the first assessment of three EP models using alchemical free energy calculations to predict ΔG_{hyd} values. We show that describing

the halogen anisotropy using some EP models can lead to a slight improvement in the prediction of the ΔG_{hyd} when compared with the models without EP, especially for the chlorinated compounds, however, this improvement is not related to the establishment of XBs but is most likely due to the improvement of the sampling of hydrogen bonds (HBs). We also highlight the importance of the choice of the EP model, especially for the iodinated molecules since a slight tendency to improve the prediction is observed for compounds with a larger σ–hole but significantly worse results were obtained for compounds that are weaker XB donors.

1 Introduction

Halogens have a prominent role in drug design due to their known capability to improve druglike properties, *e.g.* membrane permeability,[1,2] but also due to their ability to interact with biological targets such as proteins,[3–5] nucleic acids,[6] and phospholipids of the cell membrane[7] *via* halogen-bonds (XBs).[8] This type of noncovalent interaction (R–X$\cdots$B, with X = Cl, Br, I) arises from the existence of a localized region of depleted electron density at the tip of covalently bound halogens, called σ–hole,[9] which enables their interaction with a Lewis base (B). Moreover, since halogens also typically possess a large negative belt, the formation of hydrogen bonds (HB) with electropositive species is also possible. In Molecular Mechanics (MM) based techniques, various strategies can be applied to describe halogen anisotropy. Polarizable force fields (FFs) provide a more detailed depiction of electrostatics,[10] however, non-polarizable FFs are most widely used to study biological processes and are still preferred due to their lower computational cost. Nonetheless, the latter does not offer a proper description of the σ–hole since single punctual charges are used and halogen atoms are typically assigned a negative charge which leads to repulsive interactions with negative atoms. To overcome this problem, several strategies to emulate the σ–hole in empirical FFs have been explored, including electric multipole expansions, aspherical interatomic potentials, and off-center point charges.[11] The latter methodology, consisting of placing a positive charge,

often called extra point (EP), at a given distance from the halogen along the R–X bond axis to emulate the σ–hole (Figure 1), is the most simple and computationally-cheap strategy available.

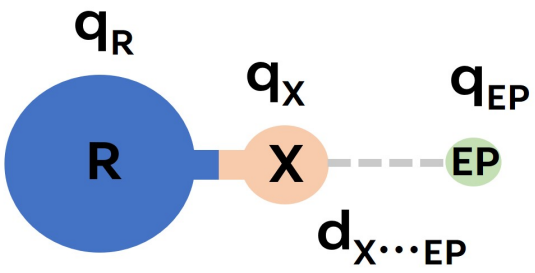

Figure 1: Simplified scheme of an extra point (EP) added to a covalently-bound halogen atom (represented as R–X). The minimal model parameters are the X$\cdots$EP distance ($d_{X\ldots EP}$) and the atomic charges of the EP, halogen, and the remaining particles (q_{EP}, q_X, and q_R, respectively).

The first attempts to emulate the σ–hole of halogen atoms using an EP were developed in the context of the General Amber Force Field (GAFF)[12–14] being later extended to other FFs such as CHARMM,[15] GROMOS,[16] and OPLS.[17,18] These earlier GAFF implementations can be divided into three basic models, herein named ***EP1***, ***EP2***, and ***EP3***. Succinctly, model ***EP1*** places the EP ($d_{X\ldots EP}$) at the value of the R_{min} Lennard-Jones (LJ) parameter of the halogen and assigns RESP partial charges, fitted to all particles.[12,19] The second model (***EP2***), proposed by Sironi and co-workers,[13] is based on a RESP fitting procedure in which $d_{X\ldots EP}$ is assigned by minimizing the error of the fit to the reference quantum-mechanical (QM) electrostatic potential (ESP). In other words, RESP charges are also assigned to the particles but $d_{X\ldots EP}$ is not constant, varying according to the quality of the ESP fit for a given molecule. The third model (***EP3***), also called "no fit" explicit σ–hole (ESH), was proposed by Hobza and co-workers[14] and uses fixed values for $d_{X\ldots EP}$ and for q_{EP}. After the RESP fitting procedure, which is performed without the EP, the value of q_{EP} is subtracted from the halogen charge, while the other atomic partial charges are not modified. A summary

of the parameters mentioned above for ***EP1***, ***EP2***, and ***EP3*** is presented in Table S1.

The hydration free energy, ΔG_{hyd}, is a fundamental property in the regulation of biochemical processes and therefore, is paramount for drug design.[20,21] Since this property is experimentally available[22] and the calculation of hydration free energies is highly sensitive to the molecular mechanical FF parameters used to describe the solvent and the solute molecules, particularly charges,[23] it is often used as a target property for FF validation. The performance of RESP[24–26] but also AM1-BCC charges[25,27,28] in reproducing the ΔG_{hyd} of small molecules has been assessed in several studies. Halogenated compounds are particularly difficult to tackle, as highlighted in reference 26, probably owing to the failure to describe the anisotropy using standard FFs. Indeed, comparing a standard charge model to multipole and hybrid point charge/multipole models for four halobenzenes (X = F, Cl, Br, I) along with the CHARMM FF indicated that a simple description of the halogenated molecules by point charges (no anisotropy) is not suitable to proper describe the thermodynamics of a halogen in water.[29] The authors, however, did not report the performance of simple EP models in their study.

Given the issues raised above, it is paramount that models such as ***EP1***, ***EP2***, and ***EP3*** are evaluated by comparing the experimental ΔG_{hyd} with the calculated ones. Surprisingly, and although this issue was marginally addressed with PBSA calculations,[30,31] such an endeavor was never systematically performed using explicit solvent simulations which are a hallmark for FF validation. Indeed, ***EP1*** was parameterized by finding the $d_{X\ldots EP}$ value that better reproduced the DFT-calculated X$\cdots$B distance on a set of 27 halogen-containing molecules complexed to various Lewis bases.[12] Hydration free energies, relative to benzene, were assessed with and without EP for chloro-, bromo-, and iodobenzene. The anisotropy description improved the accuracy of the calculated values, however, these results were limited to three molecules, and no absolute hydration free energies were reported. Model ***EP2*** was tested in MD simulations of protein complexes with halogenated ligands and reproduced the geometrical parameters obtained from both crystallographic data and hybrid quantum me-

chanics/molecular mechanics calculations[13] whereas hydration energies were not addressed. Model ***EP3*** was based on the comparison of MM dissociation curves with state-of-the-art CCSD(T)/CBS values[14] and was later shown to improve the protein-ligand geometries and XB features in a docking study with halogenated enzyme inhibitors.[32] Again, hydration energies were not estimated.

Despite the above-described parameterization strategies reproducing some experimental features, these were mainly based on structural parameters, and thus, a full assessment of these EP models in reproducing ΔG_{hyd} is yet to be performed. In this work, we evaluate the performance of GAFF-based EP models in the determination of ΔG_{hyd} using alchemical free energy calculations for a large data set of halogenated molecules comprising 107 chlorinated, 23 brominated, and 12 iodinated compounds whose experimental values are available. By assessing the performance of each charge model, we hope to provide the community with a solid ground to validate parameters for halogenated species that are relevant in computer-aided drug design and biomolecular simulations.

2 Methods

2.1 Data Sources

As previously done for PBSA calculations,[30,31] in this work, 142 halogenated molecules were studied including 107 chlorinated, 23 brominated, and 12 iodinated compounds (Figures S1-S3). Experimental ΔG_{hyd} values were retrieved from FreeSolv[22] (version 0.51[33]). Besides the experimental values, FreeSolv also provides calculated ones obtained with GAFF, TIP3P water molecules, and AM1-BCC charges (no EP) which were also used in this work (***AM1-BCC*** model). The list of molecules and the experimental ΔG_{hyd} values along with the calculated ones are provided in Supporting Information (CSV files).

