

Article

Quantum Chemical Microsolvation by Automated Water Placement

Miguel Steiner ^{1,2}, Tanja Holzknicht ¹, Michael Schauerl ¹ and Maren Podewitz ^{1,*}

¹ Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria; miguel.steiner@phys.chem.ethz.ch (M.S.); tanjaholzknicht@gmail.com (T.H.); michael.schauerl@uibk.ac.at (M.S.)

² Laboratorium für Physikalische Chemie, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

* Correspondence: maren.podewitz@uibk.ac.at

Abstract: We developed a quantitative approach to quantum chemical microsolvation. Key in our methodology is the automatic placement of individual solvent molecules based on the free energy solvation thermodynamics derived from molecular dynamics (MD) simulations and grid inhomogeneous solvation theory (GIST). This protocol enabled us to rigorously define the number, position, and orientation of individual solvent molecules and to determine their interaction with the solute based on physical quantities. The generated solute–solvent clusters served as an input for subsequent quantum chemical investigations. We showcased the applicability, scope, and limitations of this computational approach for a number of small molecules, including urea, 2-aminobenzothiazole, (+)-*syn*-benzotriboleone, benzoic acid, and helicene. Our results show excellent agreement with the available *ab initio* molecular dynamics data and experimental results.

Keywords: microsolvation; implicit-explicit solvation; density functional theory; molecular dynamics; grid inhomogeneous solvation theory; bulk phase



Citation: Steiner, M.; Holzknicht, T.; Schauerl, M.; Podewitz, M.

Quantum Chemical Microsolvation
by Automated Water Placement.

Molecules **2021**, *26*, 1793. <https://doi.org/10.3390/molecules26061793>

Academic Editor: Martin Brehm

Received: 25 February 2021

Accepted: 15 March 2021

Published: 23 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Solvents play a pivotal role in solution phase chemistry. The efficiency of many chemical transformations ranging from organic reactions to (transition-)metal catalyzed processes is intricately linked to the solvent in which they are carried out [1–4]. As a result, the chemically relevant solution phase conformations of the reactants, intermediates, and products may differ significantly from their gas phase or solid-state counterparts and can even depend on the specific solvent [5]. This is in particular true for bio- or biomimetic molecules and biochemical reactions that take place in water.

To accurately account for the reactivity in the condensed phase, a reliable characterization is necessary. Despite experimental advances, the achieved temporal and spatial resolution is often not sufficient to decipher the impact of complex solvent effects. Theoretical investigations are an attractive alternative to fill this gap, but the accurate treatment of (explicit) solvation is still an active field of research in computational chemistry [5–8].

Yet, several strategies exist. In the simplest model, only bulk properties of the solvent are considered. These so-called implicit solvation models are computationally cost-effective and have long been applied in computational chemistry [9–12]. However, they fail if explicit solute–solvent interactions are expected to play a role. At the other end of the spectrum, condensed phase approaches exist, which take a large number of solvent molecules into account. Most accurately, solute and solvent are treated quantum-mechanically. However, such *ab initio* molecular dynamics (MD) approaches are quite expensive and limited to relatively small systems [13–15]. As a third strategy, explicit–implicit solvation approaches—also referred to as microsolvation—have emerged as a low-cost alternative. Here, a small number of explicit solvent molecules are incorporated in a quantum chemical calculation [8,16–20].

However, microsolvation approaches implicate new challenges such as how many solvent molecules to include, where to put them, and how to orient them. Until today, numerous methodologies to microsolvation have been reported, which can be largely divided into two categories: bottom-up approaches, where starting from the solute alone, explicit solvent molecules are successively added, and top-down approaches, where starting from a condensed phase calculation—often a classical MD simulation—a few explicit solvent molecules are selected.

Bottom-up approaches comprise, for example, the cluster-continuum approach by Pliego and coworkers, where individual solvent molecules are added and the free energy of solvation, ΔG_{solv} , is calculated quantum-mechanically. The number of solvents that minimizes ΔG_{solv} is then taken for further investigations. Alternative approaches are based on an analysis of the molecular electrostatic potential to identify interaction sites [21–23]. Following a related idea, Sure et al. proposed a protocol based on the analysis of the screening charge densities [24]. Another idea is to start from “idealized” solvent geometries, such as the structures of clathrates, to describe microsolvated amino acids [25]. Simm and coworkers developed a stochastic approach, where interaction surfaces of solute and solvent are identified and randomly combined. Additional conformational sampling ensures generation of a structural ensemble [26]. More recently, global optimization procedures in combination with machine learning techniques have emerged to speed up the generation and identification of the most stable microsolvated clusters [27,28].

