JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

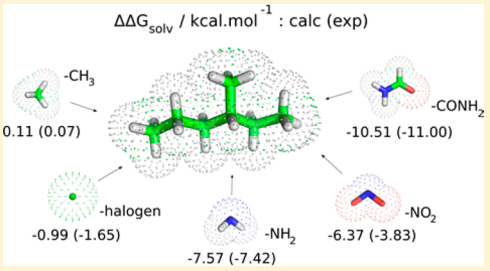
Prediction of Solvation Free Energies with Thermodynamic Integration Using the General Amber Force Field

Silvia A. Martins, Sergio F. Sousa, Maria João Ramos, and Pedro A. Fernandes*

REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal

Supporting Information

ABSTRACT: Computer-aided drug design (CADD) techniques can be very effective in reducing costs and speeding up drug discovery. The determination of binding and solvation free energies is pivotal for this process and is, therefore, the subject of many studies. In this work, the solvation free energy change ($\Delta\Delta G_{solv}$) for a total of 92 transformations in small molecules was predicted using Thermodynamic Integration (TI). It was our aim to compare experimental and calculated solvation free energies for typical and prime additions considered in drug optimizations, analyzing trends, and optimizing a TI protocol. The results showed a good agreement between experimental and predicted values, with an overestimation of the predicted values for CH_3, halogens, and NH_2, as well as an underestimation for $CONH_2$, but all fall within ±1 kcal/mol. NO_2 addition showed a larger and systematic underestimation of the predicted $\Delta\Delta G_{solv}$, indicating the need for special attention in these cases. For small molecules, if no experimental data is available, using TI as a theoretical strategy thus appears to be a suitable choice in CADD. It provides a good compromise between time and accuracy.



INTRODUCTION

Life results from a complex combination of individual chemicals and chemical reactions and one of the primary goals in computer-aided drug design (CADD) is the determination of the binding free energy and of the solvation free energy that play a role in those reactions.[1]

In drug optimization efforts, two goals are expected for the new/improved drug: increased activity and reduced dose levels and/or increased selectivity and reduced side effects. While trying to improve the affinity between a given drug and its receptor, one of the most common strategies in silico approaches is to make small changes by adding suitable substituents. With computational studies, it is possible to simulate a wide variety of substitutions and predict their effect in the protein-affinity of the compound.

Binding and solvation free energies are intricately connected, being the latter frequently required for an accurate determination of the first. The properties of the molecules present in any chemical or biological system are dependent on interactions with the environment, and therefore, special attention needs to be given to the solvation effects. When we consider biochemical reactions, water is the solvent par excellence and it often plays a critical role on them as water can act both as an hydrogen-bond donor and acceptor. The free energy of solvation has a major importance in the determination of solubilities, partition coeficients, association and disassociation, binding constants, phase equilibria, and reaction rates.[2] Free energy is one of the most important quantities in thermodynamics, but it is also a challenging task to calculate it efficiently and accurately. The solvation free energy of small molecules can help to understand desolvation of the ligand in the thermodynamic process of protein−ligand binding,[3−5] playing an important role in the process.

In this study, we have compared experimental and calculated solvation free energies differences ($\Delta\Delta G_{solv}$) for a set of 92 transformations in small molecules, divided in 5 groups according to the substitution group considered: methyl, halogen (F, Cl, Br, and I), amino, amide, and nitro. The experimental values were gathered from different source documents and the computational calculations were performed with Thermodynamic Integration (TI), using explicit solvent. TI is considered a suitable and accurate method to determine free energies. The substitutions evaluated in this study are taken as typical and in the group of the first considered in drug optimizations, with effects in the lipophilicity/hydrophilicity, steric/electronic properties, among others. Since absolute solvation free energies have been experimentally determined for a considerable amount of small molecules, the study allows for a direct comparison between the experimental and calculated values. So, we also intended to test an optimized TI protocol applied previously to HO additions[6] in order to address properly the changes in $\Delta\Delta G_{solv}$ brought by these transformations. It is our aim to present trends and provide a theoretical strategy when there are no experimental data available.

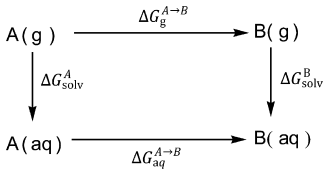
Received: April 22, 2014

   dx.doi.org/10.1021/ct500346y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

METHODOLOGY

Thermodynamic Integration was originally proposed by Kirkwood[7] and is widely used to calculate solvation free energies, with good results even though computationally expensive. It is an equilibrium method that allows the estimation of the free energy differences between two discrete states, an initial reference state (state A), and a final target state (state B). In equilibrium methods, a hybrid system is used to transform system A into B, where an average of values obtained from intermediates are used to calculate the free energy difference. In the thermodynamic integration approach, a path between the states is defined, and by using a thermodynamic cycle, it is possible to computationally measure the energy difference, provided that the path is reversible. Free energy is a state function, and therefore, the transformation process can be chemical or alchemical. The free energy difference between two molecules (A and B) can be obtained using the following thermodynamic cycle, shown in Scheme 1.