2.2 Charge models

Besides the AM1-BCC charges taken from the FreeSolv (model ***AM1-BCC***), a standard RESP charge calculation was also performed and the resultant model is named ***RESP***. Then, three different off-center point-charge implementations, termed ***EP1***, ***EP2***, and ***EP3***, were studied. In all cases, an EP is added along the C–X covalent bond axis at a given $d_{X\cdots EP}$ distance with the C–X$\cdots$EP angle being fixed at 180°. Table S1 summarizes the parameters of each model as described in the Introduction. The starting three-dimensional coordinates for charge fitting were obtained through B3LYP/6-311G(d,p) geometry optimizations, whereas restrained ESP (RESP) charges[34] were based on electrostatic potentials (ESP) generated at the HF/6-31G(d)[35–37] level of theory for all elements, with the exception of iodine, where the 6-311G(d)[38] basis set was used instead. The full procedure for generating the charges of the models is fully described elsewhere.[30,31]

2.3 MD simulations and alchemical free energy calculations

The hydration free energies (ΔG_{hyd}) were obtained through alchemical free energy calculations using a decoupling approach (starting from the solute in solution and ending with the solute in vacuum) with 20 intermediate states (λ) and a 1-1-6 form of the softcore potential as described by Mobley and co-workers.[28] The first 5 states correspond to the gradual turning-off of the electrostatic interactions while in the remaining 15 lambdas, the Lennard-Jones terms are slowly switched off. This protocol was previously used to generate the calculated ΔG_{hyd} values reported in FreeSolv (model ***AM1-BCC***). Herein we used GROMACS version 2020.6 (CPU and GPU implementation)[39] along with the new *2df* type virtual site to model the EPs, when necessary. Contrarily to previous versions, this virtual site type allows for the introduction of the EP at a C–X$\cdots$EP angle of 180° and at a fixed X$\cdots$EP distance ($\mathrm{d_{X\cdots EP}}$). In older versions, the X$\cdots$EP distance was defined as a function of the C-X value which could lead to instabilities during the simulations due to fluctuations in $\mathrm{d_{X\cdots EP}}$ that, in some cases, led to crashes of the virtual particle with the water molecules, especially in

the case of iodine. The results published in reference 28 (model ***AM1-BCC***) used GROMACS 4.6.7, however, the values are statistically similar between GROMACS versions or CPU/GPU implementation (see Table S2).

For each value of λ, energy minimization, equilibration, and production runs were conducted. For the minimization step, the steepest descent method was used with a force threshold of 100.0 kJ mol^{-1} nm^{-1} and a maximum step size of 0.01 Å. Then, Langevin dynamics during 50 ps in the canonical ensemble (NVT) at 298.15 K followed by 50 ps in the isothermal–isobaric ensemble (NPT) with the Berendsen barostat (1 bar) and by an extra 50 ns in NPT with the Parrinello-Rahman barostat, were performed. A time step of 2 fs was used in all equilibration steps. Finally, three independent production runs of 5 ns were performed using the same conditions as in the last step of equilibration, thus yielding a total of 213 μs of simulation for all systems and charge models. The solutes were solvated with TIP3P water molecules in a cubic box using three-dimensional periodic boundary conditions and the minimum image convention and the box edges were at least twice the Lennard-Jones cutoff distance, following the good practices.[40] Nonetheless, for consistency, the maximum number of water molecules used per compound corresponded to the one reported in FreeSolv.[33] The van der Waals interactions were truncated at 1.0 nm, while for electrostatics, a cutoff of 1.0 nm for direct contributions was used and long-range electrostatic interactions were treated using the fast smooth Particle-Mesh Ewald (PME) method. Bonds containing hydrogen atoms were constrained using the P-LINCS algorithm. Sample input files can be found on the FreeSolv page.[33]

2.4 Analysis

The Multistate Bennett Acceptance Ratio (MBAR)[41] method, implemented in the Alchemical Analysis python tool,[42] was used to obtain the hydration free energies. The calculated ΔG_{hyd} values are the average of the three replicates whereas the errors ($\approx$ 95% confidence) correspond to $\pm$ 2$\times$SEM (SEM = standard error of the mean). The experimen-

tal uncertainties are not reported for most of the compounds in our data set (an arbitrary default value was assigned) which impairs their use for statistical purposes.

The mean absolute error (MAE, eq. 1) and the root mean square deviation (RMSD, eq. 2) of the calculated values against the experimental ones were used to evaluate the accuracy of the ΔG_{hyd} values.

$$\mathrm{MAE} = n^{-1} \sum_{i=1}^{n} |\Delta G_{\mathrm{hyd}}(\mathrm{calc})_i - \Delta G_{\mathrm{hyd}}(\mathrm{exp})_i| \tag{1}$$

$$\mathrm{RMSD} = \sqrt{n^{-1} \sum_{i=1}^{n} (\Delta G_{\mathrm{hyd}}(\mathrm{calc})_i - \Delta G_{\mathrm{hyd}}(\mathrm{exp})_i)^2} \tag{2}$$

The number of halogen bonds (XBs) and hydrogen bonds (HBs) established/accepted by the halogens in the selected compounds were determined using a free energy criterion as described in reference 7, which was inspired by a previous work by Baptista and co-workers.[43] Briefly, this method is based on the representation of the configurational space as a free energy surface using the halogen/hydrogen bond distances ($\mathrm{d_{X\cdots O}}$) and angle (C–X$\cdots$O or O–H$\cdots$X, respectively) as coordinates. XBs/HBs are assigned as those configurations that belong to the XB/HB minimum, if existent. For this purpose, plain MD simulations using the same protocol as mentioned above were performed (3 replicates, 5 ns each).

3 Results and Discussion

The hydration free energies (ΔG_{hyd}) were calculated using explicit solvent simulations for a set of 142 halogenated molecules taken from the FreeSolv database. Five different charge setups were tested in this study, namely, two models without EP addition (models ***AM1-BCC*** and ***RESP***) and three off-center point charge implementations (***EP1***, ***EP2***, and ***EP3***). A discussion of the performance of each charge model is presented in the following sections.

3.1 Models without a EP: *AM1-BCC* and *RESP*

In order to understand the impact of the description of the halogen anisotropy *via* the usage of a EP in the determination of ΔG_{hyd}, we must first look at the performance of charge models with no halogen anisotropy representation. Thus, we calculated the ΔG_{hyd} using plain RESP charges (model ***RESP***), and the correlation of these values with the experimental ones is presented in Figure 2 along with the corresponding MAE and RMSD values. For

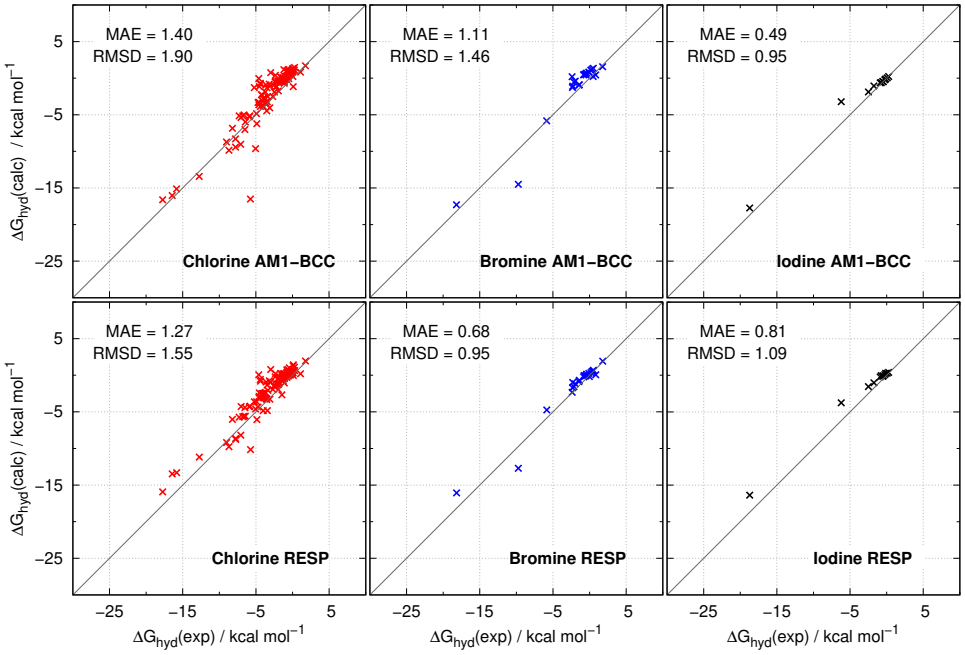

Figure 2: Correlation between calculated and experimental ΔG_{hyd} values using ***RESP*** and ***AM1-BCC*** charge models for each subset of halogenated compounds. The inset shows the MAE and the RMSD values obtained (kcal mol^{-1}).