However, a general problem arises when taking the solute as starting point to build the microsolvation model: the question is how to define a reliable solute conformation without knowing it a priori. Convergence may be challenged if the structure changes drastically with increasing number of explicitly treated solvent molecules. In addition, evaluation of many solute–solvent cluster configurations may be necessary to identify the most stable arrangement, which is computationally demanding.

Top-down approaches circumvent this issue by starting the investigation from a condensed phase structure. Here, microsolvated solute structures are often derived from snapshots of a classical MD simulation or from Monte Carlo sampling. Typically, all solvent molecules within a certain distance to the solute are included in a follow-up quantum chemical calculation (for selected examples, see Refs. [17,29–31]). Following a slightly different approach, Nocito and Baran developed an averaged condensed phase environment derived from snapshots of classical MD trajectories to be applied in quantum chemical calculations [32].

Several studies showed the practicality of such approaches to microsolvation. For example, solute–solvent cluster models were found to improve the description of electronic excitation spectra, [29,32,33] or of IR and Raman spectra [30]. However, in particular in top-down approaches, the number of explicit solvent molecules to be included in the cluster model is often justified based on the distance to the solute alone. A rigorous criterion to select the number of explicit solvent molecules to be included in the microsolvation model based on physical properties is missing. This can be attributed to the inherent difficulty in condensed phase (classical) MD simulations to define the solute–solvent interaction strength and to assign locations to the solvent molecules [34].

However, in biomolecular simulations, the quantification of solute–solvent interactions and the identification of hydration sites—that is, the location of water molecules that show slow or no movement during the simulation—are an important goal due to its relevance in drug design [35–40]. Several computational tools exist to aid this procedure [37,39–41]. For example, grid inhomogeneous solvation theory (GIST) allows the analysis and thermodynamic quantification of the solute–solvent interaction on a grid over the whole trajectory [42,43]. Prior studies have shown that GIST can be used to identify favorable interaction sites in heterocycles [44,45] or in biomolecules, [43,46,47] for example, in antifreeze proteins [47] and kinases [46]. In the latter studies, the grid points were further analyzed to identify positions with high water density to visualize locations of strongly

interacting solvent molecules. Yet, the only purpose of these initial studies was to enhance visualization, and the analysis determined oxygen positions only.

In the present study, we build on this idea to propose a rigorous protocol to quantum chemical microsolvation by automated water placement based on physical properties. We combined the GIST analysis with an evaluation of the averaged solvent density to place individual water molecules and to quantify their interaction with the solute. A crucial point was the development of an algorithm to assign hydrogen positions to obtain water orientations. Our approach is implemented in a fully automated Python script. The user can choose the number of explicit solvent molecules for subsequent quantum chemical studies based on their position and interaction strength. We chose five different small molecules, which differ in size and polarity, to test our methodology: urea and benzoic acid due to their biological relevance; (+)-*syn*-benzotribooneol because of its ability to act as a water receptor [48]; helicene because of its apolar nature; and 2-aminobenzothiazole, which can undergo water-assisted tautomerization [49]. All of these molecules have previously been investigated either experimentally by rotational spectroscopy as well as X-ray and neutron diffraction analysis or theoretically by quantum chemical or molecular dynamics studies [48–59].

After a brief introduction to GIST and to our methodology, the capabilities of our top-down approach to quantum chemical microsolvation are illustrated in the example of the previously introduced five small molecules. Thereby, we aimed at including a variety of compounds that differ in size and polarity.

2. Grid Inhomogeneous Solvation Theory (GIST)

Grid inhomogeneous solvation theory (GIST) is a method to calculate the free energy of solvation for a fixed solute conformation by evaluating (local) thermodynamics—that is, solute–solvent interactions—on a grid [42,43,46,47]. While a detailed derivation of the GIST formalism can be found elsewhere, [42,43] we briefly introduce the theoretical foundations here.

In GIST, the equations of inhomogeneous solvation theory (IST) [60] are solved numerically by sampling a constrained solute in solution via MD simulations and assigning solvent molecules to a three-dimensional grid. The result of this analysis is a three-dimensional density map of the solvent around the solute. From this map, the free energy of solvation ΔA_{solv} can be calculated as a sum over the individual contributions of all grid voxels r according to Equation (1).

$$\Delta A_{\text{solv}} = \sum_r \Delta A_r \quad (1)$$

The free energy of solvation of each grid voxel r is split into an energy and an entropy term as described in Equation (2).

$$\Delta A_r = \Delta E_r - T\Delta S_r \quad (2)$$

The energy term is further split into solute–solvent (ΔE_{sw}) and solvent–solvent (ΔE_{ww}) interactions (see Equation (3)).

