

# A robust and efficient implicit solvation model for fast semiempirical methods

Sebastian Ehlert, Marcel Stahn, Sebastian Spicher, and Stefan Grimme\*

*Mulliken Center of Theoretical Chemistry, Bonn, Germany*

E-mail: [grimme@thch.uni-bonn.de](mailto:grimme@thch.uni-bonn.de)

## Abstract

We present a robust and efficient method to implicitly account for solvation effects in modern semiempirical quantum mechanics and force-fields. A computationally efficient yet accurate solvation model based on the analytical linearized Poisson–Boltzmann (ALPB) model is parameterized for the extended tight binding (xTB) and density functional tight binding (DFTB) methods as well as for the recently proposed GFN-FF general force-field. The proposed methods perform well over a broad range of systems and applications, from conformational energies over transition-metal complexes to large supramolecular association reactions of charged species. For hydration free energies of small molecules GFN1-xTB(ALPB) is reaching the accuracy of sophisticated explicitly solvated approaches, with a mean absolute deviation of only 1.4 kcal/mol compared to experiment. Logarithmic octanol–water partition coefficients ( $\log K_{ow}$ ) are computed with a mean absolute deviation of about 0.65 using GFN2-xTB(ALPB) compared to experimental values indicating a consistent description of differential solvent effects. Overall, more than twenty solvents for each of the six semiempirical methods are parameterized and tested. They are readily available in the `xtb` and `dftb+` programs for diverse computational applications.

# 1 Introduction

Solvation is ubiquitous in biological systems and plays an important role in many aspects of chemistry. Therefore, any computational method targeting to describe structures or interactions under realistic conditions must account for solvation effects. Solute–solvent interactions for example in hydration processes<sup>1,2</sup> or binding free energy computation<sup>3</sup> can be evaluated explicitly by free energy methods like thermodynamic integration<sup>4–6</sup> or free energy perturbation.<sup>7–10</sup> Techniques like metadynamics,<sup>11</sup> umbrella sampling<sup>12</sup> or replica exchange<sup>13,14</sup> allow to enhance the efficiency of the configuration space sampling determining the precision in the free energy computation.<sup>15</sup> Beside molecular dynamic based sampling techniques, Monte-Carlo methods can be used to effectively sample the conformational landscape.<sup>1,16,17</sup> While those methods are a suitable for accurate free energy computation, they require significant computational effort, which can be prohibitive for detailed investigations of chemical reaction mechanisms or high throughput computational workflows.

A practical compromise is usually found in the application of implicit solvation models.<sup>18–24</sup> In this approach, the contributions to the solvation free energy are often partitioned into polar (electrostatic) and non-polar (dispersion and cavity) solvation free energies, which allows devising tailored models for each contribution separately. The polar contribution is mostly approximated by conductor-like screening (COSMO)<sup>25</sup> or polarizable continuum models (PCM).<sup>26–29</sup> Implicit solvation models like the conductor-like screening model for real solvents (COSMO-RS),<sup>24,30–32</sup> the solvent models based on solute electron density (SMD)<sup>33</sup> or the three-dimensional reference interaction site model (3D-RISM)<sup>34–36</sup> enable accurate computation of solvation free energies based on standard (mostly DFT) electronic structure input data. However, while low-scaling implementations of COSMO/PCM have been proposed,<sup>37</sup> they still yield a noticeable computational overhead for force-field (FF) or semiempirical quantum mechanical (SQM) methods. For example the PM6 and PM7 methods have been successfully combined with COSMO<sup>38,39</sup> in previous studies using linear scaling algorithms.

A promising alternative are generalized Born (GB) models,<sup>40–42</sup> which are used in the reaction field based solvation models (SM), like SM6,<sup>43</sup> SM8<sup>44,45</sup> or SM12<sup>46</sup> and for FFs as generalized Born and surface area model (GBSA).<sup>47–49</sup> These models allow devising computationally efficient but yet accurate schemes to include solvation effects in large scale simulations.<sup>50</sup>

For the many SQM methods available, robust solvation models are needed, as this is tightly bound to the applicability of the respective methods in computational chemistry or biology.<sup>51,52</sup> One of the most promising members of the family of SQM methods are tight binding (TB) approaches, like density functional tight binding (DFTB)<sup>53,54</sup> or extended tight binding (xTB).<sup>55</sup> While the xTB methods were from the very beginning coupled with an implicit solvation model,<sup>56,57</sup> only a few DFTB implementations account for solvation effects.<sup>58–60</sup>

In this work we revisit the implicit solvation models tied with the xTB methods and try to systematically improve the underlying theory as well as the parametrization. Both experimental data from the MNSOL database<sup>43,61,62</sup> and theoretical values based on COSMO-RS are used as reference data in the fit. To allow for a fair comparison with DFTB, we also implemented, parameterized, and tested the implicit solvation models developed for xTB with three DFTB Hamiltonians. Furthermore, we include the recently devised general force field GFN-FF<sup>63</sup> which is not equipped with a tailored solvation model yet, but employed the GFN2-xTB solvation model in Ref. 63.

We investigate the accuracy of solvation free energies using the implicit solvation models based on the curated FreeSolv database<sup>64,65</sup> for experimental hydration free energies. Since accurate experimental data are usually only available for small compounds, we devise a benchmark set of back-corrected experimental association solvation free energies for large supramolecular complexes as well, building upon the established S30L set.<sup>66</sup> Furthermore, we employ a set of experimental octanol–water partition coefficients for 26 organic compounds and investigate the differential description of two solvents by the new models.

This paper is organized as follows. In section 2 the theory and algorithms used to implement the models with the TB methods are presented. Technical details of the parametrization and generation of reference data are described in section 3. In section 4 performance comparisons for a wide range of systems and benchmarks of the investigated methods are shown. Finally, in section 5 conclusions and perspectives are presented.

## 2 Theory

To describe a solute in a given solvent or dielectric medium the solvation free energy  $\Delta G_{\text{solv}}$  is partitioned into a polar contribution  $\Delta G_{\text{polar}}$ , which depends on the electrostatic potential, a non-polar contribution  $\Delta G_{\text{npol}}$ , which depends on the shape of the solute cavity and a constant shift  $\Delta G_{\text{shift}}$  depending on the reference state for the solvation process. The solvation free energy is therefore given as

$$\Delta G_{\text{solv}} = \Delta G_{\text{polar}} + \Delta G_{\text{npol}} + \Delta G_{\text{shift}}. \quad (1)$$

This partitioning is commonly used in many popular solvation models like GBSA<sup>42,50</sup> or SMD<sup>33</sup> resulting in energy expressions of different complexity for the individual contributions. Changes in the internal free energy of the molecule (rotation, vibration and conformational partition function) upon solvation is mostly not accounted for explicitly but absorbed into the empirical parameters of the model (see section 4.1).

A suitable form for an efficient evaluation of the polar contributions is the analytical linearized Poisson–Boltzmann (ALPB) model,<sup>48</sup> where the polar contribution is given by

$$\Delta G_{\text{polar}}^{\text{ALPB}} = -\frac{1}{2} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{1}{1 + \alpha\beta} \sum_{A=1}^N \sum_{B=1}^N q_A q_B \left( \frac{1}{f(R_{AB}, a_A, a_B)} + \frac{\alpha\beta}{\mathcal{A}_{\text{det}}} \right), \quad (2)$$

and  $\epsilon_{\text{in}}$  is the dielectric constant of the solute,  $\epsilon_{\text{out}}$  the dielectric constant of the solvent,  $q_{A/B}$  are atomic partial charges,  $f$  is the interaction kernel,  $a_{A/B}$  are the atomic Born radii for

atoms  $A$  and  $B$  and  $\mathcal{A}_{\text{det}}$  is the electrostatic size of the solute. The value of  $\alpha$  is fixed at 0.571214 following the work of Sigalov et al.<sup>48</sup> and  $\beta$  is  $\epsilon_{\text{in}}/\epsilon_{\text{out}}$ . The electrostatic size  $\mathcal{A}_{\text{det}}$  is calculated from the inertia tensor as proposed in Ref. 48, while for  $\epsilon_{\text{in}}$  the dielectric constant of the vacuum, i. e., one, is chosen. Since GB models are derived in the limit of  $\epsilon_{\text{out}} \rightarrow \infty$  or  $\beta \rightarrow 0$  we can effectively cast them into the same energy expression as the ALPB model by setting  $\alpha\beta$  to zero.