comparison purposes, the same correlation is also shown for the AM1-BCC charges taken from reference 28 (model ***AM1-BCC***). The performance of both charge models is quite satisfactory (MAE $<$ 1.40 kcal mol^{-1}, RMSD $<$ 1.90 kcal mol^{-1}), however, ***RESP*** leads to lower MAE and RMSD values for the chlorinated and brominated compounds when compared with ***AM1-BCC***, while the opposite trend is observed for the iodinated molecules for which ***RESP*** led to a considerably higher MAE value (0.81 *vs* 0.49 kcal mol^{-1}), while the

RMSD is relatively similar (1.09 *vs* 0.95 kcal mol^{-1}). It is not unprecedented that the less computationally demanding AM1-BCC charges outperform RESP or more computationally expensive *ab initio* methods.[25,44,45] This might be due to the adjustments performed on some bond charge corrections (BCC) to better reproduce relative solvation energies,[46] though is not immediately clear why the same does not occur for chlorinated and brominated molecules. Regardless of the halogen, the majority of the compounds possess an absolute deviation below 2 kcal mol^{-1} (Figures S4 and S5), however, there is a tendency to overestimate ΔG_{hyd}, *i.e.*, the calculated values tend to be more positive than the experimental ones (Figure S4).

The highest deviations are observed for chlorinated compounds 1,2,3,4-tetrachloro-5-(2,3,4,6-tetrachlorophenyl)benzene ($\approx$ 5 kcal mol^{-1}, ***RESP***) and dialifor (≈ -11 kcal mol^{-1}, ***AM1-BCC***) whereas in the case of the iodinated molecules, 2-iodophenol had the highest deviation (2.45 kcal mol^{-1} and 2.98 kcal mol^{-1} for ***RESP*** and ***AM1-BCC***, respectively). Interestingly, this compound had also the highest deviations in PBSA calculations using different radii setups and EP implementations.[30,31] For the brominated compounds, ***RESP*** and ***AM1-BCC*** models yielded the highest deviation for bromacil, in this case underestimating ΔG_{hyd} (deviation of -2.97 kcal mol^{-1} and -4.77 kcal mol^{-1}, respectively). Bromacil was also an outlier in the previous PBSA studies.[30,31] A curious fact occurs for 5-iodouracil ($\Delta G_{\text{hyd}}(\text{exp}) = -18.72$ kcal mol^{-1}) and 5-bromouracil ($\Delta G_{\text{hyd}}(\text{exp}) = -18.17$ kcal mol^{-1}). Using ***RESP***, the deviations are 2.34 kcal mol^{-1} and 2.10 kcal mol^{-1}, respectively. However, the values drop to 0.98 kcal mol^{-1} and 0.87 kcal mol^{-1} with ***AM1-BCC***, suggesting that the bond charge correction describes properly this class of compounds.

3.2 Off-center point-charge model EP1

This charge model places the EP at the respective R_{min} value of the halogen and atomic partial charges for all atoms and for the EP are obtained using a RESP fitting procedure (Table S1 and Figure 1). On average, more positive charges are assigned to the EP of iodine, following the order of halogen size (I > Br > Cl), as seen in Figure S6 (bottom).

The correlation between calculated and experimental ΔG_{hyd} values, as well as the MAE and RMSD obtained using this method, are presented in Figure 3, top. Since this EP model also uses RESP charges, we can compare these results with ***RESP*** to see directly the impact of the EP in the prediction of ΔG_{hyd} values.

Considering the chlorinated compounds, a MAE of 0.93 kcal mol^{-1} and a RMSD of 1.27 kcal mol^{-1} were achieved, which improves the values obtained with ***RESP*** without EP (1.27 and 1.55 kcal mol^{-1}, respectively). The performance is also better than the ones obtained with ***AM1-BCC***. However, such improvement was not observed for the brominated compounds, for which the MAE (0.64 kcal mol^{-1}) and RMSD (0.93 kcal mol^{-1}) values are comparable to ***RESP*** (0.68 and 0.95 kcal mol^{-1}, respectively). The same behavior is observed for the iodinated molecules, for which ***EP1*** led to a MAE of 0.87 kcal mol^{-1} and RMSD of 0.97 kcal mol^{-1}, very similar to the ***RESP*** values (0.81 kcal mol^{-1} and 1.09 kcal mol^{-1}, respectively). This is surprising since the strength of the σ–hole should be larger in heavier halogens and thus, a proper description of the anisotropy should be more important. We will come back to this point later.

Overall, using this EP model, a tendency to underestimate ΔG_{hyd} (the values are too negative) of the iodinated and brominated molecules is observed while for chlorinated molecules a slight underestimation was obtained (Figure S7). For the latter set, the errors are more normally distributed around zero (Figure S8). Again, 2-iodophenol presents the largest deviation between calculated and experimental values (1.66 kcal mol^{-1}) for the iodinated set, though smaller than the one observed for both ***RESP*** and ***AM1-BCC***. This along with 5-iodouracil, with a 0.71 kcal mol^{-1} shift, are the only ones presenting positive deviations but systematically improving the prediction when compared with ***RESP*** and ***AM1-BCC***. For the remaining molecules (halogenated aliphatic compounds), the calculated ΔG_{hyd} are more negative than the experimental ones. This is highlighted by 2-iodopropane for which a deviation of -1.37 kcal mol^{-1} was found. In this case, the description of the halogen anisotropy led to a worst prediction of the ΔG_{hyd} value when compared with ***RESP*** (0.39 kcal mol^{-1}).