$$\Delta E_r = \Delta E_{\text{sw}} + \Delta E_{\text{ww}} \quad (3)$$

As our analysis is limited to water only, we refer to the solvent by the subscript “w.” The entropy is split into a translational and an orientational term (Equations (4) and (5)), which are calculated as a Shannon entropy approximated by a nearest-neighbor approach,

$$S_r^{\text{trans}} = R \left(\gamma + \frac{1}{N_r} \sum_{i=1}^{N_r} \ln \left(\frac{N_f \rho^0 4\pi \cdot d_{\text{trans}}^3}{3} \right) \right) \quad (4)$$

$$S_r^{\text{orient}} = R \left(\gamma + \frac{1}{N_r} \sum_{i=1}^{N_r} \ln \left(\frac{N_f (\Delta\omega_i)^3}{6\pi} \right) \right) \quad (5)$$

with R being the ideal gas constant, N_r the total number of solvent molecules in voxel r during the simulation, and γ Euler's constant, which corrects for an asymptotic bias. N_f is the number of frames of the simulation, ρ^0 is the reference density of the solvent model, d_{trans} is the closest distance and $\Delta\omega$ is the closest angular distance between two solvent molecules.

By quantifying the individual enthalpic and entropic contributions to the free energy of solvation, GIST allows us to identify interaction sites, or rather grid voxels, which show strong interactions with the solute. Identification of such sites is essential for the location of solvent molecules in a microsolvation approach.

3. Computational Methodology

3.1. Computational Protocol to Microsolvation

Our computational protocol to quantum chemical microsolvation is shown in Figure 1. It started with sampling the solute in solution by molecular dynamics (MD) simulations (I). Second, we performed a grid inhomogeneous solvation theory (GIST) analysis of the obtained trajectory (II) to identify favorable interaction sites with the solute. Based on the interaction sites and the water density of the trajectory, individual water molecules were placed and oriented (III) and then ranked according to their interaction strength with the solute (IV). As a last step, the user could select the number of water molecules for subsequent optimizations (V).

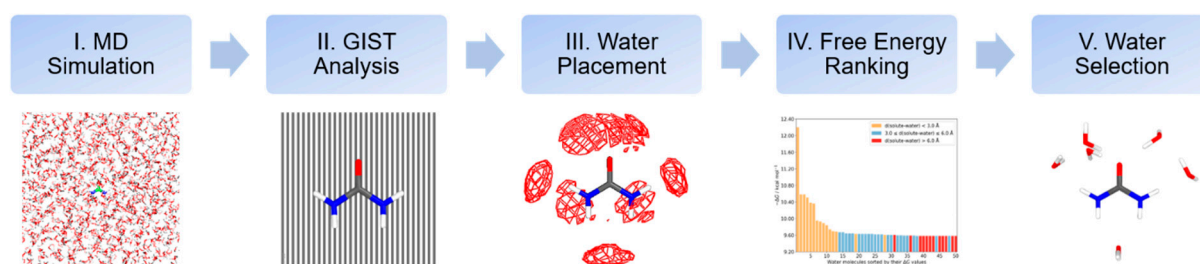


Figure 1. Workflow for the automated generation of microsolvated structures retrieved with our Free Energy Based Identification of Solvation Sites (FEBISS) algorithm from molecular dynamics (MD) data and grid inhomogeneous solvation theory (GIST) analysis.

The workflow after the simulation was automated and autonomously handled by our program Free Energy Based Identification of Solvation Sites (FEBISS).

MD Simulations and GIST Analysis. Prior to any analysis, an MD simulation of the solute in the solvent (here, in water) had to be performed. In case the solute showed significant dynamics, a clustering of the trajectory had to be performed to identify the most populated conformations. The most stable solute conformation or the cluster representatives were then simulated with the solute coordinates being constrained as required for GIST. The MD simulations can be performed with any software (any force field and water model) as long as the written topology and trajectory files can be read by the freely available open-source software Cpptraj [61].

The obtained trajectory was then analyzed with GIST to calculate the thermodynamic contributions of the solute–solvent interaction on a grid (with a spacing of 0.5 Å). As outlined above, GIST discretizes the trajectory and allows to identify solute–solvent interaction sites (grid voxels) with specific solvation thermodynamics. For the purpose of our computational protocol to microsolvation, we were solely interested in interaction sites (grid voxels) that showed a favorable (negative) free energy of solvation ΔA_r and a high water density.

Water Placement. In our protocol, solvent positions were assigned based on the averaged solvent density around the solute and the discretized free energies of solvation. First, the algorithm identified the grid point of highest water density. A water molecule was placed at this position and the density of one water molecule was subtracted from

this grid point. If the density at this grid point was lower than that of a single water molecule, neighboring voxels were included until the sum of their densities suffices. The reference density of one water molecule was taken from the applied water model, here TIP3P [62]. The assignment procedure was repeated until 95% of the total water density was allocated to individual water atoms. Second, the free energy of the selected grid voxel was density-weighted and assigned to the positioned water molecule. This ensured a more accurate representation of the free energy of solvation in the case of insufficient sampling. The water molecules were assigned the free energy value from the grid voxel, where the oxygen nucleus was placed.

