Investigation and Development of Quantum Chemical Solvation Models

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Marcel Siegfried Stahn
aus
Aachen

Bonn, 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Stefan Grimme
2. Gutachter:	Prof. Dr. Thomas Bredow
Tag der Promotion:	05.07.2024
Erscheinungsjahr:	2024

The theory of dielectrics has more booby traps than a gamma function has poles.

– Lars Onsager –

Affirmation in Lieu of an Oath

I, Marcel Stahn, hereby affirm the following in lieu of an oath:

The doctoral thesis I submitted is my own work and was prepared without unauthorized outside assistance. I have not included text passages, graphics or other materials from third parties or my own examination papers without identifying them. Only the sources and resources that I have indicated were used. All verbatim and non-verbatim citations from other works are identified in accordance with the citation rules for academic writing. The thesis that I submitted has not yet been published, or has been published in full or in part at the locations indicated. The thesis that I submitted has not yet been submitted in any form as part of an examination/qualification course. I have not failed twice in a doctoral examination procedure at a German university due to rejection of a doctoral thesis or failed examinations. I prepared the doctoral thesis that I submitted in accordance with the principles of good research practice. I am aware of the significance and criminal consequences of a false affirmation in lieu of an oath under § 156 of the Criminal Code (StGB).

My statements are true to the best of my knowledge and belief.

Bonn, 5th July 2024 Marcel Stahn

Signature

Abstract

The main topic of this thesis is the investigation and development of quantum chemical solvation models. In combination with modern quantum chemical methods, these models can predict various molecular properties for complex and chemically diverse systems in solution. A special focus lies thereby on the creation and application of automatic multi-level workflows to combine various levels of theory for an efficient and accurate calculation that is readily usable even by non-expert users. These multi-level workflows use carefully chosen combinations of semi-empirical methods for cheap screening in combination with higher-level quantum chemical methods to yield an accurate prediction for all parts of the free energy. The computed free energy, in turn, can be used to calculate molecular properties, like partition coefficients or vapor pressures from thermodynamic relationships. To give a full picture of this topic, Chapter 2 of this thesis will cover the fundamental theory behind the calculation of the total free energy. The total free energy thereby includes the electronic energy, which describes the energy of a molecule in the gas phase at a temperature of 0 K, but also additional contributions to account for finite temperature and conformational effects. Finally, the theory behind implicit solvation models, which are used to approximate solvent-solute interactions in a computationally efficient way, will be briefly introduced. Chapters 3 and 4 will deal with the application of an automated multi-level workflow for environmentally relevant and other highly complex systems (*e.g.*, frustrated Lewis pairs), which are largely dominated by non-covalent interactions.

While most of the contributions to the free energy are already rather accurate, or (in the case of electronic structure theory) at least systematically improvable, solvation contributions can still be a major source of error. Because solvation effects can have a large impact on both, the final energy, as well as the molecular structure of a system, an accurate and robust solvation description is important in all parts of a multi-level workflow, even for the cheap screening parts.

This thesis will, therefore, propose different approaches to improve the accuracy of solvation models used on all levels of theory. Chapter 5 will introduce a post-SCF solvation model for semi-empirical methods based on an efficient implementation of a polarizable continuum model (PCM) into the `xTB` program package. The calculated results are then post-processed by a combination of literature-known approaches. This combined method yields a significant improvement of the solvation description on the "low-level" side of the quantum chemical hierarchy in comparison with the analytical linearized Poisson Boltzmann (ALPB) solvation model used as default in the `xTB` program package. While this eXtended Conductor-like Polarizable Continuum Model (CPCM-X) is specifically designed for the GFN2-xTB method, it can, in theory, be used as a post-processing method with a variety of PCM class solvation models, given that enough reference data for training is available. However, despite the use of sophisticated solvation models, it is still challenging to produce enough suitable reference data given the aforementioned inaccuracies in the description of solvation effects.

In Chapter 6, therefore, an efficient approach for the dynamic adjustment of radii for continuum solvation (the DRACO approach) is presented to improve existing solvation models. This approach dynamically scales the atomic radii used for a solvation evaluation based on an atoms-in-molecules approach. To do this, it uses an interface to efficient charge models to obtain atomic partial charges and fractional coordination numbers, which are used to model the chemical environment of the atoms. Incorporating dynamic radii in existing solvation models improves their accuracy significantly, especially for highly polar and ionic solutes. The DRACO method is tested on various versatile benchmark sets and published as an open source library on GitHub. It introduces no additional computational overhead and can be used with any solvation model that allows a custom modification of solute radii.
Although this thesis already yields significant improvements to the description of solvation interactions, the development of new solvation models is an ongoing process. Nonetheless, the physical insights gained through this thesis may contribute to further improvements if incorporated into future research.

Contents

CHAPTER 1

Introduction

The pace at which scientific breakthroughs occur due to advancements in modern technology is truly astonishing. A few decades ago, chemists needed to perform hundreds or thousands of experiments to determine possible candidates for a single commercial drug. This process could take anywhere from 12 to 15 years and cost over a billion dollars.[1] However, in today's world, automated computational workflows that combine an efficient sampling of the chemical space with highly accurate computational methods have expedited a significant portion of these processes. As a result, the necessary time, cost, and even environmental impact of the chemical drug development process has been vastly reduced.
A milestone on the way to reliable and widely usable computational chemistry methods was the development of density functional theory (DFT, see section 2.1.3), and efficient semi-empirical quantum mechanical methods (SQMs, see section 2.1.4). Sophisticated quantum chemical (QC) methods can yield accurate predictions for various molecular properties, such as saturation pressures (*cf.* Chapter 3), partition coefficients, or acid dissociation constants.[2–5] These properties can be important measures for predicting a drug's environmental and biological hazards like its tendency to become airborne or to bioaccumulate.[6,7] Experimental measurements of these properties are often time-consuming and expensive and can even be quite challenging, especially for reactions involving multiple or short-lived intermediates.
All of these properties are closely related to the free energy change of a specific chemical reaction (in the case of acid dissociation constants) or a phase shift (saturation pressure, partition coefficients). If the free energy of all involved systems is known, these properties can be derived from thermodynamic relationships. The free energy accounts for the system's energy and degrees of freedom. Furthermore, not only can molecule-specific properties be predicted if the free energy of a chemical system is known, but it is also possible to determine whether a new drug is pharmacologically active by assessing the difference in free energy between the protein-ligand complex and the free monomers, the so-called binding or association free energy (ΔG_a).[8–10]
To perform a computational investigation, first, it is essential to know the molecular structure of the relevant compounds and the appropriate experimental conditions, such as solvents or temperature. However, obtaining suitable input structures for molecules can be incredibly challenging. Although the three-dimensional structure can be provided from X-ray crystallography, the measurement conditions and the absence of solvent effects in solid-state crystals add additional uncertainty, making it difficult to obtain an accurate input structure. In addition, the same molecule can have different three-dimensional shapes that only differ by the rotation of a few single bonds (conformers).[11,12] These conformers,

while belonging to the same molecule, can have significantly different free energies, and only a few of them with the highest population will impact the actual molecular property. This problem is even more enhanced when the property depends on the molecule being in separate, sometimes highly different, phases. An excellent example of a case where a phase shift is an essential inherent characteristic of a property is the saturation pressure of a compound (P_{sat}). The influence of a solvent can considerably change a molecule's three-dimensional shape by stabilizing some molecular structures and destabilizing others (see Figure 1.1). Because of this, sampling the conformational space in each respective phase is typically the first step in setting up a computational calculation. This is often done by combining highly efficient force-field (FF) and SQM methods to investigate the relevant conformational space efficiently by creating large conformer-rotamer ensembles (CRE) and screening them on higher levels of theory. Finally, calculating the saturation pressure of a compound involves considering the difference in free energy (the reaction free energy of the phase change ΔG_r) of a complete ensemble in two different phases, the gas phase and a condensed phase. This final step is typically performed by utilizing higher-level DFT methods.

The combination of efficient force-field or semi-empirical methods with more sophisticated DFT methods, *i.e.*, a multi-level workflow, allows the processing of several tens of thousands of possible conformers on standard end-user hardware. Utilizing modern cloud technology, it is possible to screen billions of conformers in just a few days, provided sufficient resources are available. A recent study demonstrated that up to 12.7 billion conformers can be screened using a modern QC software package on approximately 29,000 CPU cores, taking just 56 hours. The cost of the whole sampling process was estimated to be around 20,300 dollars.[13] While this might seem expensive initially, it is insignificant compared to the costs of an entire commercial drug development process.

Nonetheless, even if it may become possible to screen the full conformational space of a large and flexible molecule by brute force, utilizing an enormous amount of computer power, there is still room for improvement by reducing the possible candidates through better screening methods and, consequently, reducing computational effort and cost. Furthermore, sophisticated computational chemistry methods are needed for the few low-lying and, therefore, relevant conformers to predict the free energy accurately. Complicating matters further is that a molecule's free energy is not a singular quantity but rather consists of several contributions that are often approximated separately. Commonly, the free energy for a single conformer is split into at least three contributions, namely the electronic energy (E_{el}), a thermodynamic contribution (G_T), and a solvation contribution (ΔG_{solv}) which will be explained in more detail in Chapter 2 of this thesis. For some of these contributions, accurate methods are already readily available.[14]

Sophisticated DFT methods will yield an accurate prediction of the electronic structure of the system and, thus, the electronic energy (see section 2.1), while efficient SQM methods allow the sampling at a highly reduced computational cost. Moreover, the description of the electronic energy is consistently improvable using more sophisticated methods or a larger basis set, given that the necessary computer processing power is available. For the inclusion of thermodynamic effects, *i.e.*, finite temperature effects, various approximate methods are available (see section 2.2), which will yield rather accurate results for an acceptable cost.

The one thing that is still a significant cause of error for current computational workflows is the influence of solvent on molecular properties and reactions, *i.e.*, the solvation contribution ΔG_{solv}. For example, when predicting drug properties, it is vital to consider solvent-solute interactions, as these compounds are typically digested in the human body and cannot be treated as being in the gas phase. There are, in principle, two ways to fathom solvent interactions, either by explicitly considering

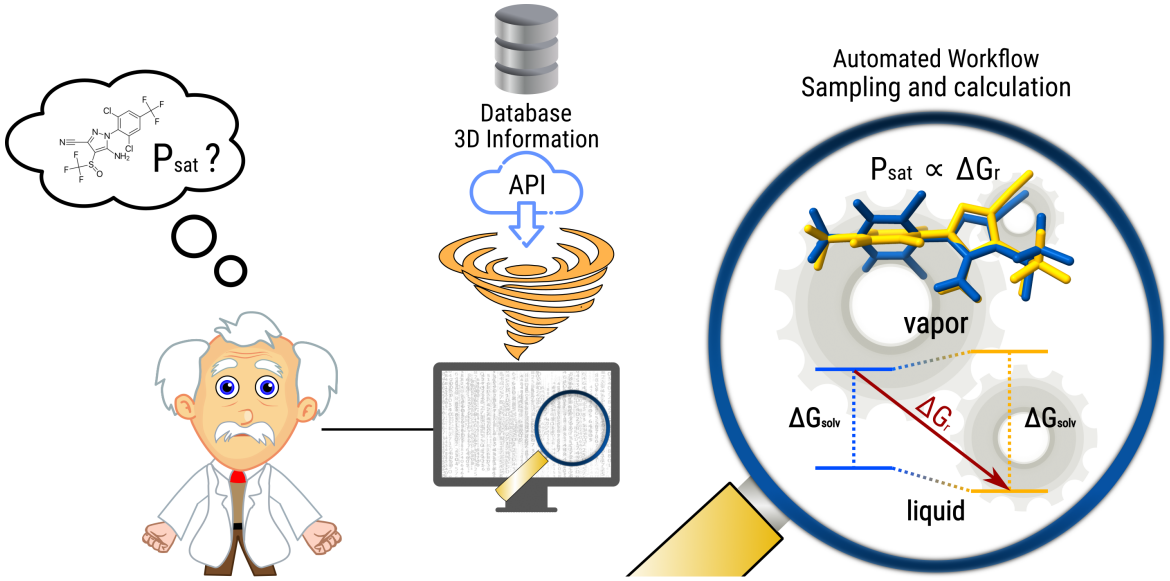


Figure 1.1: Schematic representation of a workflow that allows the calculation of saturation pressures for flexible compounds (P_{sat}).[15] First, a suitable description of the 3D information of a compound has to be obtained, *e.g.*, by performing a search in a suitable database with an identifier. Afterward, the conformational space of the molecule has to be screened individually in the gas phase and in the condensed phase to obtain a realistic description of the energy levels for each phase and the energy change during the phase transition. The saturation pressure can finally be obtained by thermodynamic relationships, using the reaction free energy of the phase change (ΔG_r) as an input.

the solvent molecules as part of the input geometry or by considering the solvent interactions as an average effect on the solute, *e.g.*, by a dielectric continuum. The latter approach is called implicit solvation. Explicitly considering the solvent molecules would vastly blow up the computational cost, which makes explicit solvent models too expensive for screening purposes.[16] But even if accuracy is important, explicit solvent models often yield worse results than their implicit counterparts,[17–19] which can be partly attributed to the difficult task of setting the models up correctly and considering full solvation shells.[20] Furthermore, adding solvent molecules to a system increases the complexity of sampling the conformational space. This is due to the additional degrees of freedom introduced by the solvent molecules. As the methods used for explicit solvation, such as molecular dynamics, often do not parallelize well, using more computer processing power can only partially compensate for the larger system size. For this reason, implicit solvation models are used in a wide range of applications, from screening purposes to highly accurate QC calculations.

Today, many different implicit solvation models are available. The basis for the theory of implicit solvation builds the Poisson-Boltzman equation, and the more simple solvation models are purely electrostatic ones[21] that treat the solvent as a uniform dielectric continuum that primarily depends on the dielectric constant (*cf.* section 2.3). More sophisticated solvation models also include extra terms to account for specific solvent-solute interactions, like hydrogen bonds or van der Waals interactions, but mostly by empirical terms. Although these models frequently produce satisfactory outcomes, they are limited by their reliance on empirical data. Consequently, investing more computer processing power, such as upgrading to higher levels of theory, does not consistently lead to improvements. This

is due to the inherent limitations of empirical data incorporated into the models,[22–24] which makes it hard to obtain suitable high-level reference data for the creation of new models or the evaluation of compounds that are outside of the common chemical space.
Moreover, most of these models are optimized for electronic structure calculation on the DFT level and are, therefore, not readily available for use with screening workflows based on SQM methods. In recent work, Ehlert *et al.* presented a solvation model for semi-empirical methods that employs the analytical linearized Poisson-Boltzmann method (ALPB).[25,26] This model can be used in combination with a rigorous conformational screening approach (CRENSO).[12] Although this model has shown reasonable agreement with experimental data, it is not as accurate as other, more sophisticated, solvation models, such as the conductor-like screening model for real solvents (COSMO-RS),[27,28] the COSMO segment activity coefficient (COSMO-SAC)[29] model, or the universal solvent model based on solute electron density (SMD).[30] In fact, while being computationally very efficient, the final accuracy of the ALPB is inferior to that of these other models.[23]
However, even these highly sophisticated empirical solvation models are still limited in terms of accuracy. This is especially true for solutes that include highly charged solutes,[31] or for solvents that exhibit strong explicit interactions,[32] that cannot be fathomed by the model in question. One possible reason for that is that the radii used to build the solute cavity for which electrostatic equations are solved are often statically determined based on the atomic number, either by empirically optimizing calculated properties to experimental data[33] or by using radii determined from, *e.g.*, crystal structures.[34,35] In reality, however, these radii should depend on the atom's atomic environment.[36–39] While some efforts have been made to include information about the system for the determination of the radii, these attempts mostly focus on water as a solvent and are either limited to certain atoms or functional groups,[40] or are developed for specific applications with a narrow range of training data.[41,42] Other, more sophisticated approaches based on an isocontour of the electron density introduce a significant additional computational effort and have other limitations, like the lack of analytical gradients, which would make an optimization of the structure geometry highly difficult.[43,44]
This work focuses on two areas of research: the investigation and the development of quantum chemical solvation models. Chapters 3 and 4 of this thesis will focus on the application of state-of-the-art quantum-chemical methods involving complete computational workflows with a rigorous sampling approach, highly efficient and accurate electronic structure methods, and sophisticated solvation models, to tackle the limit of accuracy, that is possible to reach for property calculations and computational investigations. Chapter 5 will introduce a new solvation model for the semi-empirical GFN2-xTB method based on a polarizable continuum model (PCM) that can be used to improve the low-level parts of a multi-level workflow. In Chapter 5, a new approach to obtain dynamic radii for atoms based on an atoms-in-molecule approach is introduced and investigated, which improves existing solvation models significantly, especially for the description of ionic compounds.

CHAPTER 2

Theory

This chapter introduces the theoretical foundation for calculating the various parts of the total free energy (G_{total}) that can be used to predict various properties through thermodynamic relationships. The forthcoming chapter will thus provide a summary of the existing knowledge and will be influenced by the book "Introduction to Computational Chemistry" authored by Frank Jensen.[45]
Generally, the total free energy G_{total} can be partitioned into different parts, which are usually the electronic energy E_{el}, a thermostatistical contribution G_T and a solvation contribution ΔG_{solv} as follows:

$$G_{total} = E_{el} + G_T + \Delta G_{solv} \,. \tag{2.1}$$

The electronic energy is obtained from the electronic structure of a system. It corresponds to a molecule in the gas phase at 0 K and will be introduced in section 2.1. The thermostatistical contribution (see section 2.2) introduces finite temperature effects; for example, adding the contribution $G_{T,298\,K}$ would correspond to a molecule in the gas phase at 298 K. The solvation contribution, at last, corresponds to a transfer from the gas phase to a solvated phase and is introduced in section 2.3. This contribution needs to be considered if some solvent is involved in the experiment.
All of these contributions are commonly calculated for a single molecular structure, which may not be enough for highly flexible molecules for which many different, energetically comparable conformers can exist. How and why a conformational ensemble should be treated instead will be explained in section 2.4.

2.1 Electronic Energy

The electronic energy (E_{el}) corresponds to the energy of a molecule in the gas phase at absolute zero ($T = 0$ K). The electronic energy of a system can be calculated *ab initio* ("from the beginning") without reference to experimental data or by introducing additional approximations (*e.g.*, by semi-empirical methods, see section 2.1.4).

2.1.1 The electronic Hamiltonian

The electronic energy (E_{el}) of a system depends on its electronic structure, but as electrons are very light particles ($m_e \approx 9.109 \cdot 10^{-31}$ kg),[46] they tend to behave not as we would expect them from

classical mechanics. Because of that, if we want to describe the electronic structure of a system, there is no way around quantum mechanics (QM), *i.e.*, the time-independent Schrödinger equation,[47,48] given as

$$\mathbf{H}\Psi = \mathrm{E}\Psi\,, \tag{2.2}$$

where $\mathbf{H}$ is the Hamiltonian operator, Ψ is the wave function, and E is the energy. The total, time-independent Hamiltonian operator consists of the kinetic ($\mathbf{T}$) and potential ($\mathbf{V}$) energies of electrons and nuclei and can be written as[49]

$$\mathbf{H}_{\mathrm{tot}} = \mathbf{T}_n + \mathbf{T}_e + \mathbf{V}_{ne} + \mathbf{V}_{ee} + \mathbf{V}_{nn}\,, \tag{2.3}$$

where $\mathbf{T}_n$ and $\mathbf{T}_e$ is the kinetic energy of the nuclei and electrons and $\mathbf{V}_{ne}$, $\mathbf{V}_{nn}$ and $\mathbf{V}_{ee}$ is the potential energy between nuclei-electrons, nuclei-nuclei and electrons-electrons, respectively.

Because nuclei are much heavier than electrons, the motion of the electrons can be seen as independent from the motion of the nuclei. This assumption is called the Born-Oppenheimer approximation.[50–52] The nuclei positions are treated as parameters, and therefore, the kinetic energy of the nuclei can be neglected ($\mathbf{T}_n = 0$), which leads to the electronic Hamiltonian Operator

$$\mathbf{H}_{\mathrm{el}} = \mathbf{T}_e + \mathbf{V}_{ne} + \mathbf{V}_{ee} + \mathbf{V}_{nn}\,. \tag{2.4}$$

The contributions to the electronic Hamilton operator can be calculated as[53]

$$\mathbf{T}_e = -\sum_{i}^{N_{\mathrm{elec}}} \frac{1}{2}\nabla_i^2\,, \tag{2.5}$$

$$\mathbf{V}_{ne} = -\sum_{A}^{N_{\mathrm{nuclei}}}\sum_{i}^{N_{\mathrm{elec}}} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|}\,, \tag{2.6}$$

$$\mathbf{V}_{nn} = \sum_{A}^{N_{\mathrm{nuclei}}}\sum_{B>A}^{N_{\mathrm{nuclei}}} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}\,, \tag{2.7}$$

$$\mathbf{V}_{ee} = \sum_{i}^{N_{\mathrm{elec}}}\sum_{j>i}^{N_{\mathrm{elec}}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \tag{2.8}$$

where ∇_i^2 is the Laplace operator, $\mathbf{r}_i$ are the coordinates of electron i, $\mathbf{R}_{A/B}$ are the coordinates of the nucleus A/B and $Z_{A/B}$ is the atomic number of the nucleus A/B.

These operators can be collected according to the number of electron indices for a fixed set of nuclei positions. $\mathbf{V}_{nn}$ is constant for a given system and does not depend on electron positions. $\mathbf{V}_{ne}$ and $\mathbf{T}_e$ only depend on one electron coordinate, and $\mathbf{V}_{ee}$ depends on two electron coordinates and is the most expensive to calculate. By formulating a core-Hamiltonian $\mathbf{h}_i$ that describes the motion of an electron

in the field of all nuclei and a two-electron operator $\mathbf{g}_{ij}$ for the electron-electron repulsion as

$$\mathbf{h}_i = \frac{1}{2}\nabla_i^2 - \sum_A^{N_{\text{nuclei}}} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|}\,, \tag{2.9}$$

$$\mathbf{g}_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \tag{2.10}$$

the electronic Hamilton operator can be rewritten as

$$\mathbf{H}_{\text{el}} = \sum_i^{N_{\text{elec}}} \mathbf{h}_i + \sum_{j>i}^{N_{\text{elec}}} \mathbf{g}_{ij} + \underbrace{\mathbf{V}_{nn}}_{\text{const.}}\,. \tag{2.11}$$

2.1.2 Hartree-Fock Theory

Despite that, even with the limitations introduced by utilizing the Born-Oppenheimer approximation and neglecting relativistic effects, the exact solution to the (electronic) Schrödinger equation is still overwhelmingly complex for a multi-particle system. Therefore, approximate solutions to the many-particle wave function are needed.[53] These solutions can be generated by exploiting the variational principle, which states that an approximate wave function always has an equal or larger energy than the corresponding exact wave function:

$$E_{\text{el,approx.}} \geq E_{\text{el,exact}}\,. \tag{2.12}$$

The energy of an (approximate) wave function can be calculated as the expectation value of the electronic Hamilton operator as

$$E_{\text{el}} = \langle \Psi | \mathbf{H}_{\text{el}} | \Psi \rangle\,. \tag{2.13}$$

One way to formulate an approximate wave function is a Slater determinant (SD). Using an SD is especially elegant, as this form fulfills the Pauli principle, which states that two electrons cannot have exactly the same quantum state and retains the indistinguishability of the electrons. The general case of a wave function with N electrons can be written as a Slater determinant as

$$\Psi \approx \Phi(1,2,..,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \dots & \phi_N(N) \end{vmatrix}. \tag{2.14}$$

The utilization of a single determinant for the wave function, though convenient, results in the neglect of electron correlation. In the context of the Hartree-Fock (HF) theory, electron-electron repulsion is therefore accounted for only on an average level.[54–57] To explicitly consider electron correlation, more complex methods such as multideterminant approaches (electron correlation methods) are required.[58] However, these methods are computationally intensive and not covered in this thesis.

Using the single determinant approach with the formulation of the electronic Hamiltonian in equation

2.11, the Hartree-Fock energy can be expressed as

$$E_{\text{el, HF}} = \langle \Phi | \mathbf{H}_{\text{el}} | \Phi \rangle = \sum_{i=1}^{N_{\text{elec}}} \underbrace{\langle \phi_i | \mathbf{h}_i | \phi_i \rangle}_{h_i} + \frac{1}{2} \sum_{i=1}^{N_{\text{elec}}} \sum_{j=1}^{N_{\text{elec}}} \left(\underbrace{\langle \phi_i \phi_j | \mathbf{g}_{ij} | \phi_i \phi_j \rangle}_{J_{ij}} - \underbrace{\langle \phi_i \phi_j | \mathbf{g}_{ij} | \phi_j \phi_i \rangle}_{K_{ij}} \right) + \mathbf{V}_{nn} \,. \tag{2.15}$$

Here, h_i represents the one-particle energy of the electron i, J_{ij} depicts the Coulomb interaction of electrons i and j and K_{ij} the corresponding exchange element. The $\frac{1}{2}$ factor counteracts the electron-electron interactions' double counting. The self-interaction of an electron (J_{ii}) is exactly canceled out by the exchange element K_{ii}.
According to the variational principle (see equation 2.12), finding a set of molecular orbitals that minimizes the energy is necessary. A Lagrange function is employed to maintain the normalization and orthogonality of the MOs. By ensuring that the Lagrange function remains stationary with respect to a variation in the orbitals, the minimum of the Lagrange function can be determined as follows:

$$L = E - \sum_{ij}^{N_{\text{elec}}} \lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \,, \tag{2.16}$$

$$\delta L = \delta E - \sum_{ij}^{N_{\text{elec}}} \lambda_{ij} (\langle \delta\phi_i | \phi_j \rangle - \langle \phi_i | \delta\phi_j \rangle) \stackrel{!}{=} 0 \,. \tag{2.17}$$

Using equation 2.15 for the variation of the energy (δE) and introducing a unitary transformation diagonalizing the Lagrange multiplier matrix ($\lambda_{ij} = 0$ and $\lambda_{ii} = \varepsilon_i$), the Hartree-Fock equation

$$\left(\underbrace{h_i + \sum_{j}^{N_{\text{elec}}} (J_j - K_j)}_{\mathbf{F}_i} \right) \phi_i' = \varepsilon_i \phi_i' \tag{2.18}$$

is derived. The term $h_i + \sum_j^{N_{\text{elec}}} (J_j - K_j)$ is called the Fock operator $\mathbf{F}_i$. The special molecular orbitals obtained by this unitary transformation (ϕ_i') are called canonical orbitals, with ε_i being the canonical orbital energies.

2.1.3 Density Functional Theory

Another approach to approximate the electronic energy of a system is based on the proof that the electron density of a system completely determines the electronic ground-state energy,[59] *i.e.*, there is a one-to-one mapping between electron density (ρ) and electronic ground-state energy (E_{el}).
This theory is called Density Functional Theory (DFT), with the DFT ground-state energy $E_{\text{DFT}}[\rho]$. Because, up until today, the exact functional dependence is still unknown, the development of new DFT functionals, connecting the electron density and the ground-state energy is an active area of

research.[14,60–62] The complete ground-state energy can, analogous to equation 2.4, be divided into several electron-density dependent parts: The kinetic energy $T[\rho]$, the potential energy between electrons and nuclei $V_{ne}[\rho]$ and the potential energy between two electrons $V_{ee}[\rho]$. The latter can be divided into coulomb ($J[\rho]$) and exchange ($K[\rho]$) parts according to equation 2.15, resulting in a general DFT energy expression as

$$E_{\mathrm{DFT}}[\rho] = T[\rho] + V_{ne}[\rho] + \underbrace{J[\rho] + K[\rho]}_{V_{ee}[\rho]} . \tag{2.19}$$

The classical expressions are used for $V_{ne}[\rho]$ and $J[\rho]$:

$$V_{ne}[\rho] = -\sum_{a}^{N_{\mathrm{nuclei}}} \int \frac{Z_a(\mathbf{R}_a)\rho(\mathbf{r})}{|\mathbf{R}_a - \mathbf{r}|} d\mathbf{r} , \tag{2.20}$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r'})}{|\mathbf{r} - \mathbf{r'}|} d\mathbf{r} d\mathbf{r'} . \tag{2.21}$$

Early attempts to deduce the kinetic and exchange energies of the electrons considered the electron distribution as a uniform electron gas.[63,64] Although this may be a reasonable assumption for the valence electrons in periodic systems, this assumption is highly inaccurate for molecules. With the introduction of orbitals, first suggested by Kohn and Sham in 1965,[65] DFT became usable for computational chemistry.

Following the Kohn-Sham (KS) model, the kinetic energy can be split into two parts: the major contribution to the kinetic energy, which is derived from HF theory under the assumption of non-interacting electrons ($T_S[\rho]$) as

$$T_S = \sum_{i}^{N_{\mathrm{elec}}} \left\langle \phi_i \left| -\frac{1}{2}\nabla^2 \right| \phi_i \right\rangle , \tag{2.22}$$

and an additional correction, which is called the exchange-correlation term ($E_{XC}[\rho]$). $E_{XC}[\rho]$ is defined as the missing part to the exact energy as

$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + (V_{ee}[\rho] - J[\rho]) . \tag{2.23}$$

This results in the general KS-DFT equation

$$E_{\text{KS-DFT}}[\rho] = T_S[\rho] + V_{ne}[\rho] + J[\rho] + E_{XC}[\rho] . \tag{2.24}$$

The task in KS-DFT is to derive an approximation for the energy-exchange functional only (in comparison to the total kinetic and exchange functional in orbital-free DFT). Incorporating orbitals in KS-DFT reduces the susceptibility to inaccuracies in the function. On the other hand, this also increases complexity since the number of variables increases from 3 to 3N. Moreover, it is not yet known how the exchange-correlation and electron density in KS-DFT are precisely related, similar to the relationship between the complete electronic ground-state energy and the electron density in orbital-free DFT. The functionals developed to solve this problem can be categorized according to the information used, such as the local density or its derivative. A commonly used categorization scheme

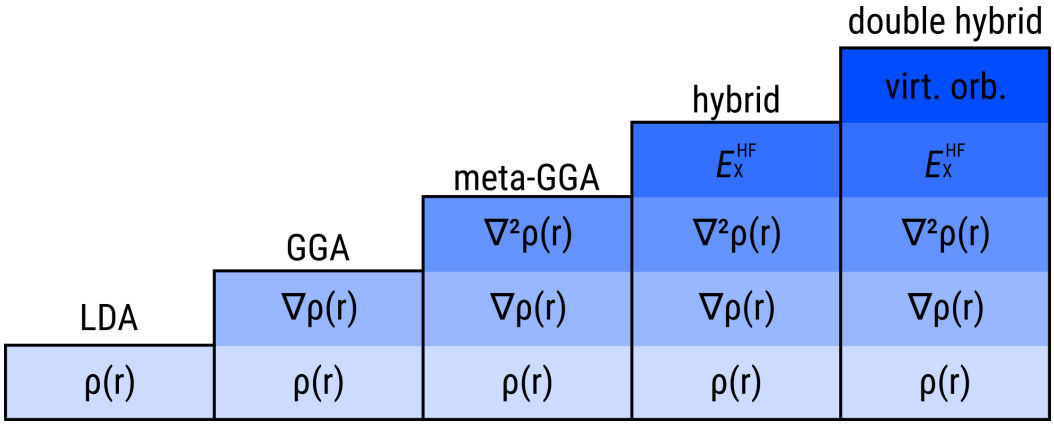


Figure 2.1: Schematic representation of the Jacob's ladder as proposed by Perdew & Schmidt.[66]

was proposed by Perdew & Schmidt.[66] The so-called Jacob's ladder categorizes the functionals into five rungs as visualized in Figure 2.1.
The lowest rung of the Jacob's ladder is called local density approximation (LDA). Functionals of this rung only use the local electron density $\rho(\mathbf{r})$, *i.e.*, the density is locally treated as a uniform electron gas.[57,67,68] By climbing to a higher rung of Jacob's ladder, the complexity of the functional and, thus, the computational cost and accuracy will increase. Generalized gradient approximation (GGA) type functionals additionally use the first gradient of the electron density $\nabla\rho(\mathbf{r})$.[69–72] The next rung, which is built by the meta-GGA functionals, also includes higher-order derivatives of the electron density $\nabla^2\rho(\mathbf{r})$, or alternatively, the orbital kinetic energy $\tau(\mathbf{r})$.[73–75] The upper two rungs of Jacob's ladder are called hybrid and double hybrid, respectively. As the name suggests, they endeavor to approximate the exchange-correlation energy by using another theory, where the exchange energy can be exactly calculated, namely the Hartee-Fock theory (see section 2.1.2).[76–78] In spite of that, because the KS orbitals are not identical to the HF orbitals, the calculated energy is not the exact solution, and therefore, only a fraction of the so-called Fock exchange is taken as an additional contribution.[79] As the last rung of Jacob's ladder, double hybrid functionals introduce virtual orbital information, *e.g.*, by perturbation theory.[80–82]

2.1.4 Density Functional Tight-Binding

While KS-DFT (section 2.1.3) is already a computationally efficient method in comparison with HF theory (section 2.1.2), it still can become too expensive for routine screening workflows or computation of large molecules with several hundred atoms (and therefore thousands of electrons). With the introduction of several approximations, semi-empirical quantum mechanical methods (SQMs) can significantly reduce this computational effort. Early SQMs introduced additional approximations to the Hartree-Fock theory (*e.g.*, Neglect of Diatomic Differential Overlap), but since nowadays DFT methods have mostly replaced HF methods, new SQMs were developed as approximations to DFT.
A common class of models is based on tight-binding, like Density Functional based Tight-Binding (DFTB)[83] or extended Tight-Binding (xTB).[84,85] Instead of calculating the DFT energy from equation 2.24, at first, a reference density ρ_0 is calculated from a superposition of neutral atomic densities, based on the assumption that only valence electrons have to be explicitly considered because the core

electron density is assumed to be mostly invariant to the environment. The energy is then expanded as a Taylor series in terms of charge density fluctuation $\delta\rho$ as[86]

$$E[\rho] = E^0[\rho_0] + E^1[\rho_0, \delta\rho] + E^2[\rho_0, (\delta\rho)^2] + E^3[\rho_0, (\delta\rho)^3] + \cdots . \tag{2.25}$$

In the case of the GFN2-xTB[85] method, the contributions to the total free energy are defined as follows:

$$E_{\text{GFN2-xTB}} = E^{(0)}_{\text{rep}} + E^{(1)}_{\text{EHT}} + E^{(2)}_{\text{ies+xc}} + E^{(3)}_{\text{ies+xc}} + E_{\text{aes}} + E_{\text{axc}} + E_{\text{D4}} + G_{\text{fermi}} . \tag{2.26}$$

The specifics of the individual contributions will be discussed below.
The zeroth-order term is the repulsion energy, which is independent of the charge fluctuations and therefore only includes geometry-dependent contributions as[85]

$$E^{(0)}_{\text{rep}} = \sum_{AB}^{N} \frac{Z^{\text{eff}}_A Z^{\text{eff}}_B}{R_{AB}} \mathrm{e}^{-(a_A a_B)^{1/2} (R_{AB})^{k_{\text{rep}}}} , \tag{2.27}$$

where R_{AB} is the distance between atom A and B and $Z^{\text{eff}}_{A/B}$, as well as $a_{A/B}$ are parameters that define the magnitude of the repulsive interaction. While $Z^{\text{eff}}_{A/B}$ can be thought of as effective nuclear charges that correspond to the term of V_{nn} (*cf.* equation 2.7), in GFN2-xTB they correlate not as strongly with the nuclear charge as in other methods, like GFN1-xTB.[84,85] k_{rep} is a parameter to differentiate between light and heavier elements and is mostly kept constant (one for H/He, 3/2 otherwise).
The first-order term is derived from extended Hückel Theory (EHT) and given as[85]

$$E^{(1)}_{\text{EHT}} = \sum_{\kappa} \sum_{\lambda} P_{\kappa\lambda} \mathrm{H}_{\kappa\lambda} , \tag{2.28}$$

with the density matrix $P_{\kappa\lambda} = P^0_{\kappa\lambda} + \delta P_{\kappa\lambda}$ and the Hamiltonian matrix H. The diagonal elements of the Hamiltonian matrix are thereby calculated based on an empirical scheme. In GFN2-xTB, they are flexible in accordance with the chemical environment by introducing the fractional coordination number CN_κ as

$$\mathrm{H}_{\kappa\kappa} = h_\kappa - h_{\text{CN}_\kappa} \text{CN}_\kappa , \tag{2.29}$$

where h_κ and h_{CN_κ} are empirical parameters. The off-diagonal elements are also empirically determined based on averages of the diagonal elements.
For charged and polar systems, the net partial charges of the individual atoms are nonzero. Therefore, the electron density ρ at these atoms will deviate from the reference density of the neutral atom ρ_0 by a charge fluctuation $\delta\rho$. This deviation is treated with higher-order terms, using shell-wise partitioned Mulliken partial charges, called the isotropic electrostatic and exchange-correlation energy ($E_{\text{ies+xc}}$). The energy consists of a second-order and a third-order term and is calculated as

$$E_{\text{ies+xc}} = \underbrace{\frac{1}{2} \sum_{A,B} \sum_{l \in A} \sum_{l' \in B} q_{A,l} q_{B,l'} \gamma_{AB,ll'}}_{E^2_{\text{ies+xc}}} + \underbrace{\frac{1}{3} \sum_{A} \sum_{l \in A} \Gamma_{A,l} q^3_{A,l}}_{E^3_{\text{ies+xc}}} , \tag{2.30}$$

where $q_{X,l}$ are the respective Mulliken charges of the shell l on atom X, $\gamma_{AB,ll'}$ is the distance-dependent Coulomb interaction and $\Gamma_{A,l}$ is a shell and element-specific parameter. The Coulomb

interaction is calculated as[87–90]

$$\gamma_{AB,ll'} = \left(\frac{1}{(R_{AB})^2 + \eta_{AB,ll'}^{-2}} \right)^{\frac{1}{2}} , \tag{2.31}$$

where R_{AB} is the distance between the atoms A and B, and $\eta_{AB,ll'}$ is the average of the effective chemical hardness of the two shells l and l' on the atoms A and B. $\eta_{AB,ll'}$ is empirically determined from element and shell-specific parameters.

In addition to these general tight-binding terms, GFN2-xTB also includes multipole electrostatics with the additional second-order anisotropic E_{aes} and E_{axc} terms. Furthermore, the DFT-D4[91,92] correction is used to include London dispersion interactions, which is an enhancement to the DFT-D3[93] model used in GFN1-xTB.[84] Static correlation due to the spin-restricted wave function used in GFN2-xTB is handled by introducing fractional occupations of the energy levels at finite electronic temperatures. This approach is called Fermi smearing with its corresponding term G_{fermi}.[85,94]

2.2 Thermostatistical Contribution

The electronic ground-state energy introduced in section 2.1 is usually calculated for a single molecule or an aggregate in the gas phase at 0 K. Experiments, on the other hand, mostly deal with finite temperatures and with macroscopic samples. The connection between a microscopic system and a macroscopic system is described by the partition function. The partition function for a single molecule (q) is defined as an exponential sum over all possible quantum states as[95]

$$q = \sum_{i=\text{states}}^{\infty} \exp\left(\frac{-\epsilon_i}{\mathrm{k}T} \right) , \tag{2.32}$$

where k is the Boltzmann constant, T is the temperature, and ϵ_i is the energy of the respective quantum state. For a macroscopic system of N non-interacting identical particles (ideal gas), the macroscopic partition function Q can be calculated as

$$Q = \frac{q^N}{N!} . \tag{2.33}$$

From the partition function, a variety of thermodynamic functions can be calculated, like the internal energy U and the Helmholtz free energy A:

$$U = \mathrm{k}T^2 \left(\frac{\delta \ln Q}{\delta T} \right)_V , \tag{2.34}$$

$$A = -\mathrm{k}T \ln Q . \tag{2.35}$$

Because the pressure (P) and the heat capacity at constant volume (C_V) can be calculated as derivatives of these thermodynamic functions with respect to the volume (V) as

$$P = -\left(\frac{\delta A}{\delta V}\right)_T , \tag{2.36}$$

$$C_V = \left(\frac{\delta U}{\delta V}\right)_V , \tag{2.37}$$

also, the enthalpy (H) and the entropy (S) can be derived as

$$H = U + PV = \mathrm{k}T^2 \left(\frac{\delta \ln Q}{\delta T}\right)_V + \mathrm{k}TV \left(\frac{\delta \ln Q}{\delta V}\right)_T \tag{2.38}$$

$$S = \frac{U - A}{T} = \mathrm{k}T \left(\frac{\delta \ln Q}{\delta T}\right)_V + \mathrm{k}T \ln Q \, . \tag{2.39}$$

Finally, the Gibbs free energy is therefore defined as

$$G = H - TS = \mathrm{k}TV \left(\frac{\delta \ln Q}{\delta V}\right)_T - \mathrm{k}T \ln Q \, . \tag{2.40}$$

Still, in order to know all possible quantum states for a system and be able to derive q (cf. equation 2.32), it would be necessary to solve the nuclear Schrödinger equation, which is not feasible for larger systems. Therefore, the Rigid-Rotor Harmonic-Oscillator (RRHO) approximation is often employed to solve this problem, which assumes that the electronic, rotational and vibrational degrees of freedom are seperable.[95,96] Using the HO approximation, vibrational frequencies are obtained by a second-order Taylor expansion around the equilibrium distance while the rotational degrees of freedom are described as a rigid-rotor. q_{tot} can therefore be written as

$$q_{\mathrm{tot}} = q_{\mathrm{rot}} q_{\mathrm{vib}} q_{\mathrm{trans}} q_{\mathrm{elec}} \, , \tag{2.41}$$

with the individual contributions

$$q_{\mathrm{rot}} = \frac{\sqrt{\pi}}{\sigma} \left(\frac{8\pi^2 \mathrm{k}T}{h^2}\right)^{\frac{3}{2}} \sqrt{I_x I_y I_z} \, , \tag{2.42}$$

$$q_{\mathrm{vib}} = \prod_i^{vib.modes} \frac{\mathrm{e}^{-h\nu_i/(2\mathrm{k}T)}}{1 - \mathrm{e}^{-h\nu_i/(\mathrm{k}T)}} \, , \tag{2.43}$$

$$q_{\mathrm{trans}} = \left(\frac{2\pi M \mathrm{k}T}{h^2}\right)^{\frac{3}{2}} V \, , \tag{2.44}$$

$$q_{\mathrm{elec}} = \sum_i^{\infty} g_i \mathrm{e}^{\frac{\epsilon_i}{\mathrm{k}T}} \, . \tag{2.45}$$

Here, $I_{x,y,z}$ denotes the moment of inertia around the principle axes, h is the Planck constant, σ is the rotational symmetry number, ν_i is the vibrational frequency for mode i, M is the mass of the molecule, g_i the degeneracy of state i and ϵ_i the energy of state i.