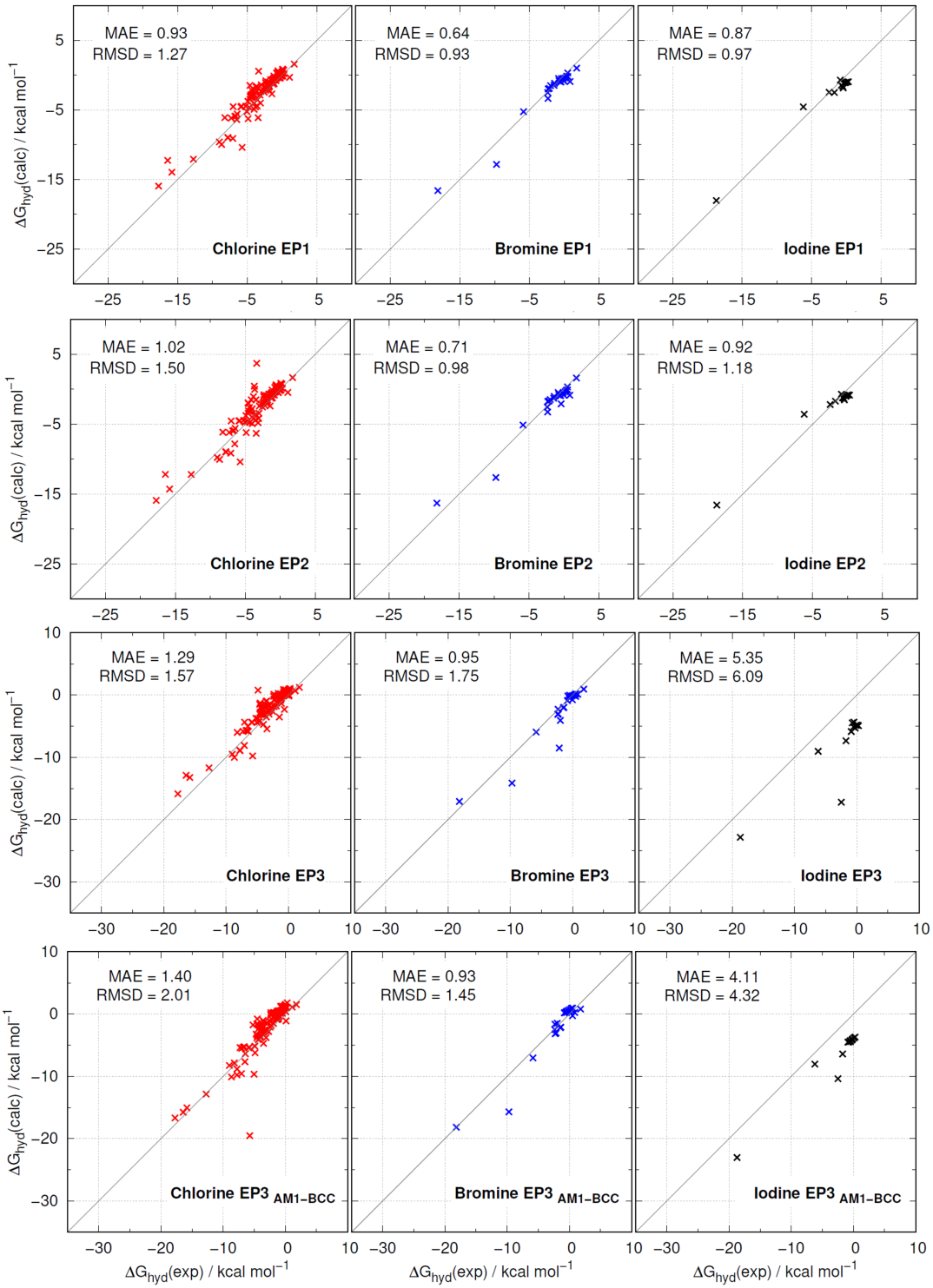

Figure 3: Correlation between calculated and experimental ΔG_{hyd} values using ***EP1***, ***EP2***, ***EP3***, and ***EP3***$_{\text{AM1-BCC}}$ for each subset of halogenated compounds. The inset shows the MAE and the RMSD values obtained (kcal mol^{-1}).

Considering the brominated compounds, bromacil was again the worst outlier leading to a deviation of -3.11 kcal mol^{-1}, very close to the one obtained with ***RESP*** (-2.97 kcal mol^{-1}). The addition of an EP to 5-bromouracil also improved the prediction when compared with ***RESP*** (1.56 and 2.1 kcal mol^{-1}, respectively). As before, dialifor had the highest deviation from the experimental value for the chlorinated compounds, with a deviation of -4.64 kcal mol^{-1}, similar to what was obtained without EP (-4.42 kcal mol^{-1}). Nonetheless, and given the size of the sample of chlorinated compounds, most predictions were improved with the addition of ***EP1*** as reflected by the lower MAE and RMSD values.

Overall, the ***EP1*** model is efficient at predicting the ΔG_{hyd} values, increasing the performance for the chlorinated molecules while keeping a similar performance for brominated and iodinated compounds when compared with the ***RESP*** model (without EP). We will further analyze this issue below.

3.3 Off-center point-charge model EP2

The ***EP2*** model is inspired by the work of Sironi and co-workers,[13] in which the EP is placed at a distance that yields the best fit to the QM electrostatic potential. Thus, for each compound, a specific X$\cdots$EP distance is assigned along with RESP atomic charges for all atoms (including the EP). In this case, the charges attributed to the EP do not follow a clear trend regarding the halogen size as observed for ***EP1*** (Figure S6).

In Figure 3, the calculated (***EP2***) versus the experimental values are presented as well as the MAE and RMSD obtained. The use of ***EP2*** model has a very similar performance to ***EP1***. As for the previous EP, there is a slight tendency to underestimate ΔG_{hyd} (Figure S9-Figure S10) but curiously, the largest outliers 2-iodophenol and 5-iodouracil have positive deviations (2.64 kcal mol^{-1} and 2.14 kcal mol^{-1}, respectively). Indeed, this method performs considerably worse than ***EP1*** for the latter compound where the deviation was only 0.71 kcal mol^{-1}. Considering the brominated compounds, bromacil is again the most difficult compound to tackle, the calculated ΔG_{hyd} being 2.90 kcal mol^{-1} more negative than the

experimental value (-9.73 kcal mol^{-1}). For the chlorinated set, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) was the worst outlier (deviation of 7.08 kcal mol^{-1}), whereas dialifor has the second largest deviation but with opposite sign (−4.66 kcal mol^{-1}).

3.4 Off-center point-charge model EP3

The ***EP3*** model, also called "no fit" explicit σ–hole (ESH), was proposed by Hobza and co-workers.[14,32] Herein, a RESP fitting procedure without the EP addition is used. An EP is added afterward using specific $d_{X\ldots EP}$ and q_{EP} values (Table S1) while subtracting q_{EP} from the halogen charge.[32] Since the addition of the EP and correction of charges is made *a posteriori*, we also tested this method using the AM1-BCC charges provided by FreeSolv (henceforth named ***EP3***$_{\text{AM1-BCC}}$).

By looking at the MAE and RMSD values obtained when using these models (Figure 3, bottom) it is clear that for the iodinated compounds they perform exceptionally worst than for the other halogens and other EP models, the MAE reaching 5.35 kcal mol^{-1} using ***EP3*** with RESP charges. Additionally, for chlorinated and brominated compounds the ***EP3*** values with RESP charges also do not outperform the other models (with and without EP). Curiously, ***EP3***$_{\text{AM1-BCC}}$ leads to higher MAE and RMSD values for the chlorinated and brominated molecules when compared to the original ***EP3*** RESP model, while for iodine, lower values were obtained. This behavior is consistent with what was observed regarding the performance of the RESP charges versus the AM1-BCC without an EP.

A substantial underestimation of ΔG_{hyd} values (too negative) is obtained for the iodinated compounds (Figures S11-S14), the largest being observed for diiodomethane using ***EP3*** and ***EP3***$_{\text{AM1-BCC}}$ charges, with a negative deviation (-14.72 kcal mol^{-1} and -7.88 kcal mol^{-1}, respectively) from its experimental value (ΔG_{hyd}(exp)= -2.49 kcal mol^{-1}). Iodobenzene had the second largest deviation (-5.59 kcal mol^{-1} and -4.95 kcal mol^{-1}) for the experimental value (ΔG_{hyd}(exp)= -1.74 kcal mol^{-1}), further highlighting the difficulty in the prediction of ΔG_{hyd} for this molecule. Such a large underestimation was not observed for the bromi-

nated compounds (Figures S11-S14) for which bromoform (-6.38 kcal mol^{-1}) and bromacil (-5.97 kcal mol^{-1}) presented the largest deviations for ***EP3*** and ***EP3***$_{\text{AM1-BCC}}$, respectively. Interestingly, the usage of both ***EP3*** and ***EP3***$_{\text{AM1-BCC}}$ leads to a slight overestimation of the calculated ΔG_{hyd} chlorinated molecules (Figures S11-S14). For ***EP3***, the largest outlier (5.65 kcal mol^{-1}) was methanesulfonylchlorid whose experimental ΔG_{hyd} value is -4.87 kcal mol^{-1} whereas for ***EP3***$_{\text{AM1-BCC}}$, dialifor was again largely miss-predicted with a deviation of -13.80 kcal mol^{-1} from the experimental value (-5.74 kcal mol^{-1}).