By this procedure, water molecules at points of high density and favorable free energy could be identified. Next, positions of the protons were retrieved by an algorithm sketched in Figure 2. Here, the positions of all protons were analyzed for each frame of the simulation. These proton positions were binned onto a smaller grid with a spacing of 0.1 Å, as shown in Figure 2A. The first proton was then placed on the voxel with the highest occupation number. For the second proton, our algorithm only considered positions resulting in a water angle that deviated from the equilibrium angle by less than 5°, as shown in Figure 2B. Within this spherical segment, the second proton was placed at the voxel with the highest occupation number (Figure 2C). The specific value of the H–O–H angle depends on the applied water model in the simulation. For our study, we used the TIP3P water model, [62] with an equilibrium H–O–H angle of 104.5°.

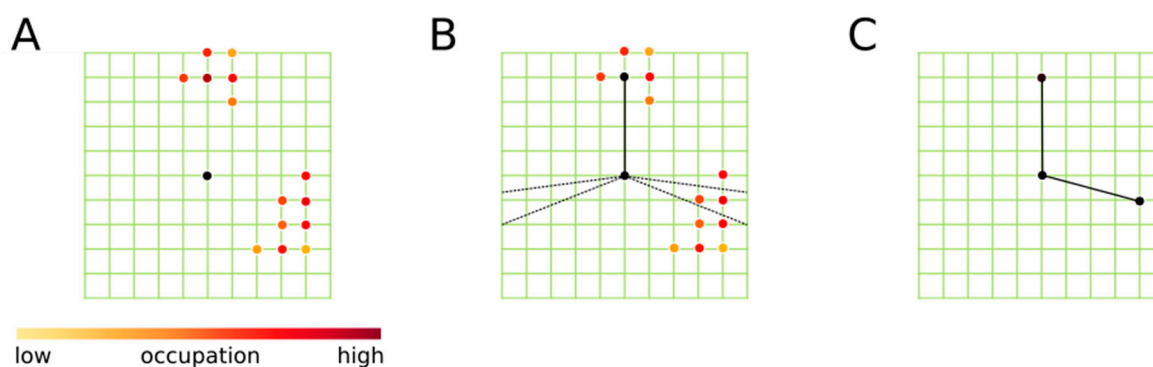


Figure 2. Schematic representation of the hydrogen-positioning algorithm. (A) The black center dot represents the oxygen atom, whereas the color-coded dots represent the occupation number of the protons on the grid. (B) Placement of the first proton at the highest grid occupation. To obtain a reasonable water geometry with an H–O–H angle of $104.5 \pm 5.0^\circ$, the position of the second proton is limited to the space in between the dotted lines, which corresponds to a spherical segment in three dimensions. (C) Resulting water orientation.

Free Energy Ranking. We assigned each placed water molecule the free energy value ΔA of the voxel where the oxygen was placed, weighted by the density of that voxel. The water molecules were then sorted by their ΔA values and displayed as a bar chart, where the height denotes their free energy and a color coding indicates their distance to the solute. We may categorize them into three groups: (i) water molecules with a distance $< 3 \text{ \AA}$, color-coded in orange: such water molecules typically describe the first solvation shell in hydrophilic regions of the solute; (ii) water molecules with distances between $3 \text{ \AA} < d < 6 \text{ \AA}$, color-coded in blue: such molecules comprise either the second solvation shell in hydrophilic regions of the solute or the first solvation shell in hydrophobic regions; and (iii) water molecules with $d > 6 \text{ \AA}$, depicted in red: they are typical “bulk-like” molecules with almost no interaction with the solute.

Water Selection within the Graphical User Interface (GUI). A bar chart displays the ranking of the placed water molecules and serves as a graphical user interface (GUI). Users can select the water molecules for the microsolvated structure based on the interaction energy (solvation free energy) and the distance to the solute by clicking on the individual

bars. Complementary radial distribution functions (RDF) plots of the simulation including integral values can further aid the user in the selection of the number of solvent molecules.

Implementation and Usage. To obtain microsolvated solute structures with our computational protocol, any MD simulation package can be used as long as it generates a trajectory and topology file that can be converted to be readable by the freely available open-source software Cpptraj [61]. Most conveniently, this can be achieved with the Amber package [63,64].

Besides the MD simulation, our Python program FEBISS handles the entire workflow as outlined above. It calls the GIST analysis and performs the newly implemented water placement—both steps are integrated in the C++ open-source software Cpptraj [61]. It generates the interactive bar charts so that the user can select the number of water molecules; it saves the solvated structures as pdb files and automatically displays them in PyMOL [65].