It is generally observed that the energy difference between the ground state and the first excited electronic state is significantly greater than the product of the Boltzmann constant and the temperature (kT), which is used in the calculation of the electronic partition function (q_{elec}). In such cases, only the ground state needs to be taken into consideration, which can be determined by utilizing an appropriate electronic structure method (as discussed in section 2.1). Given the partition functions, G_T can be calculated according to equation 2.40, as the sum of the non-electronic degrees of freedom:

$$G_T = (H_{\text{rot}} + H_{\text{vib}} + H_{\text{trans}}) - T(S_{\text{rot}} + S_{\text{vib}} + S_{\text{trans}}) \,. \tag{2.46}$$

2.3 Solvation Contribution

The last sections have dealt with electronic energy (section 2.1) and thermostatistical contributions (section 2.2). Considering these contributions, it is possible to calculate the free energy for molecules in the gas phase at finite temperatures. In contrast to that, most chemical reactions take place in a particular medium, most commonly a solvent. In this, the interactions and properties in experiments can strongly differ from the ones observed in the gas phase. For this reason, approximating the effect of solvent molecules on a solute is of utmost importance for nearly all areas of computational chemistry. Considering solvent effects can either be done explicitly, *e.g.*, by simulating individual solvent molecules and therefore explicitly simulating the solvent-solute interface,[19,97] or implicitly, *e.g.*, by a dielectric continuum.[98,99] Due to the extra solvent molecules, which can far outnumber the atoms of the actual solute, explicit solvation models typically have significant additional computational overhead and are hardly usable for high throughput screening applications. As they are also not part of this thesis, this chapter will concentrate on implicit solvation models.

For an implicit solvation model, the solvation free energy (ΔG_{solv}) is defined as an additive contribution to the total free energy in the gas phase (G_{gas}), which describes the energy change that corresponds to a transfer of a molecule from the gas phase (with the energy G_{gas}) to the solvation phase (with the energy G_{sol}):

$$G_{\text{sol}} = G_{\text{gas}} + \Delta G_{\text{solv}} \,. \tag{2.47}$$

The solvation free energy (ΔG_{solv}) is usually partitioned into a sum of contributions, which is often the electrostatic energy ΔG_{elec}, the energy required to create a cavity for the solute in a solvent ΔG_{cav}, the stabilization of the solute due to dispersion interactions with the solvent molecules ΔG_{disp} and the exchange-repulsion ΔG_{exch} as[100]

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \underbrace{\Delta G_{\text{cav}} + \Delta G_{\text{disp}} + \Delta G_{\text{exch}}}_{\Delta G_{\text{ne}}} \,. \tag{2.48}$$

The latter three parts are commonly combined into a non-electrostatic contribution to the free energy ΔG_{ne}. Depending on the method and theoretical framework, additional contributions are sometimes considered, such as a hydrogen bonding correction ΔG_{hb}.

This section will summarize the theory behind some basic implicit solvation models. These models serve as the basis for more sophisticated models that have been developed over time. For a more detailed overview of the theory behind more complex solvation models, including those created over the course of this thesis, please refer to Appendix C and D.

2.3.1 Poisson-Boltzman Methods

The electrostatic energy contribution to the solvation free energy can be calculated using the dielectric continuum theory, which assumes that a solvent can be coarse-grained by a polarizable dielectric continuum. The Poisson equation[101,102] gives the interaction of a solute with this continuum in its generalized form as

$$\nabla\left(\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})\right) = -4\pi\rho(\mathbf{r})\,, \tag{2.49}$$

where $\varepsilon(\mathbf{r})$ is the relative permittivity in space, $\nabla\varphi(\mathbf{r})$ is the gradient of the electrostatic potential and $\rho(\mathbf{r})$ is the charge density of the solute, obtained *e.g.*, from an electronic structure calculation (*cf.* section 2.1). In implicit solvation models, the relative permittivity $\varepsilon(\mathbf{r})$ is most commonly replaced by a scalar, which is the macroscopic dielectric constant ε, leading to:

$$\varepsilon\nabla^2\varphi(\mathbf{r}) = -4\pi\rho(\mathbf{r})\,. \tag{2.50}$$

However, there is still the need to differentiate between the solute itself (the inner dielectric constant ε_{in}) and the dielectric continuum surrounding it (the outer dielectric constant ε_{out}). In most quantum chemical calculations, ε_{in} is set to one, corresponding to a vacuum's relative permittivity. After obtaining $\varphi(\mathbf{r})$ from equation 2.50, this potential can be separated into two parts as

$$\varphi(\mathbf{r}) = \varphi^{\rho}(\mathbf{r}) + \varphi^{\text{reac}}(\mathbf{r})\,, \tag{2.51}$$

where $\varphi^{\text{reac}}(\mathbf{r})$ is the reaction field arising from the polarization of the dielectric continuum and $\varphi^{\rho}(\mathbf{r})$ is induced by the charge density of the solute as

$$\varphi^{\rho}(\mathbf{r}) = \int \frac{\rho(\mathbf{r'})}{|\mathbf{r'} - \mathbf{r}|} d\mathbf{r'}\,. \tag{2.52}$$

The electrostatic part of the solvation free energy can then be obtained as[103]

$$\Delta G_{\text{elec}} = \frac{1}{2}\int \varphi^{\rho}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}\,. \tag{2.53}$$

For a realistic molecular cavity, the Poisson equation has to be solved numerically, *e.g.*, by a grid representation, making this method computationally demanding.

2.3.2 The (Generalized) Born Model

If a spherical cavity is assumed, and this spherical cavity contains the entire charge density, an analytic solution to the Poisson equation exists as

$$\Delta G_{\text{elec}}(q) = -\left(1 - \frac{1}{\varepsilon}\right)\frac{q^2}{2R}\,, \tag{2.54}$$

where q is the net charge in a cavity of radius R. This equation leads to the formulation of the Born model.[104] Using partial point charges instead of a net charge and combining Coulomb interactions with the Born formula via a switching function (interaction kernel) yields the generalized Born

equation[105,106]

$$\Delta G_{\text{elec}}(Q_A, Q_B) = -\left(1 - \frac{1}{\varepsilon}\right)\frac{Q_A Q_B}{f_{AB}}\,, \tag{2.55}$$

where Q_A and Q_B are partial charges and f_{AB} is a switching function depending on the internuclear distances and Born radii for the respective atoms. A well-known interaction kernel is the one proposed by Still in 1990[105] as

$$f_{AB}^{\text{Still}} = \left(R_{AB}^2 + r_A r_B \exp\left[-\frac{R_{AB}^2}{4 r_A r_B}\right]\right)^{\frac{1}{2}}\,, \tag{2.56}$$

where R_{AB} is the distance between nucleus A and nucleus B and r_B, as well as r_A are the respective Born radii.

2.3.3 Polarizable Continuum Models

While the generalized Born model works quite well for small systems, a spherical cavity is a relatively crude approximation of a molecular cavity, especially for larger and more complex molecules. Therefore, in implicit solvation methods, the cavity is more commonly formed by overlapping spheres, which are created, for example, by utilizing van der Waals radii (*cf.* section 2.3.4). This makes it again necessary to solve the Poisson equation numerically. Because this is computationally expensive in a three-dimensional space, the problem is often reformulated into an apparent surface charge (ASC) problem on a two-dimensional solvent surface.[107] This class of models is called polarizable continuum models (PCMs).[98,99,108,109]

Instead of solving the continuum solvation problem in three dimensions, PCMs transform it into a surface charge problem that numerical methods can solve far more efficiently. This opens up the problem of defining an interface between the inner cavity and the outer dielectric continuum. In its most simple form, a "jump" boundary condition is assumed.[110] That means if a point inside the cavity undergoes an infinitesimal change in position, so it moves outside the cavity, it will experience an electric field of ε_{out} instead of ε_{in}. This leads to

$$\varepsilon_{\text{out}} \delta_{\mathbf{s}} \varphi(\mathbf{s})|_{\mathbf{s}=\mathbf{s}^+} = \varepsilon_{\text{in}} \delta_{\mathbf{s}} \varphi(\mathbf{s})|_{\mathbf{s}=\mathbf{s}^-}\,, \tag{2.57}$$

where $\mathbf{s}$ is a point on the cavity surface, $\delta_{\mathbf{s}}$ is the normal derivative, and $\mathbf{s}^{\pm}$ denotes if the one-sided derivatives need to be evaluated directly inside or directly outside the cavity.

In the following, $\sigma(\mathbf{s})$ is used as the surface charge to differentiate from the volume charge $\rho(\mathbf{r})$. In order to satisfy equation 2.57, equation 2.50 leads to[110]

$$4\pi\sigma(\mathbf{s}) = \left(\frac{\varepsilon_{\text{out}} - \varepsilon_{\text{in}}}{\varepsilon_{\text{out}}}\right)\delta_{\mathbf{s}} \varphi(\mathbf{s})|_{\mathbf{s}=\mathbf{s}^-}\,. \tag{2.58}$$

For ε_{out} set to the dielectric constant of the solvent (ε) and $\varepsilon_{\text{in}} = 1$, this yields

$$\sigma(\mathbf{s}) = \frac{1}{4\pi}\left(\frac{\varepsilon - 1}{\varepsilon}\right)\delta_{\mathbf{s}} \varphi(\mathbf{s})|_{\mathbf{s}=\mathbf{s}^-}\,, \tag{2.59}$$

which corresponds to the formulation of the "original" PCM model.[21,111,112] Analogous to equation 2.51, φ can be split into two parts. One which comes directly from the solute (φ^{ρ}) and one from the

reaction-field contribution. To calculate the reaction-field contribution $\varphi^{\sigma}(\mathbf{r})$, only the cavity surface instead of the whole space has to be discretized. The corresponding potential can be calculated as

$$\varphi^{\sigma}(\mathbf{r}) = \int \frac{\sigma(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} d\mathbf{s}\,. \tag{2.60}$$

This could also be added to the Hamiltonian operator for a self-consistent treatment. The electrostatic part of the solvation free energy can finally be calculated as an integral over the surface (*cf.* equation 2.53) as

$$\Delta G_{\text{elec}} = \frac{1}{2} \int \varphi^{\rho}(\mathbf{s})\sigma(\mathbf{s}) d\mathbf{s}\,. \tag{2.61}$$

2.3.4 Non-electrostatic Contribution

The non-electrostatic contribution is often assumed to be proportional to the surface area of the solute. As a result, the entire surface area is frequently mapped to individual solute atoms, and the proportionality factor for the non-electrostatic contribution is assumed to be solely element-specific. This surface area depends on the algorithm used to create the solvent cavity. A relatively simple approach would be to use a spherical cavity, which would also allow the analytical derivation of the electronic energy (*cf.* section 2.3.2). Because most molecules are not well defined by a spherical cavity, nowadays, primarily cavities created by overlapping van der Waals spheres are used (see also section 2.3.3). The radii used for the spheres are usually fixed per element and are either empirical parameters obtained by optimizing calculated properties to match experimental data[33] or directly deduced from crystal structures.[34] Still, even if the same radii are used for creating a cavity, several ways exist to evaluate the corresponding surface area, as seen in Figure 2.2. The sum of the surface areas of the single spheres denominates the van der Waals surface, which is the easiest way to obtain the surface. A solvent-excluded surface (SES) is created using a probe sphere with a fixed radius r_{probe}. When this radius is set to the realistic radius of the solvent (*e.g.*, about 1.3 Å for water), this yields the physically motivated picture of pockets inaccessible to the solvent molecule that should not be considered. On the other hand, in some approaches, a fixed (sometimes arbitrary) probe sphere radius is used,[30,113] invalidating the physical motivation. The solvent-accessible surface (SAS) is obtained in the same way, with the solvent's radius added to the atoms' radius first.

Having obtained the surface area of the atoms, the non-electrostatic part of the free energy is derived empirically. In their most simple form, a linear dependence is assumed, which leads to the following equation:

$$\Delta G_{\text{ne}} = \sum_{i}^{atoms} \gamma_i S_i\,, \tag{2.62}$$

where γ_i is an element-specific parameter and S_i is the corresponding surface area of the atom. However, Such a linear dependence may perform quite poorly[114], and therefore, more sophisticated methods to approximate ΔG_{ne} in dependence of the surface area have been developed.[115] Despite the progress made in developing models for non-electrostatic energy evaluation, they remain largely empirical and rely heavily on the functional and electrostatic models used. This suggests that these models may, at least partially, compensate for diverse errors arising from the underlying computational evaluation rather than accurately describing the non-electrostatic energy.

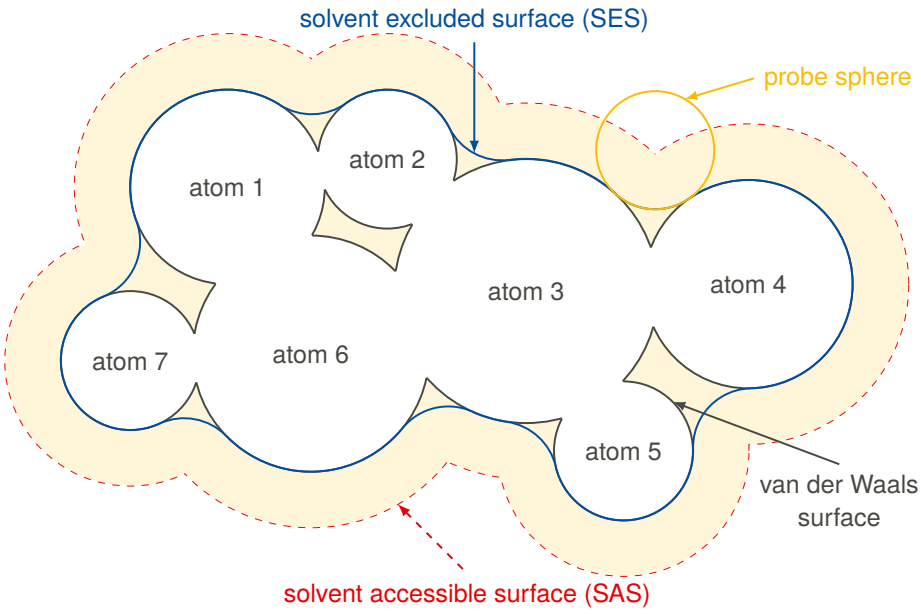


Figure 2.2: Visualization of a cavity creation using overlapping van der Waals spheres, with different approaches for the surface creation.

2.4 Conformational Contribution

The last sections comprised everything needed to calculate the total free energy for a single geometric structure of a molecule, *i.e.*, a conformer. This is usually enough for small and rather rigid molecules but not for larger and more flexible molecules, where there is not only one possible conformation but several hundred (or even more, dependent on the possible degrees of freedom). Not all of these structures are energetically equivalent and, at finite temperatures, only a few of them are occupied, which is described by the Boltzmann weight of a conformer as[12]

$$p_c = \frac{\mathrm{e}^{-G_c/\mathrm{k}T}}{\sum_j^{\text{Conf.}} \mathrm{e}^{G_j/\mathrm{k}T}}\,, \tag{2.63}$$

where k is the Boltzmann constant, $G_{c/j}$ is the free energy of the respective conformer, obtained according to equation 2.1, and T is the temperature. As a conformer's occupation strongly correlates with its energy, it is normally sufficient to investigate conformers only in a relatively narrow energy window (*e.g.*, about 2.5 kcal/mol at room temperature). For properties that depend only on energy, limiting oneself to a single conformer with the lowest energy (the "lowest-lying" conformer) may be sufficient. However, this is mostly not the case for properties that strongly depend on the geometric structure of the molecule and not only on the energy itself, like optical rotation.[116] But even for heavily energy-dependent properties, like octanol-water partition coefficients[5] or vapor pressures, a sophisticated conformer search workflow can be important, as the lowest-lying conformer in one phase may not be the lowest-lying conformer in another phase (*cf.* Figure 1.1). For more details on vapor pressures, please refer to Appendix A. Therefore, for an accurate property calculation, the average free

energy of a full conformer ensemble (CE) should be considered, which is calculated as follows

$$\overline{G}_{\text{total}} = \sum_{i}^{\text{Conf.}} p_i G_i \,. \tag{2.64}$$

To generate a suitable CE, an exploration of the potential energy surface (PES) of the system in question is necessary. The conformer ensembles used throughout this thesis were obtained using the CRENSO workflow.[5,12] A CE is thereby produced by an extensive exploration of the PES with the Conformer-Rotamer Ensemble Sampling tool (CREST)[11] in combination with the reevaluation of the obtained conformers at higher levels of theory using the Energetic Sorting (ENSO)[12] algorithm.
The underlying idea of CREST is to introduce a history-dependent bias potential to accelerate the exploration of the PES. This approach is called meta-dynamics (MTD) simulation.[117] The biasing contribution is thereby given as

$$\sum_{i}^{n} k_i \mathrm{e}^{-\alpha \Delta_i^2} \,. \tag{2.65}$$

The parameters k_i and α are used to define the shape of the bias potential, while the values of Δ_i are based on the root-mean-square deviations (RMSDs) of atomic Cartesian coordinates using n reference structures. During the MTD, structures that have a high similarity to already explored regions of the PES will yield a small RMSD and receive an energy penalty. This way, new minimum structures (conformers) are reached faster than by using an unbiased molecular dynamics simulation. Although a single run of the CREST algorithm suffices to generate a comprehensive CE, the exploration of the PES is constrained by the approximations inherent in the method used for performing the MTD simulation.
To counteract this shortcoming and ensure a more complete mapping of the PES, the default version of the CRENSO workflow involves conducting multiple searches with CREST at the GFN-FF[118] level of theory on slightly modified versions of the original PES. The modifications include scaling London dispersion contributions and artificially altering the molecular charge. By exploring all of these modified versions of the PES, an ensemble with more diverse conformations is reached, which reduces the likelihood of missing relevant conformers. To enable comparability, the differently obtained CEs are reoptimized at the same unmodified semi-empirical GFN2-xTB[85] level of theory and then condensed using a combined principal component analysis and k-means clustering algorithm.
Because the calculation of molecular properties typically requires more accurate free energies, the CE created by CREST has to be refined at the DFT level. The command line ENSO program (CENSO)[12] was designed to refine a CE on several (higher) levels of theory. According to the Boltzmann distribution (*cf.* equation 2.63), only conformers in a certain energy window from the lowest conformer contribute significantly to the final property. Therefore, to save computational time, the multilevel sorting scheme employed by the CENSO program works with several energy thresholds, starting on a low level of theory and only treats the most relevant conformers at higher levels of theory (*cf.* figure 2.1 for an overview of DFT functionals). A short overview of the complete CRENSO workflow used throughout this thesis is also given in Appendix A. A more detailed description of CREST[11], CENSO[12], and the full CRENSO[5,12] workflow can be found in the respective publications.

CHAPTER 3

Quantum Chemical Calculation of the Vapor Pressure of Volatile and Semi Volatile Organic Compounds

Marcel Stahn[*], Stefan Grimme[*], Tunga Salthammer[†], Uwe Hohm[‡], Wolf-Ulrich Palm[§]

Received 26th June 2022 , Accepted 1st October 2022, First published on 3rd October 2022

Reprinted (adapted) in Appendix A with permission[**] from
M. Stahn, S. Grimme, T. Salthammer, U. Hohm and W.-U. Palm, *Quantum chemical calculation of the vapor pressure of volatile and semi volatile organic compounds*, Environmental Science: Processes & Impacts **24**.11 (2022) 2153–2166, DOI: https://doi.org/10.1039/d2em00271j

Own manuscript contributions

- workflow development and methodology
- performing all quantum chemical calculations
- data curation, investigation, and visualization of the results
- writing of the manuscript

[*]Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[†]Department of Material Analysis and Indoor Chemistry, Fraunhofer WKI, 38108 Braunschweig, Germany
[‡]Institute of Physical and Theoretical Chemistry, University of Braunschweig, 38106 Braunschweig, Germany
[§]Institute of Sustainable and Environmental Chemistry, Leuphana University Lüneburg, 21335 Lüneburg, Germany

The vapor pressure describes the volatility of a substance and is thus a measure of its tendency to sublimate into the gas phase.[120] Substances with higher vapor pressures will more easily become airborne and, therefore, exhibit a higher environmental hazard. There are different methods for the experimental evaluation of vapor pressures, like gas chromatography (GC).[121] Still, these methods are usually associated with high levels of effort and are even rather inaccurate, especially for lower vapor pressures. For this reason, methods based on structure-activity relationships (QSAR)[122] are often used to predict vapor pressures for screening purposes. Although these methods can be effective for common compounds with a limited range of chemical diversity, they rely heavily on empirical data. As a consequence, their accuracy unsurprisingly declines significantly when dealing with compounds that fall outside of the relatively small boundary of known chemical space.[123,124]
This initial work mostly focuses on investigating common and state-of-the-art quantum chemical (QM) methods and solvation models for high throughput routine investigations of flexible compounds. The goal of this work is the application of a highly automated multi-level workflow (the CRENSO[12] workflow) for the calculation of vapor pressures of volatile and semi-volatile organic compounds with environmental relevance. This, in comparison to QSAR, much less empirical strategy, depends on the prediction of a change in free energy (ΔG) for the transition from a condensed phase to the gas phase. Because this work also deals with highly flexible compounds, the consideration of a full conformer ensemble in each phase is necessary to predict this phase shift energy accurately. The conformer ensemble is created with a metadynamics-based approach using the CREST[125] program utilizing a combination of efficient conformational screening with the general force field GFN-FF.[118] The resulting conformer ensemble is reranked with the semi-empirical method GFN2-xTB[85] in combination with the analytical linearized Poisson-Boltzmann (ALPB)[26] model to include solvation interactions. The final conformer ensemble is then reoptimized and evaluated with density functional theory, using the r^{2}SCAN-3c[126] composite method.
One key point of this workflow is the treatment of the condensed phase. Considering an ensemble of molecules explicitly for the condensed phase calculation would significantly increase the computational cost, which would make this calculation unfeasible. Therefore, treating the molecule-molecule interactions implicitly would be preferable, like a solvent in an implicit solvation model. However, implicit solvation models are generally parametrized for specific solvents and are therefore not able to use arbitrary substances in a self-solvation treatment. The conductor-like screening model for real solvents (COSMO-RS)[28] used in this workflow has a slightly different approach based on pre-performed density functional theory calculations for solutes in an ideal conductor. This enables an on-the-fly calculation of the self-solvated phase. However, this treatment is only possible for the final evaluation of the conformer ensemble in the last step of the multi-level workflow because the ALPB solvation model does not allow for a self-solvation treatment. Therefore, so-called sampling solvents with a dielectric constant similar to the molecules in question were chosen for the conformational screening part.
In this study, we examine the effectiveness of a rigorous screening process of different conformations, along with the self-solvation treatment via COSMO-RS, in comparison to a single-structure COSMO-RS approach. We also compare the results with those obtained from the QSAR model SPARC and a more advanced linear-free-energy relationship approach. The aim is to determine the accuracy of our approach in relation to experimental data. This work shows that the CRENSO workflow is superior to conventional prediction models, does not inhibit any bias based on the compound class, and provides reliable vapor pressures for liquids and sub-cooled liquids over a wide range of pressures.

CHAPTER 4

Dispersion Energy-Stabilized Boron and Phosphorus Lewis Pairs

Benedikt Sieland[†*], Marcel Stahn[‡*], Roland Schoch[†], Constantin Daniliuc[§], Sebastian Spicher[¶], Stefan Grimme[‡], Andreas Hansen[‡], Jan Paradies[†]

Received 29th June 2023, First published on 11th August 2023

Reprinted (adapted) in Appendix B with permission[**] from
B. Sieland, M. Stahn, R. Schoch, C. Daniliuc, S. Spicher, S. Grimme, A. Hansen and J. Paradies, *Dispersion Energy-Stabilized Boron and Phosphorus Lewis Pairs*, Angewandte Chemie International Edition **62**.35 (2023) e202308752, DOI: https://doi.org/10.1002/anie.202308752

Own manuscript contributions

- performing all quantum chemical calculations
- investigation and visualization of the results
- writing of the manuscript and SI, especially the computational part

[†]Department of Chemistry, Paderborn University, Warburger Strasse 100, 33098 Paderborn, Germany
[*]These authors contributed equally to this work
[‡]Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[§]Organic Chemistry Institute, Westfälische Wilhelms-Universität Münster, Corrensstraße 40, 48149 Münster, Germany
[¶]BASF SE, RGQ/SQ—B1, Carl-Bosch Straße 38, 67056 Ludwigshafen am Rhein, Germany

Quantum chemical methods have reached a stage where they can accurately predict molecular properties for many different molecules, which was demonstrated by the calculation of vapor pressures in chapter 3. In this work, an especially complicated class of structures, known as Frustrated Lewis pairs (FLPs), is investigated. While FLPs consist of monomers that are not covalently bound, they exhibit strong Lewis acid – Lewis base interactions, which should lead to the formation of a Lewis acid-base adduct. However, due to the steric demands of the monomer, this formation is not possible, which ultimately leads to an encounter complex that is dominated by non-covalent interactions.[128,129] This encounter complex exhibits an "unquenched" reactivity.[130,131] For this reason, FLPs can be used as a catalyst for many reactions that undergo a heterolysis, *e.g.*, the splitting of hydrogen.[132]
One particularly interesting subclass of FLPs is the class of London dispersion energy-stabilized boron and phosphorus Lewis pairs. Instead of being dominated by the difference in Lewis acidity of the two monomers, the encounter complex is stabilized by aliphatic substituents that function as dispersion energy donors (DEDs).[133] Crucial for the formation of the encounter complex of these pairs is the subtle interplay of many formally weak, attractive, and repulsive interactions. This makes calculating the difference in energy between the encounter complex and the monomers (the association free energy ΔG_a) especially challenging for modern quantum chemical methods and solvation models because solute-solvent interactions can play a crucial role by quenching dispersion interactions.[134]
This work investigates many functionals used for density functional theory (DFT) calculations, ranging from cheap meta-GGA functionals to highly accurate range-separated hybrids. It compares their accuracy to experimentally determined association-free energies. These methods do not perform satisfactorily, as they could not reproduce the experimental trend sufficiently. We attribute this fact to an insufficient description of the conformational space by the "crude" conformational screening. With the help of the advanced CRENSO[12] workflow, which was specially developed for flexible molecules and already showed promising results in earlier work (see also chapter 3), it was finally possible to accurately reproduce the experimental values for all investigated compounds. However, an unexpected trend between two of the investigated Lewis pairs could be observed that was not justified by the experimental data.
Further investigation into this issue led to the conclusion that the solvent models yield an unsatisfactory description of the solvation interactions. This is not surprising, as they are typically developed for small, relatively rigid organic molecules and are, therefore, of limited accuracy for the large and flexible systems investigated in this work, especially if these exhibit strong charge-transfer effects. Creating new solvation models that can better describe these kinds of effects, *e.g.*, by taking the environment of the solute atoms into account for the creation of the solute cavity instead of using static radii (*c.f.* chapter 6), should therefore be in the focus of future research. Another possible source of error is the neglect of dynamic effects, such as the B-P bond elongation of weakly bound Lewis pairs in solution, which is impossible to capture with an approach based on static equilibrium geometries. Nonetheless, by employing state-of-the-art workflows that use advanced sampling and the refinement of the conformer ensembles at finite temperature with modern DFT functionals, combined with the Boltzmann averaging of the thermally populated ensemble, a large part of these missing contributions can be fathomed, and the free energy of association of complicated systems can be accurately calculated. The full computational workflow could ultimately predict the association free energies of the boron- and phosphorus-centered Lewis pairs within an accuracy of 0.5 to 1.0 kcal/mol. The association becomes more exergonic with the increasing number of atoms in the dispersion energy donors within the Lewis pair. It is mostly unaffected by changes to the electron pair donor/acceptor properties of the phosphane and borane.

CHAPTER 5

Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods

Marcel Stahn[*], Sebastian Ehlert[†], Stefan Grimme[*]

Received 26th June 2022 , Accepted 1st October 2022, First published on 3rd October 2022

Reprinted (adapted) in Appendix C with permission[**] from
M. Stahn, S. Ehlert and S. Grimme, *Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods*, The Journal of Physical Chemistry A **127**.33 (2023) 7036–7043,
DOI: https://doi.org/10.1021/acs.jpca.3c04382

Own manuscript contributions

- development of the `CPCM-X` software library and implementation in the `xTB` software code
- parameter optimization
- investigation and visualization of the results
- writing of the manuscript

[*]Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[†]Microsoft Research AI4Science, 1118 CZ Schiphol, Netherlands
[**]Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

The previous works that are part of this thesis (see Chapters 3 and 4) showed that the utilization of state-of-the-art implicit solvation models in combination with robust quantum mechanical workflows can yield an accurate prediction of properties for a wide variety of complex systems. However, even the most advanced solvation models still have some shortcomings and are one of the main remaining sources of errors for computational calculations. This problem is even more pronounced for the less sophisticated but more efficient solvation models that are routinely used in combination with semi-empirical models in the screening part of a multi-level workflow, like the analytical linearized Poisson Boltzman (ALPB)[26] solvation model. Additionally, these models are often limited to a certain (narrow) range of solvents for which the necessary parameters are available, which reduces their applicability for innovative approaches, like self-solvation (see chapter 3) or novel solvents. Replacing the ALPB model with a more sophisticated solvation model, like the conductor-like screening model for real solvents (COSMO-RS),[28] would therefore not only possibly significantly increase the accuracy but also greatly extend the range of applicability, as there would be no need for solvent-specific parameterizations (see chapter 3). However, COSMO-RS depends on pre-performed density functional theory (DFT) calculations utilizing the conductor-like screening model (COSMO)[136] and is therefore not easily usable with a semi-empirical method, like GFN2-xTB.[85]

The introduction of the eXtended Conductor-like Polarizable Continuum Model (CPCM-X) for implicit solvation treatment solves these problems. The necessary DFT calculations are replaced with semi-empirical calculations incorporating an efficient implementation of a domain-decomposited conductor-like screening model (ddCOSMO).[137] Using the results of these calculations, CPCM-X can, therefore, utilize the COSMO-RS approach to describe the electrostatic interactions of the solute with the solvent. Additionally, the non-electrostatic interactions are approximated using the cavity-dispersion-solvent formalism introduced by the universal solvation model based on solute electron density (SMD).[30] This combination of a fast and efficient polarizable continuum model with a sophisticated post-SCF approach to evaluate the respective results makes it possible to accurately describe solute-solvent interactions for a wide variety of different solvents while still being more than two orders of magnitude faster than a corresponding approach using DFT calculations.

CPCM-X is tested on various benchmark sets and consistently outperforms the default ALPB solvation model implemented in the `xTB` program package. Especially for neutral compounds, CPCM-X significantly improves the description of the solvation free energy. This is because CPCM-X is intended to represent the subtle characteristics of solvent molecules in the immediate vicinity of the solute molecule accurately (near-field interactions). This feature is particularly important for non-ionic compounds, as opposed to ionic ones, where electrostatic interactions are expected to have a major impact on solvent molecules that are further away (far-field interactions). For large supramolecular complexes that exhibit significant non-electrostatic interactions, improvements of up to 30 % in terms of mean absolute deviation and 40 % in terms of error range were observed.

The CPCM-X solvation model is specifically built and parametrized for the GFN2-xTB semi-empirical method and natively interfaced with the `xTB` source code. Still, the theoretical framework is generally method-agnostic and could, therefore, also be used with any DFT functional and quantum chemistry software package. To make the task of parameter optimization as simple as possible, CPCM-X is built as an open source library. This also vastly decreases the effort to implement the method into different quantum chemistry software packages.

CHAPTER 6

Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)

Christoph Plett[†*], Marcel Stahn[†*], Markus Bursch[†‡], Jan-Michael Mewes[†§], Stefan Grimme[†]

Received 19th December 2023, First published on 26th February 2024

Reprinted (adapted) in Appendix D with permission[**] from
C. Plett, M. Stahn, M. Bursch, J.-M. Mewes and S. Grimme, *Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)*, en, The Journal of Physical Chemistry Letters **15**.9 (2024) 2462–2469, DOI: https://doi.org/10.1021/acs.jpclett.3c03551

Own manuscript contributions

- co-development and design of the methodology; except parameter optimization
- co-development of the DRACO software library
- investigation and visualization of the results, in particular for supramolecular complexes, dynamic bond dissociation and zwitterions
- writing of the manuscript

[†]Mulliken Center for Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany
[*]These authors contributed equally to this work
[‡]Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany
[§]beeOLED GmbH, 01257 Dresden, Germany

While the introduction of a new kind of solvation model for semi-empirical methods in Chapter 5 improved the description of solvation free energies for the low-level side of quantum chemical hierarchy, it did not solve one of the inherent problems of many state-of-the-art implicit solvation models, which is the determination of the solute cavity. This cavity is often created by overlapping van der Waals spheres (see section 2.3.4 and 2.3.3), dependent on atomic radii. In fundamental solvation models like the conductor-like polarizable continuum model (CPCM), these radii are statically allocated per element, irrespective of the specific solvent or solute. However, it is a fact that a dynamic variation in the cavity radius plays a crucial role in solute-solvent interactions.[36–39] Yet it is often disregarded for reasons of simplicity. Taking this aspect into account would certainly lead to a better understanding of the chemical processes involved and may help to improve the overall result.

To solve this issue, over the course of this work, a generally applicable and efficient approach for the determination of system-specific atom radii for solute cavity construction based on their molecular environment is developed. The final method is termed Dynamic Radii Adjustment for Continuum Solvation (DRACO). The DRACO approach uses atomic partial charges to model the atomic environment, which are calculated by an interface to several highly efficient charge models (EEQ[92,139] and CEH[140]). In addition, it uses fractional coordination numbers analogous to the D3 London dispersion model.[91] DRACO relies on three parameters per element that are fitted on experimental data from the Minnesota Solvation Database (MNSOL).[141]

While there are many different aspects of solvent-solute interactions, two main aspects are fathomed by the DRACO approach. The first one is relatively simple and describes the effect of a varying electron density on the atomic radii, *i.e.*, less electron density should lead to smaller radii. However, the second one takes the explicit solvent-solute interactions into account, which are influenced by the local electron density. These interactions should depend on the polarity of the solvent and do not necessarily align with the direct effect of the varying electron density (see Figure D.1). Therefore, the different effects of a varying electron density can even partially cancel out each other.

The effects of the DRACO approach are demonstrated for the tautomeric equilibrium reaction of glycin in water, where the introduction of dynamic radii significantly improves the prediction of the tautomer equilibrium for all tested solvation models compared to experimental data.

To perform statistical analysis of solvation free energies, the DRACO method is evaluated not only on the training data, but also on various benchmark sets. These benchmark sets include a wide range of molecules, from small neutral ones to highly charged ions, as well as large supramolecular molecules. This testing ensures the robustness of the DRACO approach. Combined with purely electrostatic solvation models, like CPCM and COSMO, DRACO reduces the mean absolute deviation (MAD) of the solvation free energy significantly by up to 4.5 kcal/mol (67 %). Even in combination with the highly empirical universal solvation model based on solute electron density (SMD), DRACO substantially reduces the MAD for charged solutes by up to 1.5 kcal/mol (39%), while neutral solutes are still slightly improved (0.2 kcal/mol or 16%). Furthermore, the incorporation of dynamic radii in implicit solvation models impacts not only static structure evaluations but also dynamic processes, which is demonstrated by an investigation of the bond dissociation of HCl in water.

The DRACO program, which was developed in this work, is available to download for free as an open source program on GitHub.[142] The program is currently compatible with the ORCA[143] and TURBOMOLE[144] program packages, but it has the potential to be used with any quantum chemistry program that permits custom radii for cavity construction. Additionally, the DRACO program can be easily extended to any radii-based solvation model and solvent in principle.

CHAPTER 7

Conclusion and Outlook

This thesis explored the potential and limitations of modern computational chemistry, with a special emphasis on sophisticated solvation models in combination with advanced quantum chemical methods. Although the chemical space is vast and its enormous complexity seems rather limitless, it is no longer unfathomable in many cases. Through the understanding of fundamental theory and with the help of modern computational infrastructure, it is already possible to predict the outcome of chemical reactions and various properties of unknown compounds for large parts of this space. The advancement of new technologies, such as artificial intelligence and machine learning, combined with the ever-increasing amount of readily available computing power through networked virtual servers ("the cloud"), will continue to push the boundaries of what is possible in terms of accuracy and processable system size. However, even with the capabilities of the cloud, it is still not possible to consistently treat every structure with highly accurate methods, as the conformational space of a single large flexible molecule can consist of hundreds of thousands of conformers. For this reason, the conformational space is often explored by force fields in combination with efficient sampling methods, like metadynamics or genetic algorithms that, while suitable enough for the creation of (crudely ordered) conformer rotamer ensembles, often lack in terms of accuracy.
Because of that, there is a dire need for efficient and robust semi-empirical quantum chemical methods (SQMs) like GFN2-xTB that allow the sampling of large conformer ensembles. While SQMs are accurate enough for sorting out relatively high-lying conformers, they are still not able to reach the accuracy needed for reliable property predictions, which are sensible to small changes in free energy. Instead, the reduced conformational ensemble after sampling needs to be treated with higher-level quantum chemical methods, like methods based on density functional theory (DFT). This combination of various levels of theory for the accurate calculation of molecular properties is called a multi-level workflow.
In chapters 3 and 4, the capabilities of such an advanced state-of-the-art multi-level quantum chemical workflow to automatically calculate properties for a large variety of flexible and complicated compounds were investigated, ranging from smaller organic molecules over typical drugs to large supramolecular complexes and frustrated Lewis pairs. Non-covalent interactions largely dominate the latter, which further increases the complexity of these compounds. Through the combination of robust methods with huge databases containing molecular structures, it was possible to set up an automatic workflow that enables even non-expert users to get accurate predictions from readily available molecular identifiers, like Chemical Abstracts Service (CAS) registry numbers or simplified molecular-input

line-entry system (SMILES) identifiers.[15] In addition, this work demonstrated that by utilizing the appropriate quantum chemical methods, it is possible to achieve comparable or even higher accuracy than experiments for certain molecular properties, such as very low saturation pressures. This approach is not only faster but also more accessible than experimental methods and less prone to errors that are introduced by additional factors like impurities in the investigated compounds, application errors, or ambient conditions. While there are methods available that allow a faster evaluation of properties based on highly empirical quantitative structure-activity relationship (QSAR) models, these methods often exhibit a strong bias if chemically diverse structures are investigated that are not well represented by the training data used to create these models.

Despite the increasing possibilities, this work also showed that there are still some significant challenges computational chemistry methods must overcome to provide an accurate and consistent description of reality. One of the remaining sources of inaccuracies is the description of solute-solvent interactions. Even the most sophisticated implicit solvation models investigated in this work, such as the Universal Solvation Model Based on Solute Electron Density (SMD), yield insufficient results for certain complex systems, especially charged or highly polar ones. This is because implicit solvation models exhibit a high amount of empiricism and are often parameterized to yield an accurate description of common small organic molecules. The underrepresentation of complicated systems in the training data and the inflexibility of the implicit solvation models to adapt to polar systems leads to a wrong description of the free energy. To further complicate the issue, the inaccuracy in solvation free energy description on higher levels of theory is not the only possible source of errors, as solvent-solute interactions can also be crucial in determining the molecular structure and a wrong geometry obtained by lower-level sampling steps will influence the final result. Therefore, it is necessary to ensure high accuracy on all levels of theory used in a multi-level workflow with enhanced conformational sampling (*cf.* chapter 3).

For this reason, over the course of this work, a new solvation model for semi-empirical methods was developed (chapter 5), which significantly enhanced the description of solvent-solute interactions and thus the overall accuracy of the solvation free energy by methods, that are on the low-level side of quantum chemical hierarchy. The extended conductor-like polarizable continuum model (CPCM-X) is thereby based on an efficient implementation of a polarizable continuum model into the extended tight-binding (`xTB`) program package. The calculated results are then post-processed based on an approach first introduced by Klamt *et al.* for the conductor-like screening model for real solvents (COSMO-RS).[27] Additionally, the description of non-electrostatic contributions is enhanced in comparison to the default analytical linearized Poisson Boltzman (ALPB) solvation model by the introduction of a cavity-dispersion-solvent formalism, formulated by Marenich *et al*.[30] CPCM-X improves the description of solvation free energies by up to 30 % in terms of mean absolute deviation and 40 % in terms of error range in comparison to the ALPB solvation model. In combination with an efficient method for electronic energy calculations, it has already been utilized to accurately and efficiently predict molecular properties.[145] Nonetheless, while the CPCM-X model is able to enhance the description of solvation effects significantly, it is inherently built for the semiempirical method GFN2-xTB and, therefore, not out-of-the-box suitable for improving the final solution description at the higher levels of theory (*e.g.* DFT) used for the final property prediction. This limitation could be tackled by a reparametrization of the method, which would open up the need for accurate reference data.

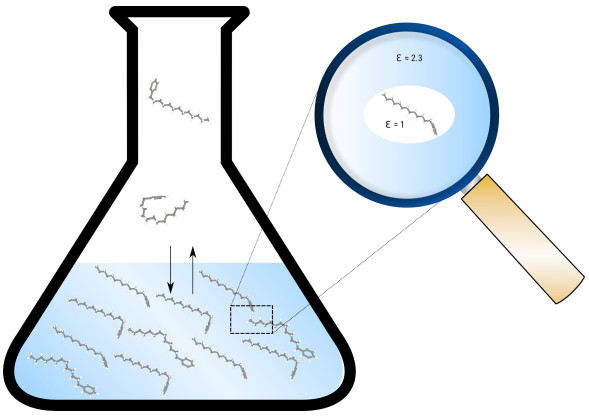
This is a problem because even highly sophisticated solvation models, like COSMO-RS,[28] still show inaccuracies, especially for solvents that have significant explicit interactions, like hydrogen bonds,

or solutes with a strong variation in local electron density, like zwitterions. Creating a sufficient amount of training data for the creation or the parameterization of solvation models becomes even more troublesome due to the fact that there is no possibility to improve the solvation description systematically, *e.g.*, by switching to a higher level of theory. One reason for these inaccuracies may be an insufficient description of the solute cavity, which plays a crucial role in how the solvent-solute interactions are approximated. This cavity is mostly determined by overlapping van der Waals spheres, built from static radii that often only depend on the element type of the atom. In reality, these radii strongly depend on the chemical environment of the atom in question.
In chapter 6, this work introduced the DRACO approach for dynamically determining solute atomic radii based on atomic partial charges and fractional coordination numbers that model this environmental dependence. The DRACO approach systematically improved the solvation description, especially for ionic and polar solvents, and was made available as an open source program package via GitHub. During the development of the DRACO approach, it was discovered that the dependence of effective atomic radii on the atomic partial charge or electron density varies between elements. The counteracting effects are that the radii increase with smaller and more negative charges, but they decrease due to shorter average solvent-solute interatomic distances.
The development of quantum chemical solvation models is still an ongoing process. Since solvent interactions are crucial in many fields of chemistry and electronic structure methods are becoming more precise, the demand for an improved and consistently improvable solvent description will continue to increase in the future. Incorporating dynamic radii obtained with the DRACO approach in current implicit solvation models can significantly advance the field. Alternatively, creating an entirely new implicit solvation model that considers the insights gained from the development of the DRACO approach may be a viable option for further improvement. With the increasing amount of computer processing power, explicit solvation models may also be worth a deeper investigation, as they have the huge advantage of explicitly modeling the near-field interactions. Combined with an implicit solvation model that accurately models the far field, this approach could yield consistently improvable results by using more sophisticated functionals for the explicit solvent calculation or by adding additional solvent shells to the calculation.
To summarize, this work outlined the huge possibilities of modern computational chemistry with state-of-the-art quantum chemical methods. In particular, one of the largest remaining sources of error, the solvation contribution, was investigated in detail, and a new solvation model for semi-empirical methods was developed that improved the description of solvent interactions for semiempirical methods. For the general improvement of implicit solvation description with DFT methods, an approach incorporating the environment of an atom in a molecule with the introduction of dynamic radii into the continuum calculation was developed, significantly enhancing the description of ionic and polar solutes. Incorporating these findings into the development of future solvation models may further enhance their accuracy and robustness over the whole chemical space.