3.5 ΔG_{hyd} and the electrostatic potential maximum on the halogen ($\mathbf{V}_{S,max}$)

In the previous section, we have seen that the inclusion of EPs typically leads to lower or similar MAE/RMSD values when compared with the ***RESP*** model without EP (with the exception of ***EP3*** models). We might expect that this improvement would be more substantial for heavier halogens due to an increase of the σ-hole, however, that was not observed. Nonetheless, it is crucial to assess the correlation between ΔG_{hyd} predictions and the strength of the σ-hole within each set. Curiously, in models lacking an EP, such as ***AM1-BCC*** and ***RESP***, there is no clear systematic increase in the error with an increase in the maximum of the electrostatic potential on the halogen ($V_{S,max}$), which measures the strength of the σ-hole (see Figure S15). Indeed, the ΔG_{hyd} of several compounds presenting large $V_{S,max}$ values are very well predicted even when no halogen anisotropy is included in the model. When an EP is introduced, a slight tendency to decrease the difference between the calculated and experimental ΔG_{hyd} values with an increase of the $V_{S,max}$ values is observed, especially for models ***EP1*** and ***EP2*** in the brominated and iodinated molecules, as depicted in Figure 4. Both ***EP3*** models systematically lead to negative deviations regardless of the σ-hole strength. These results show that, in general, the calculated ΔG_{hyd} are affected by several factors, including the presence of other substituents. While halogens (and their anisotropy) can have an impact, it is difficult to distinguish their specific effect due to the

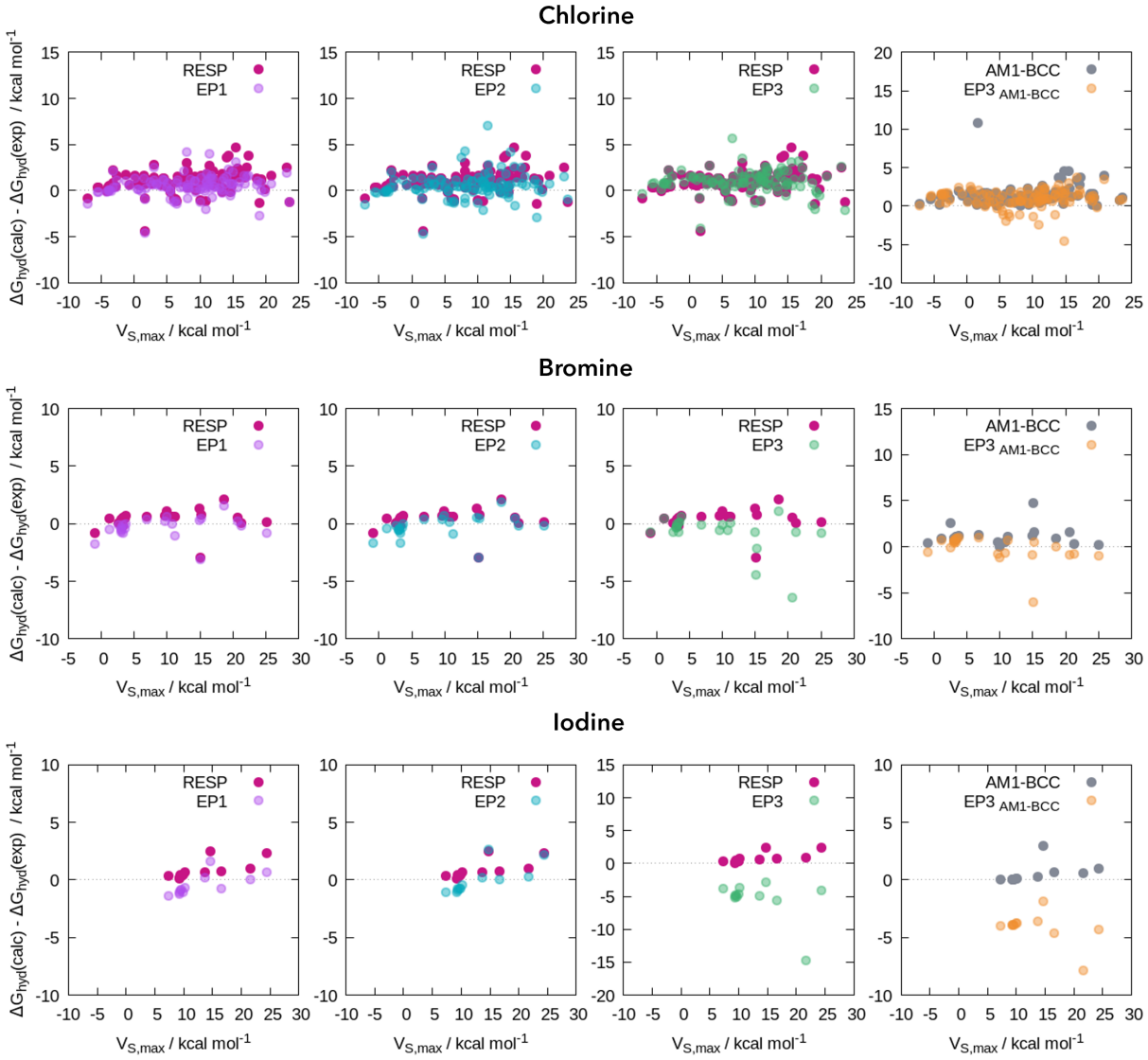

Figure 4: Difference between the calculated values (ΔG_{hyd}(calc)) and the experimental values (ΔG_{hyd}(exp)) as a function of the $V_{S,max}$ for models with EP (***EP1***, ***EP2***, ***EP3***, and ***EP3***$_{\mathrm{AM1\text{-}BCC}}$) compared with their no EP counterparts (***RESP*** or ***AM1-BCC***).

chemical diversity within the sets.

3.6 How accurate are these models for estimating $\Delta\Delta G_{\mathrm{hyd}}$ values?

Till now, we assessed the performance of each charge model on a full dataset of halogenated compounds but the rationalization of the observed trends can be challenging due

to the multiple variables at play, including the varying sizes of the individual libraries and the diversity of compounds represented within them. Fortunately, there are 11 structurally equivalent molecules (see Figure 5) allowing a direct comparison of each method by eliminating differences in sample size and structure.

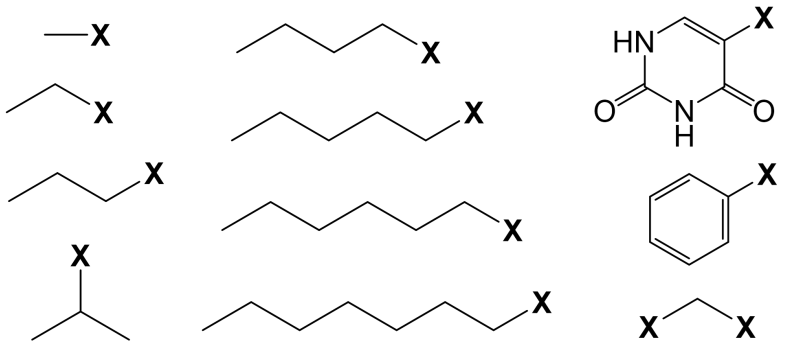

Figure 5: 11 structurally equivalent compounds taken from the main library of 142 compounds studied in this work (X = I, Br, Cl).

The correlations between the calculated and experimental ΔG_{hyd} values for this smaller library (Figure S16 for models without EP, Figures S17-S20 for EP models) show a large drop on the calculated MAE and RMSE values for the chlorinated molecules, except for models ***EP3***, for which these metrics do not decrease significantly upon the reduction of the sample size. For the other halogens, the effect of the smaller sample size is more modest, and overall, the previously observed trends for the full set were maintained, ***EP3*** models performing substantially worse than ***EP1*** and ***EP2*** and an improvement of the MAE values with these two latter models when compared with the ***RESP*** results lacking an EP.

It was previously shown that even for simple molecules such as halobenzene (X = Cl, Br, I), the experimental ΔG_{hyd} trend is not correctly predicted using simple point charge approximations. This occurs for OPLS-AA while with an EP added to this force field (OPLS-AAx) the correct order is restored.[17] The same was shown while comparing a simple point charge model to multipole and hybrid point-charge/multipole models.[29] Our smaller library of 11 compounds offers the possibility to check how accurate are these models at predicting the ΔG_{hyd} trends, i.e., the relative free energies of hydration ($\Delta\Delta G_{\text{hyd}}$) beyond halobenzene.