In this work we will test two different interaction kernels, first the canonical interaction kernel proposed by Still<sup>42</sup>

$$f_{\text{AB}}^{\text{Still}} = \left( R_{\text{AB}}^2 + a_{\text{A}}a_{\text{B}} \exp \left[ -\frac{R_{\text{AB}}^2}{4a_{\text{A}}a_{\text{B}}} \right] \right)^{\frac{1}{2}} \quad (3)$$

and the more recently proposed P16 kernel<sup>49</sup>

$$f_{\text{AB}}^{\text{P16}} = R_{\text{AB}} + \sqrt{a_{\text{A}}a_{\text{B}}} \left( 1 + \frac{1.028 \cdot R_{\text{AB}}}{16\sqrt{a_{\text{A}}a_{\text{B}}}} \right)^{16}. \quad (4)$$

The evaluation of the interaction kernel requires the atomic Born radii  $a_{\text{A/B}}$ , which are obtained by carrying out an integral over the molecular volume of the solute. For computational efficiency, we employ a pairwise approximate scheme, namely the Onufriev–Bashford–Case (OBC) corrected integrator termed GB<sup>OBC</sup>II.<sup>47</sup> The Born radius in this integrator is calculated by

$$\frac{1}{a_{\text{A}}} = \frac{1}{a_{\text{scale}}} \left( \frac{1}{R_{\text{A}}^{\text{vdw}} - R_{\text{offset}}} - \frac{\tanh[b\Psi_{\text{A}} - c\Psi_{\text{A}}^2 + d\Psi_{\text{A}}^3]}{R_{\text{A}}^{\text{vdw}}} \right), \quad (5)$$

where  $R_{\text{A}}^{\text{vdw}}$  are the D3 van-der-Waals radii,<sup>67</sup>  $a_{\text{scale}}$  is a global scaling parameter for the Born radii,  $R_{\text{offset}}$  is a global shift parameter to introduce more flexibility for the van-der-Waals radii and  $\Psi_{\text{A}}$  is the pairwise approximation to the integral over the molecular volume given by

$$\Psi_{\text{A}} = \frac{R_{\text{A}}^{\text{vdw}} - R_{\text{offset}}}{2} \sum_{\text{B} \neq \text{A}}^N \Omega(R_{\text{AB}}, R_{\text{A}}^{\text{vdw}}, s_{\text{B}} \cdot R_{\text{B}}^{\text{vdw}}), \quad (6)$$

where  $\Omega$  is the pairwise contribution to the approximate volume. To compensate for the overestimation of the molecular volume in  $\Omega$ , an element-specific de-screening parameter  $s_B$  is introduced for the van-der-Waals radius of the other atoms.

## 2.1 Non-polar Surface Area Contribution

To account for non-polar contributions to the solvation free energy we include a surface area (SA) model using the solvent-accessible surface area (SASA) to compute the free energy needed to form the solute-solvent cavity and to account for solute-solvent dispersion interactions by

$$\Delta G_{\text{npol}}^{\text{SA}} = \sum_{\text{A}}^N \gamma_{\text{A}} \sigma_{\text{A}}, \quad (7)$$

where  $\gamma_{\text{A}}$  is the surface tension of each atom and  $\sigma_{\text{A}}$  is its SASA. The surface tension is fitted as an element-specific parameter. To evaluate the complete SASA of the molecule we use

$$\sigma_{\text{total}} = \int_V \left| \nabla \left( \prod_{\text{A}} H_{\text{A}}(|\vec{R}_{\text{A}} - \vec{r}|) \right) \right| d^3 \vec{r}, \quad (8)$$

where  $H_{\text{A}}(|\vec{R}_{\text{A}} - \vec{r}|)$  is the atomic volume exclusion function,<sup>68</sup> given by

$$H_{\text{A}}(r) = \begin{cases} 0, & \text{if } r \leq R_{\text{A}}^{\text{surf}} - w \\ \frac{1}{2} + \frac{3}{4w}(r - R_{\text{A}}^{\text{surf}}) - \frac{1}{4w^3}(r - R_{\text{A}}^{\text{surf}})^3, & \text{if } R_{\text{A}}^{\text{surf}} - w < r < R_{\text{A}}^{\text{surf}} + w \\ 1, & \text{if } R_{\text{A}}^{\text{surf}} + w \leq r \end{cases} \quad (9)$$

The extent of the boundary region is defined by the smoothing parameter  $w$  and equals 0.3 Å in the here presented method. The  $R_{\text{A}}^{\text{surf}}$  are the combined van-der-Waals radii  $R_{\text{A}}^{\text{vdw}}$  with the solvent probe radius  $R_{\text{probe}}$ , the latter being a global parameter.

Since the volume exclusion function and its gradient are constant everywhere except for a narrow region around the surface area, we discretize the integral in Eq. 8 on an angular

Lebedev–Laikov grid<sup>69</sup> for an efficient numerical implementation on a sparse surface grid, as

$$\sigma_A = 4\pi \sum_g^{N_{\text{ang}}} w_g w_A \prod_{B \neq A}^N H_B(R_{Bg}), \quad (10)$$

where  $w_g$  and  $w_A$  are the angular and radial integration weights, respectively.

## 2.2 Hydrogen Bonding Correction

Specific interactions due to the molecular structure of the solvent, like hydrogen bonding (HB) between the solute and solvent molecules are not accounted for due to the implicit description of the solvent as a dielectric medium. We partition this important interaction component in atom-resolved contributions by

$$\Delta G_{\text{polar}}^{\text{HB}} = \sum_A^N \Delta G_A^{\text{HB}}. \quad (11)$$

To approximate the HB interaction we use the Keesom interaction<sup>70</sup> of hydrogen-acceptor and donor dipoles,  $\mu_{\text{AH}}$ , and  $\mu_{\text{D}}$ , respectively, with an average distance  $\bar{R}$ , which is similar to the distance of the first solvation shell. The final interaction term includes the probability  $\delta\rho_{\text{solv}}^{\text{HB}}$  that a solvent molecule interaction site is close to the respective atom. This contribution is proportional to the SASA of the interacting atom and therefore, we parameterize the HB contribution as

$$\Delta G_A^{\text{HB}} = -\frac{2}{3} \frac{\mu_{\text{AH}}^2 \mu_{\text{D}}^2}{\epsilon_{\text{out}}^2 k_{\text{B}} T \bar{R}^6} \delta\rho_{\text{solv}}^{\text{HB}} \approx -g_A^{\text{HB}} q_A^2 \frac{\sigma_A}{4\pi (R_A^{\text{surf}})^2}, \quad (12)$$

where  $g_A^{\text{HB}}$  is the HB strength, absorbing most of the constants, which is used as an element-specific parameter. Since not all methods have the dipole moment readily available we resort to approximate the quadratic dipole moment by the quadratic site atomic charge instead. Using this formula allows us to include the HB contribution with the polar electrostatic energy as a potential in the Hamiltonian or Lagrangian when minimizing the electrostatic

energy with no additional cost as we can reuse the already calculated SASA from the non-polar solvation free energy.

### 3 Computational Details

All DFT calculations were conducted with the Turbomole 7.5.1<sup>71</sup> software package. COSMORS calculations were done with COSMOtherm 19<sup>24,31,32</sup> using the BP86<sup>72</sup>/def-TZVPD<sup>73</sup> default level of theory. We employ the 2015, 2016, 2017, 2018 and 2019 versions of the BP86/def-TZVP and BP86/def-TZVPD fine parametrization in this work. The presented solvation models were implemented in the open source software packages `xtb`<sup>55,74</sup> and `DFTB+`<sup>54?</sup> and subsequently released with `xtb` version 6.3.2 and `DFTB+` version 20.1, which were used to conduct all `xtb` and `DFTB` calculations, respectively. For the third-order variant of `DFTB`, the 3ob parametrization is employed<sup>75–78</sup> and the mio parametrization is used for the second-order `DFTB` variant.<sup>79–81</sup> The ob2-base parametrization is used for the LC-`DFTB` Hamiltonian.<sup>82</sup> All `DFTB` Hamiltonians are combined with the D4(EEQ) London dispersion model without three-body contributions<sup>83</sup> using the parameters published in Ref. 54. Conformational searches and analysis were conducted with the conformer-rotamer ensemble search tool (`CREST`) version 2.10 using the `iMTD-GC` algorithm.<sup>84,85</sup> For high-level reference calculations the `CRENSO` workflow version 1.0.0 was used.<sup>86–88</sup>

#### 3.1 ALPB Training Set

The training set for the ALPB Solvation model consists of a mixture of experimental and theoretical solvation free energies. For the experimental part, we chose the Minnesota Solvation Database (`MNSOL`) version 2012<sup>43,61,62</sup> to serve as reference data. This database includes two types of solvation free energies: absolute solvation free energies and relative solvation free energies between various organic solvents and water. For the parametrization of ALPB, only the absolute solvation free energies were taken. In total the `MNSOL` Database

consists of experimental solvation free energies and optimized geometries for 520 neutral and ionic solutes, including the elements H, C, N, O, F, Si, P, S, Cl, Br, and I. The contained compounds range from small to medium sized organic solutes with a maximum atom count of 46, as well as 31 clustered ions containing a single water molecule. These experimental values were used for the parametrization of the solvents hexadecane, octanol, and water for the mentioned elements, as well as the global empirical parameters in the ALPB model.