Appendix

APPENDIX A

Quantum Chemical Calculation of the Vapor Pressure of Volatile and Semi Volatile Organic Compounds



Abstract The vapor pressure is a specific and temperature-dependent parameter that describes the volatility of a substance and thus its driving force for evaporation or sublimation into the gas phase. Depending on the magnitude of the vapor pressure, there are different methods for experimental determination. However, these are usually associated with a corresponding amount of effort and become less accurate as the vapor pressure decreases. For purposes of vapor pressure prediction, algorithms were developed that are usually based on quantitative structure–activity relationships (QSAR). The quantum mechanical (QM) approach followed here applies an alternative, much less empirical strategy, where the change in Gibbs free energy for the transition from the condensed to the gas phase is obtained from conformer ensembles computed for each phase separately. The results of this automatic, so-called CRENSO workflow are compared with experimentally determined vapor pressures for a large set of environmentally relevant compounds. In addition, comparisons are made with the single structure-based COSMO-RS QM approach, linear-free-energy relationships (LFER) as well as results from the SPARC program. We show that our CRENSO workflow is superior to conventional prediction models and provides reliable vapor pressures for liquids and sub-cooled liquids over a wide pressure range.

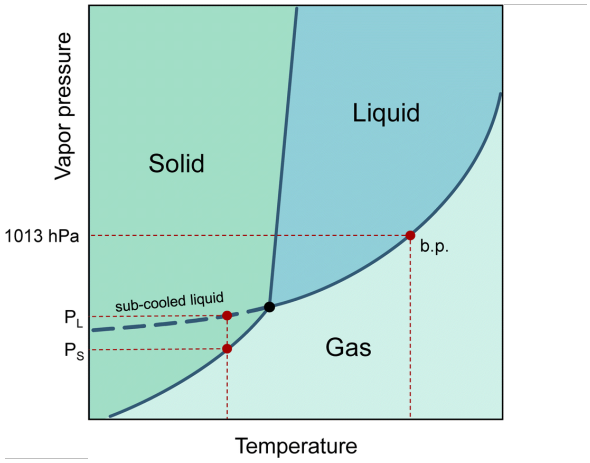


Figure A.1: Phase diagram with boiling point (b.p.) and vapor pressures for solid (P_S) and sub-cooled liquid (P_L) (see dashed curve).

A.1 Introduction

The increasing production of new chemicals makes it hard for producers to accurately determine or predict their physical properties before being released into the life cycle. The vapor pressure at a given temperature is of central importance in describing the behavior of a substance in the environment. The phase transitions from liquid to gaseous and solid to gaseous can be characterized, the saturation concentration in the gaseous phase can be calculated, the vapor pressure is linked to Henry's constant,[120] and finally, the adsorption tendency of substances on particle surfaces can be estimated.[146]

Even if a pure substance is solid at room temperature, the vapor pressure of its sub-cooled melt can still be of physical importance. This is the case, for example, when the substance is dissolved in a liquid medium. Therefore, the vapor pressure of the sub-cooled liquid is always used to describe partitioning processes.[147] In a phase diagram, the sub-cooled liquid state is the extension of the liquid phase vapor pressure line below the triple point temperature, as shown in A.1.

For many compounds, the vapor pressure can be measured very accurately.[148] However, the determination is methodologically complex. Unless it is a chromatographic method,[121] highly purified substances are needed. Even small amounts of contamination can severely disrupt the measurement. Moreover, several measurements at different temperatures are usually required. In general, the lower the vapor pressure, the more sophisticated and error-prone the measurement becomes. Experiments are therefore often carried out at higher temperatures and the result is then extrapolated to the desired temperature using the Antoine equation.[149] This is a convenient procedure as long as there is no phase transition between the measuring range and the extrapolated temperature.

Because of these experimental difficulties and uncertainties, there is a need to estimate vapor pressure using other, more readily available procedures. The simplest theoretical models are based on the Pictet–Trouton rule, which states that the vaporization entropy reaches a constant value at the boiling point. From the calculated enthalpy of vaporization at the boiling point and with the help of correction factors introduced by Fishtine,[150] the vapor pressure of a compound can then be linked to the boiling point.[151–153]

An alternative approach is chemical structure-based and calculates the excess Gibbs free energy of a compound in a liquid from the contributions of its characteristic groups.[154,155] Another possibility is to use empirical linear free energy relationships (LFER) to calculate the vapor pressure of a compound (see below for a more detailed discussion). The algorithm implemented in SPARC (Sparc Performs Automated Reasoning in Chemistry)[122] is widely used to predict vapor pressures. However, it was repeatedly found that the values calculated with SPARC for substances with low vapor pressures (approximately $< 10^{-2}$ Pa) are too small.[123,124] This may be attributed to missing data points in the calibration for low pressures or that the involved more significant intermolecular interactions cannot be fathomed with pure empiricism.

Quantum mechanical (QM) methods enable the calculation of fundamental thermodynamic properties of molecules and compute vapor pressures from the difference in Gibbs free energy between the condensed and gaseous states of a molecule. The three-dimensional structure of a molecule is explicitly considered. For several organic compounds, vapor pressures have been estimated using density functional theory (DFT) based quantum mechanics,[156–160] primarily using the COSMO-RS (conductor-like screening model for realistic solvation) method, developed by Klamt and co-workers.[28]

While this is straightforward for small, usually quite rigid molecules, open questions arise for larger, often conformationally rather flexible systems. Most important seems to be the dependence of vapor pressure on the available conformational space, which to the best of our knowledge, has not been studied systematically so far. The composition of conformational ensembles and the three-dimensional structures of individual conformers are highly dependent on the environment and can differ widely between solution and gas phase, e.g., the shortest n-alkane with a nonlinear global conformer minimum changes when solvent and temperature (entropy) effects are considered.[161] Neglecting conformational changes from the condensed to the gas-phase state in theoretical procedures could potentially lead to significant errors in the computed Gibbs free energy difference.

In a previous publication, we proposed a general multilevel QM workflow to determine liquid phase partition coefficients for molecules and tested it on compounds with environmental relevance.[5] Key aspects of this workflow are an automated, comprehensive exploration of the conformational space by artificially changing the potential energy surfaces of the compounds in the CREST[125] program to find many energetically low-lying conformers in solution, followed by re-ranking of the resulting conformer ensembles at higher DFT levels of theory using an energetic sorting and optimization algorithm (ENSO).[12] The combination of these ensemble generation and post-processing methods was dubbed CRENSO. We showed that depending on the flexibility and complexity of a system, the conformer space can have a significant impact on the thermally averaged observables. For example, the computed Kow (octanol/water) values improved by up to 1.8 log units when conformational averaging was considered. For more conformer-sensitive properties, like optical rotation,[116] the effects can be even more drastic and completely change the computed property value.

In this work, we will therefore mostly focus on flexible organic compounds with low vapor pressures for which both the theoretical methodology and the data are currently insufficient. After setting a baseline by comparing our computationally obtained values to reliable reference data from the literature, we also investigate newly emerging, environmentally relevant compounds like plasticizers, biocides and pharmaceuticals, for which no reliable vapor pressure data are available.

A.2 Methods

A.2.1 CRENSO

To accurately describe conformational effects in our workflow, we need to be able to sample the target compound independently in the condensed and gas phase to calculate the Gibbs free energy change. This is necessary because different three-dimensional molecular structures can be stabilized by enhanced or damped interactions in one of the two phases. While the idea of sampling the conformational space to obtain various solvation-dependent properties is not new,[162] widespread approaches such as COSMOconf still show large errors for flexible compounds with RMSE's (root mean square errors) of about 2.16 kcal/mol,[163] which can probably be attributed to the heuristic approach of conformer sampling and low-level DFT reranking,[164] making it easy to miss an important conformer. In addition, it can be hard to find the correct settings and conformer distributions, leading to errors of up to 4.69 log units for highly flexible systems,[165] if used without any customization. This makes sampling the chemical space of a compound a tedious work, and the CRENSO[12] workflow proved a valuable tool for automatizing such exploration without the need to adjust default settings manually.
Figure A.2 shows a short schematic overview of the applied workflow, with slight changes applied here to the solvation treatment (see below). For a more detailed description, we refer to our previous work on partition coefficients.[5]

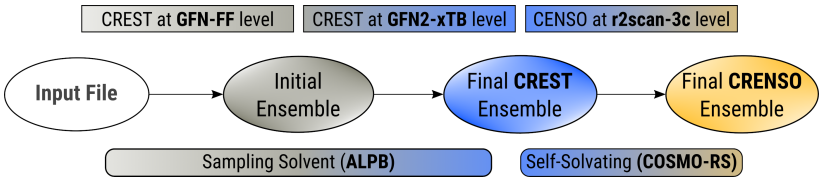


Figure A.2: Abbreviated version of the CRENSO workflow. The sampling solvent is used for the creation of the CREST ensemble, while the more accurate COSMO-RS based self-solvating procedure is used for the final CENSO ensemble and property calculations.

In short, we utilize very fast force-field (GFN-FF)[118] and semi-empirical QM (GFN2-xTB)[85] methods to extensively explore the chemical space and create an initial ensemble with conformer candidates. Then, by using carefully selected higher-level DFT methods (B97-d,[166] def2-SV(P)[167] and r2SCAN-3c[126]), this initial ensemble is energetically screened, and higher lying conformers are sorted out before the next higher level of theory is employed. In all steps, we use appropriate solvation models (ALPB,[26],[25] and COSMO-RS[27,28],), which are selected based on the respective accuracy and computational effort of the underlying method. Thermostatistical contributions are included using the single-point Hessian approach.[168]
The free energy of a molecule in each phase can then be determined by a Boltzmann weighted average over the free energies of the conformers in the ensemble giving $\overline{G}^*_{\mathrm{liq}}$ for the liquid and $\overline{G}^*_{vap}$ for the vapor state, both determined at the standard conditions of 1 mol·l^{-1}.[169] Following Ben-Naim[169] from these quantities, the vapor pressure P can be obtained from the equilibrium between the liquid and vapor state via equation A.1,

$$P = \frac{RT\rho_{\mathrm{liq}}}{MW} \cdot \exp\left(\frac{1}{RT} \cdot (\overline{G}^*_{\mathrm{liq}} - \overline{G}^*_{\mathrm{vap}})\right) \tag{A.1}$$

where ρ_{liq} is the mass density of the liquid and MW is the molar weight. The here considered vapor pressure of a neat compound refers to the situation of a molecule surrounded by other molecules of the same chemical structure in the condensed phase, which can be understood as a self-dissolved compound.[25] Note, that we cannot distinguish at this point between a (sub-cooled) liquid and solid phase, which both are treated here as a liquid. However, standard QM based solvation models like PCM or SMD[30] are usually only parameterized for a fixed set of common solvents and the necessary self-solvation treatment is impossible. The only available model which consistently can compute the necessary input data for any solvent on the fly while simultaneously calculating the properties of the same solute is the sophisticated COSMO-RS solvation approach. It is used at the very end of our workflow for high accuracy predictions of the remaining most populated conformers and is a post-processing solvation model based on the conductor-like screening theory (COSMO).[136] This makes it possible to calculate the molecular input data for the solvation calculation, specifically using the same geometry and QM data for the solute and the solvent using the default COSMO-RS molar framework. However, the state correction, which incorporates the mass density ρ_{liq} of the liquid state, see A.1, is also used in COSMO-RS and is generally not always known. For this reason, we use the density of liquid water and set $\rho_{liq} = 997\ \mathrm{kg \cdot m^{-3}}$. This would introduce an error in terms of $\Delta\overline{G}^{*} = \overline{G}^{*}_{liq} - \overline{G}^{*}_{vap}$ of not more than 0.1 kcal/mol, which is smaller than the expected standard deviation of our workflow which can be estimated from typical errors of COSMO-RS for solvation free energies of about 0.8 kcal/mol.[23] The analytical linearized Poisson Boltzmann (ALPB)[25,26] implicit solvation model used for the initial generation of conformer ensembles in our procedure (CREST step) is only parameterized for a fixed set of common solvents, and a self-solvation treatment is not feasible. However, neglecting the influence of the solvent entirely when creating the conformer ensemble can lead to significant errors. Therefore, in the first steps of the calculations, we used solvents, that were parameterized for the ALPB solvation model and have a similar dielectric constant as the target compound. These so-called sampling solvents for each compound can be found in the ESI.
With this approximation, we are still able to create differing conformer ensembles for the condensed and gas phase in the early stages of the workflow without the need to explicitly parameterize the ALPB solvation model for each new compound. The final solvation contributions are always computed at the COSMO-RS level and hence are specific for each target compound.

A.2.2 COSMO-RS

As a post-processing method, COSMO-RS uses input files created by density functional theory (DFT) calculations. These calculations were once performed as single point calculations on optimized random conformers (R) obtained from the PubChem database[170] and once on the optimized structure from the lowest lying-conformer (L) obtained after our workflow in the self-solvated phase as described above with the r2SCAN-3c[126] composite method. If there was no 3D conformer available from the PubChem database, the corresponding 2D conformer was converted to a 3D structure for the random conformer using the 3D structure converter implemented in the xTB program.[171] The COSMO-RS calculations on these single structures for comparison with the complete CRENSO treatment were also performed using COSMO-RS version '16' with fine parametrization. Vapor pressures were obtained using the intrinsic vapor pressure routine implemented in COSMOtherm. These calculations mainly serve to illustrate the effect of including extended conformer ensembles for the property calculation. Note that we use the term "COSMO-RS" throughout our work to calculate the COSMO-RS solvation free energy for a given molecular structure or a complete conformer ensemble after the CRENSO

workflow. This approach should not be confused with the result of the recommended procedure in the COSMOconf/COSMOtherm commercial software.

A.2.3 LFER and SPARC

Poly-parameter linear free energy relationships (pp-LFER) describe a set of tools to predict physical and thermodynamic properties and partition coefficients of organic compounds. The general method is based on the determination of the following compound descriptors: the molar volume (V_i), the excess molar refraction (E_i), the logarithmic hexadecane/air partitioning coefficient (L_i), the H-donor, or electron acceptor property (A_i), the H-acceptor or electron donor property (B_i) and a dipolarity/polarizability parameter (S_i).[172] Some of the parameters are accessible experimentally, V_i is usually determined using the McGowan increment method.[173] Schwarzenbach et al.[147] applied a data set of 199 apolar, monopolar, and bipolar organic compounds with known, experimentally determined vapor pressures to obtain an LFER for predicting the vapor pressure of a sub-cooled liquid with L_i, S_i, A_i, and B_i. The experimental vapor pressure range covered more than 12 orders of magnitude ($\log P_L(Pa) \approx -6 \rightarrow +6$). Using multiple regression analysis, equation A.2 was obtained (R^2 = 0.99; SD = 0.30).

$$\log P_{L,i}(298K) = -0.89 \cdot L_i - 0.44 \cdot S_i^2 - 5.43 \cdot A_i \cdot B_i + 6.51 \tag{A.2}$$

Equation A.2 was used to calculate the vapor pressures (in Pa) of the compounds discussed in this work. The required LFER coefficients L, S, A and B are listed in the ESI. All experimentally determined coefficients were taken from the UFZ-LSER database.[174] If experimental data were not available, these were calculated based on SMILES (simplified molecular input line entry specification) structure codes using a QSAR tool implemented in the UFZ-LSER database.
The SPARC algorithm uses a summation over interaction forces between molecules (dispersion, induction, dipole and H-bonding). The energies are expressed in terms of molecular-level descriptors (volume, polarizability, dipole moment and donor/acceptor properties). These are calculated from the molecular structure. The computational approach combines LFER, structure–activity relationships and molecular orbital theory.[122] SPARC needs the melting point to calculate the vapor pressure. If this was not available, or if the substance was solid at 298 K (see ESI), "assume not solid" was selected in the SPARC menu. The SPARC vapor pressure algorithm was trained with 747 experimental data for 298 K in a vapor pressure range between approximately $5 \cdot 10^{-7}$ atm and 50 atm.[175]
Both LFER and SPARC use the SMILES notation to calculate descriptors and parameters. Please note that SMILES only describes the basic structure of molecules, but not specific conformers.

A.2.4 Compounds

Reliable vapor pressure measurements are available for a large number of organic compounds. The recommended measurement method depends on the corresponding volatility. An OECD publication lists eight methods for the range between 10^{-10} Pa and 10^5 Pa.[148] All methods are experimentally complex and require highly purified substances. In comparison, the gas chromatographic method, as described by Hinckley et al.,[121] needs only small amounts of substance and is much simpler.
For testing and validation of the CRENSO method, a total of 41 volatile and semi volatile organic compounds, so-called VOCs and SVOCs,[176] were selected, for which experimentally determined vapor pressures are available. Of these, 40 data were rated as reliable, only the experimental vapor pressure for dihexyl phthalate (DHP) was doubtful. A vapor pressure range of 10^{-6} Pa to 10^2 Pa

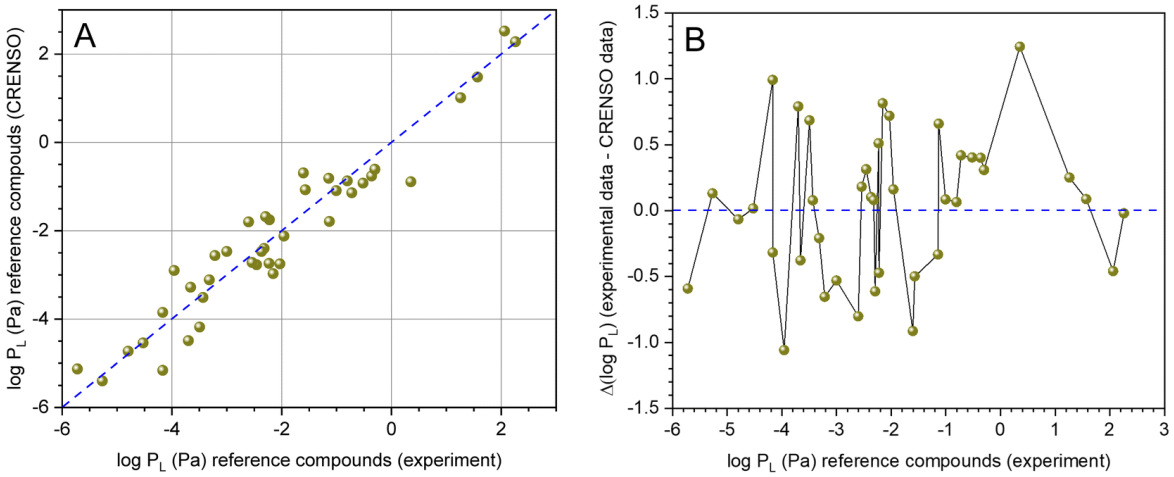


Figure A.3: (A) Correlation between experimental (see Table A.1) and calculated data (CRENSO, see Table A.2) for 40 reference compounds (DHP was not considered), (- - -) is the 1:1 line. (B) Residual analysis of experimental data versus calculated CRENSO data. The residues are normally distributed (Shapiro–Wilk test, 5 % level) with AM = 0.03 and SD = 0.54.

was covered at 298 K. The molecular flexibility of the substances varies widely, as do properties such as polarity, water solubility and partition coefficients. All compounds have no or only weak donor/acceptor properties. These substances (named "reference compounds" in the following) are compiled in table A.1. The vapor pressures relate to the liquid phase or the sub-cooled melt. If necessary, the units were converted to Pascal (Pa). In the ESI, the melting point, boiling point and enthalpy of vaporization are also provided.

For further, unbiased comparison with LFER and QSAR methods, 21 relevant compounds from the groups of plasticizers, biocides and pharmaceuticals were selected, whose vapor pressures are not sufficiently known. With some certainty, these compounds have so far not been used for the calibration of vapor pressure prediction methods. The plasticizers are compounds currently used in products that have completely or partially replaced classic additives.[204] Many of the selected pharmaceuticals were identified as emerging contaminants in wastewater.[205] The biocides are included in the list of biocidal products that may be made available on the market and used, e.g. in Germany due to an ongoing decision-making process.[206]

A.3 Results

All vapor pressure calculations for the 41 reference compounds from table A.1 are summarized in table A.2. The decimal logarithm of the vapor pressure P_L (Pa) is given in each case. If the respective substance is solid at 298 K, the value represents the vapor pressure of the sub-cooled liquid.

Figure A.3 (A) shows the direct comparison of the experimental values from Table A.1 with the values from Table A.2 calculated using CRENSO (COSMO-RS). Because our workflow contains stochastic elements, the average from three independent runs is provided. The corresponding standard deviation is usually smaller than the inherent error of the underlying QM methods. DHP was not taken into account because the experimental value appeared implausible. The data scatter around the 1:1 line,

Table A.1: Logarithmic saturation vapor pressures log $P_L^{ref.}$ of the reference compounds. All data for 298 K, SL = sub-cooled liquid

Compound	**Abbr.**	**CAS**	**log $P_L^{ref.}$**
n-Decane	C10	124-18-5	2.26[177]
n-Hexadecane	C16	544-76-3	–0.72[178]
n-Tridecylbenzene	TDB	123-02-4	–2.15[179,180]
3-Cresol	3CR	108-39-4	1.26[181]
Naphthalene **(SL)**	NAP	91-20-3	1.57[182]
Anthracene **(SL)**	ANT	120-12-7	–1.14[182]
Fluoranthene **(SL)**	FLU	206-44-0	–2.22[182]
2-Butoxy ethanol	EGBE	111-76-2	2.06[183,184]
1-Undecanol	UDC	112-42-5	–0.36[185]
Oleyl alcohol	OA	143-28-2	–3.43[186]
Glycerol	GLY	56-81-5	–1.60[187]
Benzophenone **(SL)**	BP	119-61-9	–0.80[188]
Benzophenone-3 **(SL)**	BP-3	131-57-7	–2.32[188]
Homosalate	HS	118-56-9	–1.96[189]
Dimethyl phthalate	DMP	131-11-3	–0.52[190]
Diethyl phthalate	DEP	84-66-2	–1.00[191]
Di-n-butyl phthalate	DnBP	84-74-2	–2.37[192]
Butyl benzyl phthalate	BBzP	85-68-7	–3.70[192]
Dihexyl phthalatea	DHP	84-75-3	(–2.96)[193]
Di-(2-ethylhexyl) phthalate	DEHP	117-81-7	–4.80[124,192]
Di-(2-ethylhexyl) terephthalate	DEHTP	6422-86-2	–5.27[192]
Methyl palmitoleate	MP	1120-25-8	–2.29[194]
Glutaric acid **(SL)**	GA	110-94-1	–3.00[195]
Pimelic acid **(SL)**	PA	111-16-0	–3.66[195]
Tris-(2-ethylhexyl) phosphate	TEHP	78-42-2	–4.52[189]
Tris-(2-butoxyethyl) phosphate	TBOEP	78-51-3	–4.17[189,196]
2,4,4´-Trichlorobiphenyl **(SL)**	PCB-28	7012-37-5	–1.57[197]
2,2´,4,5,5´-Pentachlorobiphenyl **(SL)**	PCB-101	37680-73-2	–2.60[197]
2,4,5,2´,4´,5´-Hexachlorobiphenyl **(SL)**	PCB-153	35065-27-1	–3.21[197]
2,2´,3,4,4´,5,5´-Heptachlorobiphenyl **(SL)**	PCB-180	35065-29-3	–3.96[197]
1,10-Dichlorodecane	C10Cl2	2162-98-3	–0.30[198]
1,2,11,12-Tetrachlorododecane **(SL)**	C12Cl4	210115-98-3	–2.46[198]
2,2´,4,4´-Tetrabromodiphenyl ether **(SL)**	BDE-47	5436-43-1	–3.49[199]
2,2´,4,4´,5-Pentabromodiphenyl ether **(SL)**	BDE-99	60348-60-9	–4.17[199]
1,2,3,4,5-Pentabromo-6-ethyl benzene **(SL)**	PBEB	85-22-3	–2.54[189]
Dodecamethylcyclohexasiloxane	D6	540-97-6	0.35[200]
Hexadecamethylheptasiloxane	L7	541-01-5	–1.13[200]
Octadecamethyloctasiloxane	L8	556-69-4	–2.03[200]
pp´-DDT **(SL)**	DDT	50-29-3	–3.32[189,201]
Diazinon	DZN	333-41-5	–2.23[202,203]
Fipronil **(SL)**	FIP	120068-37-3	–5.72[202]

Table A.2: Calculated logarithmic vapor pressures for liquids and sub-cooled liquids. CRENSO (calculated with COSMO-RS) represents the recommended method described in this work. SD is the standard deviation from three independent calculations. The LFER data were calculated from equation A.2. The SPARC data were obtained from https://archemcalc.com/sparc.html. The COSMO-RS calculations with R = Random and L = Lowest were obtained as described in the respective section

Compound	**CRENSO**	**SD CRENSO**	**LFER**	**SPARC**	**COSMO-RS (R)**	**COSMO-RS (L)**
C10	2.28	0.12	2.34	2.28	1.93	1.87
C16	–1.14	0.05	–0.41	–0.82	–1.40	–1.48
TDB	–2.97	0.23	–1.93	–2.53	–3.39	–3.44
3CR	1.01	0.07	1.28	1.62	0.78	0.81
NAP	1.48	0.00	1.54	1.52	1.14	1.14
ANT	–0.81	0.07	–1.02	–1.31	–1.21	–1.22
FLU	–1.75	0.08	–2.40	–2.07	–2.11	–2.13
EGBE	2.52	0.15	1.66	1.70	0.57	1.58
UDC	–0.76	0.01	0.01	–0.64	–1.21	–1.22
OA	–3.51	0.33	–3.50	–4.15	–3.85	–3.56
GLY	–0.69	0.03	–0.72	–1.26	–4.00	–4.33
BP	–0.87	0.02	–0.58	–0.74	–1.72	–1.30
BP-3	–2.40	0.02	–2.53	–3.79	–6.56	–2.71
HS	–2.12	0.02	–1.72	–2.44	–5.23	–2.36
DMP	–0.92	0.05	0.26	–1.25	–1.62	–1.23
DEP	–1.09	0.31	–0.40	–1.82	–2.38	–1.72
DnBP	–2.47	0.22	–2.18	–3.47	–3.14	–3.13
BBzP	–4.49	0.09	–4.12	–5.47	–6.26	–4.89
DHP	–4.49	0.35	–4.11	–5.40	–4.70	–4.64
DEHP	–4.73	0.20	–5.48	–6.85	–9.40	–7.08
DEHTP	–5.40	0.30	–4.96	–6.44	–9.40	–6.80
MP	–1.68	0.74	–1.59	–1.95	–1.92	–1.98
GA	–2.47	0.09	–2.23	–4.03	–4.40	–3.25
PA	–3.28	0.43	–3.90	–5.10	–4.92	–4.29
TEHP	–4.54	0.43	–4.39	–6.13	–6.56	–5.81
TBOEP	–3.85	0.75	–3.76	–4.80	–9.08	–6.63
PCB-28	–1.07	0.06	–1.30	–2.05	–1.51	–1.83
PCB-101	–1.80	0.01	–2.52	–3.59	–2.41	–2.60
PCB-153	–2.56	0.01	–3.35	–4.61	–2.97	–3.04
PCB-180	–2.90	0.04	–4.30	–5.71	–3.61	–3.69
C10Cl2	–0.61	0.06	0.42	–0.24	–0.77	–1.11
C12Cl4	–2.77	0.05	–1.77	–2.59	–4.06	–3.40
BDE-47	–4.18	0.01	–3.90	–5.18	–4.96	–4.95
BDE-99	–5.16	0.00	–4.92	–6.87	–6.08	–5.91
PBEB	–2.72	0.00	–3.07	–4.23	–3.21	–3.16
D6	–0.89	0.19	1.09	–0.23	–1.45	–1.80
L7	–1.79	0.01	–0.48	–0.27	–3.38	–2.44
L8	–2.75	0.16	–1.33	–1.08	–3.15	–3.57
DDT	–3.11	0.00	–3.77	–4.19	–3.58	–3.58
DZN	–2.74	0.39	–1.28	–3.10	–3.03	–3.10
FIP	–5.13	0.08	–8.35	–4.60	–9.16	–6.42

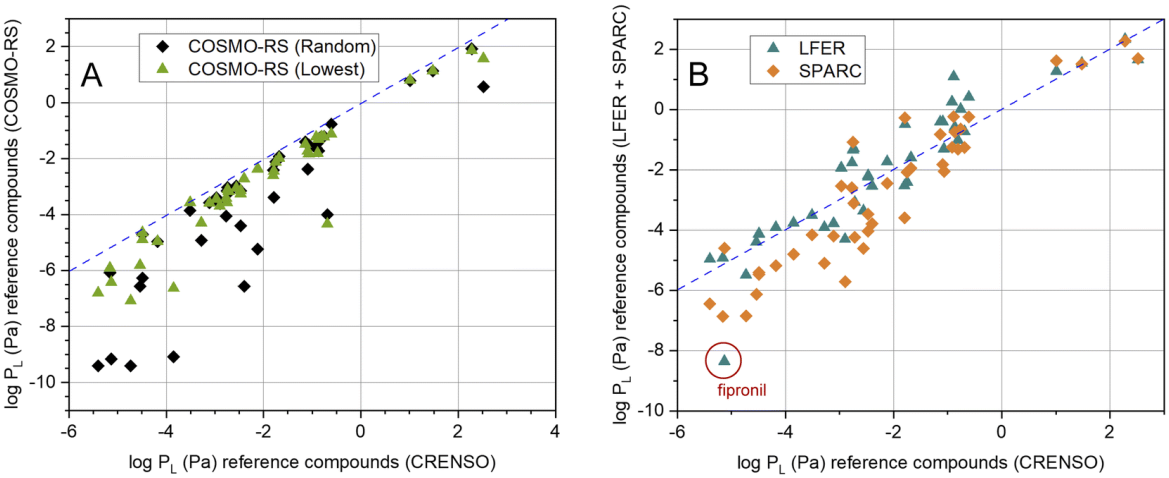


Figure A.4: (A) Correlation between calculated CRENSO data and COSMO-RS data (see Table A.2) for 41 reference compounds, (- - -) is the 1:1 line. (B) Correlation between calculated CRENSO data, calculated LFER and SPARC data (see Table A.2) for 41 reference compounds, (- - -) is the 1:1 line.

systematic deviations are not recognizable. This is also supported by the residual analysis (difference between experimental and calculated value) shown in Figure A.3 (B). The data are normally distributed at the 5 % level, and the Grubbs outlier test was negative. The arithmetic mean is AM = 0.03 and the standard deviation SD = 0.54.

It is also important which results our CRENSO method delivers in comparison to other calculation tools. In Figure A.4 (A) this is shown for COSMO-RS, based on both, the random conformer (R) and the lowest-lying conformer (L). A clear deviation from the 1:1 line is obvious, which increases with lower vapor pressures. A similar behavior can be seen for the comparison with SPARC, as shown in Figure A.4 (B). Here, too, large deviations from the 1:1 line can be observed in the area of low vapor pressures. In contrast, the correlation with the LFER method is comparatively good. However, a clear exception is the substance fipronil (FIP).

Table A.3 lists 21 relevant substances from the categories of plasticizers, biocides and pharmacologically active ingredients for which, to the best of our knowledge, no reliable vapor pressure data are available. For these substances, the vapor pressures or the vapor pressures of the sub-cooled liquid were calculated using CRENSO, LFER and SPARC. The results are also shown in Table A.3. Substances were deliberately chosen for which a low vapor pressure (< 1 Pa) was to be expected.

The graphical comparison of the CRENSO versus LFER and SPARC data is displayed in Figure A.5. For SPARC, the correlation is similar to that of the reference compounds shown in Figure A.4 (B): larger deviations with decreasing vapor pressure. However, it is immediately apparent that for these 21 compounds of emerging interest, the correlation between log P_L using CRENSO and LFER is also significantly lower than for the reference compounds.

Table A.3: Calculated vapor pressures (Pa) for liquids and sub-cooled liquids. CRENSO (calculated with COSMO-RS) represents the recommended method described in this work. The LFER data were calculated from equation A.2. The SPARC data were obtained from https://archemcalc.com/sparc.html

Compound	Abbr.	CAS	CRENSO	LFER[a]	SPARC[b]
Plasticizers					
Di-2-propylheptyl phthalate	DPHP	53306-54-0	–5.40	–6.71	–8.87
Di-isononyl phthalate[c]	DINP	28553-12-0	–5.09	–6.65	–8.04
1,2-Cyclohexane dicarboxylic acid diisononyl ester[d]	DINCH	166412-78-8	–4.50	–6.22	–6.88
Tri-(2-ethylhexyl) trimellitate	TOTM	3319-31-1	–8.22	–9.15	–12.11
Di-iso-butyl adipate	DIBA	141-04-8	–1.13	–0.61	–1.35
Di-n-butyl adipate	DnBA	105-99-7	–0.75	–1.25	–1.54
Di-2-ethylhexyl adipate	DEHA	103-23-1	–3.51	–3.92	–5.08
Di-isononyl adipate[e]	DINA	33703-08-1	–4.51	–5.16	–7.12
Biocides					
Acetamiprid **(SL)**	ACP	135410-20-7	–5.36	–1.59	–2.15
Icaridin **(SL)**	ICD	119515-38-7	–0.71	–2.26	–3.08
Cyromazine **(SL)**	CMZ	66215-27-8	–5.73	–4.76	–5.19
Diflubenzuron **(SL)**	DFB	35367-38-5	–5.17	–3.66	–6.14
Cyphenothrin	CPT	39515-40-7	–5.88	–7.07	–6.96
Methoprene **(SL)**	MTP	40596-69-8	–3.00	–1.67	–3.54
Pharmaceuticals					
Bisoprolol **(SL)**	BPL	66722-44-9	–4.04	–8.02	–6.47
Diclofenac **(SL)**	DIC	15307-86-5	–4.23	–7.11	–6.68
Dapagliflozin **(SL)**	DLF	461432-26-8	–9.61	–19.56	–17.85
Ibuprofen **(SL)**	IBU	15687-27-1	–1.91	–2.50	–2.3
Metoprolol **(SL)**	MPL	51384-51-1	–2.67	–2.22	–4.58
Naproxen **(SL)**	NPX	22204-53-1	–4.05	–5.66	–4.65
Torasemide **(SL)**	TS	56211-40-6	–8.26	–13.44	–9.66

[a] Vapor pressure for the sub-cooled liquid.[147] [b] It was assumed that the substance is not solid at room temperature. SMILES are for isomers: [c] bis(7-methyloctyl) phthalate [d] bis(7-methyloctyl) 1,2-cyclohexanedicarboxylate [e] bis(7-methyloctyl) adipate.

A.4 Discussion

A.4.1 Comparison of CRENSO calculations with experimental data

For substances that are liquid at room temperature and have sufficient volatility, the vapor pressure and enthalpy of vaporization can be determined experimentally with high accuracy. As can be seen from Figure A.3 (B), the deviations between experimental and CRENSO data have a standard deviation of approximately 0.5 log units. This means that the vapor pressure can be calculated approximately to a factor of 3. It is therefore clear that the CRENSO calculations for volatile substances deliver results, which can be regarded as reasonable trends. A typical example is 2-butoxyethanol (EGBE). Koga[183] has accurately determined temperature-dependent values for this substance. At 298 K the experimental vapor pressure is 105 Pa, while CRENSO gives 331 Pa. This difference is of practical importance since vapor pressures are used to calculate saturation concentrations for inhalation exposure tests.[207] Nevertheless, the deviation is within the predicted error of the CRENSO method.
A completely different situation arises for substances with low volatility, especially at vapor pressures

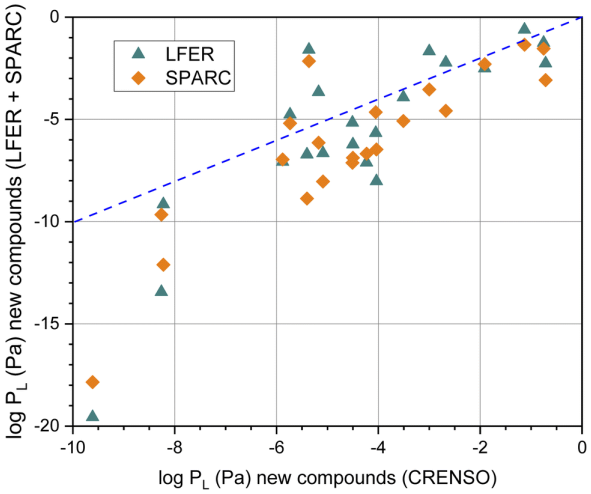


Figure A.5: Correlation of the calculated logarithmic vapor pressure using the LFER and SPARC method with the CRENSO data (see Table A.3) for the 21 compounds of emerging interest, (- - -) is the 1:1 line.

$< 10^{-2}$ Pa. The Knudsen and Langmuir methods are particularly suitable in this range, but require highly purified substances and temperature-dependent measurements and are only useful if no phase transition occurs in the temperature interval of interest. Therefore, the vapor pressure of the sub-cooled liquid P_L often has to be extrapolated from measurements of the melt using the Antoine equation.[149] Here, however, the problem can arise that organic compounds, such as fipronil, decompose at higher temperatures. This leaves only a small temperature window for the measurements, which further reduces the accuracy. Alternatively, indirect methods can be applied to determine P_L, for example via gas chromatography.[189] The vapor pressure of a sub-cooled liquid plays a role in atmospheric chemistry. On the basis of previous work, Pankow[146] states that this parameter is more important for determining a gas/particle equilibrium than the vapor pressure of the subliming solid. Below a vapor pressure of approximately 10^{-2} Pa, direct measurements are usually associated with a great deal of effort and become increasingly inaccurate. For the substance DEHP alone, Mackay et al.[208] list more than 20 values for vapor pressures at room temperature, some of which differ by orders of magnitude. Attempts were made to develop alternative measurement methods for compounds of low volatility.[124] However, these are also associated with larger experimental inaccuracies. In comparison with the experimental data, we are convinced that in the range between 10^{-2} Pa and 10^{-5} Pa, the CRENSO method does not lead to worse results than direct vapor pressure measurements. Furthermore, we believe that at vapor pressures $< 10^{-5}$ Pa, our theoretical method is even more reliable.

A.4.2 Comparison of CRENSO calculations with LFER and SPARC data

Not particularly surprising is the fundamentally good agreement between the CRENSO and LFER data. Many of the approximately 200 compounds used for the multiple regression of equation A.2 are also listed in Table A.1. Among others, these include alkanes, PCBs, PAHs, phthalates and brominated diphenyl ethers, the descriptors of which are well known.[147] The CRENSO calculations and the LFER results were at least partially evaluated using the same reference data, so the comparison is biased. It is important to note at this point, that while COSMO-RS in itself contains empirical parameters,

nothing has been specially adapted here or adjusted for the calculation of vapor pressures.
However, the differences between the methods are already clear from Figure A.3 (B) and exemplified in the case of fipronil (FIP). Here, the LFER descriptors had to be calculated using QSAR. At the same time, it is known that fluorinated compounds cannot be represented by the descriptors of hydrocarbon-based compounds.[209] The differences for the compounds of emerging interest listed in Table A.3 are even more striking (see Figure A.4 (B)). Again, most of the descriptors had to be calculated from the SMILES using QSAR. Overall, as expected, the correlation between CRENSO and LFER for these compounds is unsatisfactory. For this reason, we also refrained from comparing the reported statistical errors and standard deviations. We can assume that with the CRENSO method, there is an uncertainty of 0.5 log units for all substances. In the case of the LFER method, a standard deviation of 0.3 log units, the standard deviation of equation A.2, can only be assumed if the target compound is structurally related to the calibration compound.
With regard to SPARC the results are clear and no extensive discussion is required. It was found earlier that SPARC gives unsatisfactory results in the range of low vapor pressures.[123] Here too, both the reference compounds (see Figure A.4 (B)) and the compounds of emerging interest (see Figure A.5) clearly show that SPARC fails at vapor pressures < 1 Pa. The reason for this may be methodological. The SPARC algorithm was trained with 747 substances, but none of them had a vapor pressure $< 5 \cdot 10^{-7}$ atm (< 0.05 Pa).[175]

A.4.3 Conformational flexibility and comparison with COSMO-RS

Common programs that can predict vapor pressures are often based on qualitative structure–activity relationship methods (QSAR).[122] They mostly use group contributions methods based on a single molecular structure. In addition, empirical methods are highly dependent on the available experimental data that was used to train the respective methods. This makes these methods prone to errors for molecules that were not in the scope of the trained data or cannot be described by the same structure in both phases. This is especially true for compounds with small vapor pressure since the condensed phase will be dominated by strong interactions between the molecules, whereas these are not present in the gas phase and, in part, are replaced by intramolecular non-covalent interactions.
For this reason, the SPARC program may yield very reasonable results for compounds with higher vapor pressure while failing for compounds with lower vapor pressure, which can be seen in Figure A.4 (B) and A.5. Figure A.6 shows a statistical analysis of the errors made by LFER, SPARC and CRENSO for lower and higher vapor pressure. The CRENSO workflow and LFER show very similar statistics for the complete set of molecules, with a mean absolute deviation (MAD) between 0.44 and 0.50 log units depending on the vapor pressure of the compounds. On the other hand, SPARC significantly undershoots for compounds with small vapor pressures, reaching a MAD between 0.39 log units for higher vapor pressures and 1.23 log units for lower ones. For the entire data set, SPARC reaches a MAD of 0.87 log units, a standard deviation (SD) of 0.91 log units and a root mean square deviation (RMSD) of 1.10 log units.
Before going into a detailed analysis of the errors that would arise in quantum chemical approaches due to the neglect of conformational flexibility, let us use fipronil (FIP) as an example to explain the difference in conformational energy contributions between two phases. Figure A.7 shows the lowest conformer of fipronil found in the condensed phase by our workflow and the lowest conformer in the gas phase. For simplicity, we consider only one structure for each phase instead of a complete ensemble of conformers. The free energy diagram on the right side of the figure shows the calculated

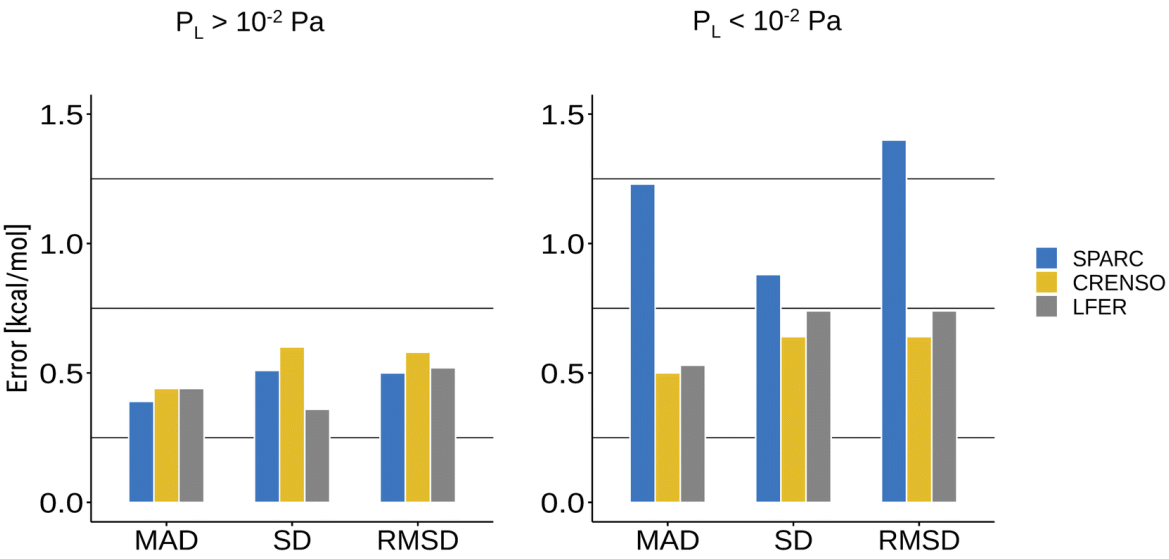


Figure A.6: Statistical errors for the tested theoretical methods. The molecule sets are splitted (depending on the reported literature value for the vapor pressure) in two parts with lower ($< 10^{-2}$ Pa) and higher ($> 10^{-2}$ Pa) vapor pressure.