Scheme 1. Schematic Representation of the Thermodynamic Cycle Involved in the Calculation of the $\Delta\Delta G_{solv}$ Associated to the Considered Transformations in the Gas-Phase and in Solution, Using TI



In this study, we performed alchemical transformations of a hydrogen atom into another functional group. TI was used here to calculate the solvation free energies changes ($\Delta\Delta G_{solv}$) of a total of 92 of those transformations. Five sets of functional group additions were considered: methyl, halogens, amino, amide, and nitro. In each transformation, a hydrogen atom was substituted by a functional group, transforming molecule A into molecule B as represented by the thermodynamic cycle (in Scheme 1). Using a coupling parameter (λ), it is possible to compute the free energy difference between two states A and B, with the equation:

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

The coupling parameter varies from 0 to 1 corresponding respectively to the initial A and final B states. Basically, the free energy difference, $\Delta G^{A\to B}$, is the integral from 0 to 1 of the expectation value of $\partial V(\lambda)/\partial\lambda$, where V is the potential energy. The integral may be evaluated numerically using a number of discrete λ points.[8−10]

The process was repeated in solution and in the gas-phase. TI calculations were performed using AMBER 10[11] with soft-core potentials and using the general Amber force field (GAFF).[12] All molecules were parametrized using the antechamber[13] module of AMBER with charges derived at the HF/6-31G(d) level of theory. These options represent typical choices for standard organic molecules composed of H, C, N, O, S, P, and halogens, and therefore, these choices are usual when handling drug-like molecules through molecular dynamics simulations, particularly in protein complexes.

For reasons of simulation stability, the transformations have been divided into three substeps each: first, the atomic partial charge on the selected hydrogen was removed ($\Delta G1$); second, the disappearance of the selected hydrogen takes place with the simultaneous appearance of the functional group (the van der Waals (vdW) interactions and radii were transformed from one to the other, $-\Delta G2$); and finally, the atomic partial charge(s) of the substituent group were switched on ($\Delta G3$). We have used soft-core potentials in substep 2, which are modified Lennard-Jones potentials that prevent simulation instabilities due to the truncation of the potential to small energy values for short, very repulsive distances.

In each step, nine λ values were considered (λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9), which is a typical option when using TI to calculate free energy differences.[14−16] For each λ value, the starting structure was minimized for 500 steps. In the substeps 1 and 3 (described above), this was performed using steepest descent followed by conjugate gradient algorithms. In substep 2, only the steepest descent minimization algorithm was used.

Next, the resultant structure was equilibrated for 50 ps, at constant pressure. Production simulations of 1 ns for each λ in the isothermal−isobaric (NPT) ensemble were performed, using the Langevin thermostat[17] with a collision frequency of 1.0 ps^{-1} at 300 K, a time step of 1 fs, and a cutoff of 9 Å for the nonbonded interactions. Final values were integrated numerically using the trapezoidal rule. TI calculations in water were performed with explicit solvent (TIP3P, minimum 12 Å to the box side) and under periodic boundary conditions with PME. This protocol has been previously tested in the determination of $\Delta\Delta G$solv upon OH addition[6] against other computational methods, including the Poisson−Boltzmann (PB) and Generalized Born[18] (GB) methods, the Polarizable Conductor Continuum model (C-PCM),[19,20] the Integral-Equation-Formalism Polarizable Continuum Model (IEF-PCM),[21−23] the Static Isodensity Polarizable Continuum Model (IPCM),[24] the Self-Consistent Isodensity Polarizable Continuum Model (SCI-PCM),[24] and the SMD model.[25]

Several studies focused in solvation free energies prediction,[26−29] creating a large amount of available experimental data. Experimental values were obtained from the literature.[25,30−38] The $\Delta\Delta G_{solv}$ experimental values result from the difference between the molecule with the added functional group (CH_3, F, Cl, Br, I, NH_2, $CONH_2$, and NO_2) and the original one (e.g., $\Delta G_{solv}^{chloroethane} - \Delta G_{solv}^{ethane}$, for the addition of a chloride atom to a primary aliphatic carbon atom).

RESULTS AND DISCUSSION

Data Set. In this study, we have assessed the performance of a thermodynamic integration protocol in the determination of the solvation free energy change ($\Delta\Delta G_{solv}$) for a total of 92 transformations, starting from typical linear, branched, cyclic, aromatic, and heterocyclic alkane molecules by replacing a hydrogen atom by a different substituent.