Finding sufficient experimental solvation free energies for other elements of the periodic table and other solvents proved to be difficult. For this reason, additionally to the compounds in the MNSOL Database, we used the fit set created for the parametrization of the xTB Hamiltonians as it covers a wide range of elements.<sup>56,57</sup> In total the set used here contains about 2500 compounds for 67 elements. The geometries were optimized with the functional PBEh-3c.<sup>89</sup> The  $\delta G_{\text{solv}}$  reference values were obtained using COSMO-RS with the default version 2016 parametrization.<sup>24,31,32</sup>

### 3.2 Parametrization

Other than the physical constants  $\epsilon$ ,  $m$ , and  $\phi$ , which describe the dielectric constant, the mass and the mass density of the solvent, ALPB uses two types of parameters: global parameters and element specific parameters. The global parameters are  $G_{\text{shift}}$  in equation 1,  $\alpha_{\text{scale}}$  and  $R_{\text{offset}}$  in equation 5 and the probe radius  $R_{\text{Probe}}$ , which is included in  $R_A^{\text{surf}}$  in equation 10. The element specific parameters are  $s_X$  in equation 6,  $\gamma_A$  in equation 7 and  $g_A^{\text{HB}}$  in equation 12. The parameterized solvents are shown in Tab. 1. The ALPB solvation model employs the P16 interaction kernel, while we use the Still interaction kernel for the GBSA solvation model in this work.

For the parameter fit, a fully automated workflow was implemented. Therefore, the compounds of the ALPB test set were split into several subgroups, depending on the contained elements. The reference solvation free energies (MNSOL database or COSMO-RS) refer to the process of transferring the molecule from the gas phase to the liquid phase, and hence,

Table 1: Parameterized solvents for ALPB in combination with GFN1-xTB, GFN2-xTB, GFN-FF, DFTB3-D4, DFTB2-D4, and LC-DFTB2-D4 or GBSA in combination with GFN1-xTB or GFN2-xTB.

solvents		
acetone	acetonitrile	aniline <sup>a</sup>
benzaldehyde <sup>a</sup>	benzene	dichlormethane
chloroform	carbendisulfid	dioxane <sup>a</sup>
dimethylformamide <sup>a</sup>	dimethylsulfoxide	ether
ethylacetate <sup>a</sup>	furan <sup>a</sup>	hexadecane <sup>a,c</sup>
hexane <sup>b</sup>	methanol	nitromethane <sup>a</sup>
octanol <sup>a,c</sup>	octanol (wet) <sup>a,c</sup>	phenol <sup>a</sup>
toluene	tetrahydrofuran	water

*a*: Not parameterized for GFN1-xTB(GBSA) and GFN2-xTB(GBSA),

*b*: Not parameterized for GFN1-xTB(GBSA),

*c*: Not parameterized for DFTB3-D4(ALPB), DFTB2-D4(ALPB), LC-DFTB2-D4(ALPB)

the total molecular energy was first calculated for the gas phase, using the respective method. Subsequently, another full self-consistent calculation for the same gas phase geometry was conducted with added implicit solvation yielding  $\Delta E$  in equation 1. In the fitting process, first the global parameters for the elements of the organic subgroup, containing H, C, N, O, F, P, S and Cl, were determined by calculating the difference between the respective free energies in the two states

$$\delta G_{\text{solv}} = G_{\text{gas}} - G_{\text{sol}}. \quad (13)$$

This value was then fitted to the reference data utilizing the Levenberg–Marquardt non-linear least squares minimization algorithm.<sup>90,91</sup> The other element-specific parameters were then fitted on a subset for each specific element with all previously determined parameters kept fixed. When developing the workflow, care had to be taken to ensure that the element specific subsets contained only the corresponding element and those that had already been considered. While the structures for the gas phase and solvation calculation are not necessarily identical, we chose to perform both calculations on the gas phase structure without relaxing the geometry. This is mainly done to reduce the computational effort in the parametrization and to avoid potential instability of the fit due to artificial intermediate pa-

rameters in the geometry relaxation with implicit solvation. The influence of the geometry relaxation will be discussed in detail in Sec. 4.1.

## 4 Results

In the following sections, we discuss the performance of the here presented new ALPB solvation models on a broad range of test sets. We also compare the parametrizations of this work with the originally published GFN1-xTB(GBSA)<sup>56</sup> and GFN2-xTB(GBSA)<sup>57</sup> models, which have been shipped with the `xtb` program distribution<sup>55</sup> but were not thoroughly benchmarked so far. As statistical measures we use the mean signed deviation (MSD), mean absolute deviation (MAD), and the standard deviation (SD) and the error range as maximum deviation minus minimum deviation for free (solvation) energies. While the discussion mainly focuses on the MAD we also investigated the MSD, SD and error range for all sets and will discuss those measures if they show deviating behavior from the trends in the MAD.

### 4.1 Influence of geometry relaxation

To assess the possible error of neglecting geometry relaxations we evaluate the water subset of the MNSOL database. The hydration free energy for all systems was evaluated in three variants of the models. First, by using the gas phase optimized geometry and ignoring geometry relaxations from the implicit solvation model. Second, by relaxing the geometry with the implicit solvation model. And third, by explicitly computing the free energy of the molecule thermostatically in the modified rigid rotor, harmonic oscillator (mRRHO) approximation<sup>92</sup> with a rotor-cutoff of  $50\text{ cm}^{-1}$  to account for the changes in rotational (structure) and vibrational (frequency) contributions. This requires two full geometry optimizations and Hessian calculations.

The results with GFN2-xTB using the parametrization for GBSA and ALPB are shown in Tab. 2. The overall error range is at 30 kcal/mol for all compounds, while only around

18 kcal/mol for neutral solutes. To better interpret different error sources the set is split in neutral solutes as well as positive and negative ions.

First, we find an overall mean MAD for the hydration free energies of 1.95 and 1.88 kcal/mol for the GBSA and ALPB solvation model, respectively. The account for geometry relaxation only leads to small changes in the MAD for neutral solutes, slightly deteriorating the MAD by less than 0.1 kcal/mol for both models.

For the charged solutes we find a larger MAD in the hydration free energies of 10.05 and 7.06 kcal/mol for GBSA and ALPB solvation models, respectively. While GBSA and ALPB show similar performance for neutral solutes, ALPB represents a significant improvement for the charged solutes due to additional charge dependent terms which are absent in GBSA. A notable observation is that GFN2-xTB(ALPB) reduces the MAD of the hydration free energies by half for cationic solutes compared to GFN2-xTB(GBSA). Furthermore, we find an overall improvement of the hydration free energies when geometry relaxations are included for charged solutes by approximately 0.2 kcal/mol for both models.

Including rotational and vibrational contributions deteriorates the performance for hydration free energies slightly but consistently for both neutral and charged solutes. Tentatively, this slightly diminished accuracy can be attributed to the translational and rotational partition functions of the ideal gas and the rigid rotor which are more approximate for the solvated system. Approaches like a harmonic solvation model<sup>93</sup> or heuristic corrections to the partition function<sup>94</sup> could improve the description but are beyond the scope of this work.