FEBISS is available on Github [66] and via pip for Python3 free of charge. To increase performance, our protocol makes use of the recently published GPU-accelerated GIST version, [67] while patch are files available on Github [68]. Our algorithm is directly implemented within the GIST analysis and can be invoked with the additional keyword *febiss*. Our Python program is able to install all dependencies automatically and performs the GIST analysis and water placement by calling Cpptraj. The analysis and the interactive bar chart can also be customized with an easy-to-use yaml input file.

3.2. Molecular Dynamics Simulation Details

All structures were generated with the Gaussian 16 program, [69] with the exception of (+)-*syn*-benzotriboorneol, which was retrieved from the available crystal structure (CDCC code: 690687) [48]. The simulation setup, including the parameterization of the systems, as well as the final production runs were performed using the Amber18 simulation package [63,64]. For the parametrization, restrained electrostatic potential (RESP) charges [70] were calculated with HF/6-31G* for each system, using the Gaussian 16 program [69]. For the sake of comparison AM1-BCC charges [71] were also calculated for urea. This step was followed by the assignment of the atom types to all structures, using the generalized Amber force field (GAFF) [72] and the antechamber module of AMBERTools19 [64]. The systems were then solvated by applying an octahedral TIP3P [62] water box, maintaining a minimum distance of 20 Å between the solute and the edges of the box. The solvated systems were subjected to an equilibration run, following a thorough protocol by Wallnöfer et al. [73]. The final production runs were carried out in NVT ensembles, where the number of particles N , the volume V and the temperature T are kept constant with $T = 300$ K, using the GPU implementation of the pmemd module of AMBER 18, [64,74] whereby the temperature was controlled by applying the Langevin thermostat [75]. To achieve a higher performance, periodic boundary conditions and the smooth particle-mesh Ewald approach [76] were applied in all simulations. Additionally, the SHAKE algorithm [77] was used to constrain bonds involving hydrogen atoms. Regarding the restrained NVT runs for the GIST analysis, the whole solute was restrained with a force constant of 100 kcal/mol Å². The time step was set to 2 fs, and snapshots of the systems coordinates were saved every 10 ps. The final trajectories comprised 30,000 frames, which corresponds to 300 ns of NVT MD simulation time.

3.3. Data Analysis and Reference Settings

For the GIST analysis, we applied a 30 Å × 30 Å × 30 Å cubic grid with a spacing of 0.5 Å to obtain a reasonable balance between accuracy and convergence [43]. As all of the thermodynamic terms are state functions, reference values corresponding to the pure solvent state were needed. Throughout this paper, the reference solvent-solvent interaction was set to $E_{\text{ww}} = -9.533$ kcal mol⁻¹ (taken from the TIP3P model), [62] whereas E_{sw} , S_{trans} , and S_{rot} were all set to zero.

All data analysis was handled by our program FEBISS, which calls the Cpptraj program [61] from the Amber tools package [64]. For the sake of comparison, we applied the

3-dimensional reference interaction site model (3D-RISM), [78] using the Solvent Analysis tool implemented in the MOE software [79]. To visualize the calculated 3D-RISM water densities, the Solvent Analysis Grid panel was used. We set a minimum distance of 15 Å between the solute and the edges of the solvent box and defined a tight convergence criterion. The remaining parameters were kept to their default values. We selected a cut-off of 2.9 Å and 1.9 Å for the RDFs of the water oxygen and hydrogen atoms, respectively, which aligns with the maximum RDFs determined in previously published results [50,52].