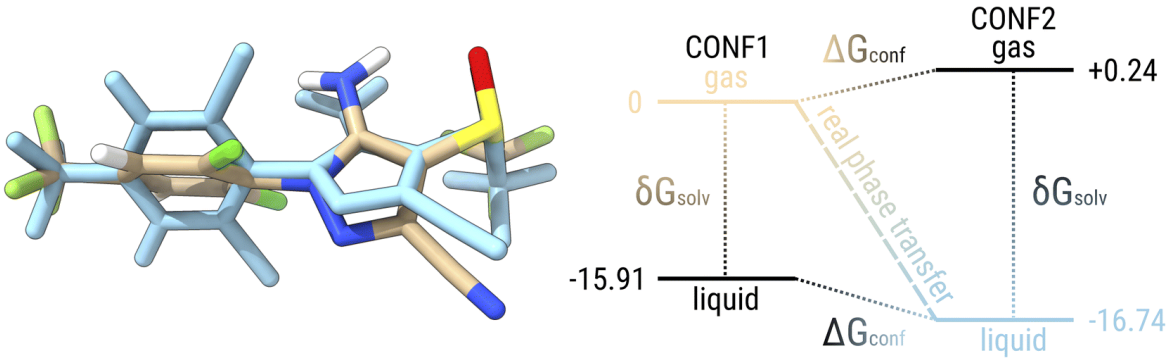


Figure A.7: Lowest conformer of fipronil (FIP) in gas phase and in condensed phase (blue). The energy diagram shows the solvation and conformational contribution to the Gibbs free energy. The energies are given in kcal/mol.

levels for the two conformers with respect to the lowest conformer in the gas phase. These energy levels are given once in the gas phase and once with the additional solvation contribution modeling the condensed phase. Considering only the lowest conformers in each phase, the "real" phase transition would correspond to the process from CONF1 in the gas phase (beige) to CONF2 in the condensed phase (blue) with an associated free energy change of –16.74 kcal/mol, which is equivalent to a saturation pressure of about –5.87 log units. On the other hand, if only the solvation contribution is added to CONF1, the energy contribution of the conformational rearrangement is absent and an artificial condensed state would be obtained (black, bottom). The energy change that can be attributed to this "frozen" phase change is –15.91 kcal/mol and thus would lead to a saturation pressure of –5.26 log units. If we consider the lowest conformer in the condensed phase and remove the contribution from solvation, this would correspond to an energy change of –16.98 kcal/mol and a saturation pressure of –6.05 log units. Thus, by neglecting conformational flexibility, we would introduce an error of about 0.24–0.83 kcal/mol or 0.18–0.61 log units, depending on which structure we assume.
To check the influence of conformations more generally, we applied the COSMO-RS method to a random and the lowest-lying conformer in the liquid, as described in section COSMO-RS, thus neglecting the conformational ensembles and their change in both phases. The results can be found in Table A.2. As expected, completely neglecting the conformational flexibility by using a random conformer leads to the worst results with a MAD of 1.87 and a SD of 1.49 log units. This shows the significant effect of a molecule's actual three dimensional molecular structure (shape) for the self-solvation free energy, which is difficult to account for by empirical QSAR or LFER models because linear combinations of intramolecular and intermolecular interactions are involved. Still significant is the error of the lowest-lying conformer-only approach with a MAD of 0.91 and a SD of 0.78, which is in the range of the errors made by the SPARC program, and larger than the best performing complete ensemble method with a MAD of only 0.47 and a SD of 0.62. We attribute the considerable improvement of about 0.4–0.5 log units mainly to the proper account of the conformational ensemble change in the gas phase. Although we have shown in our previous work[5] that there are significant variations in the conformational ensembles for flexible molecules in different solvents, they appear to be much smaller than the structural changes between the gas and condensed phases reported here.

A.5 Conclusion

The quantum mechanical workflow presented here has a decisive advantage over the usual QSAR tools: the calculated vapor pressure does not depend on structure–activity relationships, but is calculated individually for each molecule mostly based on fundamental QM and thus less susceptible to systematic errors. Our basic ansatz is non-empirical and can be systematically improved in the future by application of better procedures for the involved steps conformational search (xTB and CREST), structures and free energy ranking of conformers (DFT) and solvation free energy (COSMO-RS). The necessity of an enhanced conformer sampling workflow for predicting vapor pressures is shown by comparison to a single conformer COSMO-RS approach. From the comparison of LFER, SPARC with CRENSO (which is based on COSMO-RS), we conclude that CRENSO is currently the most reliable approach for predicting the vapor pressures of liquids and the sub-cooled liquids of solids. At high vapor pressures ($> 10^{-2}$ Pa), our method is suitable for realistic estimates but cannot compete with the accuracy of measurements. However, when looking at the variability of data in the range between 10^{-2} Pa and 10^{-5} Pa, for example for phthalates, PCBs, BDEs and many other substances,[203,208,210,211]

the quality of the CRENSO data is definitely comparable to that of experiments. Furthermore, at vapor pressures $< 10^{-5}$ Pa, which are difficult to determine experimentally, our method also opens up very reliable predictions, so that measurements might no longer be necessary in some circumstances. With further developed solvation models and/or by inclusion of explicit molecules in the self-solvation treatment in an automated cluster generation approach,[20] even higher accuracy of the predictions over the entire pressure range probably down to an MAD of 0.1–0.3 log units may be achieved.

Author Contributions

MS: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review and editing; SG: methodology, supervision, writing – review and editing; TS: conceptualization, formal analysis, investigation, validation, visualization, writing – original draft, writing – review and editing; UH: investigation, methodology, validation, writing – review and editing; WUP: investigation, methodology, validation, writing – review and editing.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgements

TS is grateful to Ms. Manuela Lingnau, Fraunhofer WKI, for the final design of Fig. A.1.

A.6 Supporting Information

This section is an adapted version of the supporting information published with the original manuscript. See DOI: https://doi.org/10.1039/d2em00271j for the original version of the electronic supporting information (ESI).

Fundamentals of vapor pressure thermodynamics

As displayed in Figure A.1 two phases of a pure substance at equilibrium are considered. In this pressure-temperature diagram the functional relationship between the equilibrium pressure P and the equilibrium temperature T is given by the Clapeyron equation A.3:

$$\frac{dP}{dT} = \frac{\Delta_{trs}S}{\Delta_{trs}V} = \frac{\Delta_{trs}H}{T\Delta_{trs}V} \,. \tag{A.3}$$

$\Delta_{trs}S$,$\Delta_{trs}H$ and $\Delta_{trs}V$ are the changes in molar entropy, molar enthalpy, and molar volume, respectively, upon phase transition. In the following we only consider the equilibrium between the condensed phase and the vapor phase. Assuming ideal gas behavior and neglecting the molar volume of the condensed phase, $\Delta_{trs}V = \Delta_{\text{vap}} \approx V_{m,g} \approx P/(RT)$, the Clausius-Clapeyron equation (A.4) results in

$$\frac{dlnP}{dT} = \frac{\Delta_{vap}H}{RT^2}\,, \tag{A.4}$$

where R is the molar gas constant and $\Delta_{vap}H$ is the molar enthalpy of vaporization. Equation A.4 can be used for extrapolation of vapor pressure data. Under the assumption that $\Delta_{vap}H$ is constant in the interval between T_1 and T_2, the vapor pressure $P_1(T_1)$ given at a temperature $T_{fus} < T_1$ can be used to calculate $P_2(T_2)$ for a temperature $T_{fus} < T_2$, T_{fus} being the melting point. If, however, $T_2 < T_{fus} < T_1$ use of $P_1(T_1)$ and $\Delta_{vap}H$ = const. in equation A.4 will result in the vapor pressure $P_2(T_2)$ of the sub-cooled liquid (see Figure A.1).

Conversion from vapor pressure P_L of the sub-cooled liquid to the saturation concentration C_0 according to the ideal gas law

$$C_0 = \frac{m}{V} = \frac{P_L M}{RT}$$

For C_0 in μg/m^3, T = 298.15K, R = 8.314 J/(K·mol), P_L in Pa and the molecular weight M in g/mol:

$$C_0 = P_L M \cdot 403.42$$

Conversion of energies between Calories and Joule

Please note that in quantum mechanical publications, the use of the unit "Calorie" for energies is still common and dominant.

$$1 \text{ cal} = 4.1868\ J, \qquad 1\ J = 0.2388 \text{ cal}\,.$$

SMILES

The SMILES (Simplified Molecular-Input Line-Entry System) notation for the selected isomer was taken from PubChem. (https://pubchem.ncbi.nlm.nih.gov)

Table A.4: Literature data of the reference compounds (SL = subcooled liquid, DCP = decomposition at higher temperatures). References in Table A.1.

Compound	Abbr.	CAS	Density kg/m^3	MP °C	BP °C	P_L Pa	Remark	$\Delta_{vap}H$ kJ/mol
n-Decane	C10	124-18-5	730	-29.7	174.1	182	298.15 K	51.4
n-Hexadecane	C16	544-76-3	773	18.2	286.9	0.19	298.15 K	81.4
N-Tridecylbenzene	TDB	0123-02-04	881	10	346	0.007	298.15 K	90.4
3-Cresol	3CR	108-39-4	1034	12.2	202	18.27	298.05 K	61.8
Naphthalene	NAP	91-20-3	1162	80.3	217.9	37	298.15 K **(SL)**	56.1
Anthracene	ANT	0120-12-07	1240	215.8	339.9	0.072	298.15 K **(SL)**	72.4
Fluoranthene	FLU	206-44-0	1252	110.2	384	$6.0{\cdot}10^{-3}$	298.15 K **(SL)**	79.3
2-Butoxy ethanol	EGBE	111-76-2	902	-75	171	115	298 K	56.6
1-Undecanol	UDC	112-42-5	835	19	243	0.2383	293.18 K	85
Oleyl alcohol	OA	143-28-2	849	7	300	$3.7{\cdot}10^{-4}$	298.15 K	101.2
Glycerol	GLY	56-81-5	1261	18.1	290	0.0249	298.75 K	86.76
Benzophenone	BP	119-61-9	1085	48.5	305.4	0.157	298.48 K **(SL)**	77.1
Benzophenone-3	BP-3	131-57-7	1320	64	315	$4.8{\cdot}10^{-3}$	298.15 K **(SL)**	76.3
Homosalate	HS	118-56-9	1045	-20	341	$1.1{\cdot}10^{-2}$	298.15 K	95.3
Dimethyl phthalate	DMP	131-11-3	1190	5.5	283.7	0.304	298.15 K	77
Diethyl phthalate	DEP	84-66-2	1120	-40.5	295	0.099	298.15 K	82.1
Di-n-butyl phthalate	DnBP	84-74-2	1049	-35	340	$4.3{\cdot}10^{-3}$	298.15 K	95.2
Butylbenzyl phthalate	BBzP	85-68-7	1120	-35	370	$2.0{\cdot}10^{-4}$	298.15 K	106.2
Dihexyl phthalate	DHP	84-75-3	995	-58	350	$(1.1{\cdot}10^{-3})^a$	298.15 K	111.2
Di-(2-ethylhexyl) phthalate	DEHP	117-81-7	990	-55	384	$1.6{\cdot}10^{-5}$	298 K	116.7
Di-(2-ethylhexyl) terephthalate	DEHTP	6422-86-2	982	-48	383	$5.4{\cdot}10^{-6}$	298.15 K	123.3
Methyl palmitoleate	MP	1120-25-8	875	0	250	$5.1{\cdot}10^{-3}$	298.15 K	96.4
Glutaricacid	GA	110-94-1	1424	98	303	$1.0{\cdot}10^{-3}$	298 K **(SL)**	118
Pimelicacid	PA	111-16-0	1280	106	400	$2.2{\cdot}10^{-4}$	298 K **(SL)**	108
Tris-(2-ethylhexyl) phosphate	TEHP	78-42-2	926	-74	DCP	$3.0{\cdot}10^{-5}$	298.15 K	106
Tris-(2-butoxyethyl) phosphate	TBOEP	78-51-3	1020	-70	400	$6.8{\cdot}10^{-5}$	298.15 K	103
2,4,4'-Trichlorobiphenyl	PCB-28	7012-37-5	1400	58	207	$2.7{\cdot}10^{-2}$	298.15 K **(SL)**	77.1

Continued on next page

Table A.4: Literature data of the reference compounds (SL = subcooled liquid, DCP = decomposition at higher temperatures). References in Table A.1. (continued)

Compound	Abbr.	CAS	Density kg/m^3	MP °C	BP °C	P_L Pa	Remark	$\Delta_{vap}H$ kJ/mol
2,2',4,5,5'-Pentachlorobiphenyl	PCB-101	37680-73-2	1522	77	400	$2.5 \cdot 10^{-3}$	298.15 K **(SL)**	86.2
2,4,5,2',4',5'-Hexachlorobiphenyl	PCB-153	35065-27-1	(1000)[b]	103	400	$6.1 \cdot 10^{-4}$	298.15 K **(SL)**	91.8
2,2',3,4,4',5,5'-Heptachlorobiphenyl	PCB-180	35065-29-3	1640	112	400	$1.1 \cdot 10^{-4}$	298.15 K **(SL)**	94.1
1,10-Dichlorodecane	C10Cl2	2162-98-3	994	15.6	284.3	0.5	298.15 K	67.3
1,2,11,12-Tetrachlorododecane	C12Cl4	210115-98-3	1100	-	354.6	$3.5 \cdot 10^{-3}$	298.15 K **(SL)**	81.9
2,2',4,4'-Tetrabromodiphenyl ether	BDE-47	5436-43-1	2161	84	DCP	$3.2 \cdot 10^{-4}$	298.15 K **(SL)**	92
2,2',4,4',5-Pentabromodiphenyl ether	BDE-99	60348-60-9	2250	92.5	DCP	$6.8 \cdot 10^{-5}$	298.15 K **(SL)**	100.3
1,2,3,4,5-Pentabromo-6-ethyl benzene	PBEB	85-22-3	(1000)[b]	138	-	$2.9 \cdot 10^{-3}$	298.15 K **(SL)**	78.3
Dodecamethylcyclohexasiloxane	D6	540-97-6	967	-3	245	2.26	298.15 K	65
Hexadecamethylheptasiloxane	L7	0541-01-05	901	-78	286.8	$7.4 \cdot 10^{-2}$	298.15 K	89
Octadecamethyloctasiloxane	L8	556-69-4	913	-63	311.6	$9.3 \cdot 10^{-3}$	298.15 K	98
pp'-DDT	DDT	50-29-3	980	109	260	$4.8 \cdot 10^{-4}$	298.15 K **(SL)**	89.3
Diazinon	DZN	333-41-5	1117	25	-	$5.9 \cdot 10^{-3}$	298.15 K	87.5
Fipronil	FIP	120068-37-3	1477	200	DCP	$1.9 \cdot 10^{-6}$	298.15 K **(SL)**	85

[a] The published reference value for DHP is doubtful. [b] Default value – density not available

Table A.5: Log P_L values of the reference compounds, calculated SPARC and LFER values (at 298 K), COSMO values from the literature and sampling solvent for the CRENSO procedure.

Abbr.	log P_L (Reference)	log P_L (SPARC)	log P_L (LFER)	log P_L (COSMO-RS)	sampling solvent (CRENSO)
C10	2.26	2.28	2.34		n-Hexadecane
C16	-0.72	-0.82	-0.41		n-Hexadecane
TDB	-2.15	-2.53	-1.93		n-Hexadecane
3CR	1.26	1.62	1.28		Phenol
NAP	1.57	1.52	1.54	1.84	Benzene
ANT	-1.14	-1.31	-1.02	-0.37	Benzene
FLU	-2.22	-2.07	-2.4	-1.33	Benzene
EGBE	2.06	1.7	1.66		Diethylether

Continued on next page

Table A.5: Log P_L values of the reference compounds, calculated SPARC and LFER values (at 298 K), COSMO values from the literature and sampling solvent for the CRENSO procedure. (Continued)

UDC	-0.36	-0.64	0.01		1-Octanol
OA	-3.43	-4.15	-3.5		1-Octanol
GLY	-1.6	-1.26	-0.72		Methanol
BP	-0.8	-0.74	-0.58		Phenol
BP-3	-2.32	-3.79	-2.53		Phenol
HS	-1.96	-2.44	-1.72		Phenol
DMP	-0.52	-1.25	0.26		Phenol
DEP	-1	-1.82	-0.4		Phenol
DnBP	-2.37	-3.47	-2.18		Phenol
BBzP	-3.7	-5.47	-4.12		Phenol
DHPd)	(-2.96)	-5.4	-4.11		Phenol
DEHP	-4.8	-6.85	-5.48		Phenol
DEHTP	-5.27	-6.44	-4.96		Phenol
MP	-2.29	-1.95	-1.59		Acetone
GA	-3	-4.03	-2.23		Acetone
PA	-3.66	-5.1	-3.9		Acetone
TEHP	-4.52	-6.13	-4.39		Acetone
TBOEP	-4.17	-4.8	-3.76		Acetone
PCB-28	-1.57	-2.05	-1.3		Phenol
PCB-101	-2.6	-3.59	-2.52		Phenol
PCB-153	-3.21	-4.61	-3.35		Dichloromethane
PCB-180	-3.96	-5.71	-4.3		Phenol
C10Cl2	-0.3	-0.24	0.42		Dichloromethane
C12Cl4	-2.46	-2.59	-1.77	-2.73c)	Dichloromethane
BDE-47	-3.49	-5.18	-3.9		Dichloromethane
BDE-99	-4.17	-6.87	-4.92		Dichloromethane
PBEB	-2.54	-4.23	-3.07	-2.21	Dichloromethane
D6	0.35	-0.23	1.09		Toluene
L7	-1.13	-0.27	-0.48		Toluene
L8	-2.03	-1.08	-1.33		Toluene
DDT	-3.32	-4.19	-3.77		Dichloromethane

Continued on next page

Table A.5: Log P_L values of the reference compounds, calculated SPARC and LFER values (at 298 K), COSMO values from the literature and sampling solvent for the CRENSO procedure. (Continued)

DZN	-2.23	-3.1	-1.28	Aniline
FIP	-5.72	-4.6	-8.35	Aniline

Table A.6: LFER parameters of the reference compounds. All data for 298 K. The data were retrieved from the UFZ-LSER database.[174]

Compound	Abbr.	L	S	A	B	log P_L (SPARC)
n-Decane	C10	4.686	0	0	0	2.34
n-Hexadecane	C16	7.771	0	0	0	-0.41
N-Tridecylbenzene	TDB	9.376	0.47	0	0.15	-1.93
3-Cresol	3CR	4.31	0.88	0.57	0.34	1.28
Naphthalene	NAP	5.161	0.92	0	0.2	1.54
Anthracene	ANT	7.568	1.34	0	0.28	-1.02
Fluoranthene	FLU	8.827	1.55	0	0.24	-2.4
2-Butoxy ethanol	EGBE	3.806	0.5	0.3	0.83	1.66
1-Undecanol	UDC	6.128	0.42	0.37	0.48	0.01
Oleyl alcohol	OA	9.565	0.58	0.36	0.69	-3.5
Glycerol	GLY	3.145	0.72	0.64	1.21	-0.72
Benzophenone	BP	6.852	1.5	0	0.5	-0.58
Benzophenone-3	BP-3	8.84	1.63	0	0.62	-2.53
Homosalate	HS	8.614	0.87	0.09	0.48	-1.72
Dimethyl phthalate	DMP	6.05	1.4	0	0.86	0.26
Diethyl phthalate	DEP	6.79	1.4	0	0.88	-0.4
Di-n-butyl phthalate	DnBP	8.97	1.27	0	0.95	-2.18
Butylbenzyl phthalate	BBzP	10.82	1.51	0	1.13	-4.12
Dihexyl phthalate	DHP	11.11	1.29	0	0.96	-4.11
Di-(2-ethylhexyl) phthalate	DEHP	12.7	1.25	0	1.02	-5.48
Di-(2-ethylhexyl) terephthalate	DEHTP	12.327	1.06	0	0.71	-4.96
Methyl palmitoleate	MP	8.853	0.71	0	0.68	-1.59
Glutari cacid	GA	4.207	1.28	1.05	0.75	-2.23

Continued on next page

Table A.6: LFER parameters of the reference compounds. All data for 298 K. The data were retrieved from the UFZ-LSER database.[174] (Continued)

Pimelic acid	PA	5.277	1.26	1.1	0.84	-3.9
Tris-(2-ethylhexyl) phosphate	TEHP	12.091	0.57	0	0.88	-4.39
Tris-(2-butoxyethyl) phosphate	TBOEP	10.565	1.4	0	2.22	-3.76
2,4,4'-Trichlorobiphenyl	PCB-28	7.904	1.33	0	0.15	-1.3
2,2',4,5,5'-Pentachlorobiphenyl	PCB-101	8.868	1.61	0	0.13	-2.52
2,4,5,2',4',5'-Hexachlorobiphenyl	PCB-153	9.587	1.74	0	0.11	-3.35
2,2',3,4,4',5,5'-Heptachlorobiphenyl	PCB-180	10.415	1.87	0	0.09	-4.3
1,10-Dichlorodecane	C10Cl2	6.466	0.87	0	0.17	0.42
1,2,11,12-Tetrachlorododecane	C12Cl4	8.861	0.94	0	0	-1.77
2,2',4,4'-Tetrabromodiphenylether	BDE-47	10.66	1.45	0	0.34	-3.9
2,2',4,4',5-Pentabromodiphenylether	BDE-99	11.71	1.51	0	0.44	-4.92
1,2,3,4,5-Pentabromo-6-ethylbenzene	PBEB	9.848	1.36	0	0	-3.07
Dodecamethylcyclohexasiloxane	D6	6.08	-0.12	0	0.74	1.09
Hexadecamethylheptasiloxane	L7	7.838	-0.2	0	0.94	-0.48
Octadecamethyloctasiloxane	L8	8.787	-0.23	0	1.08	-1.33
pp'-DDT	DDT	10.02	1.76	0	0.16	-3.77
Diazinon	DZN	8.001	0.81	0.06	1.18	-1.28
Fipronil	FIP	10.065	2.81	0.29	1.54	-8.35

Table A.7: Log P_L data of the studied compounds with unknown vapor pressure (SL = subcooled liquid). All data for 298 K.

Compounds	Abbr.	CAS	MP °C	BP °C	log P_L (SPARC)[a]	log P_L (LFER)[b]	
Plasticizers							
Di-2-propylheptylphthalate	DPHP	53306-54-0	25	400	-8.87	-6.71	
Di-isononylphthalate[c]	DINP	28553-12-0	-48	400	-8.04	-6.65	
1,2-Cyclohexane dicarboxylic acid diisononyl ester[d]	DINCH	166412-78-8	0	394	-6.88	-6.22	
Tri-2-ethylhexyltrimellitate	TOTM	3319-31-1	-46	414	-12.11	-9.15	
Di-iso-butyladipate	DIBA	0141-04-08	-20	279	-1.35	-0.61	
Di-n-butyladipate	DnBA	105-99-7	-32	305	-1.54	-1.25	
Di-2-ethylhexyladipate	DEHA	103-23-1	-67	400	-5.08	-3.92	
Di-isononyladipate[e]	DINA	33703-08-1	0	400	-7.12	-5.16	
Biocides							
Acetamiprid	ACP	135410-20-7	102	352	-2.15	-1.59	**SL**
Icaridin	ICD	119515-38-7	-	296	-3.08	-2.26	**SL**
Cyromazine	CMZ	66215-27-8	222	-	-5.19	-4.76	**SL**
Diflubenzuron	DFB	35367-38-5	232	-	-6.14	-3.66	**SL**
Cyphenothrin	CPT	39515-40-7	-25	-	-6.96	-7.07	
Methoprene	MTP	40596-69-8	25	256	-3.54	-1.67	**SL**
Pharmaceuticals							
Bisoprolol	BPL	66722-44-9	102	445	-6.47	-8.02	**SL**
Dicloflenac	DIC	15307-86-5	158	-	-6.68	-7.11	**SL**
Dapagliflozin	DLF	461432-26-8	65	609	-17.85	-19.56	**SL**
Ibuprofen	IBU	15687-27-1	76	320	-2.3	-2.5	**SL**
Metoprolol	MPL	51384-51-1	120	399	-4.58	-2.22	**SL**
Naproxen	NPX	22204-53-1	152	-	-4.65	-5.66	**SL**
Torasemide	TS	56211-40-6	164	-	-9.66	-13.44	**SL**

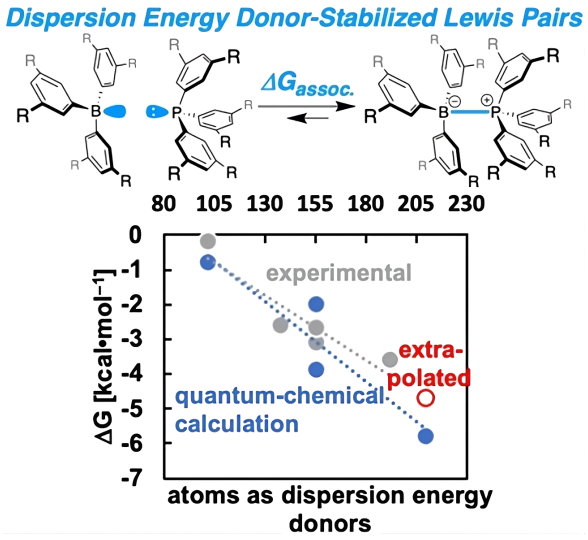
[a] It was assumed that the substance is not solid at room temperature [b] Vapor pressure for the subcooled liquid.
SMILES for [c] bis(7-methyloctyl) phthalate [d] bis(7-methyloctyl) 1,2-cyclohexanedicarboxylate [e] bis(7-methyloctyl) adipate

Table A.8: LFER parameters of the studied compounds with unknown vapor pressure. All data for 298 K. The data were retrieved from the UFZLSER database.[174]

Compounds	Abbr.	CAS	L	S	A	B	log P_L (SPARC)
Plasticizers							
Di-2-propylheptylphthalate	DPHP	53306-54-0	14.3	1.06	0	0.9	-6.71
Di-isononylphthalate	DINP	28553-12-0	13.99	1.27	0	1.02	-6.65
1,2-Cyclohexane dicarboxylic acid diisononyl ester	DINCH	166412-78-8	13.682	1.12	0	0.99	-6.22
Tri-2-ethylhexyltrimellitate	TOTM	3319-31-1	17.017	1.08	0	1.02	-9.15
Di-iso-butyladipate	DIBA	0141-04-08	7.586	0.92	0	0.97	-0.61
Di-n-butyladipate	DnBA	105-99-7	8.102	1.12	0	1.03	-1.25
Di-2-ethylhexyladipate	DEHA	103-23-1	11.297	0.92	0	0.97	-3.92
Di-isononyladipate	DINA	33703-08-1	12.478	1.13	0	1.1	-5.16
Biocide							
Acetamiprid	ACP	135410-20-7	7.51	1.57	0.05	1.21	-1.59
Icaridin	ICD	119515-38-7	7.397	1.03	0.33	0.96	-2.26
Cyromazine	CMZ	66215-27-8	6.929	1.74	0.56	1.24	-4.76
Diflubenzuron	DFB	35367-38-5	9.68	1.17	0.18	0.97	-3.66
Cyphenothrin	CPT	39515-40-7	12.326	2.29	0.05	1.11	-7.07
Methoprene	MTP	40596-69-8	8.976	0.66	0	0.97	-1.67
Pharmaceuticals							
Bisoprolol	BPL	66722-44-9	11.2	1.5	0.3	2.19	-8.02
Dicloflenac	DIC	15307-86-5	11.025	1.85	0.55	0.77	-7.11
Dapagliflozin	DLF	461432-26-8	14.465	2.58	0.97	1.95	-19.56
Ibuprofen	IBU	15687-27-1	7.184	0.7	0.56	0.79	-2.5
Metoprolol	MPL	51384-51-1	9.157	1.61	-0.06	1.72	-2.22
Naproxen	NPX	22204-53-1	9.207	2.02	0.6	0.67	-5.66
Torasemide	TS	56211-40-6	13.23	2.18	0.7	1.6	-13.44

APPENDIX B

Dispersion Energy-Stabilized Boron and Phosphorus Lewis Pairs



Abstract An isostructural series of boron/phosphorus Lewis pairs was systematically investigated. The association constants of the Lewis pairs were determined at variable temperatures, enabling the extraction of thermodynamic parameters. The stabilization of the Lewis adduct increased with increasing size of the dispersion energy donor groups, although the donor and acceptor properties of the Lewis pairs remained largely unchanged. This data was utilized to challenge state-of-the-art quantum chemical methods, which finally led to an enhanced workflow for the determination of thermochemical properties of weakly bound Lewis pairs within an accuracy of 0.6 to 1.0 kcal/mol for computed association free energies.

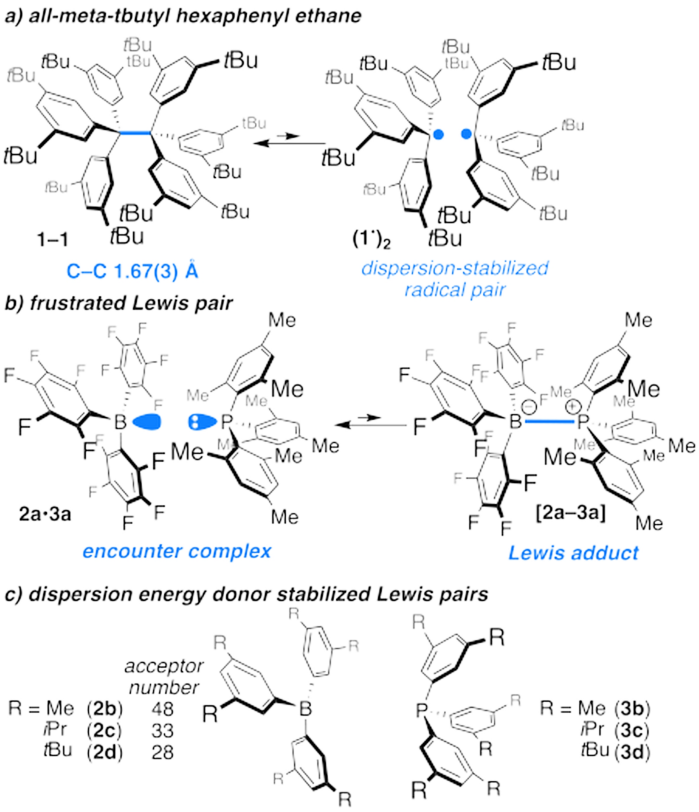


Figure B.1: a) Dispersion energy donor-stabilized all-meta-tBu-substituted hexaphenylethane; b) equilibrium of a frustrated Lewis pair and its Lewis adduct; c) conceptual approach to the stabilization of B/P-derived Lewis pairs including acceptor numbers[215] for the boranes.

B.1 Communication

Only recently, a deeper understanding of London dispersion interactions by quantum chemical and preparative experiments has provided answers to fundamental questions in molecular chemistry.[212] For example, the exceptional stability of *all-meta-tBu* hexaphenylethane (**1–1**) in contrast to the unstable parent hexaphenylethane was explained by extensive attractive dispersion interactions of the *tBu* groups.[213,214] Computational studies revealed that there is not only an ethane-type equilibrium structure as global minimum with an exceptionally long C–C bond of 1.67(3) Å, but also a local energy minimum which corresponds to the weakly dispersion energy donor-bound dimer of trityl radicals, here denoted by $(1\bullet)_2$ (Figure B.1 a).

The crucial role of the aliphatic substituents as stabilizing dispersion energy donors (DEDs) has been established for several other hexaphenylethane derivatives.[216] Structural similarities may be drawn to frustrated Lewis pairs (FLPs)[130,217–222] and their Lewis adducts, although these species mostly are of closed-shell electronic character (Figure B.1 b). Both an unquenched reactivity of the FLP and the formation of an encounter complex[128,129,131] provide the setup e.g. for the heterolytic splitting of molecular hydrogen (H_2).[132,223] Crucial for FLP reactivity is the formation of the encounter complex through the subtle interplay of many, formally weak attractive and repulsive interactions. The association of the archetype FLP consisting of $B(C_6F_5)_3$/P(2,4,6-Me_3-$C_6H_2)_3$ (**2a/3a**) was found

to be slightly endergonic characterized by the free energy of association at 298 K ($\Delta G_{a,298}$) of 0.4 ± 0.2 kcal/mol,[129] although stabilizing polar H–F interactions are present. However, when taking into account the large stabilization of (**1–1**) DEDs and the structural similarity to triaryl phosphane/triaryl borane FLPs (compare Figure B.1 a-c), the formation of an encounter complex predominantly through dispersion interactions is feasible. Only recently, Slootweg et al. introduced tBu-groups into inter¬molecular FLPs based on triarylboranes and triarylamines in order to achieve efficient electron transfer but thermodynamic parameters were not studied in detail by experimental nor computational methods.[133] However, only joined computational and experimental studies can provide the necessary interface to improve both the quantum mechanical methods and the desired molecular properties. In this light, high-level quantum chemistry in combination with reliable experimental data is of utmost importance for advancing the theoretical and molecular understanding.

We selected a series of alkyl-substituted triaryl boranes and phosphanes to study the DED stabilization of Lewis adducts (Figure B.1 c). The six compounds were characterized by NMR spectroscopy, and the Lewis acidity of the boranes and the s-character of the lone pair of the phosphorus were probed by the Gutmann-Beckett method and by the $^1J_{P-Se}$ coupling constant of the corresponding phosphane selenides, respectively (see SI for details). The fact that the $^1J_{P-Se}$ coupling constants of the phosphane selenides are nearly identical at 749 Hz shows that the donor capacity of these phosphanes is unaffected by the peripheral substituents. In contrast, the Lewis acidity of the borane (**2b-d**) decreases (smaller acceptor number) when increasing the size of the DED (compare Figure B.1 c).

The interaction of equimolar mixtures of all nine combinations of borane (**2b-d**) and phosphane (**3b-d**) were studied in solution (0.1 to 0.05 M in C_6D_6) by ^{31}P NMR spectroscopy (see SI for details), revealing Lewis acid/Lewis base interactions through differences in chemical shift in comparison to the pure compounds. This picture was substantiated by ^{1}H,^{1}H nuclear Overhauser enhancement spectroscopy (NOESY) experiments showing intermolecular cross peaks with characteristics of chemical exchange (see SI) and supporting the idea of reversible Lewis adduct formation.* Furthermore, the Lewis adducts **2b•3b**, **2b•3d** and **2d•3d** were characterized by X-ray diffraction (Figure B.2).

All three DED-stabilized Lewis adducts feature the staggered conformation in the solid state similar to **1–1**, whereas the molecular structure of $B(C_6F_5)_3/PPh_3$ exhibits an eclipsed conformation stabilized by the H–F/CH–π interaction.[225] Counterintuitively, the dative P–B bond lengths of 2.102(9) Å (**2b•3b**), 2.072(3) Å (**2b•3d**) and 2.109(9) Å (**2d•3d**) are significantly shorter than in comparable structures with fluorinated boranes and uncongested phosphanes like $Ph_3P–B(C_6F_5)_3$[226] (2.181 Å) or $(3,5\text{-}Me_2\text{-}C_6H_3)_3P–B(4\text{-}H\text{-}C_6F_4)_3$[225] (2.172 Å). This may result from the significant interaction of the DEDs, leading to smaller P–B bond lengths. Such a phenomenon is considered to be responsible for the exceptionally small intermolecular H–H distance of 1.566(5) Å in (3,5-di-*t*buthylphenyl)methane.[227]

The *i*Pr-substituted borane **2c** proved extremely difficult to synthesize, unstable and also very challenging for quantum chemical methods, so that we decided to continue our studies with the methyl and *t*Bu derivatives **2b** and **2d**. In contrast to the ^{1}H NMR spectra (resonance overlaps or severe line broadening), the ^{31}P NMR spectra were suitable for the determination of the association constant (K_a) by titration in a temperature range of 284 K to 303 K, which ultimately allowed the extraction of thermodynamic parameters by van 't Hoff analysis (Table B.1; see SI for details, Figure B.7 and B.8, Table B.10). The association is exothermic for all Lewis pairs, as indicated by the determined

*The corresponding isostructural triaryl amino derivatives displayed no intermolecular NOESY NMR interactions and were therefore not further considered in this study; see SI for details

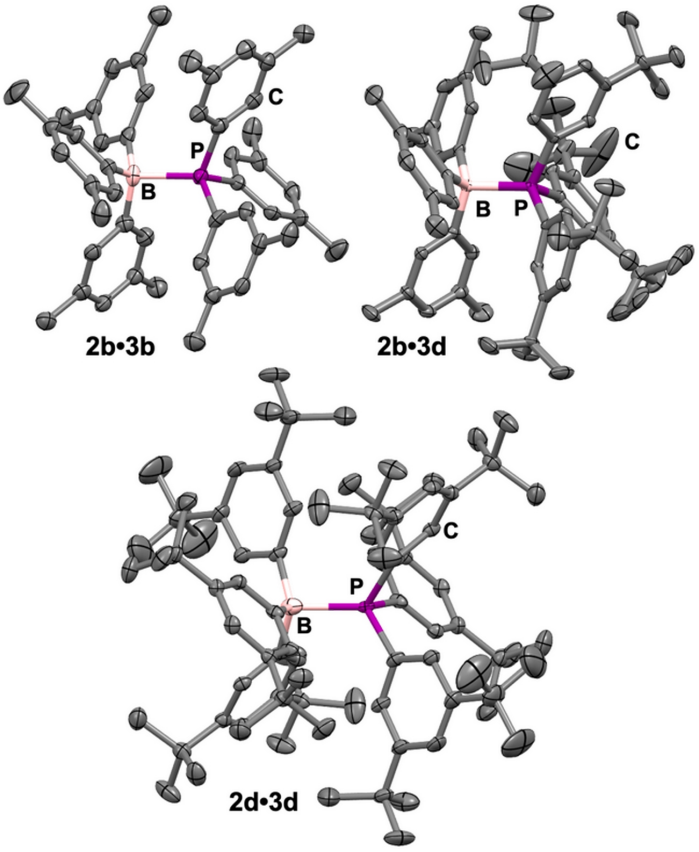


Figure B.2: Molecular structures of **2b•3b**, **2b•3d** and **2d•3d** (hydrogen atoms were omitted for clarity); selected P–B distances: **2b•3b**: 2.102(9) Å, 2b•3d: 2.072(3) Å and **2d•3d**: 2.109(9) Å.[224]

enthalpy of association (ΔH_a) ranging from -8.0 kcal/mol to -10.2 kcal/mol (Table B.1, entries 1-5). The difference in Lewis acidity of **2b** and **2d** is only marginally reflected in ΔH_a determined by titration (entries 1-5). The ΔH_a for Lewis pairs containing **2b** (acceptor number,[215] AN=48) are only marginally more negative than for Lewis pairs containing **2d** (AN=28), whereas the electron pair donor ability of the phosphane remained constant. The entropy of association (ΔS_a) is for all Lewis pairs negative and becomes continually less negative with increasing size of the DED (Table B.1, entries 1-5), counterbalancing the negative ΔH_a. Based on the calculated $\Delta G_{a,298}$, the association becomes more exergonic with increasing size of the DED, which thus must be attributed to the increasing dispersion energy donor ability of the phosphane and borane substituents. As a result of the increases in dispersion interactions, the free energies are negative and reach up to -3.6 kcal/mol for the formation of **2d•3c**. Unfortunately, the large binding constant paired with the very low solubility of **2d•3d** prohibited the analysis of this adduct by NMR titration. However, for lack of alternatives, we utilized the determination of K_a by diffusion ordered spectroscopy (DOSY) NMR, keeping in mind that this method is associated with a large error. We reevaluated all six Lewis pairs with the aim of substantiating the previously obtained data from the NMR titrations (see SI, B.10). Based on this method as well, increasing the size of the DED stabilizes the Lewis adducts and maximizes for the all-meta-tBu-system **2d•3d** in a $\Delta G_{a,298}$ of -2.8 kcal/mol.

Table B.1: Thermodynamic parameters of Lewis pair association determined by ^{31}P NMR titration.[a] Values are in kcal/mol.

Entry	DED[b]	Lewis pair	ΔH°	ΔS°	ΔG°_{298}
1	102	**2b•3b**	-8.9 ±0.3	-29.2 ±1.0	-0.2 ±0.6
2	138	**2b•3c**	-10.2 ±0.1	-25.3 ±0.5	-2.6 ±0.3
3	156	**2b•3d**	-9.6 ±0.1	-21.8 ±0.3	-3.1 ±0.2
4	156	**2d•3b**	-8.5 ±0.1	-19.4 ±0.3	-2.7 ±0.1
5	192	**2d•3c**	-8.0 ±0.3	-14.7 ±1.0	-3.6 ±0.6
6[c]	210	**2d•3d**	-16.1 ±1.0	-44.4 ±3.5	-2.8 ±2.1 (-4.7)[d]

[a] Average of two independent measurements; [b] sum of atoms in ligands of the Lewis pair; [c] determined by DOSY NMR; [d] extrapolated from plot of ΔG°298 vs. DED atom number in Lewis pair (see SI, Figure B.5).