Table 2: Mean absolute deviation in kcal/mol for the hydration free energies of GFN2-xTB(ALPB) and GFN2-xTB(GBSA) for different evaluation strategies.

subset	entries	GFN2-xTB(GBSA)			GFN2-xTB(ALPB)		
		SP only	opt.	freq.	SP only	opt.	freq.
neutral	390	1.95	2.01	2.08	1.88	1.96	2.07
positive	60	9.49	9.38	9.37	4.65	4.78	4.95
negative	83	11.32	10.92	11.05	9.39	8.98	9.13
all charged	143	10.05	9.78	9.84	7.06	6.85	7.02
all	533	4.26	4.23	4.30	3.36	3.36	3.49

To further investigate the influence of geometry relaxation we select 25 neutral solutes which show a significant change in the hydration free energy upon optimization in solution. Notable motifs with larger geometry changes are hydroxy, amid and nitro groups as well as sulfur or phosphorous containing groups, i.e. especially polar groups. To establish a benchmark set we optimized all 25 compounds using r<sup>2</sup>SCAN-3c<sup>95</sup> in the gas phase and with DCOSMO-RS<sup>96</sup> in water. To minimize the influence of the underlying electronic structure methods, we compare changes in bond lengths and angles between the gas phase and solvated structure rather than absolute bond lengths and angles. Since changes in the hydration free energy due to geometry relaxations are smaller than 0.5 kcal/mol and the root mean square deviation between the gas phase and the solvated structure is on average only 0.15 Å, small overall geometry changes are expected. In order to put the SQM results into some perspective and to establish a lower error bound we compute the geometry changes in addition with r<sup>2</sup>SCAN-3c(COSMO).

The MAD for the solvent induced geometry changes with GFN1-xTB, GFN2-xTB, DFTB3-D4 and GFN-FF and the ALPB solvation model as well as PM6-D3H4(COSMO) are shown in Tab. 3. First, because of the overall small magnitude of geometry changes upon solvation and correspondingly, the small errors involved in this test, the assessment of the methods here should be considered of qualitative nature at most. For the TB methods in combination with ALPB, we find a similar error compared to r<sup>2</sup>SCAN-3c(COSMO), which we consider as the lower bound for possible errors on this structures. PM6-D3H4(COSMO) performs only slightly worse. A notable exception is GFN-FF(ALPB) which shows doubled errors compared to the SQM methods. From this test we can conclude that the ALPB solvation model semi-quantitatively computes small geometry changes compared to more sophisticated models like DFT(DCOSMO-RS).

Solvation effects on bond lengths and bond angles are found to be small but can be much larger and chemically relevant when the entire three-dimensional structure of a molecule is considered. For conformational changes, the solvent accessibility of the polar groups in

Table 3: MAD in geometry differences for 25 neutral solutes.

	distances [ $10^{-3}$ Å]	angles [°]
r <sup>2</sup> SCAN-3c(COSMO)	1.6	0.14
GFN1-xTB(ALPB)	2.6	0.29
GFN2-xTB(ALPB)	1.7	0.28
GFN-FF(ALPB)	4.5	0.39
DFTB3-D4(ALPB)	2.0	0.20
PM6-D3H4(COSMO)	2.2	0.33

the solute as well as the SASA may change drastically and hence solvation effects can be crucial for the relative energetic ordering in conformational ensembles. As an example, the PES of the antibiotic drug erythromycin was investigated using the recently developed CRENSO<sup>86</sup> workflow with a final conformational energy threshold of 3.0 kcal/mol. The resulting optimized structure ensemble consists of six conformers within this energy window. The solvation free energies with the here presented methods were calculated as the difference in energy between the structure in the gas phase and in the liquid phase with full geometry relaxations.

Table 4 shows the deviation between r<sup>2</sup>SCAN-3c(COSMO-RS) and the here presented methods for calculated free solvation energies for these six conformers. It is apparent, that the

Table 4: Free solvation energies  $\delta G_{solv}$  in kcal/mol for the six erythromycin conformers calculated with r<sup>2</sup>SCAN-3c(COSMO-RS) and the deviation  $\delta G_{solv}(\text{model}) - \delta G_{solv}(\text{COSMO-RS})$  for the GFN methods with ALPB and GBSA.

	$\delta G_{solv}$			deviation		
	r <sup>2</sup> SCAN-3c (COSMO-RS)	GFN2-xTB (ALPB)	GFN1-xTB (ALPB)	GFN-FF (ALPB)	GFN2-xTB (GBSA)	GFN1-xTB (GBSA)
1	-27.7	1.3	-1.5	7.4	5.3	-0.1
2	-27.7	1.3	-1.5	7.3	5.3	-0.1
3	-26.5	2.1	-0.1	7.1	6.1	0.9
4	-26.4	2.2	-0.2	7.0	6.1	0.9
5	-23.9	-1.2	-2.0	6.2	2.9	-0.4
6	-28.3	0.8	-3.5	6.8	5.2	-2.7
<b>MAD</b>		1.5	1.5	7.0	5.2	0.9
<b>SD</b>		1.2	1.2	0.4	1.2	1.3

solvation free energies, as well as the deviations, significantly differ depending on the investigated conformer. Note, that changes of conformational energies on the order of 1–2 kcal/mol strongly affect thermal populations and average thermal molecular properties. With an MAD of 1.5 kcal/mol and 1.5 kcal/mol as well as an SD of 1.2 kcal/mol, GFN2-xTB(ALPB) and GFN1-xTB(ALPB) perform reasonably well. GFN-FF(ALPB) and GFN2-xTB(GBSA) produce significantly too positive solvation free energies with an MAD of 7.0 kcal/mol and 5.2 kcal/mol. However, the small SD of 0.4 kcal/mol and 1.2 kcal/mol, respectively, indicates rather systematic errors. While GFN1-xTB(GBSA) yields slightly smaller deviations than GFN2-xTB(ALPB) and GFN1-xTB(ALPB) with an MAD of 0.9 kcal/mol, the SD is a bit larger for the latter (1.3 kcal/mol) indicating a slightly lower robustness.

We have quantified the impact of geometry relaxations on the hydration free energies and observed only a minor influence for small to medium sized solutes. While we can verify that excluding geometry relaxations in the ALPB parametrization to reduce the computational effort and enhance the stability of the fit is reasonable, we also note that already for medium-sized charged solutes, neglecting geometry relaxations can increase the error in the calculated solvation free energies substantially. This also holds even more for solvation effects on conformational ensembles for flexible molecules, where we refer the reader to Ref. 86 for a more detailed discussion. Thus, for general consistency we recommend to always include geometry relaxation when calculating solvation free energies. Unless noted otherwise all solvation free energies discussed from here on include full geometry relaxations in the respective solvents.

## 4.2 Hydration Free Energies for the FreeSolv database

Water is one of the most commonly used solvents in (bio)chemistry. Yet the description of water is difficult for implicit solvation models due to its high polarity and the importance of HB as well as many-body (polarization) effects. To model chemistry in aqueous solution, an accurate description of the hydration free energies is an important test. We assessed

the performance of the here presented methods for hydration free energies by comparing our methods on the curated FreeSolv database, which contains currently 642 experimental values for neutral molecules.<sup>64,65</sup> Starting from the provided geometries the structures were optimized with the respective methods, once with the implicit solvation model, and once in gas phase. We also evaluated the contributions of rotational and vibrational thermostistical functions to the solvation energies by Hessian calculations for both, the optimized gas phase structure, and the optimized structure in implicit solvation. Yet, only minor effects on the overall statistics were obtained.

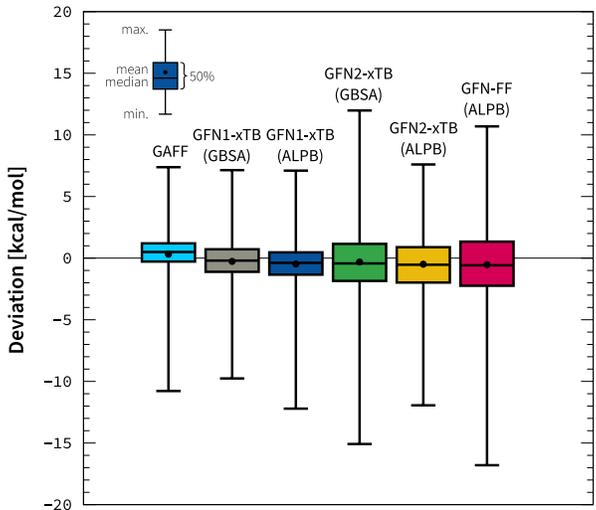


Figure 1: Statistical analysis of hydration free energies for the neutral species of the FreeSolv database. The values of the GAFF method in explicit water are taken from Ref. 64.