One transformation typically takes 20 h in an 8 processors Xeon 3.0 GHz machine, with this TI protocol.

These substitutions can be grouped into 5 categories based on the nature of the substituent group: (1) methyl substitution (40 transformations); (2) halogen substitution (29 transformations, including 1 case for fluorine, 13 cases for the chlorine, 9 cases for bromine and 6 for iodine); (3) amino substitution (8 cases); (4) amide substitution (3 cases); (5) nitro substitution (12 transformations). It was our intention to

dx.doi.org/10.1021/ct500346y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

find out if each functional group presents a definite trend or at least a well-defined range of values for the $\Delta\Delta G_{solv}$ contribution. The data were also grouped, for each substitution, according to the specific position of the substitution and the type of chain, to identify characteristic subtrends.

The distribution, together with its subdivisions is presented in detail in Figure 1. This relative number of cases per category reflects the availability of experimental solvation free energy data in the literature.

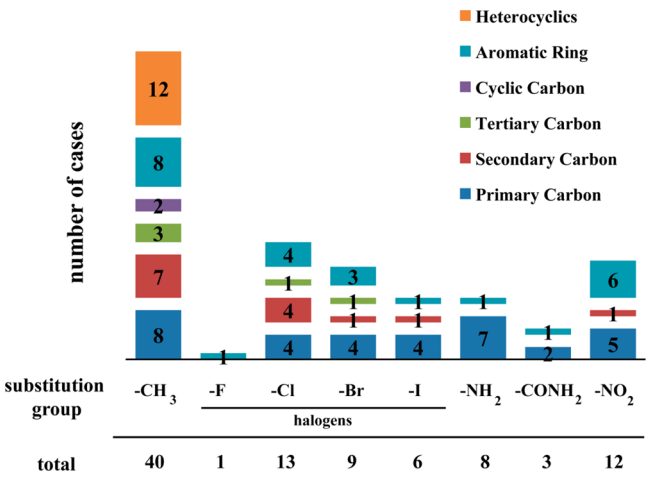


Figure 1. Number of substitutions considered in this study distributed by class.

Distribution of Experimental and Computational Values. Figure 2 presents the experimental and computational values for $\Delta\Delta G_{solv}$ free energies obtained for the different substitution types. Comparing the ranges of experimental and computational values for $\Delta\Delta G_{solv}$ free energies obtained from the addition of different groups to the different compound classes, it is possible to evaluate trends. For the nitro group addition, a displacement (shift) of the computational values to the left was observed, when comparing with the corresponding experimental data. This suggests a systematic underestimation of the predicted $\Delta\Delta G_{solv}$ energies upon NO_2 addition. For the other substitutions, however, the most noticeable feature is considerable larger amplitude in the range of computationally predicted values, when compared with the experimental ones. However, most of the individual values (indicated through vertical bars in both categories) fall in relatively narrow and similar ranges.

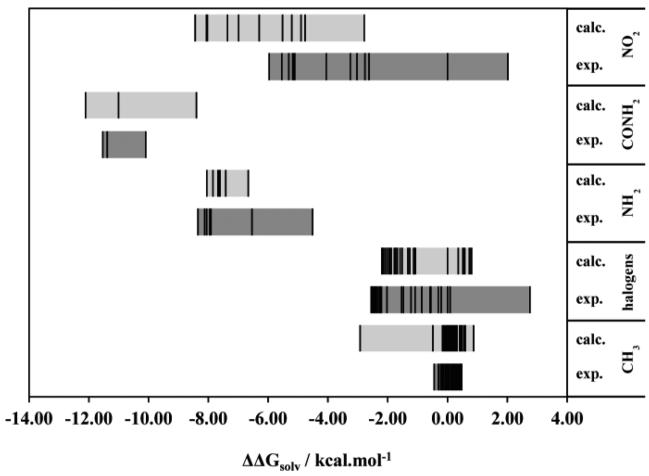


Figure 2. Ranges of experimental and computational values for $\Delta\Delta G_{solv}$ free energies obtained from the addition of different groups to different compound classes. Values expressed in kcal/mol. Lines in bold indicate the individual experimental and computational values.

In fact, in the addition of a methyl group, it is possible to see that almost all computational values are distributed in a similar range as the experimental values, except for one case. This corresponds to the pentane → hexane transformation that presents a predicted value of −2.92 kcal/mol when its experimental value is of 0.17 kcal/mol. With two other cases, halogens and amino group additions, the range of values for predicted $\Delta\Delta G_{solv}$ energies is within the range of the experimental values, and if we analyze the vertical bars, there is also a common distribution of them. There is an experimental value for the toluene → *p*-chlorotoluene transformation that expands the experimental bar to the right, and if we did not consider this value, the computational values would present only a slight shift to the right comparing to the experimental values. In amide group addition, the availability of only 3 cases precludes a substantive definition of a trend.