Table 1 shows the calculated free energies of hydration relative to the chlorinated parent molecule compared with the experimental values, whereas a graphical representation of the trends is presented in Figure S21. Curiously, all models are able to predict the Cl to Br to I substitution in Xbenzene, for which the experimental values are increasingly negative. For halobenzene, the models lacking an EP behave surprisingly well, whereas ***EP1*** exacerbates the negative trend. Model ***EP2*** gives reasonable results while ***EP3*** models, despite predicting the correct order, systematically predict too negative $\Delta\Delta G_{\mathrm{hyd}}$ values, especially for iodinated molecules, and will no longer be considered in this discussion. 5-Xuracil is another curious example. For this drug-like molecule, halogen substitution from chlorine to iodine yields increasingly negative $\Delta\Delta G_{\mathrm{hyd}}$ values, a trend respected by all methods, with ***EP1*** exacerbating the Br to I substitution. The remaining molecules are haloalkanes whose $\Delta\Delta G_{\mathrm{hyd}}$ patterns upon halogen substitution change depending on the chain size. For these molecules, the trend is not well predicted using ***RESP*** or ***AM1-BCC***, and the correct $\Delta\Delta G_{\mathrm{hyd}}$ is also not obtained using a EP, except for 1-Xbutane for which ***EP2*** produces the correct order. Overall, ***EP2*** process the less amount of mispredicted trends (values in bold in Table 1), though the difference for ***EP1*** and both ***RESP*** or ***AM1-BCC*** is not substantial.

3.7 ΔG_{hyd} and its relation with hydrogen and halogen bond formation in solution

Using the small library of 11 structurally similar compounds, we further investigated the influence of each charge model on the molecular solvation structure by investigating the formation of halogen and hydrogen bonds between the solute and the water molecules. Standard MD simulations were therefore performed and the $\mathrm{d_{X\cdots O}}$ distances along with the C–X$\cdots$O or O–H$\cdots$X were monitored, allowing to build free energy landscapes of the XB/HB configurational space, as represented in Figures S22-S24 for the XBs of 5-Xuracil. For instance, for 5-iodouracil (Figures S22), all models bearing an EP show a free energy

Table 1: Free energies of hydration relative to the chlorinated molecule ($\Delta\Delta G_{\mathrm{hyd}}$/ kcal mol^{-1}) calculated for each charge model and compared with the experimental value. In bold are highlighted the mispredicted trends. Red and green cells correspond to the worse and best calculated $\Delta\Delta G_{\mathrm{hyd}}$ when compared with the experimental value, excluding ***EP3*** models.

Compound	**X**	**exp**	***AM1-BCC***	***RESP***	***EP1***	***EP2***	***EP3*$_{\mathbf{RESP}}$**	***EP3*$_{\mathbf{AM1\text{-}BCC}}$**
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xpropane	Br	-0.23	-0.39	-0.13	-0.52	-0.42	-0.84	-0.86
	I	-0.20	**-1.41**	**-0.16**	**-1.16**	**-0.85**	**-5.51**	**-5.46**
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xbutane	Br	-0.24	-0.29	-0.16	-0.37	-1.97	-0.96	-0.83
	I	-0.09	**-0.39**	**-0.22**	**-0.95**	-0.88	**-5.91**	**-5.50**
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xheptane	Br	0.05	**-0.25**	**-0.12**	**-0.35**	**-0.29**	**-1.00**	**-0.86**
	I	-0.02	-1.24	-0.24	-1.14	-0.97	-5.95	-5.47
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xhexane	Br	0.18	**-0.18**	**-0.12**	**-0.40**	**-0.34**	**-0.90**	**-0.82**
	I	0.08	-1.22	-0.18	-0.99	-0.76	-5.65	-5.37
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xpentane	Br	0.00	**-0.26**	**-0.15**	**-0.42**	**-0.43**	**-0.86**	**-0.82**
	I	-0.04	-1.19	**-0.13**	-1.01	-0.83	-5.61	-5.45
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1-Xpropane	Br	-0.23	-0.39	-0.13	-0.52	-0.42	-0.84	-0.86
	I	-0.2	**-1.41**	**-0.16**	**-1.16**	**-0.85**	**-5.51**	**-5.46**
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5-Xuracil	Br	-0.43	-0.69	-0.14	-0.67	-0.38	-1.22	-1.49
	I	-0.98	-1.13	-0.45	-2.08	-0.67	-6.97	-6.36
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X-ethane	Br	-0.11	-0.29	-0.22	-0.47	-0.41	-0.72	-0.84
	I	-0.11	**-1.39**	-0.22	**-1.12**	**-0.88**	**-5.03**	**-5.52**
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Xbenzene	Br	-0.34	-0.48	-0.40	-0.73	-0.54	-1.54	-1.53
	I	-0.62	-0.59	-0.54	-1.74	-0.93	-6.83	-5.73
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Xmethane	Br	-0.27	-0.16	-0.30	-0.43	-0.43	-1.26	-0.87
	I	-0.34	-1.40	-0.31	-0.65	-0.66	-6.23	-5.57
	Cl	0.00	0.00	0.00	0.00	0.00	0.00	0.00
diXmethane	Br	-0.65	-0.45	-0.82	-0.81	-0.81	-3.98	-1.96
	I	-1.18	-1.92	-1.20	-1.76	-1.51	-17.11	-10.84

minimum at $d_{X\cdots O} < 3.5$ Å and C–X$\cdots$O $> 140°$, indicating the presence of XBs whereas, for models without EP, this minimum is absent (no XBs are formed). The existence of such