Fig. 1 shows the deviation of the calculated hydration free energies from the experimental values. Besides the models presented here, we include the explicit solvation approach (GAFF) from Ref. 64, which yields an MAD of 1.1 kcal/mol. In comparison, GFN1-xTB(ALPB) provides only a slightly larger MAD of 1.4 kcal/mol, which is encouraging considering the implicit nature of the solvation model and the parametric treatment of hydrogen-bonding. GFN1-xTB(GBSA) performs best from the here tested TB methods with an MAD of 1.3 kcal/mol, while GFN2-xTB(ALPB) performs slightly worse with an MAD of 1.8 kcal/mol, which is still reasonably accurate. Overall, the hydration free ener-

gies are slightly underestimated with an MSD of  $-0.5$  kcal/mol for GFN2-xTB(ALPB). The GFN2-xTB(GBSA) method performs slightly worse than the ALPB variant with an MAD of 1.9 kcal/mol. The smaller standard deviation of ALPB compared to GBSA for GFN2-xTB with 2.3 kcal/mol and 2.5 kcal/mol, respectively, indicates higher robustness and less outliers. DFTB(3ob)-D4(ALPB) yields an MAD of 1.7 kcal/mol and an MSD of  $-0.3$  kcal/mol, similar to the xTB variants.

With GFN-FF(ALPB) a respectable MAD of 2.2 kcal/mol is obtained, which is larger than for any of the tested SQM methods but still acceptable considering the about hundred-fold speed-up in typical applications. Evaluating the complete database with GFN-FF, including full geometry optimizations for solution and gas phase, takes about 36 s on one core of an Intel Xeon E5-4620 CPU, while the same calculation with GFN2-xTB takes 11 min.

### 4.3 Partition coefficients

An important property to characterize the distribution and accumulation of organic compounds and contaminants in the environment are *n*-octanol/water ( $K_{\text{OW}}$ ) partition coefficients, which correlate with observed biochemical and toxic effects<sup>97</sup> and are related to internal partitioning between biological tissues and body fluids.<sup>98</sup> While it may be used as a single descriptor in a linear free energy relationship (LFER), different  $\log K$  relationships are also of interest to form poly-parameter LFERs. Experimental partition coefficients for octanol–water are mostly determined for the transition of a compound from a wet octanol (30% water) phase to a water phase. Partition coefficients can be calculated thermodynamically from the difference in molecular free energy between these two phases. Including thermostistical contributions (see Section 4.1), the value is obtained by

$$\log K_{\text{OW}} = \frac{1}{k_{\text{B}}T \log e} \left( G_{\text{water}} + \Delta G_{\text{mRRHO, water}}^T - (G_{\text{octanol}} + \Delta G_{\text{mRRHO, octanol}}^T) \right), \quad (14)$$

where  $G$  is the total energy including solvation effects,  $G_{\text{mRRHO}}^T$  is the thermostatical contribution,  $k_B$  is Boltzmann’s constant,  $T$  is the temperature, and  $e$  Euler’s number.

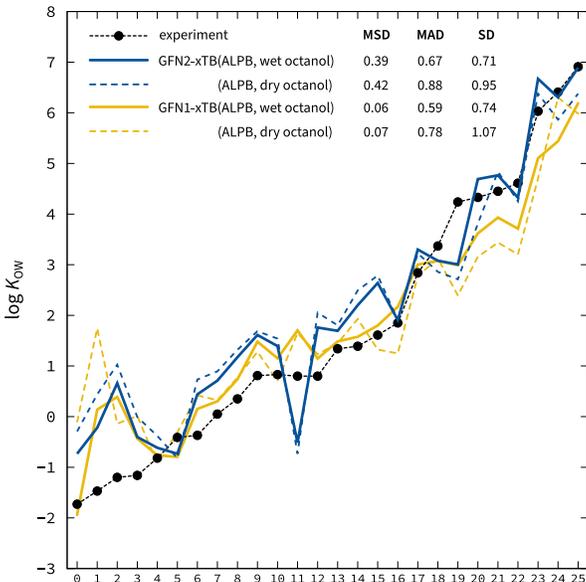


Figure 2:  $\log K_{\text{OW}}$  partition coefficients of 26 organic compounds ordered according to increasing values. The values are once calculated using wet octanol parameters and once using dry octanol. Statistical measures are given in kcal/mol.

Fig. 2 shows calculated octanol–water partition coefficients with reference to experimental values for 26 typical organic compounds. To classify the results, we calculated the  $\log K_{\text{OW}}$  values with  $\text{r}^2\text{SCAN-3c}(\text{COSMO-RS})$ , obtaining a very good MAD of 0.48 units. The calculations for the GFN methods were performed for dry octanol and wet octanol (30% water). For such a complicated property, GFN2-xTB(GBSA) shows reasonably small MAD values of 0.61 and 0.66, for dry and wet octanol, respectively. With an MAD of 0.88 and 0.67, GFN2-xTB(ALPB) performs slightly worse than GFN1-xTB(ALPB) which is in line with the results for the hydration free energy benchmark in section 4.2. Both methods show the expected slight overestimation of the  $\log K_{\text{ow}}$  values for dry compared to wet octanol. While not fully reaching the accuracy of the more sophisticated DFT based solvation model, given the semi-empirical nature of the GFN methods, the results are reasonable and useful for screening or high-throughput studies.

## 4.4 Supramolecular host–guest binding reactions

To assess the accuracy of the TB methods in combination with our parameterized solvation models for larger systems, which are the main target application, we propose a benchmark set of experimental, back-corrected solvation free energies for realistic host–guest binding reactions based on the existing S30L benchmark set.<sup>66</sup> For each of the 30 association reactions  $\text{host} + \text{guest} \rightarrow \text{guest@host}$  an experimental binding free energy  $\Delta G_{\text{a, exp}}$  was taken from the original work. The reference solvation free energies  $\Delta\delta G_{\text{solv}}$  are obtained according to

$$\Delta\delta G_{\text{solv}} = \Delta G_{\text{a, exp}} - \Delta E_{\text{a}} - \Delta G_{\text{mRRHO}}^T, \quad (15)$$

where  $\Delta E_{\text{a}}$  are accurate DLPNO-CCSD(T) reference values from Ref. 99 and  $\Delta G_{\text{RRHO}}$  are thermostistical corrections taken from Ref.<sup>100</sup> The resulting backcorrected association solvation free energies range from +90 kcal/mol for  $[\text{Ad}_2(\text{NMe}_3)_2@\text{CB7}]^{2+}$  (24) to  $-2.7$  kcal/mol for  $\text{AdOH@CB7}$  (21). The estimated accuracy of the reference values is about 3–4 kcal/mol. For the association reactions studied here, too strong solvation of the individual compounds will result in a too large association solvation energy and therefore a positive MSD. Similarly, too weak solvation of the individual compounds will result in an overall negative MSD. For comparison, results are presented for the COSMO-RS parametrization based on BP86/def2-TZVP densities from 2016 and the more recent 2019 parametrization for BP86/def2-TZVP densities (for both the BP86/def2-TZVPD fine parametrization was employed for water), and the SMD solvation model. The SMD values are based on BP86/def2-SVP calculations and were extracted from the original S30L publication.<sup>66</sup> The solvation free energy contribution to the association free energy for each complex of the S30L is plotted in Fig. 3.

For the S30L set COSMO-RS(2016) yields an MAD of 2.83 kcal/mol and a small MSD of 0.54 kcal/mol indicating somewhat too strong solvation of the reactants (guest and empty host). Nevertheless, this relatively small MAD (considering the huge range of values) indicates the reliability of the reference values. The results over the whole set differ only slightly

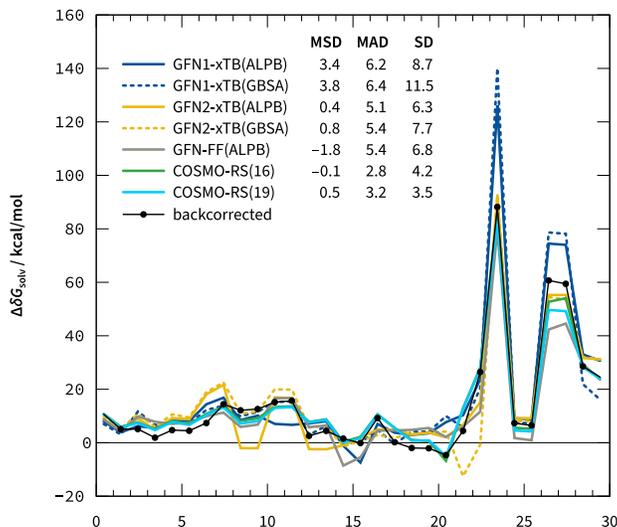


Figure 3: Deviation of COSMO-RS with 2016 and 2019 parametrization, GFN2-xTB with GBSA and ALPB, GFN1-xTB with GBSA and ALPB and GFN-FF(ALPB) to the backcorrected, experimental association solvation free energies. The statistical measures are given in kcal/mol for each of the selected methods.

compared to the 2015, 2017 and 2018 parametrizations while the 2019 version in contrast shows an increased MAD of 3.2 kcal/mol. The worse performance of the latest COSMO-RS parametrization is mainly caused by the underestimation of the reaction solvation free energies for water with the fine parametrization. For this variant, the MSD is shifted to  $-0.1$  kcal/mol. For SMD an MAD of 3.4 kcal/mol and an MSD close to zero (0.1 kcal/mol) is obtained.