The results of the NMR experiments are the basis for a detailed quantum chemical investigation. The standard approach for modeling the association of FLPs involves starting from gas phase geometries at 0 K and comparing the corresponding association energies or individual contributions thereof, such as the dispersion energy.[133] Computational studies, especially ones that include free energies of association, are rare for FLPs,[131,228–231] and thermostatistical corrections and solvation contributions are usually calculated using a protocol that starts from the gas phase geometries at 0 K. This was also our starting point, but, whereas geometries were described at the PBEh-3c[232] level after a CREST[125] conformer search with GFN2-xTB[85] that provided B–P bond lengths very similar to those in the crystal structure (see SI Table B.5), but the calculated $\Delta G_{a,298}$ with these geometries were clearly too exergonic compared to the measured values (Table B.4). Furthermore, the experimentally observed trend of stronger association with increasing ligand size, due to enhanced dispersion interactions, was not reproduced either. To resolve this apparent contradiction, the accuracy of the electronic gas phase energy calculation was first assessed by comparing the density functional theory (DFT) interaction energies for several, known well-performing functionals to accurate local coupled cluster ("DLPNO-CCSD(T1)/CBS")[233,234] values (see Table B.7). The dispersion contributions were also investigated in detail. They show the expected trend, with a high correlation of dispersion interactions and DED atoms. As expected for DED-stabilized Lewis pairs, the intermolecular D4[92] dispersion energy significantly contributes to the overall $\Delta G_{a,298}$ with values as large as -59.7 kcal/mol for **2d•3d** (depending on the tested functional). However, all of these investigations did not result in a better description of the association free energies.

We therefore suspected a wrong geometry due to an insufficient description of dynamic effects (especially a too-short B–P bond length) as the main cause for the too exergonic $\Delta G_{a,298}$ values calculated with the standard approach, using equilibrium (optimized) molecule geometries.[133,231] Due to partial quenching of the dispersion interactions by solvation interactions[134,235] and dynamic effects (298 K instead of 0 K), the actual (solvated) geometries and P–B distances at finite temperatures are expected to differ from the equilibrium and from the crystal structure geometries. Because modeling by molecular dynamics (MD) was not feasible (see SI), we had to resort to a static approach. In the course of this study, a more advanced approach for conformational sampling and efficient DFT-refinement was developed, which allowed the usage of larger initial ensembles and more accurate DFT methods (the "CRENSO" workflow[12,14]), which had already shown promising results for complicated systems and properties.[119] The CRENSO workflow, while still being a static approach, uses a combination

Table B.2: Boron-phosphorus distances in Å of the final calculated geometries with r^{2}SCAN-3c after advanced conformer sampling in comparison with the experimental values.

	2b•3b	**2d•3b**	**2b•3d**	**2d•3d**
$\overline{B-P}_{calc}$	2.103	2.084	2.103	2.077
$\overline{B-P}_{exp}$	2.102(9)	2.072(3)	-	2.109(9)

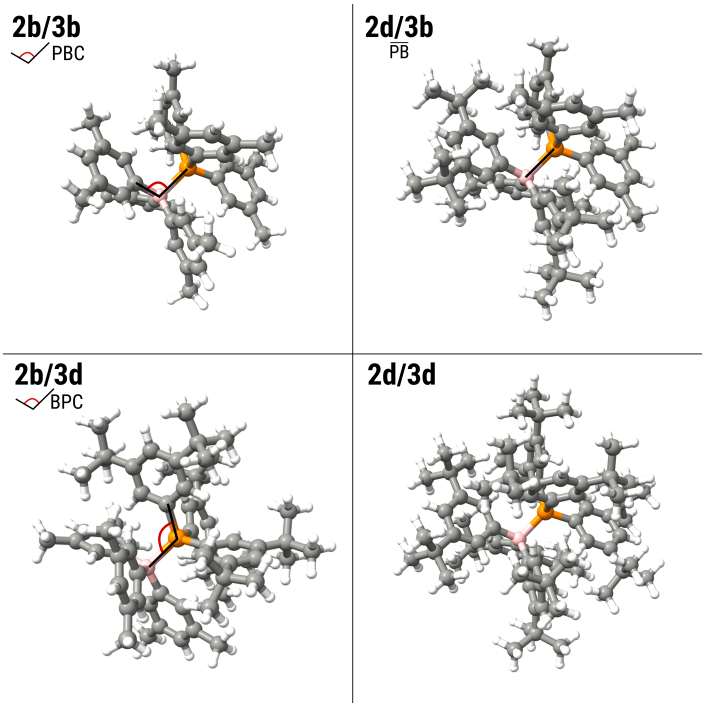


Figure B.3: Geometries with r^{2}SCAN-3c after advanced conformer sampling of the lowest-lying conformers of **2b•3b**, **2d•3b**, **2b•3d** and **2d•3d** Lewis pairs.

of advanced sampling techniques and a refinement with the r^{2}SCAN-3c[126] DFT composite method, which also performed well in our evaluation of interaction energies, to obtain thermally populated ensemble averages in solution (for more details see SI). Only with this refined workflow, we could achieve a good agreement of the calculated $\Delta G_{a,298}$ with the experimentally determined ones. This is probably mainly due to the better conformer sampling, which led to structures that better reproduce the actual geometries in solution at room temperature, even if the P–B bond lengths still tends to correspond to that in the crystal structure (cf. Table B.2, Figure B.3). For even higher accuracy, however, explicit consideration of dynamic effects via MD simulations would be required, which would then presumably lead to elongated P–B bonds. However, this approach was not feasible in terms of computational effort for these large systems. Notably, the slightly reduced bond B–P bond length in **2b•3d** compared to **2b•3b** was reproduced. This may be caused by a delicate balance between repulsive and attractive interactions in **2b•3d** because the electronic properties of the phosphanes are nearly identical as evidenced by the identical $^1J_{P\text{-}Se}$ coupling constants and the identical highest occupied molecular orbital (HOMO) energies.[236]

Table B.3: Comparison of experimentally obtained free energies of association of Lewis pairs with the calculated ones; energies in kcal/mol (*extrapolated, compare Supporting Information Figure B.4).

Contribution	**2b•3b**	**2b•3d**	**2d•3b**	**2d•3d**
DED[a]	102	156	156	210
ΔE	-24.9	-31.2	-31.7	-37.2
ΔG_{mRRHO}	19.7	20.5	20.4	19.3
ΔG_{solv}	4.4	6.8	9.2	12.1
$\Delta G_{assoc.,calc.}$	**-0.8**	**-3.9**	**-2.0**	**-5.8**
$\Delta G_{assoc.,exp.}$	**-0.2 ±0.6**	**-3.1 ±0.2**	**-2.7 ±0.1**	**-4.7***

[a] Sum of atoms in the ligands of the boron and phosphorus.

Energetically, the $\Delta G_{a,298}$ values are nearly on point with the experimental values (Table B.3), which may be attributed to the rigorous sampling and the refinement of the conformer ensembles at finite temperature in solvation, and therefore already includes significant parts of the missing contributions in contrast to the initial GFN2-xTB/PBEh-3c approach. However, the used solvation models were typically developed for small, relatively rigid organic molecules and are, therefore, of limited accuracy for the here investigated large and flexible systems, which marks these good results quite unexpected. While the overall trend is replicated almost perfectly, considering the complexity of the problem and excepting the $\Delta G_{a,298}$ of **2d•3b**, the calculated stability of the investigated Lewis pair is still slightly overestimated. We attribute this to the fact that the dynamic effects on the geometry are still not fully accounted for within the employed workflow. Only the $\Delta G_{a,298}$ of **2d•3b** shows a minor inconsistency from the expected trend using this system. To better comprehend this cause, we examined the individual contributions to the $\Delta G_{a,298}$ as a function of the number of DED atoms and found an irregularity in the solvation-free energy contribution. This can be attributed to an inaccurate description of the non-electrostatic contributions to the ΔG_{solv} by the employed solvation model.[†] Implicit solvation of bulky complexes is still a challenge for modern solvation models, for which, however, no better option is yet available.

In summary, we systematically investigated the association of six isostructural Lewis pairs that differed only by their dispersion energy donors. Joined synthetic and computational experiments determined the exergonic association of the boron- and phosphorus-centered Lewis pairs within an accuracy of 0.5 to 1.0 kcal/mol. In general, the association becomes more exergonic with the increasing number of atoms in the dispersion energy donors within the Lewis pair, and is unaffected by changes to the electron pair donor/acceptor properties of the phosphane and borane. The experimentally determined association constants and derived thermodynamic parameters were utilized for the great challenge of developing a quantum chemical workflow for the calculation of weakly bound Lewis pair complexes. Solvation contributions have a significant impact on the stability of Lewis pairs. Furthermore, dynamic effects, such as the B–P bond elongation of weakly bound Lewis pairs in solution are highly important for the accurate energy description by density functional theory, which is impossible to be captured through a solely static approach. However, by employing state-of-the-art workflows, that use advanced sampling and the refinement of the conformer ensembles at finite temperature with modern DFT functionals, combined with the Boltzmann averaging of the thermally populated ensemble, a large part of these missing contributions can be fathomed, and the free energy of association of complicated

[†] A more detailed investigation of this can be found in the supporting information.

systems were very accurately calculated. Incorporating these findings into future investigations will enhance the accuracy and robustness of studies in frustrated Lewis pair chemistry.

Acknowledgments

The German Research Foundation (DFG) is gratefully acknowledged for financial support (PA 1562/15-1). Open Access funding enabled and organized by Projekt DEAL.

B.2 Supporting Information

This section is an (adapted) excerpt of the supporting information, containing the tables and figures referenced in the main text and the computational details for completeness. See DOI: https://doi.org/10.1002/anie.202308752 for the full supporting information.

Computational Details

The association free energy, e.g. of a frustrated Lewis pair (FLP), is given by the difference of Gibbs free energies between the pair (A-B) and the acid (A) and base (B) as

$$\Delta G_{\text{assoc.}} = \Delta G_{A-B} - \Delta G_A - \Delta G_B, \tag{B.1}$$

The theoretical Gibbs free energy for a single compound can be computed by DFT as[96]

$$\Delta G_{\text{total}} = E_{\text{gas}} + G_T + \delta G_{\text{solv}}. \tag{B.2}$$

Here, E_{gas} is the electronic energy in the gas phase at 0 K, G_T is a thermostatistical contribution at temperature T including the zero-point vibrational energy and temperature-dependent vibrational frequency contributions, and δG_{solv} is the free energy of solvation, corresponding to the phase change from gas to liquid phase.
The association free energy of a frustrated Lewis pair can therefore be calculated as[237]

$$\Delta G^{\circ}_{298} = \Delta_R G_{\text{assoc.}} = \Delta E_{\text{gas}} + \Delta G_{\text{T,298}} + \Delta \delta G_{\text{solv}}. \tag{B.3}$$

$\Delta_R G_{\text{assoc.}}$ is the reaction free energy of a reaction of an acid and a base to an FLP and ΔE_{gas}, ΔG_{T}, and $\Delta \delta G_{\text{solv}}$ are the corresponding decomposed contributions.
In the present case, this single structure approach may not be adequate since the complex structure is rather flexible. This gives rise to several energetically close conformers, significantly impacting energies and properties at standard temperature. Hence, a Boltzmann-weighted ensemble over the relevant conformational space must be considered. Therefore, as a first check to evaluate the complexity of the system, a conformer search was performed with the conformer rotamer ensemble sampling tool (CREST)[125] utilizing the GFN2-xTB[85] semiempirical method combined with implicit solvation[26] in toluene and a search energy cutoff of 1.5 kcal/mol (”crude” conformational screening). The resulting geometries were then re-optimized with the PBEh-3c[232] composite method and the DCOSMO-RS[238] implicit solvation. PBEh-3c frequencies were calculated in a modified rigid rotor harmonic oscillator (mRRHO) approach.[96] The solvation-free energies were computed using the conductor-like screening

model for realistic solvation (COSMO-RS)[28] with the 2016 parametrization. Table B.4 shows the results for the various conformers and contributions. All contributions were also calculated by a Boltzmann weighted average over the conformer ensemble. The results for the total association free energy differs by about 0.1-0.3 kcal/mol if the conformer ensemble is taken into account.

However, while the geometries seemed to agree very well with the crystal structure ones (Table B.5), these theoretically obtained values did not agree very well with the experimental data (see Table B.4). To shed some light on this issue, first the accuracy of the electronic gas phase energy calculation was assessed by comparing the DFT interaction energies to accurate reference values at the "DLPNO-CCSD(T1)/CBS" level of theory.[233] Using these higher-level reference energies did not significantly improve the description of the association free energy.

As a wrong description of the electronic energy could therefore be ruled out as the leading cause of the discrepancy, after extensive investigations and detailed comparison with the experimental structures, we suspected a wrong geometry as the major source for the deviation of the calculated from the measured association free energies. Due to partial quenching of the dispersion interactions in solution and dynamic effects (298 K instead of 0 K), the solvated geometries at finite temperatures are expected to differ from the ones in the crystal structure (e.g. longer P-B bonds should lead to mitigated interactions). One approach to resolve this issue is to apply metadynamics[239,240] (MD) with a scaled tight binding Hamiltonian.[241] By manually down-scaling the P–B interaction in the GFN2-xTB Hamiltonian, we would be able to simulate this effect, which would consequently lead to a more realistic B–P bond length and this, in turn, to a less exergonic free energy. Unfortunately, down-scaling the P–B interactions in the MDs at room temperatures led to a dissociation of the FLPs. Therefore, we had to stick to the static approach and try to include the missing contributions in another way.

As further investigations and new geometry optimizations were therefore necessary, we assessed the electronic energies calculated with selected DFT functionals, which proved robust and reliable in benchmarks,[242] as single point calculations on the optimized geometries, to find a suitable one for these investigations. All investigated DFT functionals, except for the range separated ωB97X-V[243] hybrid functional, which already includes non-local long-range correlation, were evaluated together with the DFT-D4[92] dispersion correction. It is apparent from the results in Table B.6 that neglecting these dispersion interactions would introduce a huge error since the investigated FLP association reactions are highly dominated by dispersion interactions.

The D4 correction to the interaction energy reaches values of up to -59.7 kcal/mol for the **2d•3d** FLP (depending on the underlying functional). The association energies based on the electronic gas phase single point energies $\Delta E_{gas,0K}$ (including dispersion interactions) can be found in Table B.7. The hybrid functional PBE0-D4[244] as well as the range separated hybrid ωB97X-V systematically underestimate the association energy (MD = MAD). The latter error can be attributed to the missing many-body dispersion effects in the used VV10 dispersion model. This can lead to sizeable errors for larger complexes, where cooperative effects become important.[245] The more accurate double hybrid functionals PWPB95[246]-D4 and revDSD[247]-BLYP-D4 but also the efficient r2SCAN-3c[126] composite method perform very well for the investigated systems. Hence, r2SCAN-3c was considered for all further DFT calculations as the established default method in our workflow.[5,12,14] We have also checked the calculated solvation free energies (on the geometries obtained from "crude" screening) by comparing them with to the corresponding values obtained with the conductor-like screening model (CPCM)[248] and, to evaluate the importance of nonelectrostatic interactions, with the universal solvent

Table B.4: Association free energies and individual contributions calculated after "crude" conformational screening on the PBEh-3c level of theory. The final association free energy is derived by a Boltzmann weighted average over the conformer ensemble. Energies are given in kcal/mol. The lowest-lying conformer is denoted by an asterix.

	$\Delta E_{gas,0K}$	$\Delta\delta G_{solv}$	ΔG_T	ΔG_{assoc}
	2b•3b			
111	-25.0	4.8	18.3	-1.9
121*	-26.9	4.7	18.0	-4.2
112	-20.4	4.2	17.8	1.6
122	-22.3	4.1	17.7	-0.5
$\Delta\overline{G}_{assoc.}$ // *exp.*				**-4.2 // -0.2**
	2d•3b			
111	-30.9	9.6	19.1	-2.2
121*	-32.9	9.6	-19.1	-4.2
$\Delta\overline{G}_{assoc.}$ // *exp.*				**-4.5 // -2.7**
	2b•3d			
111	-30.9	6.7	18.2	-6.0
121	-31.1	6.7	19.0	-5.4
151*	-32.9	7.8	17.7	-7.4
112	-30.7	6.5	18.8	-5.4
122	-31.0	6.5	19.7	-4.8
152	-32.8	7.6	18.3	-6.9
113	-30.6	6.2	18.8	-5.6
123	-30.9	6.2	19.6	-5.1
153	-32.7	7.3	18.3	-7.1
$\Delta\overline{G}_{assoc.}$ // *exp.*				**-7.0 // -3.1**
	2d•3d			
111	-38.5	11.7	20.3	-6.3
121	-38.6	11.7	21.5	-5.4
151	-40.4	12.8	19.8	-7.8
112	-39.5	11.7	20.7	-7.1
122	-39.6	11.7	21.5	-6.4
152*	-41.5	12.8	20.2	-8.5
113	-38.3	11.8	20.7	-5.8
123	-38.5	11.8	21.5	-5.2
153	-40.3	12.9	20.2	-7.2
114	-37.9	11.3	21.7	-4.9
124	-38.1	11.3	21.7	-5.1
154	-39.9	12.4	20.4	-7.1
$\Delta\overline{G}_{assoc.}$ // *exp., extrapolated*				**-8.1 // -4.7**

Table B.5: B-P bond lengths of the original lowest-lying PBEh-3c geometries (after "crude" conformational screening) in comparison to experimental ones in Å.

	2b•3b	**2d•3b**	**2b•3d**	**2d•3d**
$B–P_{calc}$	2.107	2.078	2.100	2.073
$B–P_{exp}$	2.102(9)	2.072(3)	-	2.109(9)

Table B.6: DFT-D4 dispersion contribution to the association energies for selected functionals in kcal/mol.

Conf.	PBE0-D4	B3LYP-D4	PWPB95-D4	revDSD-BLYP-D4
		2b•3b		
111	-25.2	-37.6	-12.0	-12.1
121	-25.4	-37.8	-12.1	-12.1
112	-25.7	-38.4	-12.1	-12.0
122	-25.8	-38.6	-12.1	-12.1
		2d•3b		
111	-31.5	-45.7	-14.9	-14.1
121	-31.6	-45.9	-14.9	-14.1
		2b•3d		
111	-29.1	-42.4	-13.9	-13.4
121	-29.2	-42.6	-13.9	-13.4
151	-29.9	-43.4	-14.0	-13.4
112	-29.3	-42.8	-14.1	-13.6
122	-29.5	-43.1	-14.2	-13.7
152	-30.2	-43.9	-14.2	-13.6
113	-29.3	-42.8	-14.1	-13.6
123	-29.5	-43.1	-14.2	-13.7
153	-30.2	-43.9	-14.2	-13.6
		2d•3d		
111	-38.9	-55.6	-18.7	-17.2
121	-39.1	-55.9	-18.8	-17.2
151	-39.8	-56.7	-18.9	-17.2
112	-40.2	-57.8	-19.6	-18.2
122	-40.4	-58.0	-19.6	-18.2
152	-41.0	-58.8	-19.7	-18.2
113	-40.3	-57.9	-19.6	-18.2
123	-40.5	-58.1	-19.7	-18.2
153	-41.1	-58.9	-19.8	-18.2
114	-40.8	-58.7	-20.0	-18.6
124	-40.9	-58.9	-20.0	-18.6
154	-41.6	-59.7	-20.1	-18.5

model based on density (SMD)[30] (see Table B.8). As expected, SMD agrees much better with COSMO-RS than CPCM. However, the differences between SMD and COSMO-RS are more pronounced for FLPs that include **3d** monomers, which can probably be attributed to a different description of the solvent accessible surface area (SASA, see below). In contrast to SMD and COSMO-RS, the purely electrostatic approach utilized in CPCM misses the important non-electrostatic contributions leading to an even stronger overbinding of the FLPs.

Still, the investigated systems are complicated and the problem to include all relevant contributions with a static approach remained. Fortunately, while this investigation took place, a more advanced, and rigorous approach for conformational sampling consisting of a combination of CREST and an energetic sorting algorithm (ENSO)[12] employing the efficient r2SCAN-3c composite method was developed in our group. Since r2SCAN-3c reproduced the FLP association energies exceptionally well (see above), it seemed particularly promising to try this new workflow for the systems in question. Other than this promising composite method, the CRENSO[12] workflow employs an advanced sampling technique, adopting a force-field approach with GFN-FF[118] to sample conformations on artificially modified potential energy surfaces. This approach is far superior to using a single potential energy surface obtained with a semi-empirical method and has already proven to deliver accurate and robust results for complicated and flexible molecules[5]. Solvation contributions were included using the ALPB[26] solvation model. The thermostatic contributions were obtained from a single point hessian (SPH) approach[168] ("fine" conformational screening). The general applicability of the SPH approach was already established by comparison to DFT values and saves hours of computational time, while also being able to calculate frequencies for solvated compounds.[5] Solvation free energies for the r2SCAN-3c/DCOSMO-RS[238] optimized geometries were again obtained using the COSMO-RS method with the 2016 parametrization and benzene as a solvent, and all contributions were calculated as Boltzmann weighted average. Using this advanced workflow, we obtained more realistic association free energies (see Table B.3, Figure B.4).

Still, there remains a slight overestimation of the association free energies, which may be due to dynamic effects on the geometries not accessible with this static approach. Interestingly, the association free energy is now slightly underestimated for the **2d•3d** FLP (see Figure B.4). Hence, we looked at the various contributions to the association free energy via an energy decomposition analysis (EDA) dependent on the number of atoms in the ligands (see Figure B.5). The EDA was performed using the supermolecular approach on the B3LYP-D4[69,70,76]/def2-TZVP level of theory with the TURBOMOLE[144] V7.3.1 program package.

The relative contribution of the intermolecular dispersion energy ΔE_{disp} increases linearly with the number of atoms (R^2 = 0.98) and the thermostatistical contribution ΔG_T is nearly constant, which is the expected behavior. These contributions are also practically identical for the medium-size FLPs **2d•3b** and **2b•3d**. The energy differences between these FLPs are negligible for the electronic gas phase energy $\Delta E_{\text{gas,no disp}}$. Therefore, we conclude that these contributions are not the source of the observed error. However, the solvation contribution ΔG_{solv} shows an irregularity: it is significantly larger for **2d•3b** compared to**2b•3d**. We attribute this mainly to an overestimation of the solvation free energies of the monomers. Attractive non-electrostatic interactions with the solvent are usually approximated by considering the compound's solvent-accessible surface area (SASA). The SASA describes a modified van der Waals surface areas taking into account that, due to steric reasons, not all parts of the surface of a compound are accessible for a given solvent. Hence, the SASA depends not only on the elemental distribution of the compound but also on its geometry. Due to the free electron

Table B.7: Reference values for the gas phase association energies at 0 K calculated with "DLPNO-CCSD(T1)[234]/CBS" in comparison with energies calculated with several DFT functionals after basic conformational screening in kcal/mol. Furthermore the mean deviation (MD), the mean absolute deviation (MAD), the standard deviation (SD), and the absolute maximum error (AMAX) of the presented data

Conf.	PBE0-D4	B3LYP-D4	ωB97X-V	PWPB95-D4	revDSD-BLYP-D4	r^{2}SCAN-3c	**ref.**
				2b•3b			
111	-30.4	-26.4	-27.2	-26.9	-25.8	-25.1	**-26.0**
121	-32.1	-28.1	-28.8	-27.7	-27.4	-26.8	**-27.7**
112	-25.3	-22.6	-22.3	-22.3	-21.1	-20.0	**-20.7**
122	-27.1	-24.	-23.9	-23.1	-22.7	-21.7	**-22.3**
				2d•3b			
111	-36.3	-32.5	-36.0	-31.4	-30.9	-33.6	**-31.4**
121	-38.1	-34.2	-37.6	-32.2	-32.5	-35.3	**-33.1**
				2b•3d			
111	-35.8	-31.7	-34.5	-30.8	-30.6	-30.9	**-30.7**
121	-36.0	-31.9	-34.8	-31.0	-30.8	-31.0	**-30.9**
151	-38.1	-33.6	-36.8	-32.1	-32.2	-33.2	**-32.8**
112	-35.9	-32.0	-34.6	-30.6	-30.7	-30.8	**-30.6**
122	-36.1	-32.2	-34.9	-30.8	-30.9	-30.9	**-30.8**
152	-38.2	-33.9	-36.8	-31.9	-32.3	-33.1	**-32.7**
113	-35.9	-32.0	-34.4	-30.7	-30.6	-30.7	**-30.4**
123	-36.1	-32.2	-34.7	-30.9	-30.8	-30.8	**-30.6**
153	-38.2	-33.9	-36.7	-32.0	-32.3	-32.9	**-32.5**
				2d•3d			
111	-42.8	-38.9	-45.5	-37.2	-36.6	-37.1	**-37.5**
121	-43.0	-39.1	-45.8	-37.4	-36.8	-38.2	**-37.7**
151	-45.1	-40.8	-47.7	-38.5	-38.2	-39.4	**-39.6**
112	-43.6	-40.1	-46.4	-38.7	-37.8	-38.0	**-38.2**
122	-43.8	-40.2	-46.7	-38.9	-38.0	-38.2	**-38.4**
152	-46.0	-42.0	-48.7	-40.0	-39.5	-40.3	**-40.3**
113	-42.7	-39.0	-45.3	-38.1	-36.5	-37.0	**-37.1**
123	-42.8	-39.1	-45.6	-38.4	-36.7	-37.2	**-37.7**
153	-45.0	-40.8	-47.6	-39.4	-38.2	-39.3	**-39.2**
114	-42.2	-38.6	-44.9	-38.2	-36.0	-36.7	**-36.8**
124	-42.4	-38.7	-45.2	-38.4	-36.2	-36.8	**-37.0**
154	-44.5	-40.5	-47.1	-39.5	-37.7	-39.0	**-38.9**
MD	**-5.27**	**-1.41**	**-5.54**	**-0.22**	**0.42**	**-0.10**	
MAD	**5.27**	**1.41**	**5.54**	**0.59**	**0.53**	**0.45**	
SD	**0.38**	**0.44**	**2.61**	**0.73**	**0.48**	**0.72**	
AMAX	**5.8**	**1.9**	**8.4**	**1.6**	**1.4**	**2.2**	

Table B.8: Comparison of implicit solvation models after "crude" conformational screening. Reaction solvation free energies $\Delta\delta G_{solv}$ refer to averages over conformer ensembles and are given in kcal/mol.

	COSMO-RS	SMD	CPCM
2b•3b	4.64	5.72	1.90
2d•3b	9.56	6.68	2.11
2b•3d	7.50	6.32	2.12
2d•3d	12.57	9.21	2.40

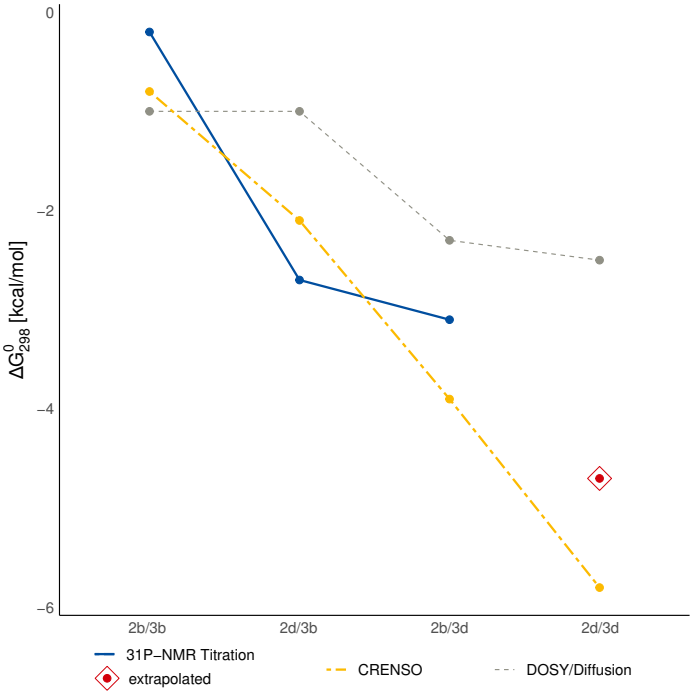


Figure B.4: Comparison of experimentally obtained association free energies with the calculated ones for **2b/3b**, **2d/3b**, **2b/3d**, **2d/3d**.

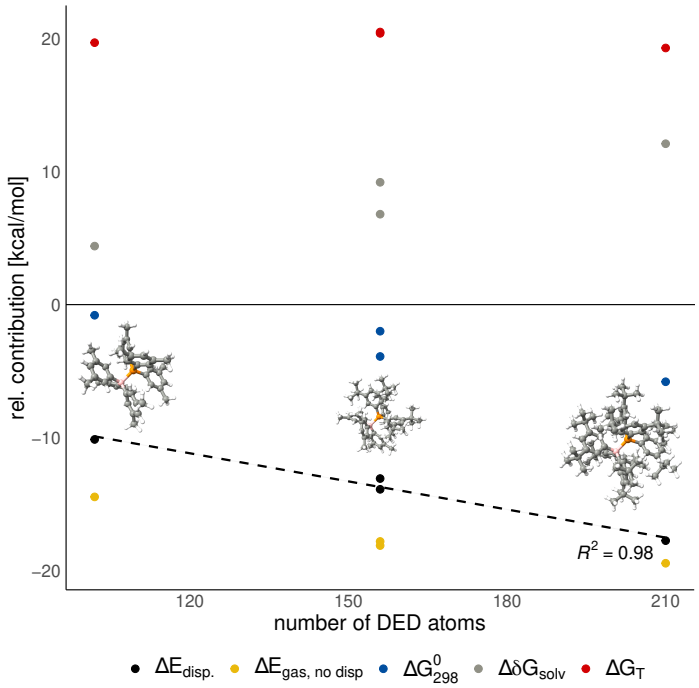


Figure B.5: Various contributions to the association free energy obtained via an energy decomposition analysis. The energies are given with respect to the number of ligand atoms. For the dispersion contribution, a linear regression is performed.

Table B.9: Comparison of the SASA of the **2d** and the **3d** monomer calculated with a solvent radius of 1.3 Å and with a probe radius of 0.4 Å in Å^2.

r_{probe}	1.3 Å	0.4 Å
2d	858	799
3d	901	795

pair of the **3d** monomer, the ligands are oriented trigonal pyramidal, making some parts of the surface inaccessible to the solvent. This is not the case for the 2d monomer, leading to a significant difference in the SASA of 43 Å^2 between these two monomers when calculated with a probe radius of 1.3 Å. A visualization of this effect is depicted in Figure B.6. Hence, a significant difference in the attractive solvent interactions destabilizing the **3d** monomer compared to the **2d** monomer is found, which in turn stabilizes the **2b•3d** complex.

This also explains the more pronounced differences in the solvation free energy between SMD and COSMO-RS when the calculation involves a **3d** monomer. The SMD model uses a solvent radius of just 0.4 Å.[30] As can be seen from the calculations of the SASA in Table B.9 changing the probe radius leads to a nearly identical size of the SASA for both, the **2d** monomer and the **3d** monomer, which can also be observed from the total solvation free energies in table B.8.

Therefore, we attribute the major source for remaining deviations of the calculated association free energies from the measured ones to the implicit solvation treatment, for which, however, no better

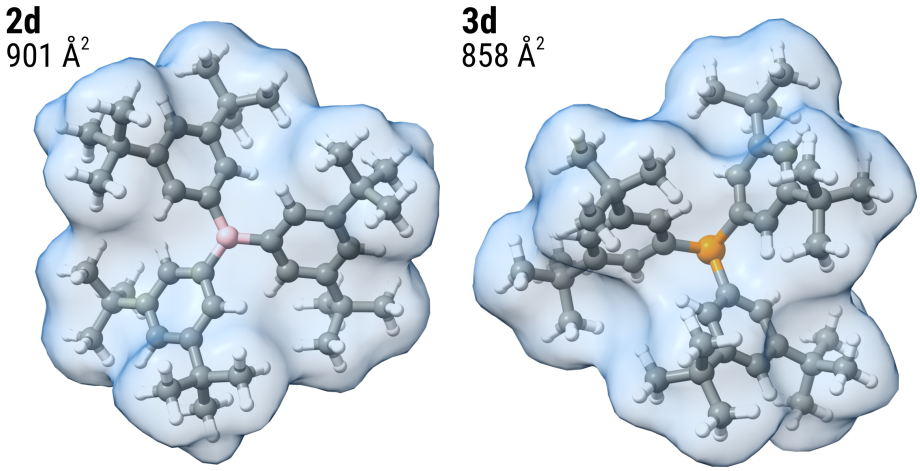


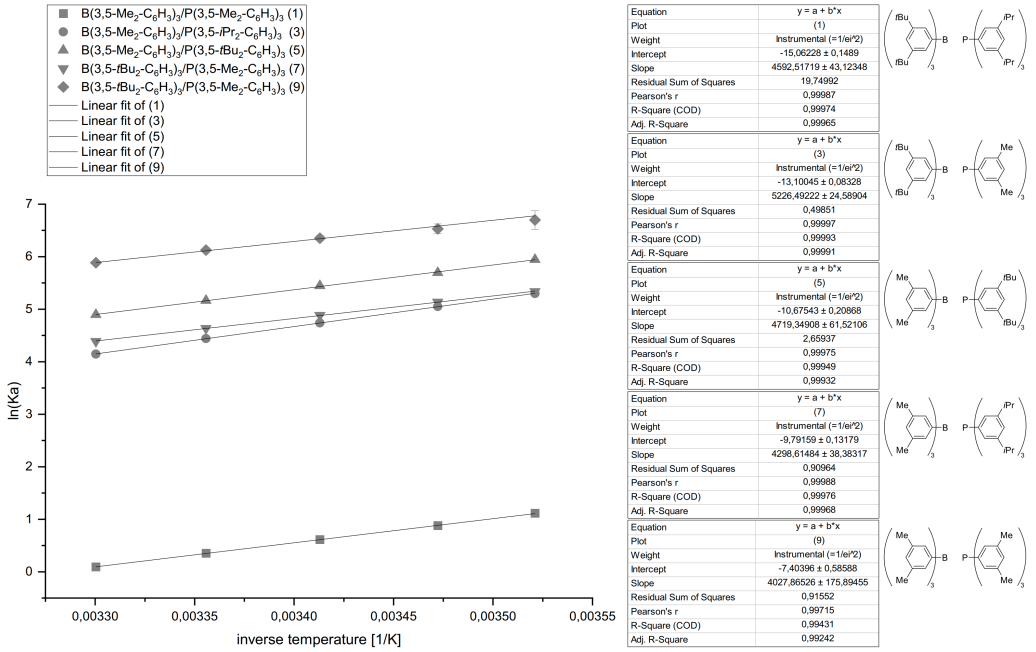
Figure B.6: SASA of the **2d** and the **3d** Monomer calculated with a probe radius of 1.3 Å.

alternative is available at the moment.

Supporting Tables and Figures

Table B.10: Comparison of thermodynamic parameters determined by 31P NMR titration and DOSY experiments (values in kcal/mol, free energy calculated for 298K)

P NMR titration						
Lewis Pair	ΔH	ΔΔH [10^{-2}]	ΔS [10^{-3}]	ΔΔS [10^{-3}]	ΔG	ΔΔG
2b•3b	-8.9	±29.4	-29.1	±1.0	-0.2	±0.6
2b•3c	-10.2	±14.0	-25.3	±0.5	-2.6	±0.3
2b•3d	-9.6	±7.3	-21.8	±0.3	-3.1	±0.2
2c•3b			n.a.			
2d•3b	-8.5	±7.0	-19.4	±0.2	-2.7	±0.1
2d•3c	-8.0	±29.5	-14.7	±1.0	-3.6	±0.6
2d•3d			n.a.			
DOSY experiments						
Lewis Pair	ΔH	ΔΔH [10^{-2}]	ΔS [10^{-3}]	ΔΔS [10^{-3}]	ΔG	ΔΔG
2b•3b	-7.6	±0.29	-21.8	±1.0	-1.2	±0.6
2b•3c	-21.2	±1.21	-67.5	±4.1	-1.1	±2.4
2b•3d	-22.0	±2.67	-0.1	±0.0	-1.6	±5.4
2c•3b	-38.3	±5.1	-121.4	±17.3	-2.1	±10.2
2d•3b	-28.9	±2.2	-93.0	±7.7	-1.2	±4.5
2d•3c	-23.1	±0.4	-71.1	±1.5	-1.9	±0.9
2d•3d	-16.1	±1.0	-44.3	±3.5	-2.8	±2.1



Equation	y = a + b*x
Plot	(1)
Weight	Instrumental (=1/ei^2)
Intercept	-15,06228 ± 0,1489
Slope	4592,51719 ± 43,12348
Residual Sum of Squares	19,74992
Pearson's r	0,99987
R-Square (COD)	0,99974
Adj. R-Square	0,99965

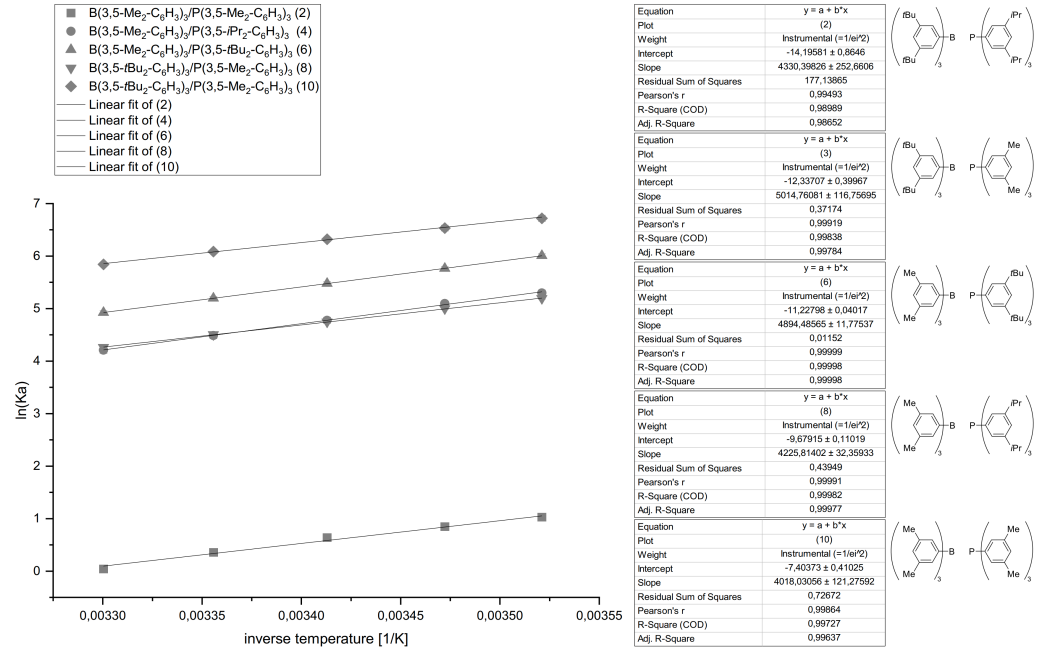
Equation	y = a + b*x
Plot	(3)
Weight	Instrumental (=1/ei^2)
Intercept	-13,10045 ± 0,08328
Slope	5226,49222 ± 24,58904
Residual Sum of Squares	0,49851
Pearson's r	0,99997
R-Square (COD)	0,99993
Adj. R-Square	0,99991

Equation	y = a + b*x
Plot	(5)
Weight	Instrumental (=1/ei^2)
Intercept	-10,67543 ± 0,20868
Slope	4719,34908 ± 61,52106
Residual Sum of Squares	2,65937
Pearson's r	0,99975
R-Square (COD)	0,99949
Adj. R-Square	0,99932

Equation	y = a + b*x
Plot	(7)
Weight	Instrumental (=1/ei^2)
Intercept	-9,79159 ± 0,13179
Slope	4298,61484 ± 38,38317
Residual Sum of Squares	0,90964
Pearson's r	0,99988
R-Square (COD)	0,99976
Adj. R-Square	0,99968

Equation	y = a + b*x
Plot	(9)
Weight	Instrumental (=1/ei^2)
Intercept	-7,40396 ± 0,58588
Slope	4027,86526 ± 175,89455
Residual Sum of Squares	0,91552
Pearson's r	0,99715
R-Square (COD)	0,99431
Adj. R-Square	0,99242

Figure B.7: Van't Hoffs plot of Lewis pairs from ^{3}1P NMR titration and their linear fit.



Equation	y = a + b*x
Plot	(2)
Weight	Instrumental (=1/ei^2)
Intercept	-14,19581 ± 0,8646
Slope	4330,39826 ± 252,6606
Residual Sum of Squares	177,13865
Pearson's r	0,99493
R-Square (COD)	0,98989
Adj. R-Square	0,98652

Equation	y = a + b*x
Plot	(4)
Weight	Instrumental (=1/ei^2)
Intercept	-12,33707 ± 0,39967
Slope	5014,76081 ± 116,75695
Residual Sum of Squares	0,37174
Pearson's r	0,99919
R-Square (COD)	0,99838
Adj. R-Square	0,99784

Equation	y = a + b*x
Plot	(6)
Weight	Instrumental (=1/ei^2)
Intercept	-11,22798 ± 0,04017
Slope	4894,48565 ± 11,77537
Residual Sum of Squares	0,01152
Pearson's r	0,99999
R-Square (COD)	0,99998
Adj. R-Square	0,99998

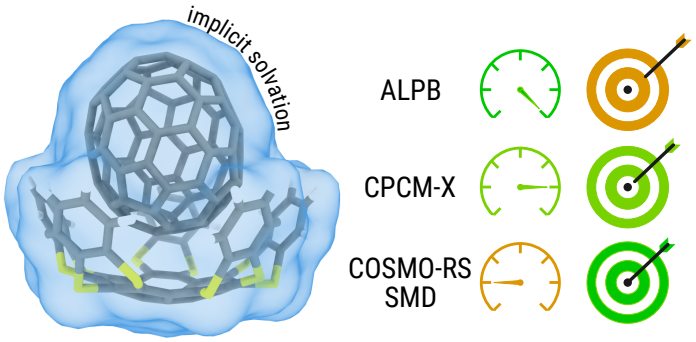
Equation	y = a + b*x
Plot	(8)
Weight	Instrumental (=1/ei^2)
Intercept	-9,67915 ± 0,11019
Slope	4225,81402 ± 32,35933
Residual Sum of Squares	0,43949
Pearson's r	0,99991
R-Square (COD)	0,99982
Adj. R-Square	0,99977

Equation	y = a + b*x
Plot	(10)
Weight	Instrumental (=1/ei^2)
Intercept	-7,40373 ± 0,41025
Slope	4018,03056 ± 121,27592
Residual Sum of Squares	0,72672
Pearson's r	0,99864
R-Square (COD)	0,99727
Adj. R-Square	0,99637

Figure B.8: Van't Hoffs plot of Lewis pairs from ^{3}1P NMR titration and their linear fit.