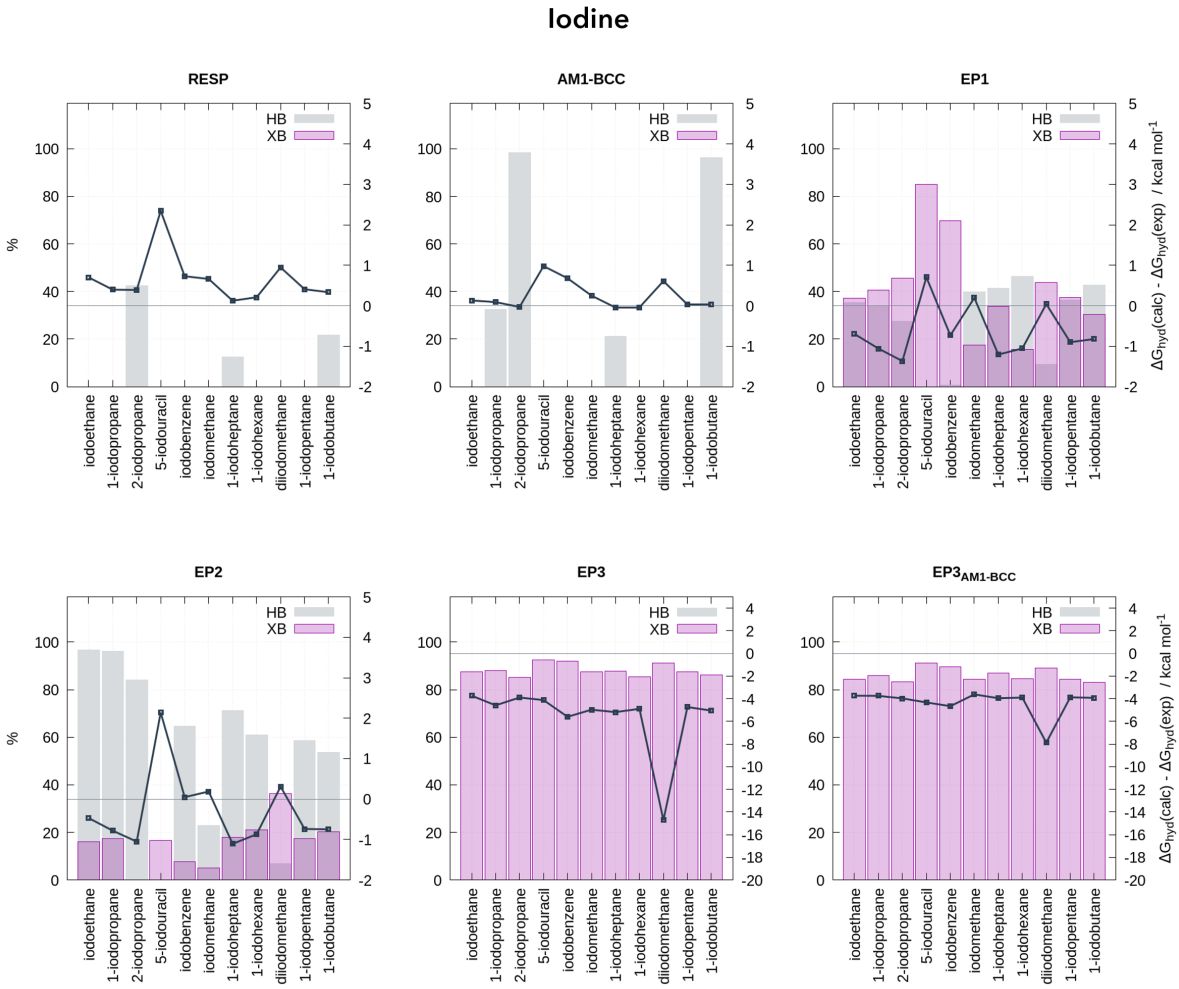

Figure 6: Percentage of halogen bonds (XB, purple) and hydrogen bonds (HB, gray) sampled during the MD trajectory for each iodinated compound in the smaller set of structurally identical halogenated compounds. The difference between the experimental and calculated ΔG_{hyd} value is also shown as line points (black). The scale for the ΔG_{hyd}(calc)-ΔG_{hyd}(exp) in the ***EP3*** models is different from the remainder plots.

minima on these landscapes allows us to quantify the percentage of XBs/HBs formed during the simulation time as these are assigned based on the configurations that fall into that basin. The percentage of XBs and HBs formed by the smaller set of iodinated molecules is presented in Figure 6 along with the difference between calculated and experimental ΔG_{hyd} values. The brominated and chlorinated counterparts are represented in Figures S25-S26. As expected, models without an EP (***RESP*** and ***AM1-BCC***) do not sample XBs regardless

of the halogen type but surprisingly, they are also not very prone to accepting O–H$\cdots$X HBs with water in spite of the negative charges assigned to the halogen atoms. With this regard, ***AM1-BCC*** charges tend to sample much less HBs than ***RESP*** for brominated and chlorinated molecules (Figures S25-S26) whereas the inverse trend is observed in the iodinated set (Figure 6). This enhanced capacity for hydrogen bonding could explain the best performance of ***AM1-BCC*** for iodinated compounds and ***RESP*** for the lighter halogens, but such tendency is only observed for chlorinated compounds for which a large increase in HBs leads to a smaller difference between the calculated and experimental ΔG_{hyd} values and concomitantly, lower MAEs.

We now turn our attention to the EP models. Regarding the formation of XBs, all EP models follow the expected trend: few or no XBs sampled for chlorinated molecules, moderate to high number of XBs present for the brominated ones while iodinated molecules present a very large number of XBs sampled. The low to nonexistent formation of stable XBs in the chlorinated molecules upon EP addition indicates that the improvement in the prediction of ΔG_{hyd} when going from a simple point charge model to the EP model may not be related to the actual sampling of stable XBs but rather seems caused by a better sampling of HBs.

For iodinated compounds the introduction of the electronic anisotropy *via* the EP leads to the establishment of XBs, however, for compounds with lower $V_{S,max}$ and thus, with a lower capacity to perform these interactions, the prediction of the ΔG_{hyd} got significantly worse and led to more negative ΔG_{hyd} values than the experimental ones. This is especially true for ***EP1*** (Figure 6) for which the XB population is $> 20\%$ for all compounds. Notice that most compounds from this smaller library correspond to apolar aliphatic molecules and thus, the σ–hole is not activated. Therefore, the strength of the XBs is expected to be lower than that found in drug-like aromatic compounds such as 5-iodouracil. This molecule is indeed particularly interesting as it is more similar to those expected to be found in drug design studies. When both ***RESP*** and ***AM1-BCC*** models are used, a deviation (peak) is

seen between the calculated and experimental values. No HBs are sampled in this scenario. With ***EP1***, XBs are highly sampled whereas HBs were not, leading to a smaller peak. With ***EP2***, the peak increases again which seems related to a poor sampling of both HBs and XBs. Curiously, for 5-bromouracil and 5-chlorouracil (Figures S25-S26), both the sampling and lack of sampling of XBs with ***EP1*** and ***EP2***, respectively, lead to a peak in the deviation. Again, the lack of proper sampling of HB seems relevant.

Finally, we address the correlation between the charges assigned to the EP and halogen with the number of XBs sampled (Figure S27). For ***EP1***, an increase in q_{EP} leads to a concomitant increase in the % of XBs. Such effect is less visible when using ***EP2*** for wich such relation is not observed. This indicates that besides the EP charge, its distance ($d_{X\cdots EP}$) is crucial for the behavior of the halogen in solution. Indeed, for certain $d_{X\cdots EP}$ values, even when the q_X is lower and q_{EP} is higher, the sampling of XB is significantly lower. Using again 5-iodouracil as an example, a lower % of XBs was obtained with ***EP2*** even though this model attributes a significantly lower halogen charge (-0.41 e) and higher EP charge (0.17 e) when compared with ***EP1*** (-0.25 e and 0.08 e, respectively). Notice however that extremely different $d_{X\cdots EP}$ values are used in these models: 2.15 Å *vs* 1.41 Å for ***EP1*** and ***EP2***, respectively. The very large overestimation of the XBs in iodinated compounds when using ***EP3*** seems related to the high EP charge (0.3 e) since, as the EP charge drops in bromine and chlorine, the % of XBs seems more reasonable and comparable to the other models. Curiously, using ***EP3*** PBSA calculations,[31] an MAE of 0.87 kcal mol^{-1} indicating that the explicit interactions with the water molecules are overestimated and that the representation of the solvent by a dielectric constant attenuates the problem (no explicit sampling of XBs and HBs are possible). All the above results seem to point that if one wishes to use this simple approach of adding off-center point-charges to represent the halogen anisotropy, not only the charge/distance dependence of the EP must be carefully taken into account but also the Lennard-Jones parameters of the specific pair (in this case X$\cdots$O) should be tuned, for instance, to tackle the oversampling of XBs which typically lead to an underestimation of

ΔG_{hyd}. However, the lack of microscopic experimental data, *e.g.* the water structure around the halogenated molecules, might impair a proper assessment of the competing features (XBs *vs* HBs).

4 Conclusions

In this work, we focused on assessing the behavior of off-center point-charge models (EPs), used to describe halogen anisotropy in FF simulations, in the calculation of ΔG_{hyd} values using explicit solvent simulations. Such a task had never been performed in the literature and is crucial for the critical validation of the parameters. By using a dataset of 142 halogenated molecules (107 chlorinated, 23 brominated, and 12 iodinated) we compared the calculated ΔG_{hyd} values obtained without an EP and with EP implementations from the literature with the experimental values taken from the FreeSolv database.[33]

Standard RESP charges without the addition of an EP (***RESP*** model) led to smaller MAE values for the brominated and chlorinated molecules when compared with AM1-BCC charges (also without EP, ***AM1-BCC*** model) whereas for iodinated molecules, the opposite was observed. Apart from ***EP3*** models which seem to highly overestimate XBs, the description of the halogen anisotropy using ***EP1*** and ***EP2***, typically leads to an improvement of the MAE values for the chlorinated molecules whereas for brominated and iodinated compounds, the errors are equivalent to those found with the plain ***RESP*** model. For the full set of molecules, ***EP1*** yielded slightly better results, however, no relation with the electrostatic potential maximum on the halogen ($V_{S,max}$) was observed.

In an attempt to deepen the understanding of the behavior of these models, a smaller set of 11 structurally equivalent compounds was used for further investigation. $\Delta\Delta G_{\text{hyd}}$values show that ***EP1*** or ***EP2*** do not clearly outperform the standard model without EP (***RESP***). By measuring the population of both HB and XBs, we showed that the systematic improvement of the MAE values observed for chlorinated molecules is most likely due to a better

sampling of HBs when an EP is added. On the other hand, the excessive sampling of XBs for the heavier halogen leads to an underestimation of the calculated ΔG_{hyd} values (too negative).