With the here presented GFN2-xTB(ALPB) solvation model a very good MAD of 5.1 kcal/mol is achieved for an SQM method. GFN2-xTB(ALPB) slightly overestimates (MSD of 0.4 kcal/mol) the association solvation free energy. The DFTB(3ob)-D4(ALPB) model shows a rather poor performance with an overall MAD of 7.2 kcal/mol, which can be attributed to the poor description of the mainly dispersion bound complex 1–14, while the remaining systems are described sufficiently well. We also evaluate the DFTB(mio)-D4(ALPB) model for the 27 systems, which can be described with the base mio-parametrization, excluding systems 4, 15 and 16 because of missing parameters for halogens. With an MAD of 7.4 kcal/mol it performs similar to the 3ob-parametrization. Both versions significantly underestimate the values with

an MSD of  $-4.5$  kcal/mol and  $-5.1$  kcal/mol, respectively. The LC-DFTB(ob2)-D4(ALPB) model was excluded from this test set due to missing parameters.

The GFN2-xTB(GBSA) model yields an MAD of 5.4 kcal/mol and is slightly worse compared to the new GFN2-xTB(ALPB) model. The trend of overestimating the association solvation energies is also present in the GFN2-xTB(GBSA) model as seen from the MSD of 0.8 kcal/mol. We note that GFN2-xTB(ALPB) reduces the error range significantly compared to GFN2-xTB(GBSA) from 38.8 kcal/mol to only 25.3 kcal/mol.

For GFN1-xTB(GBSA) we exclude systems 15 and 16 due to missing parametrization data for the solvent cyclohexane, for the remaining 28 systems the MAD is 6.4 kcal/mol. GFN1-xTB(ALPB) performs slightly better with an MAD of 6.2 kcal/mol on the 28 systems and somewhat worse with an MAD of 6.2 kcal/mol on the complete set compared to GFN2-xTB. Both ALPB and GBSA employed together with GFN1-xTB yield a systematic overestimation of the association solvation free energies with an MSD of 3.4 and 3.8 kcal/mol, respectively. Again, we find that the error range with GFN1-xTB(ALPB) is significantly reduced compared to its GBSA variant from 60.3 to 47.1 kcal/mol. GFN-FF(ALPB) performs with an MAD of 5.4 kcal/mol almost as good as GFN2-xTB(ALPB) on this set. Even the error range for GFN-FF(ALPB) is similarly small with a value of 26.0 kcal/mol compared to the SQM method.

Overall, the the ALPB solvation model together with the GFN methods yields a good description for the solvation contributions for this challenging supramolecular reactions.

## 4.5 Transition Metal Chemistry

The here presented solvation models have been thoroughly investigated for neutral and charged solutes comprised of main group elements. Here we extend the investigation to neutral and charged solutes containing transition metal elements as well.

We evaluate all reaction solvation free energies for the reactions in the MOR41 benchmark set<sup>101</sup> using COSMO-RS for the three representative solvents, water, acetonitrile (ACN)

and tetrahydrofuran (THF). The MOR41 benchmark set consists of metal-organic reactions featuring 3d and late transition metals and covers a wide range of possible d-block elements. Since we are comparing with COSMO-RS, we neglect geometry relaxations consistently in the reference method and in the tested solvation models. Due to missing parametrization data the DFTB methods cannot be considered here.

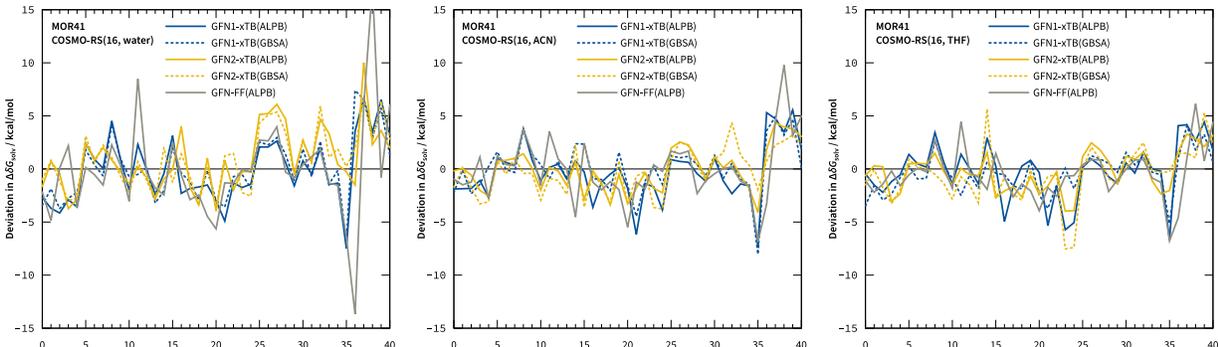


Figure 4: Left: deviation in the MOR41 reaction solvation free energies from the COSMO-RS reference for water. Center: deviation in the MOR41 reaction solvation free energies from the COSMO-RS reference for acetonitrile (ACN). Right: deviation in the MOR41 reaction solvation free energies from the DCOSMO-RS reference for tetrahydrofuran (THF). PM6-D3H4(COSMO) and PM7(COSMO) were not included in the graphic due the large error range of 25 and 48 kcal/mol, respectively.

The deviation of the tested methods for each of the 41 reaction solvation free energies is shown in Fig. 4 (left panel). Overall, we find MAD values in the range of 2.3 to 3.0 kcal/mol for the here presented solvation models. The best performing methods are GFN2-xTB, with both ALPB and GBSA, and GFN1-xTB(GBSA) all with an MAD of 2.3 kcal/mol. Only the ALPB solvation model for GFN1-xTB gives a slightly larger MAD of 2.5 kcal/mol. The GFN-FF(ALPB) method yields a slightly worse MAD of 3.0 kcal/mol. For comparison we included PM6-D3H4(COSMO) and PM7(COSMO), which perform badly for this kind of systems with an MAD of 3.7 and 5.3 kcal/mol, respectively.

Additionally, we have investigated the same systems for ACN and THF with COSMO-RS and the statistical data are shown in Fig. 4 (center and right panel). The overall trend of the deviation in the reaction solvation free energies is similar compared to the reaction hydration

free energies, while the magnitude of the overall deviation is reduced with the polarity of the solvent. For ACN we find an MAD ranging from 1.6 to 2.1 kcal/mol. GFN2-xTB(ALPB) is performing best for this solvent with an MAD of 1.6 kcal/mol, while the GBSA variant yields a larger MAD of 1.8 kcal/mol. For GFN1-xTB we find a similar good agreement using the GBSA solvation model with an MAD of 1.7 kcal/mol and a slightly deteriorated performance with GFN1-xTB(ALPB) (MAD 1.9 kcal/mol. PM6-D3H4(COSMO) and PM7(COSMO) yield a rather large MAD compared to this of 3.4 and 5.1 kcal/mol. In case of THF as solvent deviations are further reduced with MADs ranging from 1.5 to 1.8 kcal/mol for the here presented methods.

The overall performance of the semiempirical methods is reasonable, considering that they inherently yield a much larger error for the reaction energies, as seen in the MAD for the MOR41 set with is 13.2 kcal/mol and 11.8 kcal/mol for GFN1-xTB and GFN2-xTB, respectively, compared to <5 kcal/mol for well performing DFT methods.

Furthermore, we investigated the tetrakis(isonitrile)rhodium(I) cation,<sup>102</sup> which has been previously analyzed under different aspects in theoretical studies.<sup>83,103,104</sup> Due to its relatively high charge it is an interesting and challenging example for the computation of solvation free energies. Here, we focus on the formation of the dication complex from two (mono)cations as shown in Fig. 5.