In general, however, it can be concluded that both alternatives yield characteristic and comparable values for each substitution type, although the computational approach can lead to some outliers.

Comparison of the Mean Averages. Table 1 presents an overview of the mean averages for the experimental and computational values of $\Delta\Delta G_{solv}$ free energies, resulting from the addition of different groups (CH_3, F, Cl, Br, I, NH_2, $CONH_2$, and NO_2). It presents also the corresponding standard deviation, and minimum and maximal values. Figure 3 presents a comparison of the mean average and standard deviation of the experimental and computational values for $\Delta\Delta G_{solv}$ free energies, obtained for the different transformations. It complements the observations made concerning Figure 2. In particular, it confirms the relatively good agreement in the average values for the different transformations, with a higher standard deviation for the computational values.

In CH_3, NH_2, and NO_2 additions the experimental average is above the computational average, whereas with halogens and $CONH_2$ addition the experimental average is below the computational one.

In methyl group addition, the experimental values for this transformation are smaller than 0.5 kcal/mol, therefore leading to a small mean, 0.07 kcal/mol, with an also small variation, 0.21 kcal/mol. We can point out as a reason for that, the negligible effect of this group addition. The predicted values, although presenting a slight higher mean and deviation values (0.11 and 0.55 kcal/mol), also remain in acceptable proximity to the 0.5 kcal/mol.

When considering the halogen group addition, the difference between experimental and computational averages is a little over 0.5 kcal/mol, with a similar standard deviation for both: 1.17 and 1.10 kcal/mol, respectively. Nevertheless, this does not occur with all four different additions within this group. Considering bromide addition, the difference of computational and experimental averages reaches almost 1 kcal/mol, with both standard deviations over 0.5 kcal/mol. With the iodide addition, the differences are larger, nearly 1.5 kcal/mol, with a smaller variation on the experimental values than on the computational ones.

dx.doi.org/10.1021/ct500346y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Experimental and Computational Values of $\Delta\Delta G_{solv}$ Free Energies Resulting from the Addition of Different Groups (CH_3, F, Cl, Br, I, NH_2, $CONH_2$, and NO_2) to Different Compound Classes[a]

	$\Delta\Delta G_{solv}$ exp			$\Delta\Delta G_{solv}$ calc		
addition to	mean avg.	min	max	mean avg.	min	max
CH_3	0.07 ± 0.21	−0.45	0.48	0.11 ± 0.55	−2.92	0.87
primary carbon	0.15 ± 0.12	−0.14	0.25	−0.13 ± 1.10	−2.92	0.87
secondary carbon	0.28 ± 0.09	0.18	0.43	0.02 ± 0.25	−0.49	0.30
tertiary carbon	0.25 ± 0.07	0.19	0.36	0.00 ± 0.04	−0.03	0.05
cyclic carbon	0.44 ± 0.04	0.40	0.48	0.22 ± 0.01	0.21	0.23
aromatic ring	−0.04 ± 0.11	−0.31	0.04	0.34 ± 0.20	−0.07	0.59
heterocyclics	−0.12 ± 0.15	−0.45	0.07	0.19 ± 0.20	−0.17	0.48
halogens	−1.65 ± 1.17			−0.99 ± 1.10		
F	0.08			0.80		
aromatic ring	0.08			0.80		
Cl	−1.40 ± 1.42	−2.46	2.75	−1.24 ± 1.16	−2.20	0.53
primary carbon	−2.35 ± 0.09	−2.46	−2.24	−2.09 ± 0.11	−2.20	−1.92
secondary carbon	−2.19 ± 0.10	−2.27	−2.03	−1.95 ± 0.12	−2.08	−1.75
tertiary carbon	−1.22			−1.88		
aromatic ring	0.28 ± 1.46	−1.09	2.75	0.49 ± 0.07	0.36	0.53
Br	−1.84 ± 0.79	−2.54	−0.56	−0.91 ± 1.08	−1.78	0.75
primary carbon	−2.50 ± 0.05	−2.54	−2.42	−1.71 ± 0.07	−1.78	−1.59
secondary carbon	−2.45			−1.68		
tertiary carbon	−1.47			−1.52		
aromatic ring	−0.89 ± 0.46	−1.54	−0.56	0.61 ± 0.10	0.51	0.75
I	−2.20 ± 0.08	−2.56	−0.86	−0.86 ± 0.72	−1.33	0.73
primary carbon	−2.47 ± 0.08	−2.56	−2.35	−1.19 ± 0.11	−1.33	−1.07
secondary carbon	−2.43			−1.13		
aromatic ring	−0.86			0.73		
NH_2	−7.42 ± 1.21	−8.35	−4.52	−7.57 ± 0.38	−8.05	−6.67
primary carbon	−7.84 ± 0.55	−8.35	−6.54	−7.56 ± 0.41	−8.05	−6.67
aromatic ring	−4.52			−7.65		
$CONH_2$	−11.00 ± 0.65	−11.53	−10.10	−10.51 ± 1.56	−12.11	−8.40
primary carbon	−11.46 ± 0.07	−11.53	−11.38	−11.56 ± 0.55	−12.11	−11.01
aromatic ring	−10.10			−8.40		
NO_2	−3.83 ± 2.09	−5.97	2.01	−6.37 ± 1.68	−8.44	−2.79
primary carbon	−5.43 ± 0.30	−5.97	−5.15	−7.78 ± 0.75	−8.44	−6.30
secondary carbon	−5.11			−7.36		
aromatic ring	−2.29 ± 1.98	−4.06	2.01	−5.03 ± 1.24	−6.99	−2.79