APPENDIX C

Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods



Abstract We have developed a new method to accurately account for solvation effects in semiempirical quantum mechanics (SQM) based on a polarizable continuum model (PCM). The extended conductor-like polarizable continuum model (CPCM-X) incorporates a computationally efficient domain decomposition conductor-like screening model (ddCOSMO) for extended tight binding (xTB) methods and uses a post-processing approach based on established solvation models, like the conductor-like screening model for real solvents (COSMO-RS) and the universal solvent model based on electron density (SMD). According to various benchmarks, the approach performs well across a broad range of systems and applications, including hydration free energies, non-aqueous solvation free energies, and large supramolecular association reactions of neutral and charged species. Our method for computing solvation free energies is much more accurate than the current methods in the `xtb` program package. It improves the accuracy of solvation free energies by up to 40% for larger supramolecular association reactions to match even the accuracy of higher-level DFT-based solvation models like COSMO-RS and SMD while being computationally more than two orders of magnitude faster. The proposed method and the underlying ddCOSMO model are readily available for a wide variety of solvents and are accessible in `xtb` for use in various computational applications.

C.1 Introduction

In the past decades, quantum chemical (QC) methods transitioned from a niche topic to an essential component of modern research. Starting from fundamental wave function theory (WFT), the development of approximate methods, like density functional theory (DFT), led to enhanced efficiency enabling the treatment of bigger, more complex molecules. However, even the most efficient DFT methods are still limited by costly integral calculations. For this reason, over the last years, semi-empirical quantum chemical methods (SQM), like GFN2-xTB[85] and DFTB,[83] were developed, introducing drastic integral approximations to treat larger systems in a reasonable amount of computational time.

Most current QC methods calculate the investigated compounds first without any thermostatistical contribution (i.e., at 0 K) and in the gas phase ($E_{\text{gas,0 K}}$). Naturally, this is far from reality, where experiments predominantly occur in laboratories at room temperature and in a solvent. Therefore, additional contributions must be considered to compare theoretical calculations to experimental data or to use calculations for screening purposes.

The mostly relevant total free energy G_{total} includes the solvation free energy (ΔG_{solv}), which corresponds to the transition of a compound from the gas phase to the solvent phase and a thermostatistical contribution ΔG_T

$$G_{\text{total}} = E_{\text{gas,0 K}} + \Delta G_T + \Delta G_{\text{solv}}. \tag{C.1}$$

While the ΔG_T term is typically approximated by a modified rigid rotor harmonic oscillator (mRRHO)[96] scheme, there are several choices to model the solvation part.[14,98] The adequacy of these choices strongly depends on the used quantum chemistry package, the QC methods, and the investigated system. Solvation effects are essential for various properties and should always be considered when conducting a theoretical study referring to a solvated state. Furthermore, some properties, which depend on fluid phase equilibria (e.g., partition coefficients or vapor pressures), cannot even be fathomed without considering these effects.

In recent work, we introduced a solvation model for semiempirical methods based on the analytical polarizable Poisson-Boltzmann method (ALPB).[25,26] While we showed that reasonable agreement with experimental data can be obtained with this approach, there are more accurate solvation models out there, like the conductor-like screening model for real solvents (COSMO-RS),[27,28] the COSMO segment activity coefficient (COSMO-SAC)[29] model or the universal solvent model based on solute electron density (SMD),[30] which are based on polarizable continuum models (PCMs),[21] which can reach an accuracy far out of range of ALPB.[23]

However, these models are based on DFT calculations and are therefore not readily available for use with GFN2-xTB or SQM methods developed in the future, which limits their usability for very large systems due to computational cost. Moreover, we showed that even the less sophisticated ALPB method can be an essential part of a computational workflow that combines rigorous conformational screening and multi-level refinement (CRENSO).[12] By utilizing more sophisticated PCM-type solvation models in the later parts of the workflow, it allowed us to accurately calculate vapor pressures[119] and partition coefficients[5] for a large variety of different compounds, even for very flexible ones.

In this work, we describe the implementation of a PCM-type domain-decomposition conductor-like screening model (ddCOSMO)[137,249] in our `xtb` code and develop a post-processing solvation model, that is based on established solvation models for DFT methods, to further refine the description of the solvation free energy with SQM methods.

C.2 Theory

The standard state solvation free energy (ΔG_{solv}) can usually be partitioned as

$$\Delta G_{\text{solv}} = \Delta G_{\text{ES}} + \Delta G_{\text{NE}} + \Delta G_{\text{corr.}}, \tag{C.2}$$

where ΔG_{ES} is the electrostatic contribution and ΔG_{NE} is the non-electrostatic contribution to the solvation free energy. $\Delta G_{\text{corr.}}$ is a standard state correction factor, which is dependent on the solvent's temperature and density. The standard state correction is further dependent on the theoretical reference framework of the method, which is a stoichiometric ratio of solute and solvent, and the final reference state, chosen to be a concentration of 1 $\frac{mol}{L}$. The standard state correction can then be calculated as

$$\Delta G_{\text{corr}} = RT \ln \frac{\rho_S V_{\text{IG}}}{M_s}, \tag{C.3}$$

where ρ_s is the density of the solvent, V_{IG} is the ideal gas molar volume and M_s is the molecular weight of the solvent.
The extended conductor-like polarizable continuum model (CPCM-X) proposed in this work is based on an approach developed by Klamt et al.[27,28] for the electrostatic part of the energy and therefore further splits this contribution into an ideal screening part (ΔG_{IS}) and a restoring free energy part (ΔG_{res}) as

$$\Delta G_{\text{ES}} = \Delta G_{\text{IS}} + \Delta G_{\text{res}}. \tag{C.4}$$

ΔG_{IS} can be obtained from the difference of a GFN2-xTB calculation in the gas phase and in an ideal conductor using a specially parametrized version of a computationally efficient COSMO implementation (ddCOSMO). $\Delta G_{\text{res.}}$ is calculated in a post-SCF approach using COSMO-RS theory (see below).
Because early versions of our model showed insufficient agreement with experimental data, we replaced part of the gas phase term introduced in the original version of COSMO-RS with another, more sophisticated term introduced by Marenich et al.,[30] which is called the Cavity-Dispersion-Solvent-Structure (CDS) Formalism (ΔG_{CDS}). Staying in the framework of splitting the solvation free energy into electrostatic and non-electrostatic parts, ΔG_{CDS} replaces the non-electrostatic contributions, which leads to the following equation for the solvation free energy

$$\Delta G_{\text{solv}} = \Delta G_{\text{IS}} + \Delta G_{\text{res}} + \Delta G_{\text{CDS}} + \Delta G_{\text{corr}}. \tag{C.5}$$

C.2.1 Ideal Solvation

The free energy contribution in the domain decomposition COSMO model[137] is defined as

$$\Delta G_{\text{IS}} = f(\varepsilon) \sum_{\text{A}} \sum_{\ell=0}^{L_{\text{max}}} \sum_{m=-\ell}^{\ell} [\Psi_{\text{A}}]_\ell^m [X_{\text{A}}]_\ell^m \tag{C.6}$$

where $f(\varepsilon)$ is the dielectric function of the dielectric constant ε, L_{max} is the maximum angular momentum for the spherical harmonics basis, $\vec{X}$ is the solution of the direct ddCOSMO equation (cf. Eq. C.8) and $\vec{\Psi}$ the charge distribution mapped on the spherical harmonic basis functions with angular momentum (ℓ, m) centered on atom A. For `xtb`, a point charge distribution is assumed,

leading to

$$[\Psi_{\mathrm{A}}]_{\ell}^{m} = \sqrt{\pi} q_{\mathrm{A}} \delta_{l0} \delta_{m0}. \tag{C.7}$$

with q_{A} being the atomic partial charge obtained by Mulliken population analysis in the self-consistent solution of the xTB Hamiltonian and δ_{ab} being the Kronecker delta. We obtain $\vec{X}$ by solving the direct ddCOSMO equation given by

$$\mathbf{L}\vec{X} = \vec{g} \tag{C.8}$$

with $\mathbf{L}$ being the block sparse matrix for the interaction in spherical harmonic basis (cf. eq. C.13) and $\mathbf{g}$ is the spherical harmonics expansion of the molecular potential, given as

$$[g_{\mathrm{A}}]_{\ell}^{m} = -\sum_{g}^{N_{\mathrm{g}}} w_g Y_l^m(\vec{y}_n) U_g^{\mathrm{A}} \Phi_g^{\mathrm{A}} \tag{C.9}$$

where we use the molecular potential $\vec{\Phi}$, the spherical harmonics Y_l^m, the number of angular integration points N_g, the point on the radial integration grid on a unit sphere $\vec{y}_g$, the integration weight of the angular grid w_g, and a switching function $\vec{U}$ to ensure smoothness when the number of buried grid points changes. The molecular potential is given as

$$\Phi_g^{\mathrm{A}} = \sum_{\mathrm{B}}^{N_{\mathrm{at}}} \frac{q_{\mathrm{B}}}{|\vec{R}_{\mathrm{A}} + r_{\mathrm{A}}\vec{y}_g - \vec{R}_{\mathrm{B}}|}, \tag{C.10}$$

with $\vec{R}_{\mathrm{A}}$ being the Cartesian coordinates of atom A and $\vec{r}_{\mathrm{A}}$ the van-der-Waals radius of atom A. For the switching function U is given by

$$U_g^{\mathrm{A}} = 1 - \sum_{\mathrm{B}} \frac{\min(1, f_g^{\mathrm{A}})}{f_g^{\mathrm{A}}} \cdot \chi_g^{\mathrm{AB}} \qquad \text{with} \quad f_g^{\mathrm{A}} = \sum_{\mathrm{B}} \chi_g^{\mathrm{AB}} \tag{C.11}$$

and the function χ_η is defined as

$$\chi_g^{\mathrm{AB}} = \begin{cases} 1 & \text{if } t_g^{\mathrm{AB}} \leq 1 - \eta \\ \eta^{-5}(1 - t_g^{\mathrm{AB}})^3 \left(6(t_g^{\mathrm{AB}})^2 + (15\eta - 12)t_g^{\mathrm{AB}} + 10\eta^2 - 15\eta + 6\right) & \text{if } 1 - \eta < t_g^{\mathrm{AB}} < 1 \\ 0 & \text{if } t_g^{\mathrm{AB}} \geq 1 \end{cases} \tag{C.12}$$

where η is a regularization parameter provided as adjustable numeric parameter and t_g^{AB} is a distance dependent metric (cf. eq. C.14). The ddCOSMO matrix is given by

$$[L_{\mathrm{AB}}]_{\ell\ell'}^{mm'} = -\sum_{g}^{N_g} w_g Y_\ell^m(\vec{y}_g) Y_{\ell'}^{m'}(\vec{s}_g^{\mathrm{AB}}) \cdot \omega_g^{AB} \frac{4\pi}{2\ell' + 1} (t_g^{\mathrm{AB}})^{\ell'}. \tag{C.13}$$

The distance metric t and s are given as

$$t_g^{\mathrm{AB}} = \frac{|\vec{R}_{\mathrm{A}} + r_{\mathrm{A}}\vec{y}_g - \vec{R}_{\mathrm{B}}|}{r_{\mathrm{B}}} \quad \text{and} \quad \vec{s}_g^{\mathrm{AB}} = \frac{\vec{R}_{\mathrm{A}} + r_{\mathrm{A}}\vec{y}_g - \vec{R}_{\mathrm{B}}}{|\vec{R}_{\mathrm{A}} + r_{\mathrm{A}}\vec{y}_g - \vec{R}_{\mathrm{B}}|}. \tag{C.14}$$

To obtain the potential contribution $\mathbf{F}^{\mathrm{IS}}$ of the ddCOSMO to the xTB Hamiltonian, we evaluate the derivative of the molecular potential with respect to the density matrix $\mathbf{P}$ as

$$F^{\mathrm{IS}}_{\lambda\kappa} = \sum_{\mathrm{A}}^{N_{\mathrm{at}}} \sum_{g}^{N_g} \zeta^{\mathrm{A}}_g \frac{\partial \Phi^{\mathrm{A}}_g}{\partial P_{\lambda\kappa}} \tag{C.15}$$

where we construct the potential contribution using with ζ being defined as

$$\zeta^{\mathrm{A}}_g = \sum_{l=0} \sum_{m=-l}^{l} w_g U^{\mathrm{A}}_g Y^m_l(\vec{y}_g)[S_{\mathrm{A}}]^m_\ell . \tag{C.16}$$

Here the $\mathbf{S}$ is obtained by solving the adjoint ddCOSMO equation given by

$$\mathbf{L}^* \vec{S} = \vec{\Psi} \tag{C.17}$$

This way, the ddCOSMO contribution to the electronic energy can be included self-consistently in the xTB Hamiltonian.

C.2.2 Realistic Solvation

The ddCOSMO model falls under the category of PCM-type models which assume that the solvent is a uniform dielectric continuum. This is a reasonable assumption for non-polar or alkane solvents. However, it may not hold true for solvents with permanent dipole moments (polar solvents) as the solvent can hardly be considered a homogeneous dielectric continuum due to the orientation and explicit interactions, such as hydrogen bonding. This, in turn, significantly impacts the interactions between solvent and solute, leading to changes in the solvation free energy. Klamt et al. introduced an improvement to the COSMO[136] solvation model known as the conductor-like screening model for real solvents (COSMO-RS).[28] COSMO-RS employs a reliable approach by initiating a DFT+COSMO calculation for a molecule in a perfect conductor ($\epsilon = \infty$), eliminating the requirement for empirical scaling of the dielectric screening energy. The original COSMO-RS theory remains unchanged for the chemical potential of the condensed phase, but it has been reparametrized for use with the semiempirical method GFN2-xTB. The following section provides an overview of the underlying theory, and for a more detailed explanation, we refer to the original publications by Klamt et al.[27,28,250,251]

In a perfect conductor, all molecules are ideally screened, i.e., the molecules do not interact electrostatically. The electrostatic energy gained in the complete screening process can be directly taken from the COSMO calculation and is depicted as ΔG_{IS} in equation C.5 and C.24. However, the molecules would still be frozen in their ideal gas state. Because there are no interactions, we could, in theory, rearrange the molecules to their condensed state without changing the system's energy[29].

To finally get a realistic description of the energy of the condensed phase, we need to remove this hypothetical ideal conductor by adding compensation charges to each molecular surface segment that are of opposite charge to their ideal screening ones. However, these new compensation charges will interact, adding an energy contribution to the system called the misfit energy $\Delta E_{\mathrm{misfit}}(\sigma, \sigma')$. It is

dependent on the two surface charges σ and σ' according to

$$E_{\text{misfit}}(\sigma, \sigma') = \frac{\alpha'}{2} \sum_{v}^{\text{seq.}} (\sigma_{v1} + \sigma'_{v2})^2, \tag{C.18}$$

where α' is dependent on an effective area of the surface segments and the response of the charges to the removal of the compensation patches and is treated as a free parameter. Because, in reality, the surface segments will not be frozen in position but move in a fluid state, we will not use the ideal screening charges σ_{id} in equation C.18, but averaged screening charges over a region of an averaging radius r_{av}. The averaging is done as

$$\sigma_v = \frac{\sum_{i}^{\text{seg.}} \sigma_i^* f(r_i, r_{\text{av}}, d_{\text{iv}})}{\sum_{i}^{\text{seg.}} f(r_i, r_{\text{av}}, d_{\text{iv}})}, \tag{C.19}$$

where $f(r_i, r_{\text{av}}, d_{\text{iv}})$ is a function, that depends on the distance between the segments i and v (d_{iv}), the averaging radius r_{av}, which is treated as a free parameter for the reparametrization of the method, and the mean radius of the segment (r_i).[28,29]

To calculate the misfit energy in equation C.18, the complete geometric information of an ensemble of condensed molecules would have to be known. This is generally not possible because, in a real solvent, the molecules are not frozen in time. Again, Klamt et al. introduced an approximation also employed here. Instead of considering an ensemble of molecules, an ensemble of surface patches is treated, and their corresponding ideal screening charges are normalized to a single molecule.

After normalizing these so-called sigma profiles to the area of the surface segments as

$$p'_S(\sigma) = \frac{p_S(\sigma)}{A^X}, \tag{C.20}$$

the chemical potential $\mu'_S(\sigma)$ of the a single additional patch σ in one mole of patches can be obtained as

$$\mu'_S(\sigma) = -kT \ln \int d\sigma' p'_S(\sigma') \exp\left(-\frac{E(\sigma, \sigma') - \mu_S(\sigma')}{kT}\right), \tag{C.21}$$

where $E(\sigma, \sigma')$ is the interaction energy between the patches with the screening charge density σ and σ'. This interaction energy is composed of the misfit energy from equation C.18, as well as an additional hydrogen bonding term, as

$$E_{\text{hb}}(\sigma, \sigma') = c_{\text{hb}} \max\left[0, \sigma_{\text{acc}} - \sigma_{\text{hb}}\right] \min\left[0, \sigma_{\text{don}} + \sigma_{\text{hb}}\right], \tag{C.22}$$

where c_{hb} and σ_{hb} are parameters that need to be fitted to the method and σ_{acc} and σ_{don} denotes the larger and the smaller values of σ and σ', respectively.

The chemical potential of a solute X in the solvent S can finally be obtained by integrating the σ-potential of the solvent weighted by the σ-profile of the solute as

$$\mu_S^X = \int d\sigma p^X(\sigma) \mu_S(\sigma). \tag{C.23}$$

We call this chemical potential μ_S^X the restoring free energy ΔG_{res} in equation C.5.

C.2.3 Gas Phase

The realistic description of the solvated phase may already be sufficient for pure solvent-dependent properties, like partition coefficients between various liquid phases. However, to calculate a solvation free energy ΔG_{solv}, additional information about the chemical potential of the solute molecule in the gas phase is needed. Following the COSMO-RS[28] approach we used an independent function for calculating the gas phase chemical potential as

$$\mu_{\text{gas}}^X = \Delta G_{\text{IS}} - \sum_k^{\text{atoms}} \gamma_k A_k^X - \omega n_r a^X - \eta RT, \tag{C.24}$$

where ΔG_{IS} is the ideal screening energy directly from the ideal conductor calculation with ddCOSMO (see above), $\omega n_r a^X$ is an effective ring atom correction, and ηRT is a thermodynamic correction based on the temperature. Here, $\gamma_k A_k^X$ represents the van-der-Waals (vdW) term, where γ_k is an element-based parameter and A_k^X is the exposed surface area of element k in molecule X. However, using this term in combination with GFN2-xTB did not yield satisfying results. For this reason, after extensive literature research, we decided to replace the vdW term with the enhanced cavity-dispersion-solvent (CDS) formalism, introduced by Marenich et al. in their SMD[30] model. The ΔG_{CDS} contribution is given by

$$\Delta G_{\text{CDS}} = \sum_k^{\text{atoms}} \gamma_k A_k^X + \gamma^{[M]} \sum_k^{\text{atoms}} A_k^X, \tag{C.25}$$

where γ_k is the surface tension of atom k in molecule X and $\gamma^{[M]}$ is the molecular surface tension of the solvent. Like in equation C.24, A_k is the exposed surface area, or solvent-accessible surface area (SASA), of the atom k. However, in contrast to the more straightforward approach in equation C.24, γ_k and $\gamma^{[M]}$ are not just simple parameters. Instead, γ_k depends on a parameter for a specific element γ_{Z_k} but also on parameters for neighboring elements $\gamma_{Z_k,Z_{k'}}$ and a geometry-dependent switching function T_k

$$\gamma_k = \gamma_{Z_k} + \sum_{k'}^{atoms} \gamma_{Z_k,Z_{k'}} T_k. \tag{C.26}$$

These parameters, as well as the molecular surface tension $\gamma^{[M]}$, are further dependent on solvent-specific descriptors, e.g., the refractive index of the solvent or the macroscopic surface tension at air/solvent interface at 298.15 K. More detailed information on this can be found in the respective publication.[30]

C.3 CPCM-X Training Set

While CPCM-X is based on a combination of existing implicit solvation models, the empirical parameters used must be refitted entirely to the underlying method (GFN2-xTB). Because obtaining sufficient experimental data for a robust fit proved to be difficult, we resorted to theoretically obtained reference data using standard COSMO-RS in the '19 parametrization. This is a reasonable approach,

as, due to the limited accuracy of the underlying GFN2-xTB method used to generate the COSMO files compared to DFT, we cannot expect the highest accuracy possible. Rather, we aim for a computationally efficient and robust approximation of the solvation free energy with no additional computational overhead.
The training set was established in the parametrization of several semiempirical methods, with the latest version created to develop a new density matrix tight binding scheme (PTB).[252] From this versatile set, only compounds containing the elements H,C,N,O,P,F,Cl and Br were used. Due to the questionable performance of implicit solvation models (and thus uncertainty in the reference data) for highly charged molecules, the scope of the set was reduced by removing compounds that were more than doubly charged.
The fit itself was performed using the Levenberg-Marquardt least-squares fitting algorithm. Higher-charged compounds tend to have much larger solvation free energies than neutral ones. Hence, to prevent a bias in the parametrization in favor of charged molecules, the deviations with respect to their reference data were Gaussian weighted.

C.4 Results

The results shown here were calculated with the CPCM-X module, implemented in the `xtb` code, which is freely available on GitHub. The module can be invoked by using the command line keyword `-cpcmx`, followed by the respective solvent name. `xtb` supports all solvents included in the Minnesota solvation database (MNSOL)[141] by default. Additional solvents can be included by defining their solvent properties. The GFN2-xTB/ALPB and GFN2-xTB/GBSA solvation free energies shown for comparison were also calculated with `xtb` using the difference in energy of a gas phase and a solvation phase single-point calculation. The COSMO-RS results were calculated using COSMO files obtained with TURBOMOLE 7.5.1[144] on a BP-86[69]/def2-TZVP level of theory and post-processed using the program `COSMOtherm` in the '19 parametrization. For evaluation, all solvation free energies are transferred to the Ben-Naim[169] standard state ($\frac{mol}{L}$), to match the experimental reference state used in the MNSOL database.

C.4.1 Hydration free energies

hydration free energies were evaluated on the FreeSolv[253] benchmark set and on the water subset of the MNSOL database. The water subset was divided into neutral and ionic compounds to better compare the method's accuracy with the implicit solvation models used as default in `xtb`. As geometry optimizations tend to have only a slight influence on solvation free energies for smaller, relatively rigid organic compounds[26], and to enable a direct comparison of the methods, all calculations were performed on GFN2-xTB optimized gas phase optimized geometries. A visual representation of the results for the FreeSolv Benchmark set can be found in figure C.1.
GFN2-xTB/GBSA and GFN2-xTB/ALPB already perform reasonably well on this benchmark set, with a MAD of 1.79 kcal/mol and 1.65 kcal/mol, and a SD of 2.34 kcal/mol and 2.14 kcal/mol, respectively. For comparison, the theoretical reference data (explicit solvation using a generalized Amber force field (GAFF)) that was published together with the benchmark set yields a MAD of 1.11 kcal/mol and an SD of 1.51 kcal/mol, while more sophisticated solvation models, like COSMO-RS and SMD, can yield even better results. GFN2-xTB/CPCM-X yields an excellent MAD of 1.38 kcal/mol and an SD

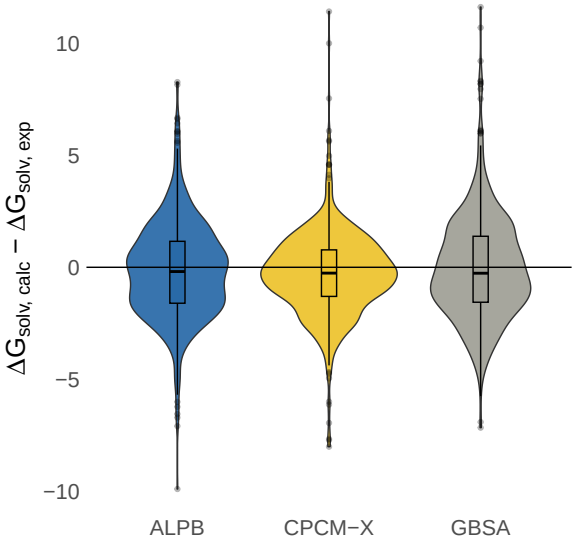


Figure C.1: The distribution of deviations from the reference hydration free energies, contained in the FreeSolv benchmark set, when compared with theoretical values obtained with ALPB, CPCM-X, and GBSA, in conjunction with GFN2-xTB.

of 1.88 kcal/mol, corresponding to an improvement of about 20 % in both statistics over the default solvation models used in the `xtb` code.

The results for the Minnesota solvation database can be found in Table C.1. For the neutral compounds, this trend in favor of CPCM-X is even more pronounced with MADs of 1.95, 1.88, and 1.46 for GFN2-xTB/GBSA, GFN2-xTB/ALPB, and GFN2-xTB/CPCM-X. However, for ionic compounds, its advantage becomes less apparent. GFN2-xTB/CPCM-X still performs best of the three methods in terms of MAD (6.93), while GFN2-XTB/ALPB yields a MAD of 7.40 kcal/mol and GFN2-xTB/GBSA yields an even higher MAD of 10.55 kcal/mol. However, the SD for CPCM-X is worse compared to the older models.

Splitting the ionic compound set into cations and anions, the reason for this unexpected result becomes apparent (Table C.2). CPCM-X systematically underestimates the solvation free energy for cations with an MD of 10.01 kcal/mol. The same trend is observable for GBSA, which seems less pronounced for ALPB with an MD of 4.23. The reason for this observation may be attributed to the choice of reference method for the parametrization of GBSA and CPCM-X. While these methods were purely fitted on theoretical data, the ALPB method was partially fitted on experimental reference data from the Minnesota solvation database, emphasizing the improvement of CPCM-X over ALPB for anionic and neutral compounds even more. Comparing results for the same dataset obtained with COSMO-RS in the '19 parametrization shows that the same systematic underestimation can be seen for cationic compounds (MD=9.75 kcal/mol). This may be partially attributed to varying solvation free energies of the proton utilized in the fitting procedure for COSMO-RS and the determination of the experimental data.[254] However, this could not be evaluated due to the undisclosed nature of this procedure.

Table C.1: MD, MAD and SD for the water subset of the Minnesota Solvation Database for the methods GBSA, ALPB and CPCM-X. The statistic is additionally calculated for a split subset depending on the total charge of the investigated molecules.

	GBSA	**ALPB**	**CPCM-X**
neutral compounds			
MD	0.05	-0.21	-0.24
MAD	1.95	1.88	1.46
SD	2.56	2.47	2.16
ionic compounds			
MD	10.46	6.77	4.50
MAD	10.55	7.40	6.93
SD	5.15	5.71	7.32
full subset			
MD	2.85	1.67	1.03
MAD	4.26	3.36	2.93
SD	5.76	4.77	4.71

Table C.2: MD, MAD and SD for the ionic water subset of the Minnesota Solvation Database for the methods GBSA, ALPB, CPCM-X and the reference method COSMO-RS. The statistic is splitted into anionic and cationic compounds.

	GBSA	**ALPB**	**CPCM-X**	**COSMO-RS**
anionic compounds				
MD	11.16	8.61	0.51	-4.12
MAD	11.32	9.39	4.68	6.18
SD	5.77	6.21	6.37	5.66
cationic compounds				
MD	9.49	4.23	10.01	9.75
MAD	9.49	4.65	10.03	9.75
SD	3.98	3.70	4.39	2.75

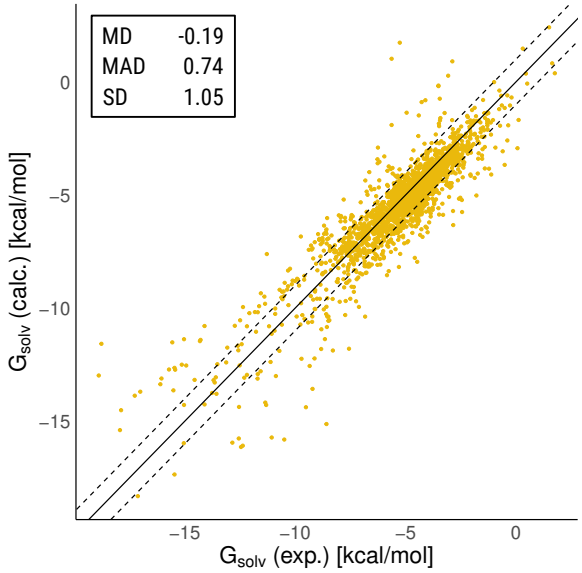


Figure C.2: Theoretical results calculated with GFN2-xTB/CPCM-X in respective to experimental data for the full non-ionic, non-aqueous MNSol Database.

C.4.2 Non-aqueos solvation free energies

As there are no special solvent parametrizations for CPCM-X (see above), it can be easily used for any solvent for which the needed solvent descriptors are available. This is true for all 91 solvents, that are present in the Minnesota database and for this reason, we can fully evaluate with reference to the experimental data given. The results are visualized in figure C.2. Because of the mentioned deviations of the reference method for the ionic cases in this database, we limit ourselves to neutral solvation free energies for this evaluation. Most of the results are in excellent agreement with experimental data with errors < 1 kcal/mol. With a MAD of 0.74 kcal/mol and a SD of 1.05 kcal/mol, CPCM-X performs better for the non-aqueous solvents than for the hydration free energies, which is expected, since more complicated noncovalent interactions, like hydrogens bonds, are less pronounced in non-polar solvents, than in water.

However, the ALPB and GBSA solvation models, which are routinely used with GFN2-xTB, need to be parametrized for each solvent individually and are therefore only available for a limited number of solvents, which makes their full evaluation compared to CPCM-X impossible. Excluding water, GBSA is available for eleven additional solvents, while ALPB is available for 22 additional solvents. As ALPB performed better than GBSA in all our previous tests, we limit the comparison to ALPB and the solvents contained in the MNSOL Database, for which ALPB was parametrized. The results can be found in table C.3. GFN2-xTB/ALPB reaches a MAD of 1.00 kcal/mol and a SD of 1.41 kcal/mol for this limited data set, which is a very good result for a semi-empirical method. However, with a MAD of 0.86 kcal/mol and a SD of 1.26 kcal/mol CPCM-X performs even better on this limited dataset, with an improvement that falls in line with previous results for the hydration free energies ($\approx 15\%$). The statistics are also summarized in table C.3.

We also calculated COSMO-RS solvation free energies for this limited data set to assess the to be expected accuracy. The results show an enhanced performance over the semiempirical solvation

Table C.3: MD, MAD, SD and error range for a total of 15 non-aqueous subsets of the Minnesota solvation database for the methods ALPB, CPCM-X, and the reference method COSMO-RS. Further information about the subsets can be found in the supporting information.

	ALPB GFN2-xTB	**CPCM-X** GFN2-xTB	**COSMO-RS** BP86/def2-TZVP
MD	-0.26	-0.14	0.02
MAD	1.00	0.86	0.52
SD	1.41	1.26	0.91
error range	16.2	13.7	13.0

models, probably due to the enhanced description of the electron density by the underlying DFT method. Nevertheless, these results validate the choice of this reference method for the parametrization of the CPCM-X method. However, the difference between GFN2-xTB/CPCM-X and DFT/COSMO-RS is rather small considering the significantly reduced computational cost of the former method.

C.4.3 Solvation free energies of larger systems

A critical aspect of the new CPCM-X solvation model is its usability with semiempirical methods regarding the enhanced computational cost of the ddCOSOMO compared to the underlying SCF procedure, especially for larger molecules. In order to test the computational performance also compared to a (usually much slower) DFT-based treatment on a statistically solid basis, we used the well-established S30L[255] supramolecular benchmark set. For this set, back-corrected solvation free energies, as well as theoretical solvation free energies calculated GFN2-xTB/GBSA, GFN2-xTB/ALPB, SMD (BP86/def2-SVP) and COSMO-RS (BP86/def2-TZVP) are available from previous works.[26,255] Because ALPB and GBSA are limited by the parametrized solvents, for some of the compounds, an alternative solvent was used in the original publication of the ALPB solvation model.[26] As CPCM-X does not have this restriction, we use the correct solvents for this method. In addition, we compare our results also to the DFT-based models to assess the loss of accuracy we suffer in the trade of enhanced computational efficiency. We restrict this investigation to non-aqueous compounds because of the previously identified irregularity for ionic hydration free energies (see above). The results are given in table C.4.

With a MAD of 3.56 kcal/mol, CPCM-X performs clearly better than ALPB or GBSA with a MAD of 5.04 kcal/mol and 5.25 kcal/mol, respectively. With a SD of 4.53 kcal/mol and an error range of 18.89 kcal/mol, it is also more robust in comparison to ALPB (SD = 6.51 kcal/mol, range = 25.37 kcal/mol) and GBSA (SD = 8.07 kcal/mol, range = 38.75 kcal/mol). As expected, SMD and COSMO-RS still show slightly better performance with MADs of 2.82 kcal/mol and 2.65 kcal/mol, respectively.

However, the better accuracy of COSMO-RS or SMD has to be set in relation to the increased performance of the new GFN2-xTB/CPCM-X method. To evaluate this, the wall computation time for calculating the whole data set was determined on an Intel(R) Core(TM) i7-10700K CPU on a single core. The results are displayed in figure C.3. As expected, the ALPB method is the fastest, with a wall time of 45 seconds, followed by the CPCM-X method, with a wall time of 196 seconds. The increased computation time can be attributed to the relatively costly ddCOSMO calculation. SMD and COSMO-RS are significantly slower than the semi-empirical solvation models with wall times of

Table C.4: Statistics for the S30L subset using the methods ALPB and CPCM-X with GFN2-xTB, as well as COSMO-RS (BP86/def2-TZVP) and SMD (BP86/def2-SVP).

	GBSA GFN2-xTB	**ALPB** GFN2-xTB	**CPCM-X** GFN2-xTB	**COSMO-RS** BP86/def2-TZVP	**SMD** BP86/def2-SVP
MD	0.52	-0.04	0.77	0.76	0.94
MAD	5.25	5.04	3.56	2.65	2.82
SD	8.07	6.51	4.53	3.22	3.66
Range	38.75	25.37	18.89	13.13	15.38

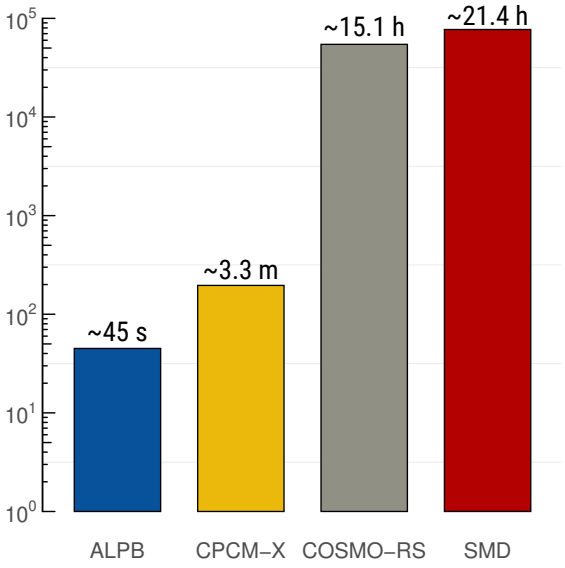


Figure C.3: Timings for the non-aqueous reaction solvation free energies ($\Delta_R G_{solv}$) of the S30L subset for the methods ALPB, CPCM-X, COSMO-RS, and SMD.

~ 21.4 hours and ~ 15.1 hours, respectively, which is due to the higher level DFT treatment, that is used for the COSMO/CPCM solvation model. This translates to an increase of more than two orders of magnitude (factor ~ 280) in computational efficiency for CPCM-X (GFN2-xTB) in comparison to COSMO-RS (BP86/def2-TZVP).

C.5 Conclusion

In this work, we introduced the extended conductor-like polarizable continuum model (CPCM-X) in combination with the semiempirical method GFN2-xTB for the computation of solvation free energies of molecules. This model consistently outperforms the default solvation models ALPB and GBSA for a wide range of compounds. However, we did observe a slight systematic overestimation of cationic hydration free energies, which may be attributed to the parametrization with the reference method COSMO-RS. Nevertheless, CPCM-X has shown a significant improvement for neutral compounds. This is because CPCM-X is designed to accurately reproduce the fine structure of solvents in the near

field of the solute molecule, which is more important for neutral compounds than for ionic ones where the electrostatic interactions also affect distant solvent molecules.
Additionally, our model has demonstrated increased performance and robustness for larger supramolecular complexes of the S30L benchmark set. We observed improvements of about 30% in terms of mean absolute deviation and up to 40% in terms of standard deviation and error range. While more sophisticated DFT-based solvation models may be even more accurate, CPCM-X is computationally more efficient, with computation times two orders of magnitude faster than the reference method.
Furthermore, our CPCM-X model and its implementation into the `xtb` program suite is open-source and not tied to any specific theory level or software package. This allows it to be refit for use with other methods and packages, providing a consistent solvation treatment over multiple levels of theory. This, in turn, provides a computationally efficient multi-level approach to calculate solvation-dependent properties.
In summary, our implementation of the CPCM-X model in `xtb` provides a significant improvement in the description of solvation at the 'low-level' side of a quantum chemistry method hierarchy. We are optimistic that this model will be a valuable addition to the GFN family, particularly for cases where an accurate solvation description is necessary.

Author Contributions

MS: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing - original draft, visualization, and software; SE: software and writing - original draft; SG: resources, writing - review & editing, supervision, project administration, and funding acquisition.

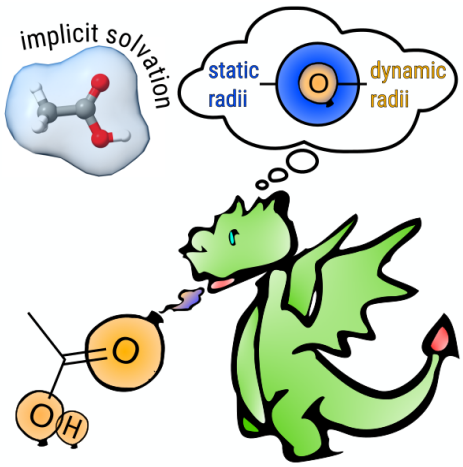
C.6 Supporting Information

The following files are available free of charge (DOI: https://doi.org/10.1021/acs.jpca.3c04382).

- Excel spreadsheet containing the statistical evaluations in kcal/mol.

APPENDIX D

Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)



Abstract. We present the Dynamic Radii Adjustment for COntinuum solvation (DRACO) approach, which employs precomputed atomic partial charges and coordination numbers of the solute atoms to improve the solute cavity. As such, DRACO is compatible with major solvent models, improving their performance significantly and robustly at virtually no extra cost, especially for charged solutes. Combined with the purely electrostatic CPCM and COSMO models, DRACO reduces the mean absolute deviation (MAD) of the solvation free energy by up to 4.5 kcal/mol (67 %) for a large dataset of polar and ionic solutes. Even in combination with the highly empirical universal solvent model (SMD), DRACO substantially reduces the MAD for charged solutes by up to 1.5 kcal/mol (39 %), while neutral solutes are slightly improved (0.2 kcal/mol or 16 %). We present an interface of DRACO with two computationally efficient atomic charge models that enables fully automated, out-of-the-box calculations with the widely used program packages ORCA and TURBOMOLE.

D.1 Introduction

A variety of complex chemical problems that could once only be studied through costly and lengthy laboratory experiments can nowadays be addressed with computational studies that use advanced quantum mechanical (QM) methods like density functional theory (DFT) at a much-reduced effort. However, without further modification, standard QM methods are often limited to the calculation of molecules in the gas phase at a temperature of 0 K. To be able to describe chemical processes at typical reaction conditions, additional factors such as thermodynamic effects at finite temperature, or the influence of a solvent have to be included.

Incorporation of solvent effects into computational studies has been facilitated by the development of various solvation models.[109,256] This also gave rise to the class of implicit solvation models, which are frequently used because they are fast and easily accessible via various QM program packages. Prominent methods in this regard are polarizable continuum models (PCMs)[99] such as the conductor-like screening model (COSMO)[136] or the conductor-like polarizable continuum model (CPCM)[248]. They are based on continuum electrostatics[21] and, as such, treat the solvent as a uniform dielectric continuum that primarily depends on the dielectric constant (ε). This approximation is generally appropriate for solvents that are only weakly interacting with the solute, but, in addition to the complete lack of non-electrostatic interactions, it introduces even larger errors for systems with very strong electrostatic solute–solvent interactions causing a significant solvent response to the solute.[251] To overcome this limitation, various solvation models evolved that introduce additional correction terms, such as the conductor-like screening model for real solvents (COSMO-RS)[27,28] or the universal solvation model based on solute electron density (SMD).[30] Nevertheless, all of these PCM models are based on the same principle: they endeavor to solve the generalized Poisson equation by reformulating it into an apparent surface charge method based on a three-dimensional solute cavity.[98] Its construction is most prominently done by overlapping atom spheres that are based on fixed atom radii. As the cavity of the solute greatly influences the performance of the models,[257] determining suitable radii is rather important, which, however, proved difficult.[33] For example, experimentally determined radii[34,35] can produce unsatisfactory results,[257] because they are often derived from crystal structures[34] and are expected to differ from effective solution values.

Consequently, the applied radii are commonly regarded as empirical parameters and are fine-tuned using experimental reference data, such as vapor pressures or solvation free energies.[258] In fundamental solvation models like CPCM and COSMO, these radii are statically allocated per element, irrespective of the specific solvent or solute, despite the fundamental variability of the cavity radius being an inherent aspect of solute-solvent interactions.[36–39] This static assignment renders them incapable of adapting to the electronic structure of the solute, potentially limiting the maximum achievable accuracy.[259,260]

Previous efforts to include information about the system for the determination of the solute cavity have been made. However, these attempts are mostly focusing on water as solvent and are either limited to certain atoms or functional groups,[40] or are developed for specific applications with a narrow range of training data.[41,42] Other, more sophisticated approaches based on an isocontour of the electron density introduce a significant additional computational effort and typically lack analytical energy gradients.[43,44]

Here, we present a generally applicable and efficient approach for determining system-specific atom radii for solute cavity construction based on their molecular environment. To do so, we employ atomic partial charges computed with the highly efficient charge models EEQ[92,139] and CEH[140] as

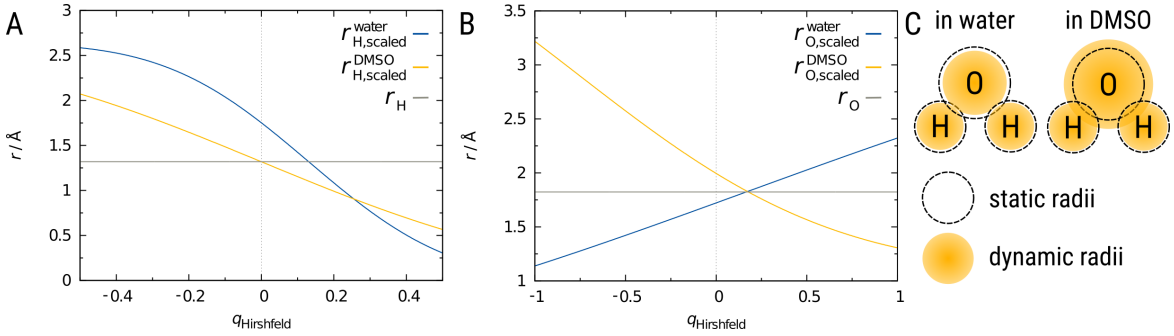


Figure D.1: A, B: Comparison of the scaled radii $r_{i,\text{scaled}}$ and default radii r_i in the CPCM model for hydrogen and oxygen using water and DMSO as a solvent. The varying electron density is modeled by atomic partial charges using the Hirshfeld analysis. C: Visualized static and dynamic radii for the water molecule.

well as (fractional) atom coordination numbers, which allow a fast on-the-fly solute evaluation and radii computation. While in principle possible for all elements, DRACO presented in this work is limited to elements that are well represented in the training data. Analytical energy gradients for the dynamic radii are also provided, if the derivatives of the partial charges are available, which is currently the case for *e.g.,* the EEQ model. Both charge models were tested against Hirshfeld charges derived from DFT and work robustly for all elements up to $Z = 86$ (radon). This so-called Dynamic Radii Adjustment for COntinuum solvation (DRACO) approach is implemented in a freely available open-source program on GitHub,[142] including readily available parameters for CPCM, SMD, and COSMO. To ensure easy applicability, only a single command line call is needed before performing the actual DFT calculation in order to add the new radii to the input of the supported QC program (currently ORCA[143] and TURBOMOLE[144]).