To obtain a backcorrected reaction solvation free energy  $\Delta\delta G_{\text{solv}}$  we use the DLPNO-CCSD(T)/CBS\* electronic reaction energy  $\Delta E_{\text{elec}}$  of 8.8 kcal/mol taken from Ref. 83 and calculate the reaction at the r<sup>2</sup>SCAN-3c level to obtain a thermostatistical correction to the reaction free energy  $\Delta G_{\text{mRRHO}}^T$  of 16.3 kcal/mol. With the experimental association free energy  $\Delta G_a$  of -2.1 kcal/mol<sup>102,104</sup> we obtain a backcorrected reaction solvation free energy  $\Delta\delta G_{\text{solv}}$  of -27.2 kcal/mol as our benchmark value. The results are shown in Tab. 5.

For this example, clear differences are observed between the GBSA and ALPB solvation models. In general, we find that ALPB provides generally less negative reaction solvation free energies compared to GBSA. GFN2-xTB(ALPB) yields a reaction solvation free energy

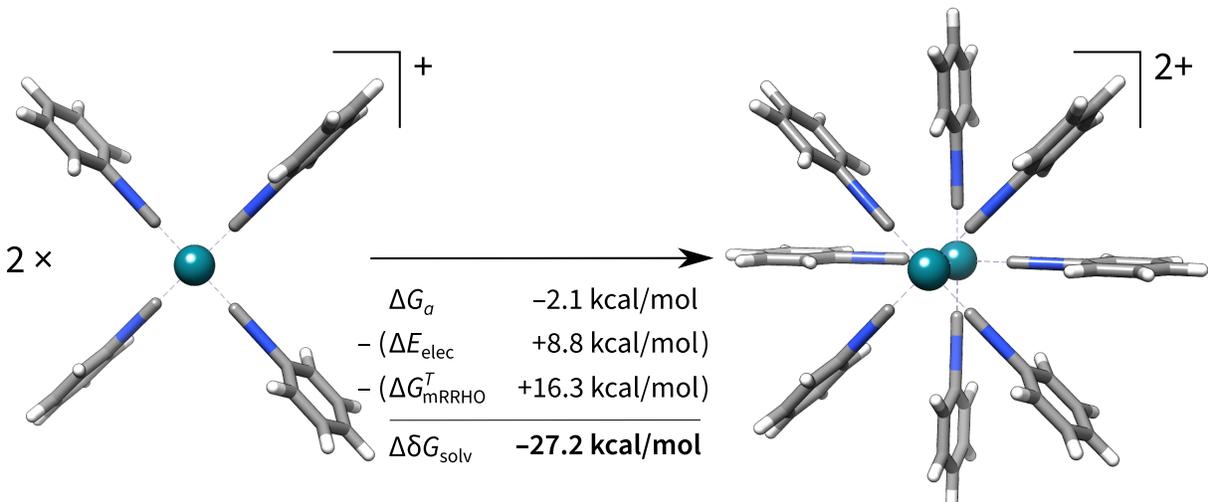


Figure 5: Formation of the Rhodium dication complex. The experimental association energy is backcorrected using the electronic association energy  $\Delta E_{\text{elec}}$  at the DLPNO-CCSD(T)/CBS\* level of theory and thermostistical correction to the reaction free energy  $\Delta G_{\text{mRRHO}}^T$  at the r<sup>2</sup>SCAN-3c level of theory.

of  $-26.2$  kcal/mol very close to the reference value, while GBSA is slightly over-shooting with  $-29.8$  kcal/mol. A similar trend between GBSA and ALPB is observed for GFN1-xTB, which results in more positive values compared to GFN2-xTB. GFN1-xTB with GBSA solvation model also reaches a quite good agreement of  $-28.3$  kcal/mol, while the ALPB model gives an overall to positive reaction solvation free energy of  $-23.8$  kcal/mol. We mainly attribute this difference of 4–5 kcal/mol to the additional charge dependent contributions in the ALPB solvation model, which are absent in most other implicit solvation models.

Table 5: Reaction solvation free energies in kcal/mol for the formation of the rhodium dication complex.

method	$\Delta \delta G_{\text{solv}}$
reference	$-27.2$
GFN1-xTB(GBSA)	$-28.3$
GFN2-xTB(GBSA)	$-29.8$
GFN1-xTB(ALPB)	$-23.8$
GFN2-xTB(ALPB)	$-26.2$
GFN-FF(ALPB)	$-24.5$
PM6-D3H4(COSMO)	$-33.3$
PM7(COSMO)	$-34.0$

Furthermore, GFN-FF(ALPB) yields a very reasonable value of  $-24.5$  kcal/mol while PM6-D3H4(COSMO) as well as PM7(COSMO) perform rather badly. Overall, the ALPB based solvation models provide a decent description of the solvation effects in this challenging transition metal reaction.

## 5 Conclusion

We presented a fast and computationally efficient solvation model suitable for combination with various tight binding Hamiltonians and even general force fields. A broad range of twenty nonpolar and polar as well as protic and aprotic solvents are readily available. In combination with the GFN family of methods all elements of the periodic table up to Radon ( $Z \leq 86$ ) are covered. For Slater–Koster based DFTB the implicit nature of the solvation model enables the description of systems which are unavailable with the respective parametrizations in an explicit approach.

The resulting methods yield consistent and reasonably accurate solvation free energies for small and large molecules with various solvents. Hydration free energies for a wide range of solutes from the FreeSolv database are in good agreement to the experimental values and close to the accuracy of explicitly solvated approaches which are clearly more elaborate and computationally expensive. Additionally, the consistent description of different solvents has been demonstrated for the accurate computation of partition coefficients, e. g.,  $K_{ow}$  for octanol and water. For the association energies in the supramolecular S30L benchmark set, also good results close to the backcorrected experimental values were obtained with the xTB models.

The effect of geometry relaxations with implicit solvation models was investigated qualitatively and semi-quantitatively and their importance for medium-sized charged solutes was evaluated. For properties depending on the description of a structural ensemble of flexible solutes such as conformational free energies, the inclusion of solvation effects is indispensable.

The ALPB and GBSA models parameterized here are implemented in the freely available `xtb` and `dftb+` program packages. Based on our tests we can recommend the ALPB solvation model in combination with GFN2-xTB as well as the GBSA solvation model in conjunction with GFN1-xTB as routinely and consistently applicable methods for energy calculations, geometry optimizations, molecular dynamics simulations, and vibrational frequency calculations. We are optimistic that the here presented solvation models will, together with current and future SQM methods, be valuable in many computational chemistry studies and workflows.

## Acknowledgement

The authors thank P. Shushkov for the initial implementation of the GBSA model in `xtb` and A. Hansen for proofreading the manuscript. S.E. thanks B. Hourahine for helpful discussions and code review on the DFTB+ implementations. S.S. thanks the “Fond der chemischen Industrie” (FCI) for financial support. This work was supported by the DFG in the framework of the “Gottfried Wilhelm Leibniz Prize” awarded to S.G.

## Supporting Information Available

The following files are available free of charge.

- LibreOffice spreadsheet containing the statistical evaluations: `ESI.ods`
- Tarball containing all optimized geometries: `geometries.tar.xz`
- Tarball containing all solvation model parameter files: `parameters.tar.xz`

## References

- (1) Jorgensen, W. L.; Nguyen, T. B. Monte Carlo simulations of the hydration of substituted benzenes with OPLS potential functions. *J. Comput. Chem.* **1993**, *14*, 195–205.

- (2) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design* **2014**, *28*, 711–720.
- (3) Wong, C. F.; McCammon, J. A. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
- (4) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (5) Berens, P. H.; Mackay, D. H.; White, G. M.; Wilson, K. R. Thermodynamics and quantum corrections from molecular dynamics for liquid water. *J. Chem. Phys.* **1983**, *79*, 2375–2389.
- (6) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–1069.
- (7) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- (8) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *The Journal of chemical physics* **1985**, *83*, 3050–3054.
- (9) Wereszczynski, J.; McCammon, J. A. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Quarterly reviews of biophysics* **2012**, *45*, 1.
- (10) Hansen, N.; Van Gunsteren, W. F. Practical aspects of free-energy calculations: a review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
- (11) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Nat. Acad. Sci., USA* **2002**, *99*, 12562–12566.
- (12) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.
- (13) Hansmann, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (14) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (15) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012**, *137*, 230901.
- (16) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689–1700.
- (17) Cabeza de Vaca, I.; Qian, Y.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Enhanced Monte Carlo methods for modeling proteins including computation of absolute free energies of binding. *J. Chem. Theory Comput.* **2018**, *14*, 3279–3288.
- (18) 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.
- (19) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.
- (20) Roux, B.; Simonson, T. Implicit solvent models. *Biophysical chemistry* **1999**, *78*, 1–20.
- (21) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (22) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187–4226.
- (23) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (24) Klamt, A. The COSMO and COSMO-RS solvation models. *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709.
- (25) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2* **1993**, 799–805.