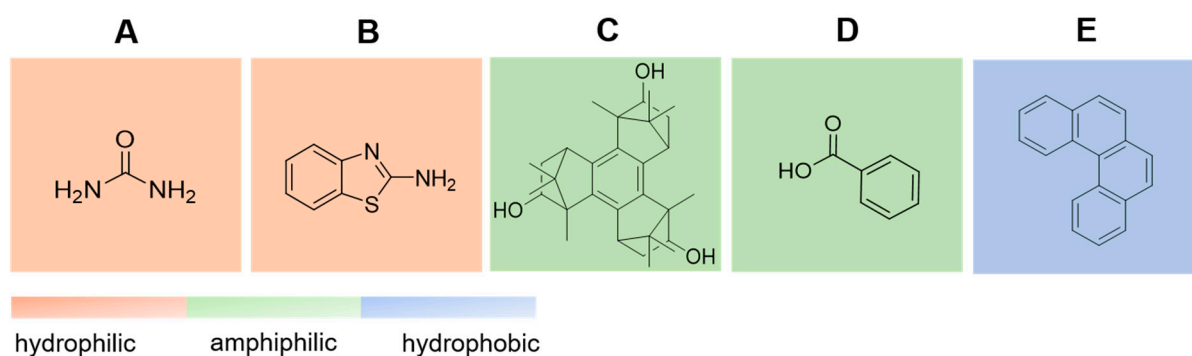
3.4. Quantum Chemical Calculations

Initial microsolvated structures retrieved from the FEBISS analysis were subjected to an optimization using density functional theory (DFT). All calculations were performed using the Turbomole electronic structure package (Version: 7.3) [80,81]. To comply with the already published data, all systems were optimized with the hybrid density functional B3LYP [82–84] and the triple-zeta basis set def2-TZVP [85]. Dispersion contributions in the DFT calculations were invoked by empirical atom-pairwise dispersion corrections with the Becke–Johnson damping [86,87]. In addition to the explicitly described water molecules, bulk solvent effects were treated implicitly, using the conductor-like screening model (COSMO) [88,89] as implemented in Turbomole. A dielectric constant of $\epsilon = 78.5$ was predefined for all calculations. In this approach, the solute is placed in a cavity of a dielectric continuum. The cavity is defined by the solvent accessible surface obtained by applying increased van der Waals radii around each atom. In the case of explicit–implicit solvation models, this is done for each explicitly described molecule (solute and solvents). It has to be ensured that the cavity is meaningful and closed, and this was the case in all our calculations. Analysis of the Hessian verified that all optimized structures only had positive eigenvalues and were indeed energy minima.

The optimized geometries of the microsolvated clusters were depicted with the molecular visualization program PyMOL [65].

4. Results

We tested the applicability and accuracy of our automated solvent placement protocol on the following five systems: urea, 2-aminobenzothiazole, (+)-*syn*-benzotriboorneol, benzoic acid, and helicene (see Scheme 1). The results of each investigated compound are discussed in order of decreasing polarity.



Scheme 1. Lewis formulae of the molecules included in our test set. The structures are ordered according to their polarity, starting with the most hydrophilic and ending with the most hydrophobic system: urea (A), 2-aminobenzothiazole (B), (+)-*syn*-benzotriboorneol (C), benzoic acid (D), and helicene (E).

4.1. Urea–Water Complexes

We started the analysis with the urea molecule, which is the smallest compound of our test library. A 300 ns restrained NVT simulation was performed and the trajectory was subjected to a GIST analysis. The main steps of the FEBISS algorithm are depicted in Figure 3A: This process includes the identification of areas with high water density

(Figure 3A (left)), the placement of water oxygen atoms (Figure 3A (middle)), and the assignment of favorable water hydrogen orientations (Figure 3A (right)). At the end of the FEBISS calculation, thermodynamic parameters for all water molecules were retrieved and the solvent molecules were ranked according to their free energy values. The resulting bar chart is depicted in Figure 3B. This graph allows the identification of thermodynamically favorable water molecules and, at the same time, permits the selection of water molecules with a predefined solute–water distance cut-off. The FEBISS graph of the urea system depicts 13 consecutive water molecules with a solute–water distance below 3 Å (Figure 3B, orange bars). Six of these water molecules have a ΔA value lower than -10.00 kcal/mol. Additionally, we can point out one strongly bound water molecule with a ΔA value of -12.20 kcal/mol.