D.2 Theory and Implementation

DRACO relies on a dynamic scaling of the original static radii r_i[28,30,261] for atom i of the respective implicit solvation model used for electrostatic evaluation. The radii are dynamically scaled according to

$$r_{i,\text{scaled}} = f_{i,\text{scale}} \cdot r_i \tag{D.1}$$

with the scaling factor determined by the effective partial charges $q_{\text{eff},i}$ as

$$f_{i,\text{scale}} = \text{erf}(a_Z \cdot (q_{\text{eff},i} - b_Z)) + 1, \tag{D.2}$$

where, a_Z, and b_Z are element-specific parameters. q_{eff} is defined as

$$q_{\text{eff},i} = q_i + k_Z \cdot q_i \cdot CN_i, \tag{D.3}$$

where q_i is the atomic partial charge and k_Z an additional parameter. CN_i is the corresponding fractional coordination number and whose construction is adapted from the D3 model[93]. The error function in Equation D.2 describes a continuous increase/decrease of the scaled radii depending on the partial charges and CNs with plateaus at large absolute values. This prevents artificially large or small radii. Thereby, a_Z defines the sensitivity of an atomic radius towards the change in the

environment while the parameter b_Z adapts the scaled radii relative to the original values. The cross term and its corresponding parameter k_Z in Equation D.3 allows to differentiate between various bonding motifs. The parameters were determined through least-squares optimization of computed solvation free energies against experimental reference data from the Minnesota solvation (MNSOL) database that have an average uncertainty of 0.2 kcal/mol for neutral and 3 kcal/mol for ionic solutes for the single experiment.[141] However, due to the large number of data points, the uncertainty for the whole data set is expected to be lower. Solvation free energies are based on the r^{2}SCAN-3c composite DFT method[126] in conjunction with the CPCM, COSMO, and SMD solvation models. Similar to the procedure used for the creation of the SMD model,[30] we optimized a separate set of parameters for water as a solvent and a combined parameter set for all other solvents. The latter includes acetonitrile, dimethylsulfoxide (DMSO), and methanol for which solvation free energies of ionic solutes are included in the database. For oxygen atoms, following the same procedure, we introduced an additional correction for solvents using Abraham's hydrogen bond acidity (α) based on a parameter c_O according to

$$f^O_{i,\text{scale}} = f_{i,\text{scale}} + c_o \cdot (0.43 - \alpha). \tag{D.4}$$

This correction yields larger oxygen radii for less hydrogen bond donating solvents and only applies to solvents that have an α value lower than 0.43 like DMSO or acetonitrile.
In principle, there are two main effects that are described by DRACO: On the one hand, a varying electron density should directly be reflected in the atomic radii, *i.e.*, less electron density results smaller effective radii. On the other hand, the local electron density influences the solute–solvent interaction and thus the equilibrium solute–solvent distance, which should depend on the polarity of the solvent. These two effects do not necessarily align and can partially cancel each other. Their influence on the radii used for the construction of the solute cavity is represented by the DRACO scaling function, which is exemplified for H and O in Figure D.1. The variation in electron density is modeled by atomic partial charges. Thereby, oxygen is representative for other electron donors, like nitrogen and fluorine, that show similar behavior. For hydrogen, the above-described effects align as the attraction to a polar solvent increases with less local electron density and a more positive atomic partial charge. In contrast, as the electron density increases for oxygen, the expected enlargement of the atomic radius is counteracted by the enhanced attraction to a polar solvent. This leads to opposing trends depending on the polarity of the solvent as seen for water and DMSO in Figure D.1.

D.3 Results and Discussion

The demonstrated effect of DRACO does not only impact static structure evaluation, *e.g.*, equilibrium structures but is also visible for dynamic processes like bond dissociation. As an example, the radii change during the heterolytic bond cleavage of HCl in water is depicted in Figure D.2. With increasing bond distance, both atoms become more strongly charged (positive in the case of hydrogen and negative in the case of chlorine), and the fractional coordination numbers decrease. As a result, the respective radii decrease towards ion formation, which matches the expected picture of stronger hydrogen bonding interaction between the solvent water and ions than for neutral molecules. The resulting potential energy curves are given in the SI.
For a more complex example, we demonstrate the impact of DRACO on the neutral-to-zwitterionic equilibrium of the amino acid glycine in water, specifically the equilibrium constant (K_D) of the

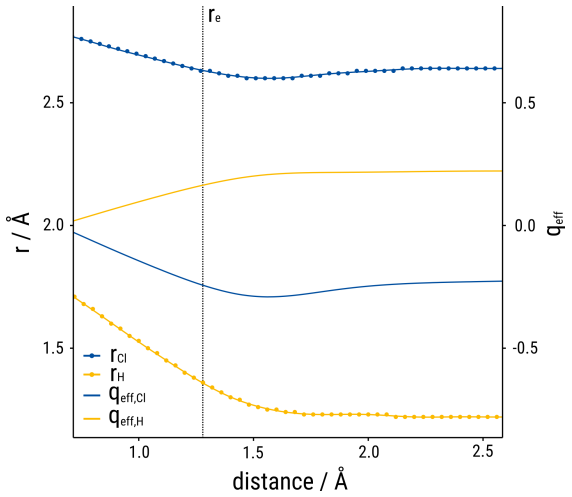


Figure D.2: Dynamic radii (r) and effective atomic partial charge (q_{eff}) for the heterolytic bond cleavage of HCl in water. The equilibrium distance (r_e) is 1.28 Å. For energies, see SI.

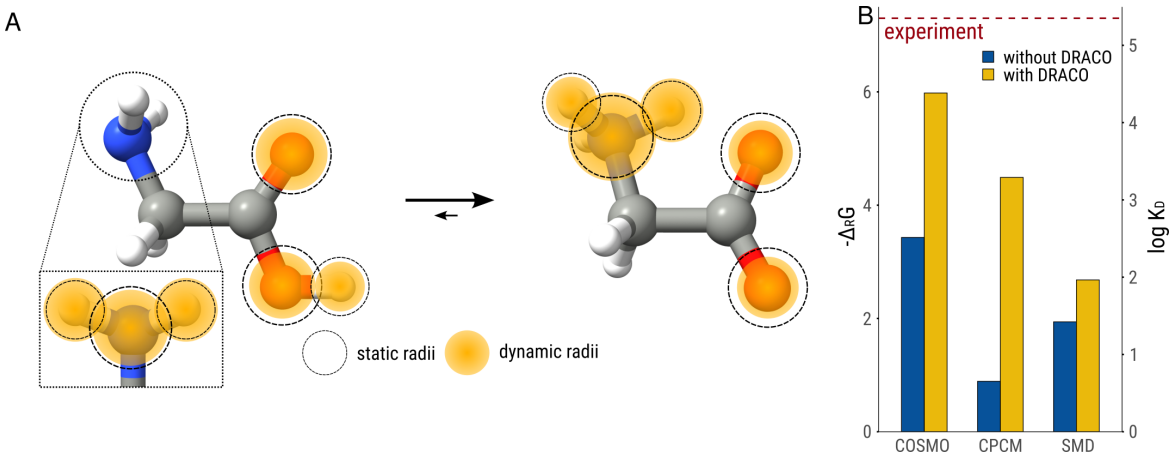


Figure D.3: A: Neutral-zwitterionic equilibrium reaction between glycine in its normal state and as an inner salt with static radii used by CPCM as a default and DRACO radii. For clarity, only radii with a significant dynamic change during the tautomeric reaction are shown. B: Calculated reaction free energies ($\Delta_R G$) in kcal/moland equilibrium constants (K_D) with and without DRACO.

reaction. The DRACO radii obtained with Hirshfeld charges from DFT calculations are visualized in Figure D.3A. The experimental K_D of this reaction is $2.4 \cdot 10^5$ at room temperature,[262] corresponding to a standard Gibbs free energy of reaction ($\Delta_r G$) of −7.3 kcal/mol. For the computation of $\Delta_r G$, we follow a multi-level workflow[14], which combines high-level DFT for electronic energies, with efficient DFT for solvation and thermostatistical contributions (see Computational Methods section). Since the reaction occurs in aqueous solution, the solvation free energy ΔG_{solv} for each tautomer must be considered. $\Delta_r G$ is then calculated by the difference of all contributions as

$$\Delta_r G = \Delta E_{gas} + \Delta G_T + \Delta\Delta G_{solv}. \tag{D.5}$$

where $\Delta\Delta G_{solv}$ is the reaction solvation free energy. CPCM with standard static radii provides a $\Delta_r G$ of −0.9 kcal/mol, which translates to an equilibrium constant of about 4.5, i.e., five orders of magnitude below the experimental value. This means that only about 82 % would be in zwitterionic form, which is qualitatively wrong compared to the experiment (99.999 % zwitterion). In stark contrast, CPCM with DRACO radii based on Hirshfeld charges provides $K_D = 2.0 \cdot 10^3$ ($\Delta_r G = -4.5$ kcal/mol or 99.95 % zwitterionic form, Figure D.3B), which is in much better agreement with the experiment. Similarly, DRACO radii also improve the $\Delta_r G$ values of other solvation models like SMD (−1.9 to −2.7 kcal/mol) and COSMO (−3.4 to −6.0 kcal/mol).

This non-trivial example illustrates that DRACO is physically sound in combination with DFT-based Hirshfeld charges. Nevertheless, obtaining charges on such high levels of theory is computationally costly. Accordingly, the applied method can be more expensive than the solvation calculation itself, especially if DRACO is combined with efficient semi-empirical solvation models such as CPCM-X[135]. Therefore, it is desirable to use a computationally more efficient charge model. Two promising candidates for this task are the EEQ[92,139] and CEH[140] charge models. To evaluate their usability, we calculated ωB97M-V/def2-TZVPPD[167,263,264] Hirshfeld charges for the molecules contained in the water subset of the MNSOL database and compared them with EEQ and CEH partial charges. The EEQ model already shows a reasonable agreement with the DFT charges, with a standard deviation (SD) of about 0.12 electrons, while the CEH model performs even better with an SD of only 0.04 electrons. The overall superior description of atomic charges by the CEH model is illustrated by the distribution of the Pearson correlation coefficients per molecule that accumulate in the region of 0.9-1.0 for CEH (cf. SI, Figure S2). Regarding computational timings, the EEQ model is typically a factor of about 50 times faster than CEH (0.05 s vs. 2.5 s for a small protein with about 900 atoms at one Intel® Core™ i7-10700K CPU). However, both offer a speed-up on the order of $10^3 - 10^4$ compared to a DFT/Hirshfeld calculation and thus add only a negligible overhead to a typical DFT calculation, which makes both models (EEQ and CEH) suitable for DRACO.

First, the parameters a_Z, b_Z, c_Z, and k_Z according to Equations D.2 and D.4 were determined by a least-squares optimization against the MNSOL database for both charge models and the atoms sufficiently represented by the MNSOL database (H, C, N, O, F, S, Cl, and Br). For atoms of other elements, default radii of the respective model are used. The evaluation of DRACO on this training data is shown in Figure D.4 divided into the water set and the combined set consisting of the solvents acetonitrile, DMSO, and methanol.

DRACO generally improves the description of the solvation free energies compared to those obtained with unscaled radii. Both the mean absolute deviations (MADs) and standard deviations (SDs) decrease significantly in all cases, which indicates that DRACO does not merely introduce a shift but generally improves the physical description. The charged solutes are more strongly affected by the adjusted radii due to their high atomic partial charges. However, also neutral species are improved. In case of water, the MADs are reduced from 1.1 kcal/mol to 0.9 kcal/mol for SMD, 2.1 kcal/mol to 1.6 kcal/mol for CPCM, and 2.3 kcal/mol to 1.8 kcal/mol for COSMO. The choice of the charge model has a relatively minor effect (0.1 − 0.2 kcal/mol in terms of MAD), as the residual errors introduced by either CEH or EEQ are small enough to be compensated upon parameter optimization. Looking at the individual models, COSMO and CPCM behave similarly, as they differ only in details and the specific implementation. Their MADs with default radii for water are 3.4 and 3.5 kcal/mol and for the other solvents 7.5 and 7.2 kcal/mol. With DRACO, MADs are reduced to about 2 kcal/mol for water and 2.5 − 2.7 kcal/mol for the other solvents. This corresponds to an improvement of 44 % for water and 67 % for the other solvents.

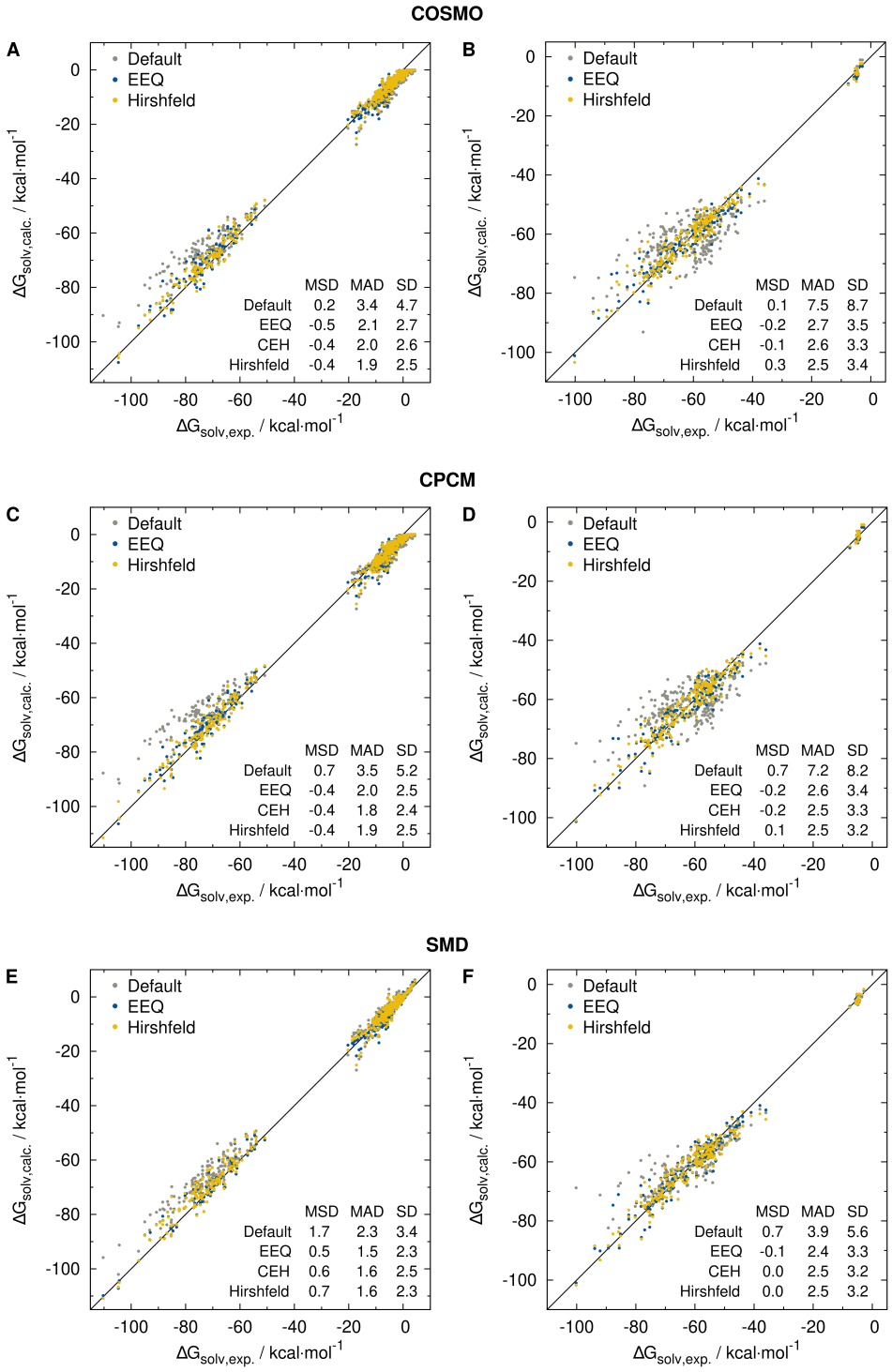


Figure D.4: Results for COSMO, CPCM, and SMD on the MNSOL database. Shown are ΔG_{solv} for the solvent water (A, C, E) and the combined acetonitrile, DMSO, and methanol sets (B, D, F). MSD = mean signed deviation.

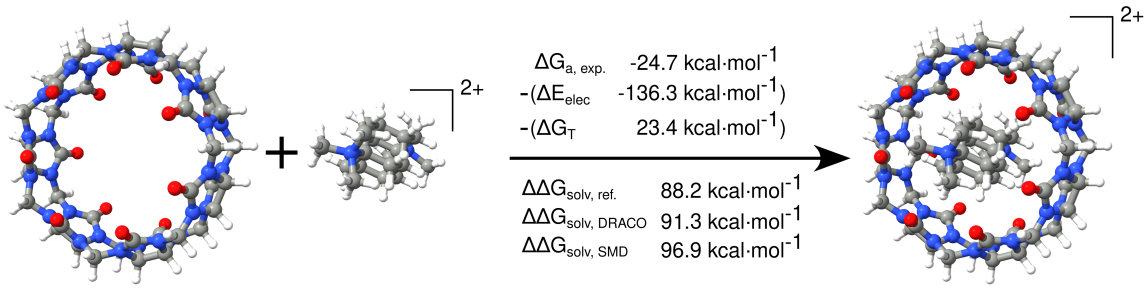


Figure D.5: Host-guest complex $Ad_2(NMe_3)_2$@CB7 with the experimental association free energy, electronic as well as thermostatistical contributions, taken from the S30L Benchmark set[255]. The association solvation free energy ($\Delta\Delta G_{solv}$) is given for SMD with and without DRACO.

For SMD, DRACO consistently improves the MAD from 2.3 kcal/mol to 1.5 kcal/mol for water and from 3.9 kcal/mol to 2.5 kcal/mol for the other solvents, corresponding to about 26-39 % reduction (depending on the solvent and charge model). This improvement is especially remarkable since SMD was originally optimized on the MNSOL database that is used here. This further corroborates that DRACO improves the physical description of the solute-solvent interaction. Moreover, it suggests that further improvement can be achieved by re-parameterizing SMD with DRACO included.
To evaluate DRACO beyond the training sets, we tested the C10 set[265] containing solvation free energies of 10 charged organic molecules in water. To further verify its robustness, we investigated the 227 neutral organic molecules of the Freesolv database[253]. The results are given in Table D.1 Here, DRACO increases the accuracy of all tested implicit solvation models significantly, also for systems beyond the training set (Table D.2). Especially for the charged solutes of the C10 set, the SDs

Table D.1: Statistical evaluation of the DRACO approach for the C10 dataset and the Freesolv systems not contained in the Minnesota database calculated with COSMO, CPCM, and SMD using the r^2SCAN-3c composite DFT method. Values are given in kcal/mol.

	C10						**Freesolv**					
	COSMO		**CPCM**		**SMD**		**COSMO**		**CPCM**		**SMD**	
	MAD	**SD**	**MAD**	**SD**	**MAD**	**SD**	**MAD**	**SD**	**MAD**	**SD**	**MAD**	**SD**
Default	5.0	3.7	5.7	4.9	3.5	2.6	2.4	1.6	2.1	1.7	0.8	1.2
EEQ	1.7	1.3	1.4	1.0	1.3	1.5	1.8	1.6	1.6	1.6	0.8	1.4
CEH	1.3	1.4	1.1	1.2	1.2	1.5	1.7	1.5	1.6	1.6	0.8	1.3

and MADs are drastically improved. The latter are even reduced by more than 50 %, approaching chemical accuracy (MAD< 1 kcal/mol). Considering the neutral species contained in the Freesolv set, the already accurate result of SMD, which is almost within chemical accuracy (MAD 0.8 kcal/mol), is kept with DRACO, whereas the MADs for COSMO and CPCM are improved.
For additional validation on a more application-oriented set containing larger molecules, we used back-corrected solvation free energy differences for association reactions ($\Delta\Delta G_{solv}$) of host-guest complexes of the S30L benchmark set.[255]) These reactions are especially challenging because the solvation contributions are large, particularly for the charged systems (up to about 90 kcal/mol, see

Table D.2: Statistical evaluation of the DRACO approach for back-corrected experimental association solvation free energies ($\Delta\Delta G_{solv}$) of the S30L dataset calculated with SMD using the r^{2}SCAN-3c composite method. Values are given in kcal/mol.

	Total	Ionic	Neutral
		MAD	
Default	3.7	4.8	3.3
EEQ	3.2	3.7	3.1
CEH	3.5	4.0	3.3
		SD	
Default	4.7	5.8	4.2
EEQ	4.3	4.0	4.0
CEH	4.4	4.3	4.0

Figure D.5 for an example). Because the non-electrostatic contributions to the solvation are also rather large here, we limit our test to SMD, which, in contrast to COSMO and CPCM, includes the corresponding terms.
The results are depicted in Table D.2. For the whole S30L set containing charged and neutral species, the MAD in $\Delta\Delta G_{solv}$ for SMD with default static radii is 3.7 kcal/mol , which is reduced significantly by 0.5 kcal/mol (about 14 %) with DRACO radii using the EEQ model. As noted before, the description improves more for the ionic systems than for the neutral ones (0.8 – 1.1 vs. 0.0 – 0.2 kcal/mol in MAD). Surprisingly, the more sophisticated quantum mechanical CEH charge model performs slightly worse for these larger and more complicated systems than the EEQ model. We tentatively attribute this to fortuitous error compensation with the empirical SMD model, which suggests that re-fitting SMD with DRACO(CEH) radii might yield even better results.

D.4 Summary

We presented DRACO: a general, robust, and efficient approach to improve implicit solvation models by including molecular information in the atomic radii determination for the construction of the solute cavity. DRACO employs fractional coordination numbers and atomic partial charges, the latter of which are obtained with the efficient semi-empirical charge models EEQ or CEH.
With DRACO, we significantly reduce the errors of the CPCM, COSMO, and SMD models. The biggest impact of DRACO was observed for charged solutes, but also neutral systems were improved considerably, even when DRACO is combined with empirical models like SMD. DRACO itself introduces only three physically motivated additional parameters per element.
The most important physical insight obtained in the course of this study is that the effective atomic radii dependence on the atomic partial charge (electron density) differs qualitatively between elements. Counteracting effects are radii increase with smaller (more negative) charge and corresponding decrease due to shorter average solvent-solute interatomic distances.
We would like to highlight that DRACO represents the third instance of successful implementation of the atoms-in-molecules approach developed in our group, which makes use of rapid semi-empirical models to derive atom-specific properties from molecules prior to the actual (DFT or semi-empirical) calculation. These properties are then applied to refine previously static (element-specific) parameters.

The foremost example of this idea is the D4 dispersion correction, where atomic charges and coordination numbers modulate precomputed C_6 coefficients.[91,92] It was followed by the development of adaptive "breathing" minimal basis sets, of which q-vSZP has been published recently.[140] Finally, this work extends this concept to the adjustment of atomic radii in implicit solvation models.

The DRACO program developed in the course of this work is made available free of charge as an open-source program on GitHub.[142] It comes with an interface to the computationally efficient EEQ/CEH charge model implementations to evaluate the chemical environment of the solute atoms at almost no computational overhead compared to the required DFT calculation. It is currently interfaced to the program packages ORCA and TurboMole but could, in principle, be used with any quantum chemistry program that allows custom radii for the cavity construction. Further, it is in principle easily extendable to any radii-based solvation model and any solvent.

D.5 Computational Methods

Dynamic radii were calculated using the DRACO program available on GitHub.[142] Hirshfeld charges, as well as energies for the glycine reaction, were computed at the ωB97M-V/def2-TZVPPD level of theory. Geometries, as well as frequencies, for the tautomeric equilibrium of glycine were computed with r^{2}SCAN-3c and the default CPCM solvation model using water as a solvent. SMD and CPCM solvation free energies, as well as high-level gas-phase energies, geometry optimizations and Hirshfeld charges were obtained using the Orca program package[143] (version 5.0.4). The default radii used for CPCM in the ORCA program package are obtained using a scheme proposed by Lange & Herbert[261]. COSMO solvation free energies are obtained using TurboMole[144] 7.6 using the default out-lying charge correction[266]. Except for the scaled radii for solvation, the default settings were employed in both program packages. If not stated otherwise, original geometries from the benchmark sets were used.

Author Contribution

CP: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing - original draft; MS: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing - original draft; MB: conceptualization, writing - review & editing; JMM: conceptualization, writing - review & editing; SG: conceptualization, methodology, resources, writing - review & editing, supervision, project administration, funding acquisition.

Acknowledgements

We are grateful to Thomas Gasevic for insightful conversations regarding implicit solvation and general chemistry. This work was financially supported by the Merck KGaA. S.G. and M.B. gratefully acknowledge financial support by the Max-Planck-Society.

D.6 Supporting Information

The following files are available free of charge.

- Excel spreadsheet containing the statistical evaluations in kcal/mol.
- PDF document containing electronic supporting information, also adapted in this section.

The full Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c03551

Statistical error measures

In this work, the following statistical measures were used.
Statistical measure for a set $x_1, \cdots, x_n$ of data points with references $r_1, \cdots, r_n$ are:

- Mean deviation (MSD):

$$MD = \frac{1}{n}\sum_{i}^{n}(x_i - r_i) \tag{D.6}$$

- Mean absolute deviation (MAD):

$$MAD = \frac{1}{n}\sum_{i}^{n}\left|x_i - r_i\right| \tag{D.7}$$

- Standard deviation (SD):

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\left(x_i - r_i\right) - MD\right)^2} \tag{D.8}$$

- Pearson correlation coefficient (ρ_p)

$$\rho_p = \frac{\sum_i^n (x_i - \overline{x})(r_i - \overline{r})}{\sqrt{\sum_i^n (x_i - \overline{x})^2 (r_i - \overline{r})^2}} \tag{D.9}$$

Supporting Figures

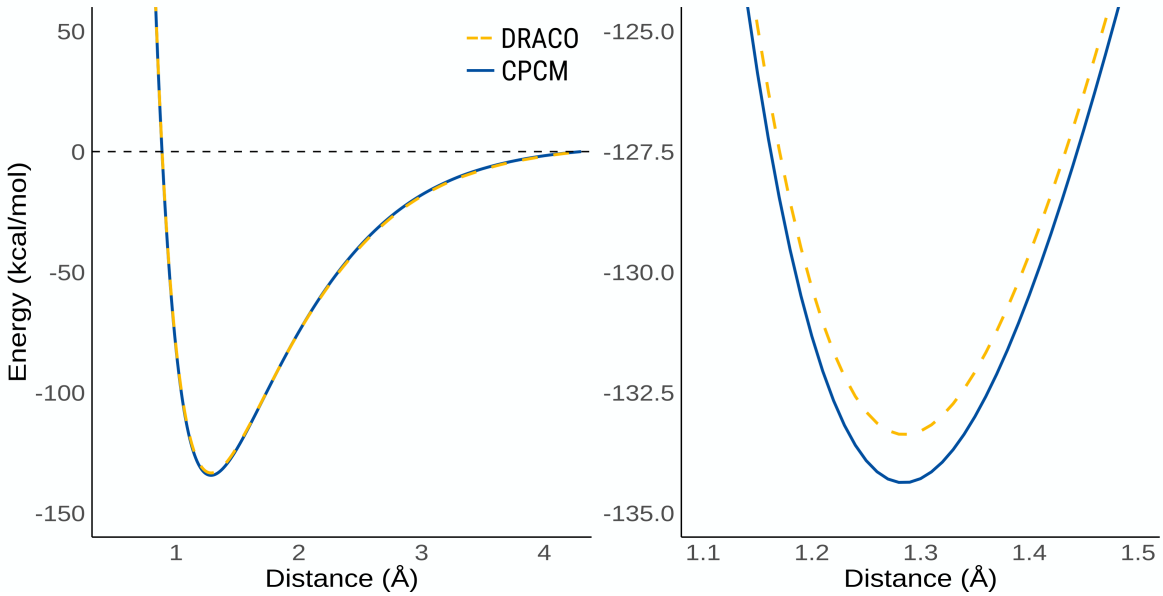


Figure D.6: Total energies for the heterolytic bond cleavage of HCl in water in respect to the dissociation limit in kcal/mol.

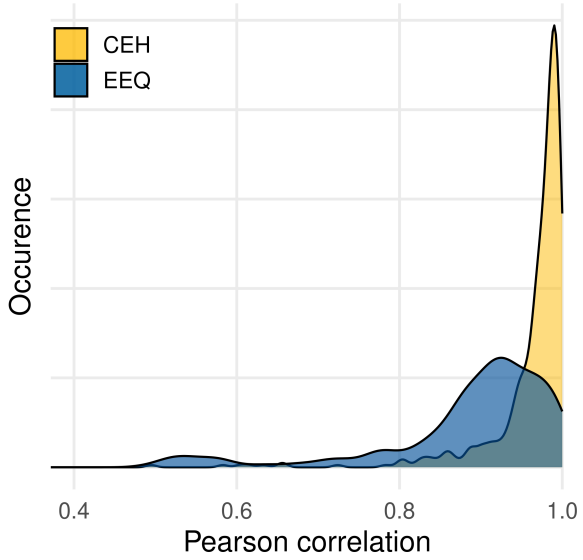


Figure D.7: Density plot of Pearson correlation coefficients per molecule between EEQ/CEH charges and Hirshfeld charges obtained at the ωB97M-V/def2-TZVPPD level of theory for all molecules in the water subset of the MNSOL database.

Bibliography

[1] J. Hughes, S. Rees, S. Kalindjian and K. Philpott, *Principles of early drug discovery*, British Journal of Pharmacology **162** (2011) 1239–1249.

[2] A. Klamt, F. Eckert, M. Hornig, M. E. Beck and T. Bürger, *Prediction of aqueous solubility of drugs and pesticides with COSMO-RS*, Journal of Computational Chemistry **23** (2002) 275–281.

[3] D. Van Der Spoel, S. Manzetti, H. Zhang and A. Klamt, *Prediction of Partition Coefficients of Environmental Toxins Using Computational Chemistry Methods*, ACS Omega **4** (2019) 13772–13781.

[4] P. Pracht and S. Grimme, *Efficient Quantum-Chemical Calculations of Acid Dissociation Constants from Free-Energy Relationships*, The Journal of Physical Chemistry A **125** (2021) 5681–5692.

[5] T. Salthammer, S. Grimme, M. Stahn, U. Hohm and W.-U. Palm, *Quantum Chemical Calculation and Evaluation of Partition Coefficients for Classical and Emerging Environmentally Relevant Organic Compounds*, Environmental Science & Technology **56** (2021) 379–391.

[6] C. J. Weschler, T. Salthammer and H. Fromme, *Partitioning of phthalates among the gas phase, airborne particles and settled dust in indoor environments*, Atmospheric Environment **42** (2008) 1449–1460.

[7] H. Yang, G. Lu, Z. Yan, J. Liu, H. Dong, X. Bao, X. Zhang and Y. Sun, *Residues, bioaccumulation, and trophic transfer of pharmaceuticals and personal care products in highly urbanized rivers affected by water diversion*, Journal of Hazardous Materials **391** (2020) 122245.

[8] A. Cammarata and K. S. Rogers, "The Interpretation of Drug Action through Linear Free Energy Relationships", *Advances in Linear Free Energy Relationships*, ed. by N. B. Chapman and J. Shorter, Boston, MA: Springer US, 1972 401–444, ISBN: 978-1-4615-8662-3.

[9] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp and P. C. Biggin, *Accurate calculation of the absolute free energy of binding for drug molecules*, Chemical Science **7** (2016) 207–218.

[10] Z. Cournia, B. Allen and W. Sherman, *Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations*, Journal of Chemical Information and Modeling **57** (2017) 2911–2937.

[11] P. Pracht, E. Caldeweyher, S. Ehlert and S. Grimme,
A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules,
ChemRxiv (2019), This content is a preprint and has not been peer-reviewed.

[12] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules*,
The Journal of Physical Chemistry A **125** (2021) 4039–4054.

[13] C. Grebner, E. Malmerberg, A. Shewmaker, J. Batista, A. Nicholls and J. Sadowski,
Virtual Screening in the Cloud: How Big Is Big Enough?,
Journal of Chemical Information and Modeling **60** (2020) 4274–4282.

[14] M. Bursch, J.-M. Mewes, A. Hansen and S. Grimme,
Best-Practice DFT Protocols for Basic Molecular Computational Chemistry,
Angewandte Chemie **134** (2022) e202205735.

[15] T. Salthammer, U. Hohm, M. Stahn and S. Grimme, *Proton-transfer rate constants for the determination of organic indoor air pollutants by online mass spectrometry*,
RSC Advances **13** (2023) 17856–17868.

[16] M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande,
Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins,
The Journal of Chemical Physics **119** (2003) 5740–5761.

[17] J. Zhang, B. Tuguldur and D. Van Der Spoel,
Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation,
Journal of Chemical Information and Modeling **55** (2015) 1192–1201.

[18] J. Zhang, B. Tuguldur and D. Van Der Spoel,
Correction to Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation,
Journal of Chemical Information and Modeling **56** (2016) 819–820.

[19] J. Zhang, H. Zhang, T. Wu, Q. Wang and D. Van Der Spoel, *Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents*,
Journal of Chemical Theory and Computation **13** (2017) 1034–1043.

[20] S. Spicher, C. Plett, P. Pracht, A. Hansen and S. Grimme, *Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods*,
Journal of Chemical Theory and Computation **18** (2022) 3174–3189.

[21] S. Miertuš, E. Scrocco and J. Tomasi, *Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects*,
Chemical Physics **55** (1981) 117–129.

[22] R. Franke and B. Hannebauer, *On the influence of basis sets and quantum chemical methods on the prediction accuracy of COSMO-RS*,
Physical Chemistry Chemical Physics **13** (2011) 21344.

[23] J. Reinisch, M. Diedenhofen, R. Wilcken, A. Udvarhelyi and A. Glöß,
Benchmarking Different QM Levels for Usage with COSMO-RS,
Journal of Chemical Information and Modeling **59** (2019) 4806–4813.

[24] L. Tomaník, L. Rulíšek and P. Slavíček,
Redox Potentials with COSMO-RS: Systematic Benchmarking with Different Databases,
Journal of Chemical Theory and Computation **19** (2023) 1014–1022.

[25] G. Sigalov, A. Fenley and A. Onufriev,
Analytical electrostatics for biomolecules: Beyond the generalized Born approximation,
The Journal of Chemical Physics **124** (2006) 124902.

[26] S. Ehlert, M. Stahn, S. Spicher and S. Grimme,
Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods,
Journal of Chemical Theory and Computation **17** (2021) 4250–4261.

[27] A. Klamt, *Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena*,
The Journal of Physical Chemistry **99** (1995) 2224–2235.

[28] A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz,
Refinement and Parametrization of COSMO-RS,
The Journal of Physical Chemistry A **102** (1998) 5074–5085.

[29] S.-T. Lin and S. I. Sandler,
A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model,
Industrial & Engineering Chemistry Research **41** (2002) 899–913.

[30] A. V. Marenich, C. J. Cramer and D. G. Truhlar,
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,
The Journal of Physical Chemistry B **113** (2009) 6378–6396.

[31] J. W. Zheng and W. H. Green,
Experimental Compilation and Computation of Hydration Free Energies for Ionic Solutes,
The Journal of Physical Chemistry A **127** (2023) 10268–10281.

[32] K. Paduszyński,
An overview of the performance of the COSMO-RS approach in predicting the activity coefficients of molecular solutes in ionic liquids and derived properties at infinite dilution,
Physical Chemistry Chemical Physics **19** (2017) 11835–11850.

[33] M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer and D. G. Truhlar,
Consistent van der Waals Radii for the Whole Main Group,
The Journal of Physical Chemistry A **113** (2009) 5806–5812.

[34] A. Bondi, *van der Waals Volumes and Radii*,
The Journal of Physical Chemistry **68** (1964) 441–451.

[35] R. S. Rowland and R. Taylor, *Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii*,
The Journal of Physical Chemistry **100** (1996) 7384–7391.

[36] R. A. Pierotti, *A scaled particle theory of aqueous and nonaqueous solutions*,
Chemical Reviews **76** (1976) 717–726.

[37] H. J. Kim, *Electronically adiabatic reaction field approach to solvation. I. Theoretical formulation via multipole expansion in a fluctuating cavity*, The Journal of Chemical Physics **105** (1996) 6818–6832.

[38] M. V. Basilevsky and D. F. Parsons, *Nonlocal continuum solvation model with oscillating susceptibility kernels: A nonrigid cavity model*, The Journal of Chemical Physics **108** (1998) 9114–9123.

[39] G. Graziano, *Cavity thermodynamics and surface tension of water*, Chemical Physics Letters **442** (2007) 307–310.

[40] B. Ginovska, D. M. Camaioni, M. Dupuis, C. A. Schwerdtfeger and Q. Gil, *Charge-Dependent Cavity Radii for an Accurate Dielectric Continuum Model of Solvation with Emphasis on Ions: Aqueous Solutes with Oxo, Hydroxo, Amino, Methyl, Chloro, Bromo, and Fluoro Functionalities*, The Journal of Physical Chemistry A **112** (2008) 10604–10613.

[41] G. Hou, X. Zhu and Q. Cui, *An Implicit Solvent Model for SCC-DFTB with Charge-Dependent Radii*, Journal of Chemical Theory and Computation **6** (2010) 2303–2314.

[42] E. R. Kuechler, T. J. Giese and D. M. York, *VR-SCOSMO: A smooth conductor-like screening model with charge-dependent radii for modeling chemical reactions*, The Journal of Chemical Physics **144** (2016) 164115.

[43] A. Pomogaeva and D. M. Chipman, *Hydration Energy from a Composite Method for Implicit Representation of Solvent*, Journal of Chemical Theory and Computation **10** (2014) 211–219.

[44] Z.-Q. You and J. M. Herbert, *Reparameterization of an Accurate, Few-Parameter Implicit Solvation Model for Quantum Chemistry: Composite Method for Implicit Representation of Solvent, CMIRS v. 1.1*, Journal of Chemical Theory and Computation **12** (2016) 4338–4346.

[45] F. Jensen, *Introduction to computational chemistry*, eng, Third edition, Chichester, West Sussex Hoboken, NJ Oxford: Wiley, 2017, ISBN: 978-1-118-82599-0.

[46] *CODATA Value: electron mass*, URL: `https://physics.nist.gov/cgi-bin/cuu/Value?me` (visited on 18/12/2023).

[47] E. Schrödinger, *Quantisierung als Eigenwertproblem*, Annalen der Physik **384** (1926) 489–527.

[48] E. Schrödinger, *Quantisierung als Eigenwertproblem*, Annalen der Physik **386** (1926) 109–139.

[49] F. Jensen, "The Adiabatic and Born-Oppenheimer Approximations", eng, *Introduction to computational chemistry*, Third edition, Chichester, West Sussex Hoboken, NJ Oxford: Wiley, 2017 90–94, ISBN: 978-1-118-82599-0.

[50] J. Simons, *An experimental chemist's guide to ab initio quantum chemistry*, The Journal of Physical Chemistry **95** (1991) 1017–1029.

[51] M. Born and W. Heisenberg, “Zur Quantentheorie der Molekeln”,
Original Scientific Papers Wissenschaftliche Originalarbeiten,
ed. by W. Blum, H. Rechenberg and H.-P. Dürr,
Berlin, Heidelberg: Springer Berlin Heidelberg, 1985 216–246, ISBN: 978-3-642-64900-4.

[52] W. Kolos and L. Wolniewicz, *Vibrational and Rotational Energies for the B 1Σu + , C 1Πu, and a 3Σg + States of the Hydrogen Molecule*,
The Journal of Chemical Physics **48** (1968) 3672–3680.

[53] F. Jensen, “Hartree-Fock Theory”, eng, *Introduction to computational chemistry*,
Third edition, Chichester, West Sussex Hoboken, NJ Oxford: Wiley, 2017 94–100,
ISBN: 978-1-118-82599-0.

[54] D. R. Hartree, *The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods*,
Mathematical Proceedings of the Cambridge Philosophical Society **24** (1928) 89–110.

[55] V. Fock, *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*,
Zeitschrift für Physik **61** (1930) 126–148.

[56] D. R. Hartree, W. Hartree and Swirles, *Self-consistent field, including exchange and superposition of configurations, with some results for oxygen*,
Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **238** (1939) 229–247.

[57] J. C. Slater, *A Simplification of the Hartree-Fock Method*,
Physical Review **81** (1951) 385–390.

[58] K. Raghavachari and J. B. Anderson, *Electron Correlation Effects in Molecules*,
The Journal of Physical Chemistry **100** (1996) 12960–12973.

[59] P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*,
Physical Review **136** (1964) B864–B871.

[60] E. S. Kryachko and E. V. Ludeña, *Density functional theory: Foundations reviewed*,
Physics Reports **544** (2014) 123–239.

[61] N. Mardirossian and M. Head-Gordon, *Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals*,
Molecular Physics **115** (2017) 2315–2372.

[62] P. Verma and D. G. Truhlar, *Status and Challenges of Density Functional Theory*,
Trends in Chemistry **2** (2020) 302–318.

[63] F. Bloch,
Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit,
Zeitschrift für Physik **57** (1929) 545–555.

[64] P. A. M. Dirac, *Note on Exchange Phenomena in the Thomas Atom*,
Mathematical Proceedings of the Cambridge Philosophical Society **26** (1930) 376–385.

[65] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*,
Physical Review **140** (1965).

[66] J. P. Perdew,
Jacob's ladder of density functional approximations for the exchange-correlation energy,
AIP Conference Proceedings **577** (2001) 1–20.