[a]Mean average values for each group are presented in bold. Values expressed in kcal/mol.

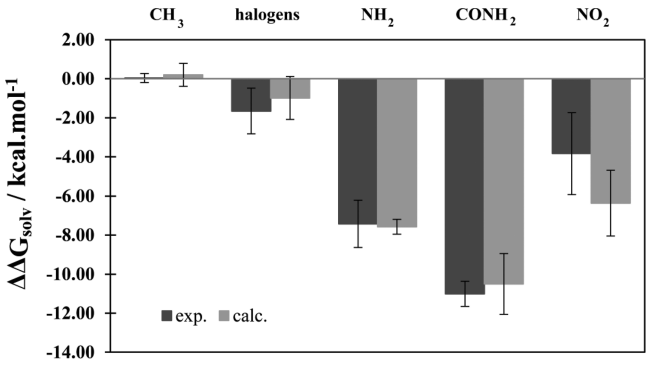


Figure 3. Comparison of averages and deviations of experimental and computational values for $\Delta\Delta G_{solv}$ free energies obtained from the addition of different groups to different compound classes. Values expressed in kcal/mol.

The addition of an amino group ($-NH_2$) presents very similar experimental and calculated averages, −7.42 and −7.57 kcal/mol respectively, with a higher standard deviation in the experimental values. The transformation benzene → aniline exhibits the greater discrepancy between experimental and computational values −4.52 and −7.65 kcal/mol, respectively.

The addition of a nitro substituent ($-NO_2$) originates the greater difference between averages (average difference of 2.5 kcal/mol). Comparing the values, in 9 of the 12 cases considered, the calculated values are underestimated in more than 2 kcal. The phenol → 2-nitrophenol transformation presents the higher difference found in this study for experimental and predicted values: almost 5 kcal/mol. For a nitro addition in a primary position, the experimental values are around −5.4 kcal/mol, but if the addition is in an aromatic ring, the magnitude of the values rises to around −3 kcal/mol. The predicted values in a primary position have an average of −7.78 kcal/mol, and an average of −5.03 kcal/mol when the addition is made in an aromatic ring. These results can justify the application of an empirical correction factor of −2.0/2.5 kcal/mol in future calculations, when predicting the $\Delta\Delta G_{solv}$ free energies for NO_2 addition. Another option could be to redetermine the molecular mechanical parameters associated with this group in order to better reproduce the solvation free energies, if NO_2 solvation assumes a leading role in the calculations that we want to perform.

dx.doi.org/10.1021/ct500346y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

In the 3 cases of amide group addition, both averages differ by 0.5 kcal/mol, but the variation is more evident in the calculated values, because the minimum and maximum are more apart. More cases would be necessary for a more grounded analysis.

It is important to notice that in aromatic scaffold addition, in some cases (CH_3, Br and I), the experimental and calculated values albeit similar present different signs, that is, while the experimental value indicates unfavorable to desolvate (negative), the calculated appear as favorable to desolvate (positive). This should be taken into account in the MSE and MUE analysis.

Figure 4 presents the averages of mean signed error (MSE) and mean unsigned error (MUE) in the calculation of $\Delta\Delta G_{solv}$