- (26) Miertuś, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics* **1981**, *55*, 117–129.
- (27) 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. C* **1996**, *100*, 16098–16104.
- (28) 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.
- (29) 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.
- (30) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Chem. Phys.* **1995**, *99*, 2224–2235.
- (31) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- (32) Klamt, A. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design*; Elsevier: Amsterdam Boston, 2005.
- (33) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (34) Ten-no, S.; Hirata, F.; Kato, S. A hybrid approach for the solvent effect on the electronic structure of a solute based on the RISM and Hartree-Fock equations. *Chemical physics letters* **1993**, *214*, 391–396.
- (35) Kovalenko, A.; Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- (36) Heil, J.; Kast, S. M. 3D RISM theory with fast reciprocal-space electrostatics. *J. Chem. Phys.* **2015**, *142*, 114107.
- (37) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. Fast domain decomposition algorithm for continuum solvation models: Energy and first derivatives. *J. Chem. Theory Comput.* **2013**, *9*, 3637–3648.
- (38) Stewart, J. J. Application of the PM6 method to modeling proteins. *Journal of molecular modeling* **2009**, *15*, 765–805.
- (39) Kríž, K.; Řezáč, J. Reparametrization of the COSMO solvent model for semiempirical methods PM6 and PM7. *J. Chem. Inf. Model.* **2019**, *59*, 229–235.
- (40) Born, M. Volumen und hydrationswärme der ionen. *Zeitschrift für Physik* **1920**, *1*, 45–48.
- (41) Tucker, S. C.; Truhlar, D. G. Generalized Born fragment charge model for solvation effects as a function of reaction coordinate. *Chemical physics letters* **1989**, *157*, 164–170.
- (42) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (43) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *Journal of chemical theory and computation* **2005**, *1*, 1133–1152.
- (44) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011–2033.
- (45) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on the generalized born approximation with asymmetric descreening. *Journal of chemical theory and computation* **2009**, *5*, 2447–2464.
- (46) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized born solvation model SM12. *Journal of Chemical Theory and Computation* **2013**, *9*, 609–620.

- (47) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.
- (48) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
- (49) Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999–2011.
- (50) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (51) Pecina, A.; Meier, R.; Fanfrlík, J.; Lepšík, M.; Řezáč, J.; Hobza, P.; Baldauf, C. The SQM/COSMO filter: reliable native pose identification based on the quantum-mechanical description of protein–ligand interactions and implicit COSMO solvation. *Chemical Communications* **2016**, *52*, 3312–3315.
- (52) Kříž, K.; Řezáč, J. Benchmarking of Semiempirical Quantum-Mechanical Methods on Systems Relevant to Computer-Aided Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 1453–1460.
- (53) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.
- (54) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshayé, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.
- (55) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **2020**, *X*, Z.
- (56) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ( $Z = 1–86$ ). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- (57) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (58) Hou, G.; Zhu, X.; Cui, Q. An Implicit Solvent Model for SCC-DFTB with Charge-Dependent Radii. *J. Chem. Theory Comput.* **2010**, *6*, 2303–2314.
- (59) Nishimoto, Y. DFTB/PCM applied to ground and excited state potential energy surfaces. *J. Phys. Chem. A* **2016**, *120*, 771–784.
- (60) Kromann, J. C.; Steinmann, C.; Jensen, J. H. Improving solvation energy predictions using the SMD solvation method and semiempirical electronic structure methods. *J. Chem. Phys.* **2018**, *149*, 104102.
- (61) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. New universal solvation model and comparison of the accuracy of the SM5. 42R, SM5. 43R, C-PCM, D-PCM, and IEF-PCM continuum solvation models for aqueous and organic solvation free energies and for vapor pressures. *The Journal of Physical Chemistry A* **2004**, *108*, 6532–6542.
- (62) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database-version 2012. *University of Minnesota, Minneapolis* **2020**,
- (63) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.
- (64) 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.
- (65) Mobley, D. L.; Shirts, M.; Lim, N.; Chodera, J.; Beauchamp, K.; Lee-Ping, FreeSolv: Version 0.52. <https://doi.org/10.5281/zenodo.1161245>.

- (66) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.
- (67) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (68) Im, W.; Lee, M. S.; Brooks III, C. L. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
- (69) Lebedev, V. I.; Laikov, D. N. A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Doklady Mathematics* **1999**, *59*, 477–481.
- (70) Keesom, W. The second virial coefficient for rigid spherical molecules, whose mutual attraction is equivalent to that of a quadruplet placed at their centre. *Proc. R. Acad. Sci* **1915**, *18*, 636–646.
- (71) TURBOMOLE, V7.5.1. University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 2021.
- (72) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. B* **1988**, *38*, 3098–3100.
- (73) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- (74) “Semiempirical Extended Tight-Binding Program Package xtb”, <https://github.com/grimme-lab/xtb>. Accessed: 2021-05-03.
- (75) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- (76) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014**, *10*, 1518–1537.
- (77) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* **2015**, *119*, 1062–1082.
- (78) Kubillus, M.; Kubar, T.; Gaus, M.; Rezac, J.; Elstner, M. Parameterization of the DFTB3 method for Br, Ca, Cl, F, I, K, and Na in organic and biological systems. *Journal of chemical theory and computation* **2015**, *11*, 332–342.
- (79) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (80) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. Application of an approximate density-functional method to sulfur containing compounds. *J. Mol. Struct. (Theochem)* **2001**, *541*, 185–194.
- (81) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *Journal of Chemical Theory and Computation* **2011**, *7*, 931–948.
- (82) Vuong, V. Q.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J. J.; Mast, T.; Niehaus, T. A.; Irle, S.; Elstner, M. Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 115–125.
- (83) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.
- (84) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.
- (85) “Conformer-Rotamer Ensemble Sampling Tool”, <https://github.com/grimme-lab/crest>. Accessed: 2021-05-03.
- (86) Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **0**, 0, null.
- (87) “Commandline ENergetic SOrting of Conformer Rotamer Ensembles”, <https://github.com/grimme-lab/censo>. Accessed: 2021-05-03.

- (88) "Scripts to automate CREST and CENSO for calculating various properties", <https://github.com/grimme-lab/crenso>. Accessed: 2021-05-03.
- (89) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015**, *143*, 54107.
- (90) Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168.
- (91) Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441.
- (92) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.
- (93) Nakai, H.; Ishikawa, A. Quantum chemical approach for condensed-phase thermochemistry: Proposal of a harmonic solvation model. *J. Chem. Phys.* **2014**, *141*, 174106.
- (94) Izato, Y.-i.; Matsugi, A.; Koshi, M.; Miyake, A. A simple heuristic approach to estimate the thermochemistry of condensed-phase molecules based on the polarizable continuum model. *Physical Chemistry Chemical Physics* **2019**, *21*, 18920–18929.
- (95) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A "Swiss army knife" composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.
- (96) Klamt, A.; Diedenhofen, M. Calculation of Solvation Free Energies with DCOSMO-RS. *J. Phys. Chem. A* **2015**, *119*, 5439–5445.
- (97) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H. *Handbook of chemical property estimation methods*; Washington, DC (USA); American Chemical Society, 1990.
- (98) Reddy, M.; Yang, R.; Andersen, M. E.; Clewell III, H. J. *Physiologically based pharmacokinetic modeling: science and applications*; John Wiley & Sons, 2005.
- (99) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *J. Chem. Phys.* **2018**, *148*, 064104.
- (100) Spicher, S.; Grimme, S. Efficient Computation of Free Energy Contributions for Association Reactions of Large Molecules. *J. Phys. Chem. Lett.* **2020**, *11*, 6606–6611.
- (101) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Chęcinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2596–2608.
- (102) Mann, K. R.; Gordon, J.; Gray, H. B. Characterization of oligomers of tetrakis (phenyl isocyanide) rhodium (I) in acetonitrile solution. *J. Am. Chem. Soc.* **1975**, *97*, 3553–3555.
- (103) Bursch, M.; Caldeweyher, E.; Hansen, A.; Neugebauer, H.; Ehlert, S.; Grimme, S. Understanding and quantifying London dispersion effects in organometallic complexes. *Acc. Chem. Res.* **2018**, *52*, 258–266.
- (104) Grimme, S.; Djukic, J.-P. Cation–Cation "Attraction": When London Dispersion Attraction Wins over Coulomb Repulsion. *Inorganic Chemistry* **2011**, *50*, 2619–2628.

## Graphical TOC Entry