[67] W. J. Carr, *Energy, Specific Heat, and Magnetic Properties of the Low-Density Electron Gas*,
Physical Review **122** (1961) 1437–1446.

[68] W. J. Carr and A. A. Maradudin, *Ground-State Energy of a High-Density Electron Gas*,
Physical Review **133** (1964) A371–A374.

[69] A. D. Becke,
Density-functional exchange-energy approximation with correct asymptotic behavior,
Physical Review A **38** (1988) 3098–3100.

[70] C. Lee, W. Yang and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Physical Review B **37** (1988) 785–789.

[71] J. P. Perdew, K. Burke and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*,
Physical Review Letters **77** (1996) 3865–3868.

[72] B. Hammer, L. B. Hansen and J. K. Nørskov, *Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals*,
Physical Review B **59** (1999) 7413–7421.

[73] A. D. Becke and M. R. Roussel,
Exchange holes in inhomogeneous systems: A coordinate-space model,
Physical Review A **39** (1989) 3761–3767.

[74] A. D. Becke, *Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing*,
The Journal of Chemical Physics **104** (1996) 1040–1046.

[75] J. P. Perdew, S. Kurth, A. Zupan and P. Blaha, *Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation*,
Physical Review Letters **82** (1999) 2544–2547.

[76] A. D. Becke, *Density-functional thermochemistry. III. The role of exact exchange*,
The Journal of Chemical Physics **98** (1993) 5648–5652.

[77] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch,
Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, The Journal of Physical Chemistry **98** (1994) 11623–11627.

[78] V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes*,
The Journal of Chemical Physics **119** (2003) 12129–12137.

[79] A. D. Becke, *A new mixing of Hartree–Fock and local density-functional theories*,
The Journal of Chemical Physics **98** (1993) 1372–1377.

[80] S. Grimme,
Semiempirical hybrid density functional with perturbative second-order correlation,
The Journal of Chemical Physics **124** (2006) 034108.

[81] E. Brémond and C. Adamo,
Seeking for parameter-free double-hybrid functionals: The PBE0-DH model,
The Journal of Chemical Physics **135** (2011) 024106.

[82] L. Goerigk and S. Grimme, *Double-hybrid density functionals*,
WIREs Computational Molecular Science **4** (2014) 576–600.

[83] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Physical Review B **58** (1998) 7260–7268.

[84] S. Grimme, C. Bannwarth and P. Shushkov, *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)*,
Journal of Chemical Theory and Computation **13** (2017) 1989–2009.

[85] C. Bannwarth, S. Ehlert and S. Grimme, *GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions*,
Journal of Chemical Theory and Computation **15** (2019) 1652–1671.

[86] M. Gaus, Q. Cui and M. Elstner, *DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)*,
Journal of Chemical Theory and Computation **7** (2011) 931–948.

[87] *Electronic Structure and Spectra of Some Nitrogen Heterocycles*,
Zeitschrift für Physikalische Chemie **12** (1957) 335–338, ed. by K. Nishimoto and N. Mataga.

[88] K. Ohno, *Some remarks on the Pariser-Parr-Pople method*,
Theoretica Chimica Acta **2** (1964) 219–227.

[89] G. Klopman, *A Semiempirical Treatment of molecular Structures. II. Molecular Terms and Application to diatomic Molecules*,
Journal of the American Chemical Society **86** (1964) 4550–4557.

[90] S. Grimme, *A simplified Tamm-Dancoff density functional approach for the electronic excitation spectra of very large molecules*,
The Journal of Chemical Physics **138** (2013) 244104.

[91] E. Caldeweyher, C. Bannwarth and S. Grimme,
Extension of the D3 dispersion coefficient model,
The Journal of Chemical Physics **147** (2017) 034112.

[92] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *A generally applicable atomic-charge dependent London dispersion correction*,
The Journal of Chemical Physics **150** (2019) 154122.

[93] S. Grimme, J. Antony, S. Ehrlich and H. Krieg,
A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,
The Journal of Chemical Physics **132** (2010) 154104.

[94] N. D. Mermin, *Thermal Properties of the Inhomogeneous Electron Gas*,
Physical Review **137** (1965) A1441–A1443.

[95] "Statistical Mechanics", eng, *Introduction to computational chemistry*, Third edition, Chichester, West Sussex Hoboken, NJ Oxford: Wiley, 2017 452–464, ISBN: 978-1-118-82599-0.

[96] S. Grimme, *Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory*, Chemistry - A European Journal **18** (2012) 9955–9964.

[97] R. M. Levy and E. Gallicchio, *COMPUTER SIMULATIONS WITH EXPLICIT SOLVENT: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects*, Annual Review of Physical Chemistry **49** (1998) 531–567.

[98] J. M. Herbert, *Dielectric continuum methods for quantum chemistry*, WIREs Computational Molecular Science **11** (2021) e1519.

[99] J. Tomasi, B. Mennucci and R. Cammi, *Quantum Mechanical Continuum Solvation Models*, Chemical Reviews **105** (2005) 2999–3094.

[100] "Solvation Models", eng, *Introduction to computational chemistry*, Third edition, Chichester, West Sussex Hoboken, NJ Oxford: Wiley, 2017 502–514, ISBN: 978-1-118-82599-0.

[101] G. Lamm, "The Poisson–Boltzmann Equation", en, *Reviews in Computational Chemistry*, ed. by K. B. Lipkowitz, R. Larter and T. R. Cundari, 1st ed., vol. 19, Wiley, 2003 147–365, ISBN: 978-0-471-23585-9.

[102] N. A. Baker, "Biomolecular Applications of Poisson–Boltzmann Methods", en, *Reviews in Computational Chemistry*, ed. by K. B. Lipkowitz, R. Larter and T. R. Cundari, 1st ed., vol. 21, Wiley, 2005 349–379, ISBN: 978-0-471-68239-4.

[103] M. P. Coons and J. M. Herbert, *Quantum chemistry in arbitrary dielectric environments: Theory and implementation of nonequilibrium Poisson boundary conditions and application to compute vertical ionization energies at the air/water interface*, The Journal of Chemical Physics **148** (2018) 222834.

[104] M. Born, *Volumen und Hydratationswärme der Ionen*, Zeitschrift für Physik **1** (1920) 45–48.

[105] W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics*, Journal of the American Chemical Society **112** (1990) 6127–6129.

[106] D. Bashford and D. A. Case, *Generalized born models of macromolecular solvation effects*, Annual Review of Physical Chemistry (2000).

[107] J. Tomasi and M. Persico, *Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent*, Chemical Reviews **94** (1994) 2027–2094.

[108] J. Tomasi, *Selected features of the polarizable continuum model for the representation of solvation*, WIREs Computational Molecular Science **1** (2011) 855–867.

[109] B. Mennucci, *Polarizable continuum model*, WIREs Computational Molecular Science **2** (2012) 386–404.

[110] D. M. Chipman, *Charge penetration in dielectric models of solvation*, The Journal of Chemical Physics **106** (1997) 10194–10206.

[111] S. Miertus and J. Tomasi, *Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes*, Chemical Physics **65** (1982) 239–245.

[112] R. Bonaccorsi, R. Cimiraglia and J. Tomasi, *Ab initio evaluation of absorption and emission transitions for molecular solutes, including separate consideration of orientational and inductive solvent effects*, Journal of Computational Chemistry **4** (1983) 567–577.

[113] A. V. Onufriev and B. Aguilar, *Accuracy of continuum electrostatic calculations based on three common dielectric boundary definitions*, Journal of Theoretical and Computational Chemistry **13** (2014) 1440006.

[114] R. C. Harris and B. M. Pettitt, *Examining the Assumptions Underlying Continuum-Solvent Models*, Journal of Chemical Theory and Computation **11** (2015) 4593–4600.

[115] C. J. Cramer and D. G. Truhlar, *A Universal Approach to Solvation Modeling*, Accounts of Chemical Research **41** (2008) 760–768.

[116] F. Bohle, J. Seibert and S. Grimme, *Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules*, The Journal of Organic Chemistry **86** (2021) 15522–15531.

[117] S. Grimme, *Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations*, Journal of Chemical Theory and Computation **15** (2019) 2847–2862.

[118] S. Spicher and S. Grimme, *Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems*, Angewandte Chemie International Edition **59** (2020) 15665–15673.

[119] M. Stahn, S. Grimme, T. Salthammer, U. Hohm and W.-U. Palm, *Quantum chemical calculation of the vapor pressure of volatile and semi volatile organic compounds*, Environmental Science: Processes & Impacts **24** (2022) 2153–2166.

[120] R. Sander, *Compilation of Henry's law constants (version 4.0) for water as solvent*, Atmospheric Chemistry and Physics **15** (2015) 4399–4981.

[121] D. A. Hinckley, T. F. Bidleman, W. T. Foreman and J. R. Tuschall, *Determination of vapor pressures for nonpolar and semipolar organic compounds from gas chromatograhic retention data*, Journal of Chemical & Engineering Data **35** (1990) 232–237.

[122] S. H. Hilal, S. W. Karickhoff and L. A. Carreira, *Prediction of the Vapor Pressure Boiling Point, Heat of Vaporization and Diffusion Coefficient of Organic Compounds*, QSAR & Combinatorial Science **22** (2003) 565–574.

[123] P. Schossler, T. Schripp, T. Salthammer and M. Bahadir, *Beyond phthalates: Gas phase concentrations and modeled gas/particle distribution of modern plasticizers*, Science of The Total Environment **409** (2011) 4031–4038.

[124] Y. Wu, C. M. A. Eichler, S. Chen and J. C. Little,
Simple Method To Measure the Vapor Pressure of Phthalates and Their Alternatives,
Environmental Science & Technology **50** (2016) 10082–10088.

[125] P. Pracht, F. Bohle and S. Grimme,
Automated exploration of the low-energy chemical space with fast quantum chemical methods,
Physical Chemistry Chemical Physics **22** (2020) 7169–7192.

[126] S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes,
r²SCAN-3c: A "Swiss army knife" composite electronic-structure method,
The Journal of Chemical Physics **154** (2021) 064103.

[127] B. Sieland, M. Stahn, R. Schoch, C. Daniliuc, S. Spicher, S. Grimme, A. Hansen and J. Paradies, *Dispersion Energy-Stabilized Boron and Phosphorus Lewis Pairs*,
Angewandte Chemie International Edition **62** (2023) e202308752.

[128] G. C. Welch and D. W. Stephan,
Facile Heterolytic Cleavage of Dihydrogen by Phosphines and Boranes,
Journal of the American Chemical Society **129** (2007) 1880–1881.

[129] L. Rocchigiani, G. Ciancaleoni, C. Zuccaccia and A. Macchioni, *Probing the Association of Frustrated Phosphine–Borane Lewis Pairs in Solution by NMR Spectroscopy*,
Journal of the American Chemical Society **136** (2014) 112–115.

[130] J. Paradies, *Structure–Reactivity Relationships in Borane-Based FLP-Catalyzed Hydrogenations, Dehydrogenations, and Cycloisomerizations*,
Accounts of Chemical Research **56** (2023) 821–834.

[131] A. R. Jupp, *Evidence for the encounter complex in frustrated Lewis pair chemistry*,
Dalton Transactions **51** (2022) 10681–10689.

[132] T. A. Rokob, I. Bakó, A. Stirling, A. Hamza and I. Pápai,
Reactivity Models of Hydrogen Activation by Frustrated Lewis Pairs: Synergistic Electron Transfers or Polarization by Electric Field?,
Journal of the American Chemical Society **135** (2013) 4425–4437.

[133] F. Holtrop, C. Helling, M. Lutz, N. P. Van Leest, B. De Bruin and J. C. Slootweg,
Enhancement of London Dispersion in Frustrated Lewis Pairs: Towards a Crystalline Encounter Complex, Synlett **34** (2023) 1122–1128.

[134] R. Pollice, F. Fleckenstein, I. Shenderovich and P. Chen,
Compensation of London Dispersion in the Gas Phase and in Aprotic Solvents,
Angewandte Chemie International Edition **58** (2019) 14281–14288.

[135] M. Stahn, S. Ehlert and S. Grimme, *Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods*,
The Journal of Physical Chemistry A **127** (2023) 7036–7043.

[136] A. Klamt and G. Schüürmann, *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.

[137] F. Lipparini, B. Stamm, E. Cancès, Y. Maday and B. Mennucci, *Fast Domain Decomposition Algorithm for Continuum Solvation Models: Energy and First Derivatives*, Journal of Chemical Theory and Computation **9** (2013) 3637–3648.

[138] C. Plett, M. Stahn, M. Bursch, J.-M. Mewes and S. Grimme, *Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)*, The Journal of Physical Chemistry Letters **15** (2024) 2462–2469.

[139] W. J. Mortier, S. K. Ghosh and S. Shankar, *Electronegativity-equalization method for the calculation of atomic charges in molecules*, Journal of the American Chemical Society **108** (1986) 4315–4320.

[140] M. Müller, A. Hansen and S. Grimme, *An atom-in-molecule adaptive polarized valence single-ζ atomic orbital basis for electronic structure calculations*, The Journal of Chemical Physics **159** (2023) 164108.

[141] A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database - version 2012*, University of Minnesota, Minneapolis, 2020.

[142] M. Stahn and C. Plett, *DRACO*, URL: https://github.com/grimme-lab/DRACO.

[143] F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *The ORCA quantum chemistry program package*, The Journal of Chemical Physics **152** (2020) 224108.

[144] R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Electronic structure calculations on workstation computers: The program system turbomole*, Chemical Physics Letters **162** (1989) 165–169.

[145] C. Wagen and A. Wagen, *Efficient and Accurate pKa Prediction Enabled by Pre-Trained Machine-Learned Interatomic Potentials*, ChemRxiv (2024), This content is a preprint and has not been peer-reviewed.

[146] J. F. Pankow, *Review and comparative analysis of the theories on partitioning between the gas and aerosol particulate phases in the atmosphere*, Atmospheric Environment (1967) **21** (1987) 2275–2283.

[147] R. P. Schwarzenbach, P. M. Geschwend and D. M. Imboden, *Environmental Organic Chemistry*, 3rd Edition, Hoboken, NJ: John Wiley & Sons, 2017, ISBN: 978-1-118-76723-8.

[148] OECD, *Test No. 104: Vapour Pressure*, en, Paris: Organisation for Economic Co-operation and Development, 2006, URL: https://www.oecd-ilibrary.org/environment/test-no-104-vapour-pressure_9789264069565-en.

[149] C. L. Yaws, *The Yaws Handbook of Vapor Pressure: Antoine Coefficients*, English, 2nd edition, Gulf Professional Publishing, 2015.

[150] S. H. Fishtine, *RELIABLE LATENT HEATS OF VAPORIZATION*, Industrial & Engineering Chemistry **55** (1963) 20–28.

[151] C. F. Grain, "Vapor Pressure", *Handbook of Chemical Property Estimation Methods*, eds. W. J. Lyman, W. F. Reehl and D. H. Rosenblatt, Washington D.C.: American Chemical Society, 1990 14/11–14/20.

[152] E. Baum, *Chemical Property Estimation: Theory and Application*, Boca Raton: CRC Press, 1997, ISBN: 978-1-315-13915-9.

[153] M. L. Sage and G. W. Sage, "Vapor Pressure", *Handbook of Property Estimation Methods*, eds. R. S. Boethling and D. Mackay, Boca Raton, FL: Lewis Publishers, 2000 53–65.

[154] W. E. Asher, J. F. Pankow, G. B. Erdakos and J. H. Seinfeld, *Estimating the vapor pressures of multi-functional oxygen-containing organic compounds using group contribution methods*, Atmospheric Environment **36** (2002) 1483–1498.

[155] W. E. Asher and J. F. Pankow, *Vapor pressure prediction for alkenoic and aromatic organic compounds by a UNIFAC-based group contribution method*, Atmospheric Environment **40** (2006) 3588–3600.

[156] T. Dupeux, T. Gaudin, C. Marteau-Roussy, J.-M. Aubry and V. Nardello-Rataj, *COSMO-RS as an effective tool for predicting the physicochemical properties of fragrance raw materials*, Flavour and Fragrance Journal **37** (2022) 106–120.

[157] J. Hammer, H. Matsukami, H. Kuramochi and S. Endo, *Direct measurements and modeling of congener group specific vapor pressure for chlorinated paraffins*, Chemosphere **281** (2021) 130909.

[158] S.-T. Lin, J. Chang, S. Wang, W. A. Goddard and S. I. Sandler, *Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model*, The Journal of Physical Chemistry A **108** (2004) 7429–7439.

[159] K. Yui, T. Motoki, H. Kato, H. Kuramochi, T. Tsuji, S.-i. Sakai and F. Wania, *Measurement of Vapor Pressures and Melting Properties of Five Polybrominated Aromatic Flame Retardants*, Journal of Chemical & Engineering Data **63** (2018) 2578–2585.

[160] L. Jiřiště and M. Klajmon, *Predicting the Thermodynamics of Ionic Liquids: What to Expect from PC-SAFT and COSMO-RS?*, The Journal of Physical Chemistry B **126** (2022) 3717–3736.

[161] J. M. Goodman, *What Is the Longest Unbranched Alkane with a Linear Global Minimum Conformation?*, Journal of Chemical Information and Computer Sciences **37** (1997) 876–878.

[162] S. Jakobtorweihen, T. Ingram and I. Smirnova, *Combination of COSMOmic and molecular dynamics simulations for the calculation of membrane–water partition coefficients*, Journal of Computational Chemistry **34** (2013) 1332–1340.

[163] M. Kolář, J. Fanfrlík, M. Lepšík, F. Forti, F. J. Luque and P. Hobza, *Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches*, The Journal of Physical Chemistry B **117** (2013) 5950–5962.

[164] A. Klamt, F. Eckert and M. Diedenhofen, *Prediction of the Free Energy of Hydration of a Challenging Set of Pesticide-Like Compounds*, The Journal of Physical Chemistry B **113** (2009) 4508–4510.

[165] P. Yamin, R. Isele-Holder and K. Leonhard, *Predicting Octanol/Water Partition Coefficients of Alcohol Ethoxylate Surfactants Using a Combination of Molecular Dynamics and the Conductor-like Screening Model for Realistic Solvents*, Industrial & Engineering Chemistry Research **55** (2016) 4782–4789.

[166] S. Grimme, *Semiempirical GGA-type density functional constructed with a long-range dispersion correction*, Journal of Computational Chemistry **27** (2006) 1787–1799.

[167] F. Weigend and R. Ahlrichs, *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*, Physical Chemistry Chemical Physics **7** (2005) 3297–3305.

[168] S. Spicher and S. Grimme, *Single-Point Hessian Calculations for Improved Vibrational Frequencies and Rigid-Rotor-Harmonic-Oscillator Thermodynamics*, Journal of Chemical Theory and Computation **17** (2021) 1701–1714.

[169] A. Ben-Naim and Y. Marcus, *Solvation thermodynamics of nonionic solutes*, The Journal of Chemical Physics **81** (1984) 2016–2027.

[170] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *PubChem in 2021: new data content and improved web interfaces*, Nucleic Acids Research **49** (2021) D1388–D1395.

[171] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Extended tight-binding quantum chemistry methods*, WIREs Computational Molecular Science **11** (2021) e1493.

[172] K.-U. Goss, *Predicting the equilibrium partitioning of organic compounds using just one linear solvation energy relationship (LSER)*, Fluid Phase Equilibria **233** (2005) 19–22.

[173] M. H. Abraham and J. C. McGowan, *The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography*, Chromatographia **23** (1987) 243–246.

[174] N. Ulrich, S. Endo, T.N. Brown, N. Watanabe, G. Bronner, M.H. Abraham and K.-U. Goss, *UFZ-LSER database v 3.2 [Internet]*, (2017), Place: Leipzig, Deutschland, Helmholtz Zentrum für Umweltforschung - UFZ.

[175] S.H. Hilal, S.W. Karickhoff and L.A. Carreira, *Verification and Validation of the Sparc Model.*, tech. rep. EPA/600/R-03-033, Environmental Protection Agency, 2003.

[176] T. Salthammer, *TVOC - Revisited*, Environment International **167** (2022) 107440.

[177] K. Růžička and V. Majer, *Simultaneous Treatment of Vapor Pressures and Related Thermal Data Between the Triple and Normal Boiling Temperatures for n -Alkanes C5–C20*, Journal of Physical and Chemical Reference Data **23** (1994) 1–39.

[178] J. S. Chickos and W. Hanshaw, *Vapor Pressures and Vaporization Enthalpies of the n -Alkanes from C_{21} to C_{30} at T = 298.15 K by Correlation Gas Chromatography*, Journal of Chemical & Engineering Data **49** (2004) 77–85.

[179] P. M. Sherblom, P. M. Gschwend and R. P. Eganhouse, *Aqueous solubilities, vapor pressures, and 1-octanol-water partition coefficients for C9-C14 linear alkylbenzenes*, Journal of Chemical & Engineering Data **37** (1992) 394–399.

[180] W.V. Wilding, T.A. Knotts, N.F. Giles and R.L. Rowley,
DIPPR Data Compilation of Pure Chemical Properties,
Design Institute for Physical Properties - AIChE, 2020.

[181] D. P. Biddiscombe and J. F. Martin, *Vapour pressures of phenol and the cresols*,
Transactions of the Faraday Society **54** (1958) 1316.

[182] Y. D. Lei, R. Chankalal, A. Chan and F. Wania,
Supercooled Liquid Vapor Pressures of the Polycyclic Aromatic Hydrocarbons,
Journal of Chemical & Engineering Data **47** (2002) 801–806.

[183] Y. Koga, *Vapor pressures of aqueous 2-butoxyethanol solutions at 25.degree.C: transitions in mixing scheme*, The Journal of Physical Chemistry **95** (1991) 4119–4126.

[184] W. Acree and J. S. Chickos,
Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies From 1880 to 2015. Part 1. C_1 - C_{10},
Journal of Physical and Chemical Reference Data **45** (2016) 033101.

[185] J. N'Guimbi, H. Kasehgari, I. Mokbel and J. Jose,
Tensions de vapeur d'alcools primaires dans le domaine 0,3 Pa à 1,5 kPa,
Thermochimica Acta **196** (1992) 367–377.

[186] B. Koutek, M. Hoskovec, P. Vrkočová, K. Konečný and L. Feltl, *Gas chromatographic determination of vapour pressures of pheromone-like compounds II. Alcohols*,
Journal of Chromatography A **679** (1994) 307–317.

[187] H. K. Cammenga, F. W. Schulze and W. Theuerl,
Vapor pressure and evaporation coefficient of glycerol,
Journal of Chemical & Engineering Data **22** (1977) 131–134.

[188] V. Štejfa, M. Fulem, K. Růžička and P. Morávek, *New Static Apparatus for Vapor Pressure Measurements: Reconciled Thermophysical Data for Benzophenone*,
Journal of Chemical & Engineering Data **61** (2016) 3627–3639.

[189] J. O. Okeme, T. F. M. Rodgers, J. M. Parnis, M. L. Diamond, T. F. Bidleman and L. M. Jantunen, *Gas Chromatographic Estimation of Vapor Pressures and Octanol–Air Partition Coefficients of Semivolatile Organic Compounds of Emerging Concern*,
Journal of Chemical & Engineering Data **65** (2020) 2467–2475.

[190] V. Roháč, J. E. Musgrove, K. Ružička, V. Ružička, M. Zábranský and K. Aim,
Thermodynamic properties of dimethyl phthalate along the (vapour + liquid) saturation curve,
The Journal of Chemical Thermodynamics **31** (1999) 971–986.

[191] V. Roháč, K. Růžička, V. Růžička, D. H. Zaitsau, G. J. Kabo, V. Diky and K. Aim,
Vapour pressure of diethyl phthalate,
The Journal of Chemical Thermodynamics **36** (2004) 929–937.

[192] C. Gobble, J. Chickos and S. P. Verevkin, *Vapor Pressures and Vaporization Enthalpies of a Series of Dialkyl Phthalates by Correlation Gas Chromatography*,
Journal of Chemical & Engineering Data **59** (2014) 1353–1365.

[193] V. A. Tsendrovskaia,
Determination of the vapor tension and volatility of some higher esters of phthalic acid,
Gigiena I Sanitariia **37** (1972) 101–103.

[194] D. Lipkind, Y. Kapustin, P. Umnahanant and J. S. Chickos,
The vaporization enthalpies and vapor pressures of a series of unsaturated fatty acid methyl esters by correlation gas chromatography, Thermochimica Acta **456** (2007) 94–101.

[195] M. Bilde, K. Barsanti, M. Booth, C. D. Cappa, N. M. Donahue, E. U. Emanuelsson, G. McFiggans, U. K. Krieger, C. Marcolli, D. Topping, P. Ziemann, M. Barley, S. Clegg, B. Dennis-Smither, M. Hallquist, Å. M. Hallquist, A. Khlystov, M. Kulmala, D. Mogensen, C. J. Percival, F. Pope, J. P. Reid, M. A. V. Ribeiro Da Silva, T. Rosenoern, K. Salo, V. P. Soonsin, T. Yli-Juuti, N. L. Prisle, J. Pagels, J. Rarey, A. A. Zardini and I. Riipinen,
Saturation Vapor Pressures and Transition Enthalpies of Low-Volatility Organic Molecules of Atmospheric Relevance: From Dicarboxylic Acids to Complex Mixtures,
Chemical Reviews **115** (2015) 4115–4156.

[196] S. Brommer, L. M. Jantunen, T. F. Bidleman, S. Harrad and M. L. Diamond,
Determination of Vapor Pressures for Organophosphate Esters,
Journal of Chemical & Engineering Data **59** (2014) 1441–1447.

[197] N. Li, F. Wania, Y. D. Lei and G. L. Daly,
A Comprehensive and Critical Compilation, Evaluation, and Selection of Physical–Chemical Property Data for Selected Polychlorinated Biphenyls,
Journal of Physical and Chemical Reference Data **32** (2003) 1545–1590.

[198] K. G. Drouillard, G. T. Tomy, D. C. G. Muir and K. J. Friesen,
Volatility of chlorinated n -alkanes (C_{10} - C_{12}): Vapor pressures and Henry's law constants,
Environmental Toxicology and Chemistry **17** (1998) 1252–1260.

[199] A. Wong, Y. D. Lei, M. Alaee and F. Wania,
Vapor Pressures of the Polybrominated Diphenyl Ethers,
Journal of Chemical & Engineering Data **46** (2001) 239–242.

[200] Y. D. Lei, F. Wania and D. Mathers, *Temperature-Dependent Vapor Pressure of Selected Cyclic and Linear Polydimethylsiloxane Oligomers*,
Journal of Chemical & Engineering Data **55** (2010) 5868–5873.

[201] L. Shen and F. Wania, *Compilation, Evaluation, and Selection of Physical-Chemical Property Data for Organochlorine Pesticides*,
Journal of Chemical & Engineering Data **50** (2005) 742–768.

[202] A. Goel, L. L. McConnell and A. Torrents, *Determination of vapor pressure-temperature relationships of current–use pesticides and transformation products*,
Journal of Environmental Science and Health, Part B **42** (2007) 343–349.

[203] D. Mackay, W.-Y. Shiu, W.-Y. Shiu and S. C. Lee,
Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals Volume III: Oxygen Containing Compounds, en, 2nd ed., CRC Press, 2006,
ISBN: 978-0-429-15007-4.

[204] T. Salthammer, *Emerging indoor pollutants*,
International Journal of Hygiene and Environmental Health **224** (2020) 113423.

[205] B. Petrie, R. Barden and B. Kasprzyk-Hordern,
A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring,
Water Research, Occurrence, fate, removal and assessment of emerging contaminants in water in the water cycle (from wastewater to drinking water) **72** (2015) 3–27.

[206] *Liste der Biozidprodukte, die in Deutschland aufgrund eines laufenden Entscheidungsverfahrens auf dem Markt bereitgestellt und verwendet werden dürfen*, 2022,
URL: https://www.baua.de/DE/Themen/Chemikalien-Biostoffe/Chemikalienrecht/Biozide/pdf/Biozidprodukte-im-Entscheidungsverfahren.html (visited on 20/02/2022).

[207] OECD, *Test No. 403: Acute Inhalation Toxicity*, en,
OECD Guidelines for the Testing of Chemicals, Section 4, OECD, 2009,
ISBN: 978-92-64-07060-8,
URL: https://www.oecd-ilibrary.org/environment/test-no-403-acute-inhalation-toxicity_9789264070608-en.

[208] D. Mackay, W.-Y. Shiu, W.-Y. Shiu and S. C. Lee,
Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals Volume I: Introduction and Hydrocarbons, en, 2nd ed., CRC Press, 2006,
ISBN: 978-0-429-15007-4.

[209] S. Endo and K.-U. Goss,
Predicting Partition Coefficients of Polyfluorinated and Organosilicon Compounds using Polyparameter Linear Free Energy Relationships (PP-LFERs),
Environmental Science & Technology **48** (2014) 2776–2784.

[210] D. Mackay, W.-Y. Shiu, W.-Y. Shiu and S. C. Lee,
Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals Volume IV: Nitrogen and Sulphur Containing Compounds and Pesticides, en, 2nd ed.,
CRC Press, 2006, ISBN: 978-0-429-15007-4.

[211] D. Mackay, W.-Y. Shiu, W.-Y. Shiu and S. C. Lee,
Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals Volume II: Halogenated Hydrocarbons, en, 2nd ed., CRC Press, 2006,
ISBN: 978-0-429-15007-4.

[212] J. P. Wagner and P. R. Schreiner,
London Dispersion in Molecular Chemistry - Reconsidering Steric Effects,
Angewandte Chemie International Edition **54** (2015) 12274–12296.

[213] S. Rösel, C. Balestrieri and P. R. Schreiner, *Sizing the role of London dispersion in the dissociation of all-meta tert-butyl hexaphenylethane*, Chemical Science **8** (2017) 405–410.

[214] S. Grimme and P. R. Schreiner,
Steric Crowding Can Stabilize a Labile Molecule: Solving the Hexaphenylethane Riddle,
Angewandte Chemie International Edition **50** (2011) 12639–12642.

[215] U. Mayer, V. Gutmann and W. Gerger, *The acceptor number – A quantitative empirical parameter for the electrophilic properties of solvents*,
Monatshefte für Chemie **106** (1975) 1235–1257.

[216] S. Rösel, J. Becker, W. D. Allen and P. R. Schreiner, *Probing the Delicate Balance between Pauli Repulsion and London Dispersion with Triphenylmethyl Derivatives*, Journal of the American Chemical Society **140** (2018) 14421–14432.

[217] G. C. Welch, R. R. S. Juan, J. D. Masuda and D. W. Stephan, *Reversible, Metal-Free Hydrogen Activation*, Science **314** (2006) 1124–1126.

[218] J. Lam, K. M. Szkop, E. Mosaferi and D. W. Stephan, *FLP catalysis: main group hydrogenations of organic unsaturated substrates*, Chemical Society Reviews **48** (2019) 3592–3612.

[219] J. Paradies, *From structure to novel reactivity in frustrated Lewis pairs*, Coordination Chemistry Reviews **380** (2019) 170–183.

[220] J. Paradies, *Mechanisms in Frustrated Lewis Pair-Catalyzed Reactions*, European Journal of Organic Chemistry **2019** (2019) 283–294.

[221] D. W. Stephan, *Diverse Uses of the Reaction of Frustrated Lewis Pair (FLP) with Hydrogen*, Journal of the American Chemical Society **143** (2021) 20002–20014.

[222] M. Ghara, H. Mondal, R. Pal and P. K. Chattaraj, *Frustrated Lewis Pairs: Bonding, Reactivity, and Applications*, The Journal of Physical Chemistry A **127** (2023) 4561–4582.

[223] S. Grimme, H. Kruse, L. Goerigk and G. Erker, *The Mechanism of Dihydrogen Activation by Frustrated Lewis Pairs Revisited*, Angewandte Chemie International Edition **49** (2010) 1402–1405.

[224] *CDC numbers: 2247791 (2b/3b); 2255773 (2d/3d); 2255774 (2b/3d); 2255775 (2b) and 2255776 (2d).*

[225] H. Jacobsen, H. Berke, S. Döring, G. Kehr, G. Erker, R. Fröhlich and O. Meyer, *Lewis Acid Properties of Tris(pentafluorophenyl)borane. Structure and Bonding in L-B(C_6F_5)$_3$ Complexes*, Organometallics **18** (1999) 1724–1735.

[226] M. Ullrich, A. J. Lough and D. W. Stephan, *Dihydrogen Activation by B(p-C_6F_4H)$_3$ and Phosphines*, Organometallics **29** (2010) 3647–3654.

[227] S. Rösel, H. Quanz, C. Logemann, J. Becker, E. Mossou, L. Cañadillas-Delgado, E. Caldeweyher, S. Grimme and P. R. Schreiner, *London Dispersion Enables the Shortest Intermolecular Hydrocarbon H···H Contact*, Journal of the American Chemical Society **139** (2017) 7428–7431.

[228] G. Skara, B. Pinter, J. Top, P. Geerlings, F. De Proft and F. De Vleeschouwer, *Conceptual Quantum Chemical Analysis of Bonding and Noncovalent Interactions in the Formation of Frustrated Lewis Pairs*, Chemistry – A European Journal **21** (2015) 5510–5519.

[229] G. Bistoni, A. A. Auer and F. Neese, *Understanding the Role of Dispersion in Frustrated Lewis Pairs and Classical Lewis Adducts: A Domain-Based Local Pair Natural Orbital Coupled Cluster Study*, Chemistry – A European Journal **23** (2017) 865–873.

[230] T. Wang, G. Kehr, L. Liu, S. Grimme, C. G. Daniliuc and G. Erker, *Selective Oxidation of an Active Intramolecular Amine/Borane Frustrated Lewis Pair with Dioxygen*, Journal of the American Chemical Society **138** (2016) 4302–4305.

[231] L. L. Zeonjuk, N. Vankova, A. Mavrandonakis, T. Heine, G.-V. Röschenthaler and J. Eicher, *On the Mechanism of Hydrogen Activation by Frustrated Lewis Pairs*, Chemistry – A European Journal **19** (2013) 17413–17424.

[232] S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*, The Journal of Chemical Physics **143** (2015) 054107.

[233] J. Gorges, S. Grimme and A. Hansen, *Reliable prediction of association (free) energies of supramolecular complexes with heavy main group elements – the HS13L benchmark set*, Physical Chemistry Chemical Physics **24** (2022) 28831–28843.

[234] C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *Natural triple excitations in local coupled cluster calculations with pair natural orbitals*, The Journal of Chemical Physics **139** (2013) 134101.

[235] T. T. Duignan, C. J. Mundy, G. K. Schenter and X. S. Zhao, *Method for Accurately Predicting Solvation Structure*, Journal of Chemical Theory and Computation **16** (2020) 5401–5409.

[236] H. Senn, D. Deubel, P. Blöchl, A. Togni and G. Frenking, *Phosphane lone-pair energies as a measure of ligand donor strengths and relation to activation energies*, Journal of Molecular Structure: THEOCHEM **506** (2000) 233–242.

[237] C. Bannwarth, A. Hansen and S. Grimme, *The Association of Two "Frustrated" Lewis Pairs by State-of-the-Art Quantum Chemical Methods*, Israel Journal of Chemistry **55** (2015) 235–242.

[238] S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen and F. Neese, *Calculation of Solvent Shifts on Electronic* $\boldsymbol{g}$ *-Tensors with the Conductor-Like Screening Model (COSMO) and Its Self-Consistent Generalization to Real Solvents (Direct COSMO-RS)*, The Journal of Physical Chemistry A **110** (2006) 2235–2245.

[239] M. Pu and T. Privalov, *Ab Initio Molecular Dynamics with Explicit Solvent Reveals a Two-Step Pathway in the Frustrated Lewis Pair Reaction*, Chemistry – A European Journal **21** (2015) 17708–17720.

[240] I. Bakó, A. Stirling, S. Bálint and I. Pápai, *Association of frustrated phosphine–borane pairs in toluene: molecular dynamics simulations*, Dalton Transactions **41** (2012) 9023.

[241] J.-M. Mewes, A. Hansen and S. Grimme, *Comment on "The Nature of Chalcogen-Bonding-Type Tellurium–Nitrogen Interactions": Fixing the Description of Finite-Temperature Effects Restores the Agreement Between Experiment and Theory*, Angewandte Chemie International Edition **60** (2021) 13144–13149.

[242] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme,
A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions,
Physical Chemistry Chemical Physics **19** (2017) 32184–32215.

[243] N. Mardirossian and M. Head-Gordon,
ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy,
Physical Chemistry Chemical Physics **16** (2014) 9904.

[244] C. Adamo and V. Barone,
Toward reliable density functional methods without adjustable parameters: The PBE0 model,
The Journal of Chemical Physics **110** (1999) 6158–6170.

[245] A. M. Reilly and A. Tkatchenko,
van der Waals dispersion interactions in molecular materials: beyond pairwise additivity,
Chemical Science **6** (2015) 3289–3301.

[246] L. Goerigk and S. Grimme, *A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions*,
Physical Chemistry Chemical Physics **13** (2011) 6670.

[247] G. Santra, N. Sylvetsky and J. M. L. Martin,
Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4,
The Journal of Physical Chemistry A **123** (2019) 5129–5143.

[248] V. Barone and M. Cossi, *Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model*,
The Journal of Physical Chemistry A **102** (1998) 1995–2001.

[249] E. Cancès, Y. Maday and B. Stamm, *Domain decomposition for implicit solvation models*,
The Journal of Chemical Physics **139** (2013) 054111.

[250] A. Klamt,
COSMO-RS : from quantum chemistry to fluid phase thermodynamics and drug design,
Amsterdam Boston: Elsevier, 2005, ISBN: 0-444-51994-7.

[251] A. Klamt, *The COSMO and COSMO-RS solvation models*,
WIREs Computational Molecular Science **1** (2011) 699–709.

[252] S. Grimme, M. Müller and A. Hansen, *A non-self-consistent tight-binding electronic structure potential in a polarized double-ζ basis set for all spd-block elements up to $Z = 86$*,
The Journal of Chemical Physics **158** (2023) 124111.

[253] D. L. Mobley and J. P. Guthrie,
FreeSolv: a database of experimental and calculated hydration free energies, with input files,
Journal of Computer-Aided Molecular Design **28** (2014) 711–720.

[254] L. C. Kröger, S. Müller, I. Smirnova and K. Leonhard,
Prediction of Solvation Free Energies of Ionic Solutes in Neutral Solvents,
The Journal of Physical Chemistry A **124** (2020) 4171–4181.

[255] R. Sure and S. Grimme,
Comprehensive Benchmark of Association (Free) Energies of Realistic Host-Guest Complexes,
Journal of Chemical Theory and Computation **11** (2015) 3785–3801.

[256] J. R. Pliego Jr and J. M. Riveros, *Hybrid discrete-continuum solvation methods*,
WIREs Computational Molecular Science **10** (2020) e1440.

[257] S. Mirzaei, M. V. Ivanov and Q. K. Timerghazin,
Improving Performance of the SMD Solvation Model: Bondi Radii Improve Predicted Aqueous Solvation Free Energies of Ions and pKa Values of Thiols,
The Journal of Physical Chemistry A **123** (2019) 9498–9504.

[258] J. M. J. Swanson, S. A. Adcock and J. A. McCammon,
Optimized Radii for Poisson-Boltzmann Calculations with the AMBER Force Field,
Journal of Chemical Theory and Computation **1** (2005) 484–493.

[259] C. J. Cramer and D. G. Truhlar,
Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics,
Chemical Reviews **99** (1999) 2161–2200.

[260] S. W. Rick and B. J. Berne, *The Aqueous Solvation of Water: A Comparison of Continuum Methods with Molecular Dynamics*,
Journal of the American Chemical Society **116** (1994) 3949–3954.

[261] A. W. Lange and J. M. Herbert, *A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach*,
The Journal of Chemical Physics **133** (2010) 244111.

[262] G. Wada, E. Tamura, M. Okina and M. Nakamura, *On the Ratio of Zwitterion Form to Uncharged Form of Glycine at Equilibrium in Various Aqueous Media*,
Bulletin of the Chemical Society of Japan **55** (1982) 3064–3067.

[263] N. Mardirossian and M. Head-Gordon, *ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation*,
The Journal of Chemical Physics **144** (2016) 214110.

[264] D. Rappoport and F. Furche,
Property-optimized Gaussian basis sets for molecular response calculations,
The Journal of Chemical Physics **133** (2010) 134105.

[265] K. Kříž and J. Řezáč,
Reparametrization of the COSMO Solvent Model for Semiempirical Methods PM6 and PM7,
Journal of Chemical Information and Modeling **59** (2019) 229–235.

[266] A. Klamt and V. Jonas, *Treatment of the outlying charge in continuum solvation models*,
The Journal of Chemical Physics **105** (1996) 9972–9981.

List of Figures

List of Tables

Acknowledgements

This thesis would not have been possible without the help of many people, to whom I am truly grateful. I want to thank my supervisor, Stefan Grimme, for giving me the freedom to pursue my ideas while providing me with guidance when needed. I also want to thank him for giving me a fresh perspective on things when I believed I had reached a dead end and for the financial support. I also want to thank Thomas Bredow for being the second reviewer of this thesis.

I want to thank my colleagues, who made working at the Mulliken Center an absolute pleasure, and thanks to whom, sometimes, work did not feel like work. First, I want to thank my office partners, Abylay Katbashev and Marcel Müller, who needed to bear with me when I was thinking out loud over some problem again. I also want to thank my colleagues and friends Christoph Plett and Thomas Gasevic, who, while they were not my office partners, sometimes felt like they were. I also need to thank Christoph for the beautiful memories made at the ACS Fall 2022 in Chicago, where we should have taken the green line way more often to get home and for creating dragons together. Furthermore, I want to thank Benedikt Baedorf, Fabian Bohle, Markus Bursch, Marvin Friede, Thomas Froitzheim, Johannes Gorges, Andreas Hansen, Christian Hölzer, Julius Kleine Büning, Julia Kohn, Lukas Kunze, Joachim Laun, Sarah Löffelsender, Jan Mewes, Hagen Neugebauer, Philipp Pracht, Thomas Rose, Sebastian Spicher, and Lukas Wittmann for scientific and non-scientific discussions and the fantastic breaks on the roof terrace. I want to especially thank my mentor, Sebastian Ehlert, who taught me a lot about Fortran programming and gave me the opportunity to collaborate on my first solvation project, which paved the way for this thesis. Thanks to Thomas Gasevic, Abylay Katbashev, and Christoph Plett for proofreading this thesis.

I thank my collaborators Constantin Daniliuc, Uwe Hohm, Wolf-Ulrich Palm, Jan Paradies, Roland Schoch, and Benedikt Sieland. It was great working together. Many thanks also go to Tunga Salthammer for allowing me a peek into the world of environmental chemistry and for the great collaboration on many topics. I am grateful to Claudia Kronz for the administrative support throughout my work at the Mulliken Center and to Jens Mekelburger for technical support.

My friends and family deserve thanks for supporting me outside of work, even though I had little spare time. Miriam, thank you for putting up with me when I was desperate because science wouldn't cooperate. Last but not least, I want to thank my parents, who were always there when I needed them, without any conditions.