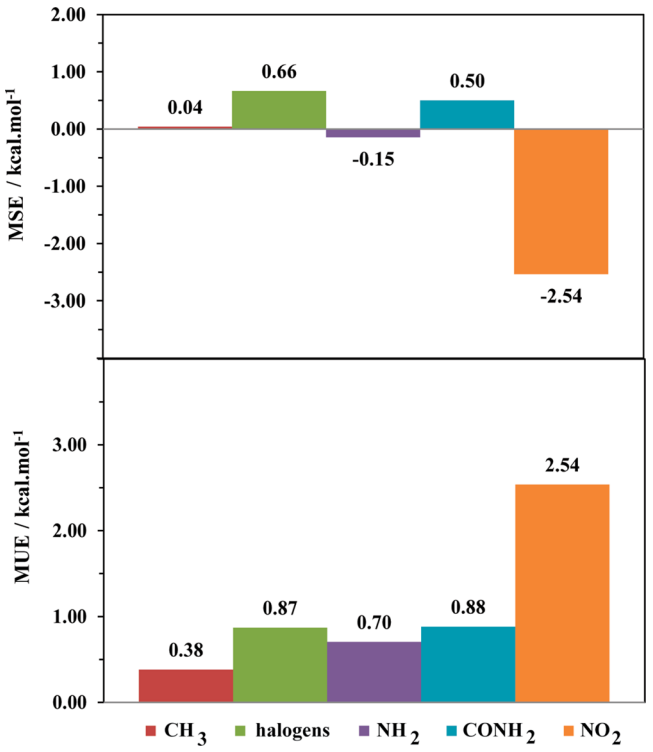


Figure 4. Average MUE and MSE in the determination of $\Delta\Delta G_{solv}$ free energies from the addition of different groups to different compound classes. Values expressed in kcal/mol.

free energies for the 5 additions tested in this study. Table 2 further decomposes these values taking into consideration the carbon atom that is subjected to the addition. The results of MSE show that the calculated values are overestimated for methyl, halogen and amide additions and underestimated for amino and nitro additions.

There is a small underestimation for methyl addition to primary, secondary and cyclic carbons and a small overestimation for tertiary, aromatic, and heterocyclic carbons.

For the addition of halogens, an MSE of 0.66 kcal/mol was obtained. However, for the four different additions within this group, the MSE varies significantly as present in Table 2. Since in the fluorine addition there is only one case considered, limiting the representativeness of the result, it is possible to say that as we descend along the halogen group in the periodic table, the MSE increases. That is, the predicted values are successively more overestimated.

Table 2. Mean Unsigned Error (MUE) and Mean Signed Error (MSE) in the Calculation of $\Delta\Delta G_{solv}$ Free for the Addition of CH_3, F, Cl, Br, I, NH_2, $CONH_2$, and NO_2 to Different Compound Classes[a]

addition to	no. cases	avg. MSE	avg. MUE
CH_3	40	0.04	0.38
primary carbon	8	−0.27	0.63
secondary carbon	7	−0.27	0.29
tertiary carbon	3	−0.26	0.26
cyclic carbon	2	−0.22	0.22
aromatic ring	8	0.38	0.40
heterocyclics	12	0.31	0.31
halogens	29	0.66	0.87
F	1	0.72	0.72
aromatic ring	1	0.72	0.72
Cl	13	0.16	0.62
primary carbon	4	0.25	0.25
secondary carbon	4	0.25	0.27
tertiary carbon	1	−0.66	0.66
aromatic ring	4	0.20	1.31
Br	9	0.93	0.94
primary carbon	4	0.79	0.79
secondary carbon	1	0.77	0.77
tertiary carbon	1	−0.04	0.04
aromatic ring	3	1.50	1.50
I	6	1.34	1.34
primary carbon	4	1.28	1.28
secondary carbon	1	1.30	1.30
aromatic ring	1	1.59	1.59
NH_2	8	−0.15	0.70
primary carbon	7	0.28	0.36
aromatic ring	1	−3.13	3.13
$CONH_2$	3	0.50	0.88
primary carbon	2	−0.10	0.47
aromatic ring	1	1.70	1.70
NO_2	12	−2.54	2.54
primary carbon	5	−2.35	2.35
secondary carbon	1	−2.25	2.25
aromatic ring	6	−2.74	2.74

[a]Values expressed in kcal/mol.

The MSE average for amino addition is small and negative (−0.15 kcal/mol) but it is greatly influenced by the MSE value for the benzene → aniline transformation. This is the only case of amino addition to an aromatic carbon and the predicted value is underestimated by 3 kcal/mol. Considering the remaining cases analyzed, corresponding to additions to a primary carbon, the MSE value is positive and lower than 0.5 kcal/mol, which indicates a small overestimation of the calculated values.

In the amide addition, the analysis of just three cases leads to an MSE of 0.5 kcal/mol although for the aromatic carbon amino addition the predicted value is overestimated by almost 2 kcal/mol.

The MSE value for the nitro addition is the higher for all the transformations in this study, −2.54 kcal/mol, indicating also the larger underestimation of the calculated values. Even if we analyze the MSE for the 3 different compound classes, it is always over 2 kcal/mol. This uniformity suggests the existence of a systematic error for this class of compounds, which could be the result of a limitation of the force field employed.

dx.doi.org/10.1021/ct500346y | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The results of MUE in the calculation of $\Delta\Delta G_{solv}$ free energies, presented also in Figure 4, show that for 4 of the 5 additions tested in this study, the predicted values differ for less than 1 kcal/mol. Only in the nitro group addition, the calculated and experimental values are apart for about 2.5 kcal/mol as suggested from the previous analysis.