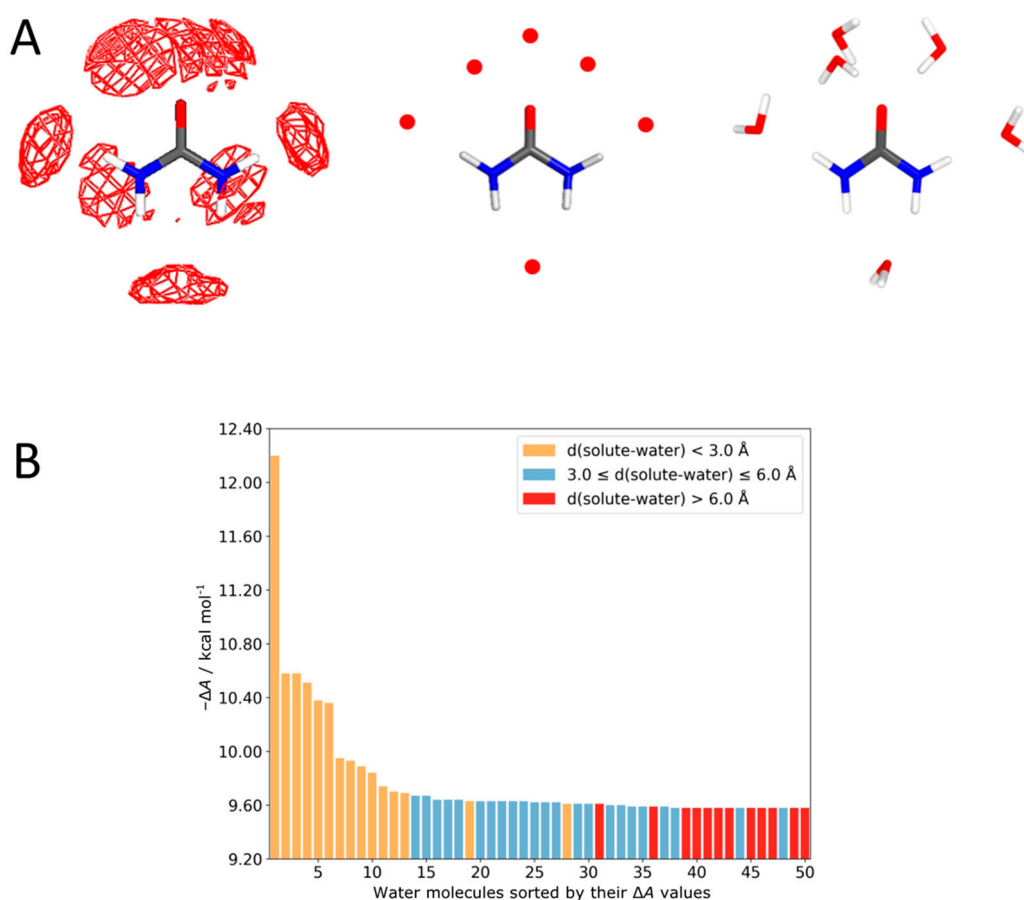


Figure 3. (A) Representation of the automated water placement algorithm. Left: Urea’s water density map. Middle: Ascertained water oxygen positions. Right: Final solvated complex with appropriate water hydrogen orientations. (B) Ranking of the automatically placed water molecules around the urea molecule. The free energy of the first 50 water molecules is depicted. The distances between solute and water are color-coded, whereby orange denotes a solute–solvent distance of less than 3 Å; blue, a solute–water distance between 3 and 6 Å; and red, a solute–water distance of more than 6 Å.

Based on the FEBISS bar chart, we selected three urea–(H₂O)_{*n*} complexes, with *n* corresponding to 1, 6, and 13 water molecules. Both the urea–(H₂O)_{*n*} complexes directly retrieved from the FEBISS calculation and the urea–(H₂O)_{*n*} complexes after DFT optimization with the COSMO solvation model are represented in Figure 4. The lowest ΔA value (FEBISS graph in Figure 3) matches a single water molecule located in between the two urea amide groups (Figure 4A and Figure S2), with $\Delta A = -12.20$ kcal/mol. The interaction is characterized by a bifurcated hydrogen bond, formed by the water oxygen atom and the amide hydrogen atoms of urea. Regarding the urea–(H₂O)₆ complex (Figure 4B), we can observe three additional water molecules coordinating urea’s carbonyl oxygen in an

elliptical configuration with $\Delta A = -10.50$ kcal/mol (compare also Figure S2). All three water molecules interact directly with the urea oxygen atom, whereby the average distance for the $O_{\text{urea}}-H_{\text{water}}$ atom pair is 1.87 Å and 1.79 Å in the non-optimized and optimized water cluster, respectively. Furthermore, we can observe two more water molecules in the region that encompasses the urea oxygen atom and the amide groups. Concerning the urea-(H₂O)₁₃ complex (Figure 4C), we can see that seven additional water molecules are distributed around the urea molecule; these do not directly interact with the solute, but form hydrogen bonds with the already placed water molecules.