It is clear that for biomolecular simulations, the description of the halogen anisotropy is essential, otherwise, XBs will not be sampled.[5] In this scope, the EP is a simple and versatile way to describe XBs in force field methods. However, in this paper, we showed that caution must be taken, especially for compounds with weaker σ–holes, since the use of an EP can lead to an excessive sampling of XBs. Our work thus provides the first systematic study on the effect of adding off-center point-charges taken from the literature on the calculation of ΔG_{hyd}, thus offering the community an excellent starting point for the parametrization of specific molecules. For instance, the most efficient EP for a given molecule could be chosen as the one yielding the best results for an analog of our dataset. Our results also point out that EPs are a simple, yet limited way to describe a complex electrostatic problem in FF-based simulations, and some investment could be made in the systematic study of more complex models such as quadrupolar electrostatics.[29] In parallel, some investment in the tunning of halogen Lennard-Jones parameters should also be performed. Nonetheless, for a proper parametrization, microscopic experimental data, *e.g.* the water structure around the halogenated molecules, allowing to infer the formation of both XBs and HBs, is fundamental.

Supporting Information Available

Supporting Tables and Figures (PDF)

A list of molecules contained in each set, the experimental ΔG_{hyd} values along with the calculated ones (CSV)

Acknowledgement

The authors thank Fundação para a Ciência e a Tecnologia (FCT), Portugal, for a doctoral grant SFRH/BD/146447/2019 (AF), and strategic projects UIDB/04046/2020–UIDP/04046/2020 (BioISI), and UID/DTP/04138/2019 (iMed.ULisboa). FCT is also acknowledged for the Individual Call to Scientific Employment Stimulus contract 2021.00381.CEECIND (PJC). This work was also funded by the European Union (TWIN2PIPSA, GA 101079147). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

References

(1) Gerebtzoff, G.; Li-Blatter, X.; Fischer, H.; Frentzel, A.; Seelig, A. Halogenation of drugs enhances membrane binding and permeation. *ChemBioChem* **2004**, *5*, 676–684.

(2) Ford, M. C.; Ho, P. S. Computational tools to model halogen bonds in medicinal chemistry. *J. Med. Chem.* **2016**, *59*, 1655–1670.

(3) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16789–16794.

(4) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J. Med. Chem.* **2013**, *56*, 1363–1388.

(5) Costa, P. J.; Nunes, R.; Vila-Viçosa, D. Halogen bonding in halocarbon-protein complexes and computational tools for rational drug design. *Expert Opin. Drug Discov.* **2019**, *14*, 805–820.

(6) Frontera, A.; Bauzá, A. Halogen bonds in protein nucleic acid recognition. *J. Chem. Theory Comput.* **2020**, *16*, 4744–4752.

(7) Nunes, R. S.; Vila-Viçosa, D.; Costa, P. J. Halogen bonding: an underestimated player in membrane–ligand interactions. *J. Am. Chem. Soc.* **2021**, *143*, 4253–4267.

(8) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013**, *85*, 1711–1713.

(9) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the σ-hole. *J. Mol. Model.* **2007**, *13*, 291–296.

(10) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annu. Rev. Biophys.* **2019**, *48*, 371–394.

(11) Kolář, M. H.; Hobza, P. Computer modeling of halogen bonds and other σ-hole interactions. *Chem. Rev.* **2016**, *116*, 5155–5187.

(12) Ibrahim, M. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564–2574.

(13) Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M. Halogen bonding in ligand–receptor systems in the framework of classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508–19516.

(14) Kolář, M.; Hobza, P. On extension of the current biomolecular empirical force field for the description of halogen bonds. *J. Chem. Theory Comput.* **2012**, *8*, 1325–1333.

(15) Gutiérrez, I. S.; Lin, F.-Y.; Vanommeslaeghe, K.; Lemkul, J. A.; Armacost, K. A.; Brooks III, C. L.; MacKerell Jr, A. D. Parametrization of halogen bonds in the

CHARMM general force field: Improved treatment of ligand–protein interactions. *Bioorg. Med. Chem* **2016**, *24*, 4812–4825.

(16) Nunes, R.; Vila-Vicosa, D.; Machuqueiro, M.; Costa, P. J. Biomolecular simulations of halogen bonds with a GROMOS force field. *J. Chem. Theory Comput.* **2018**, *14*, 5383–5392.

(17) Jorgensen, W. L.; Schyman, P. Treatment of halogen bonding in the OPLS-AA force field: application to potent anti-HIV agents. *J. Chem. Theory Comput.* **2012**, *8*, 3895–3901.

(18) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L., et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

(19) Ibrahim, M. A. A. Molecular mechanical perspective on halogen bonding. *J. Mol. Model.* **2012**, *18*, 4625–4638.

(20) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(21) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.

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

(23) Jämbeck, J. P.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. Partial atomic charges and their impact on the free energy of solvation. *J. Comp. Chem.* **2013**, *34*, 187–197.

(24) Mecklenfeld, A.; Raabe, G. Comparison of RESP and IPolQ-mod partial charges for solvation free energy calculations of various solute/solvent pairs. *J. Chem. Theory Comput.* **2017**, *13*, 6266–6274.

(25) Shivakumar, D.; Deng, Y.; Roux, B. Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.

(26) Martins, S. A.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Prediction of solvation free energies with thermodynamic integration using the general amber force field. *J. Chem. Theory Comput.* **2014**, *10*, 3570–3577.

(27) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(28) 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.

(29) El Hage, K.; Bereau, T.; Jakobsen, S.; Meuwly, M. Impact of quadrupolar electrostatics on atoms adjacent to the sigma-hole in condensed-phase simulations. *J. Chem. Theory Comput.* **2016**, *12*, 3008–3019.

(30) Nunes, R.; Vila-Viçosa, D.; Costa, P. J. Tackling Halogenated Species with PBSA: Effect of Emulating the σ-hole. *J. Chem. Theory Comput.* **2019**, *15*, 4241–4251.

(31) Fortuna, A.; Costa, P. J. Optimized Halogen Atomic Radii for PBSA Calculations Using Off-Center Point Charges. *J. Chem. Inf. Model.* **2021**, *61*, 3361–3375.

(32) Kolář, M.; Hobza, P.; Bronowska, A. K. Plugging the explicit σ-holes in molecular docking. *Chem. Commun.* **2013**, *49*, 981–983.

(33) Mobley, David L. `http://www.escholarship.org/uc/item/6sd403pz`, 2013; Experimental and Calculated Small Molecule Hydration Free Energies. UC Irvine: Department of Pharmaceutical Sciences, UCI.

(34) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(35) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(36) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

(37) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* basis set for third-row atoms. *J. Comput. Chem.* **2001**, *22*, 976–984.

(38) Glukhovtsev, M. N.; Pross, A.; McGrath, M. P.; Radom, L. Extension of Gaussian-2 (G2) theory to bromine-and iodine-containing molecules: Use of effective core potentials. *J. Chem. Phys.* **1995**, *103*, 1878–1885.

(39) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(40) Gapsys, V.; de Groot, B. L. On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size. *eLife* **2020**, *9*, e57589.

(41) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(42) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.* **2015**, *29*, 397–411.

(43) Campos, S. R.; Baptista, A. M. Conformational analysis in a multidimensional energy landscape: study of an arginylglutamate repeat. *J. Phys. Chem. B* **2009**, *113*, 15989–16001.

(44) Riquelme, M.; Lara, A.; Mobley, D. L.; Verstraelen, T.; Matamala, A. R.; Vohringer-Martinez, E. Hydration free energies in the FreeSolv database calculated with polarized iterative Hirshfeld charges. *J. Chem. Inf. Model.* **2018**, *58*, 1779–1797.

(45) Mobley, D. L.; Dumont, É.; 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.

(46) 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.

TOC Graphic