General Trends. Figure 5 presents a correlation plot of the calculated and experimental $\Delta\Delta G_{solv}$ free energy values for the different classes of transformations tested. The solid line represents the ideal 1:1 correlation and the dashed lines represent acceptable deviations of ±1 kcal/mol (i.e., overestimation or underestimation by the computational method up to 1 kcal/mol). The outliers identified in Figure 2 were not included in the preparation of Figure 5.

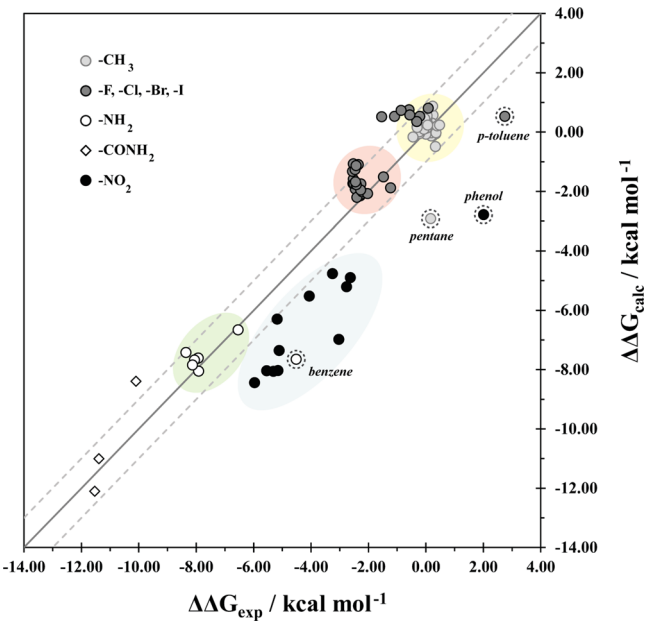


Figure 5. Correlation plot of the calculated and experimental $\Delta\Delta G_{solv}$ free energy values, in kcal/mol. The highlighted symbols (dashed) represent the outliers.

We can see that all the methyl addition cases are concentrated in the same area and within the acceptable region. It is possible to point out the pentane → hexane as an outlier as it can be easily seen. For the halogens group addition, two sets can be distinguished. One, however, is more populated and close to the ideal line although almost all cases are located in the upper acceptable region (i.e., overestimated). An outlier is once more noticeable, the *p*-toluene → *p*-chlorotoluene. The value for this transformation, places it well below the acceptable region (underestimated), thus contradicting the general trend for this group.

The amino group addition set presents a correlated distribution, with all the values being within acceptable region and close to the ideal line. Once again an outlier is present, benzene → aniline, more than 5 kcal away from the perfect correlation. Although the 3 cases for the amide addition, as already noticed, diminish the ability to a consistent analysis, these transformations are placed near the 1:1 correlation.

The nitro group addition gives the worst results, with all the points placed below the acceptable region line. Although the phenol → nitrophenol transformation is also of the same trend, it is distant from the other points, being therefore an outlier. In spite of the differences between the 5 group additions, it is possible to define for each an area that will help in future predictions.

CONCLUSIONS

The estimation of solvation free energies represents an important piece of the complex puzzle when studying drug-receptor affinity. A compromise between computational cost and accuracy for a reliable prediction of $\Delta\Delta G_{solv}$ energies can lead to a consequent boost in binding free energies calculations, so necessary in drug design.

In this study, we have calculated the solvation free energy change ($\Delta\Delta G_{solv}$) for a total of 92 transformations in small molecules, using Thermodynamic Integration. Based on the functional group considered for the additions (CH_3, F, Cl, Br, I, NH_2, $CONH_2$, and NO_2), the position of the substitution and the type of chain, we searched for trends and evaluated the standard TI protocol.

The distribution of experimental and computational values reveals larger amplitudes in the range of computationally predicted values, when compared with the experimental ones. Nevertheless, the analysis of the individual values indicates the presence of some outliers.

A good agreement between the average values for the different transformations is present, with a higher standard deviation for the computational values. In 3 categories (CH_3, halogens, and NH_2 additions), the calculated values are overestimated, whereas with the other two ($CONH_2$ an NO_2 addition) there is an underestimation. According to an ideal 1:1 correlation, it is evident that 4 of the 5 additions tested are within the acceptable deviation. The results of Table 1 help the reader to be aware of the possible errors that might be expected in his/her calculations. The solvation free energy value obtained computationally can be corrected using the differences between MD and experiment presented.

The nitro group addition values suggest a systematic underestimation of the predicted $\Delta\Delta G_{solv}$, more than 1 kcal/mol. The mean average evaluation indicates the need of a correction factor (such as 2.5 kcal/mol) or perhaps a different parametrization for the force field, in those cases.

The position of the substitution and the type of chain that serves as scaffold has a medium but not negligible effect, present both in experimental as computational values and it must be taken into account too.

The TI methodology can be computationally very intensive, especially for large systems and/or large number of structural changes. However, due to the accurate and precise results, which it ensures, is a valuable help for this kind of studies, helping defining trends to future predictions.

ASSOCIATED CONTENT

Supporting Information

Experimental solvation free energies of neutral compounds (Table S1) and experimental and computational values of $\Delta\Delta G_{solv}$ free energies for the 92 transformations considered (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Email: pafernan@fc.up.pt.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been funded by FEDER/COMPETE and Fundação para a Ciência e a Tecnologia through projects EXCL/QEQ-COM/0394/2012, SFRH/BD/46867/2008 and PEst-C/EQB/LA0006/2011.

■ REFERENCES

(1) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.
(2) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. Free-Energy of Hydrophobic Hydration—A Molecular-Dynamics Study of Noble Gases in Water. *J. Chem. Phys.* **1986**, *85*, 6720−6727.
(3) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(4) Michel, J.; Essex, J. Prediction of Protein−Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls, and Expectations. *J. Comput. Aid. Mol. Des.* **2010**, *24*, 639−658.
(5) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. In *Advances in Protein Chemistry*; Valerie, D., Ed.; Academic Press: New York, 2003; Vol. *66*, pp 123−158.
(6) Martins, S. A.; Sousa, S. F. Comparative Assessment of Computational Methods for the Determination of Solvation Free Energies in Alcohol-Based Molecules. *J. Comput. Chem.* **2013**, *34*, 1354−1362.
(7) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(8) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.
(9) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.
(10) Knight, J. L.; Brooks, C. L. λ-Dynamics Free Energy Simulation Methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.
(11) Case, D. A.; Darden, T. A.; Cheatham, I., T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, 2008.
(12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(13) Wang, J. M.; 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.
(14) Marcial, B. L.; Sousa, S. F.; Barbosa, I. L.; Dos Santos, H. F.; Ramos, M. J. Chemically Modified Tetracyclines as Inhibitors of MMP-2 Matrix Metalloproteinase: A Molecular and Structural Study. *J. Phys. Chem. B* **2012**, *116*, 13644−13654.
(15) Genheden, S.; Nilsson, I.; Ryde, U. Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51*, 947−958.
(16) Khavrutskii, I. V.; Wallqvist, A. Computing Relative Free Energies of Solvation Using Single Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *J. Chem. Theory Comput.* **2010**, *6*, 3427−3441.
(17) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides—The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-Methylamide. *Biopolymers* **1992**, *32*, 523−535.
(18) Constanciel, R.; Contreras, R. Self Consistent Field Theory of Solvent Effects Representation by Continuum Models: Introduction of Desolvation Contribution. *Theoret. Chim. Acta* **1984**, *65*, 1−11.
(19) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(20) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669−681.
(21) Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032−3041.
(22) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.
(23) Tomasi, J.; Mennucci, B.; Cancès, E. The IEF Version of the PCM Solvation Method: An Overview of a New Method Addressed to Study Molecular Solutes at the QM Ab Initio Level. *J. Mol. Struct.* **1999**, *464*, 211−226.
(24) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. Solvent Effects. 5. Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation on Ab Initio Reaction Field Calculations. *J. Phys. Chem.* **1996**, *100*, 16098−16104.
(25) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538−4543.
(26) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, *51*, 769−779.
(27) 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.
(28) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509−1519.
(29) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 2553−2558.
(30) Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation Model Based on Weighted Solvent Accessible Surface Area. *J. Phys. Chem. B* **2001**, *105*, 5055−5067.
(31) Lee, S.; Cho, K.-H.; Lee, C. J.; Kim, G. E.; Na, C. H.; In, Y.; No, K. T. Calculation of the Solvation Free Energy of Neutral and Ionic Molecules in Diverse Solvents. *J. Chem. Inf. Model.* **2010**, *51*, 105−114.
(32) Purisima, E.; Corbeil, C. R.; Sulea, T. Rapid Prediction of Solvation Free Energy. 3. Application to the SAMPL2 Challenge. *J. Comput. Aid. Mol. Des.* **2010**, *24*, 373−383.
(33) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial, Charge Models, and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128−139.
(34) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free Energies of Hydration from a Generalized Born Model and an ALL-Atom Force Field. *J. Phys. Chem. B* **2004**, *108*, 16264−16270.
(35) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23*, 517−529.
(36) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. Prediction of Solvation Free Energies of Small Organic Molecules: Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 405−412.

(37) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. Group Contributions to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution. *J. Solution Chem.* **1981**, *10*, 563−595.
(38) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Affinities of Amino-Acid Side-Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849−855.