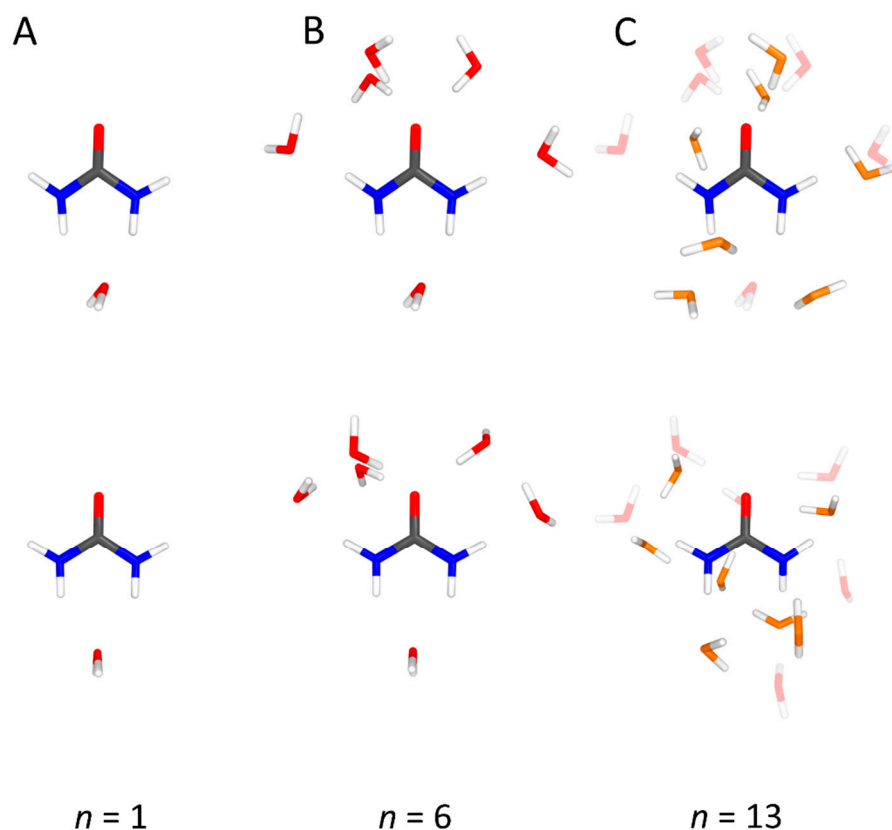


Figure 4. Representation of the urea-(H₂O)_n complexes before (upper row) and after (lower row) quantum chemical optimization with B3LYP/def2-TZVP/D3 in implicit solvent, with $n = 1$ (A), $n = 6$ (B), and $n = 13$ (C) water molecules. The six water molecules with the lowest ΔA value are colored in red, while the additional seven water molecules in the urea-(H₂O)₁₃ complex are colored in orange.

It is noteworthy that the FEBISS chart and the water positioning looked different when AM1-BCC [71] rather than RESP [70] charges were used (Figure S3). With AM1-BCC charges, four water molecules rather than three were found to be coordinated to the oxygen and one was localized between the amide groups.

When comparing radial distribution functions (RDFs) from our simulation to previously published *ab initio* molecular dynamics data, [52] we see a good agreement (see Figure S1 in Supplementary Materials). This fact proved that the sampling time of our simulation was sufficient, and that classical MD simulations are justified as input for our water placement algorithm FEBISS.

For further validation, we compared our results with the urea-water interaction sites determined by the 3D reference interaction site model (3D-RISM). In Figure 5, an overlay of the optimized urea-(H₂O)₆ complex and the solvation sites obtained with the 3D-RISM methodology is shown. We find a very good agreement between the water positions obtained with our FEBISS algorithm and the high-density water locations determined with the 3D-RISM calculation. Both methodologies agree on the existence of a water molecule in between the two amide groups, as well as on water molecules surrounding the urea

oxygen atom in a hemispherical manner. However, with our FEBISS algorithm, we were able to identify two more explicit water positions, which span the region between the urea oxygen and urea amide groups, which are less recognizable in the 3D-RISM results.

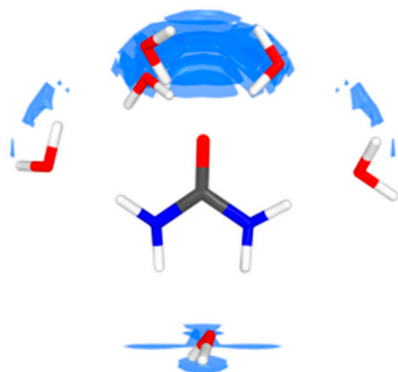


Figure 5. Comparison of the water placements around urea obtained with the FEBISS algorithm and the 3D reference interaction site model (3D-RISM) approach. The urea–(H₂O)₆ complex is represented with sticks, while the water densities obtained with the 3D-RISM approach are shaded in blue.

In addition, we investigated the dipole moments of the urea–water clusters. While urea in implicit solvent has a dipole moment of 6.10 D, the addition of one water molecule increased this value to 9.24 D, whereas for the urea–(H₂O)₆ cluster, a dipole moment of 9.57 D was found. On further increasing the number of explicitly treated waters to 13, the dipole moment decreased to 2.98 D.

4.2. Aminobenzothiazole–Water Complexes

2-Aminobenzothiazole (ABT) served as another test system in our study (see Scheme 1). It exists in two tautomer conformations, the amino (T1) and the trans-imino (T2) structural isomers, depicted in Figure 6. Previous studies have shown that water assists the tautomerization process and the conversion from T1 to T2. A microsolvation model was required to calculate meaningful transition state structures and reaction barriers, while the number and location of the explicitly treated water molecules was determined by trial and error [49]. Our goal was to investigate whether our microsolvation protocol is able to yield reasonable microsolvated structures. Therefore, we analyzed both tautomeric forms with our FEBISS algorithm and assessed the number, location, and orientation of individual solvent molecules as well as their impact on the proton transfer.

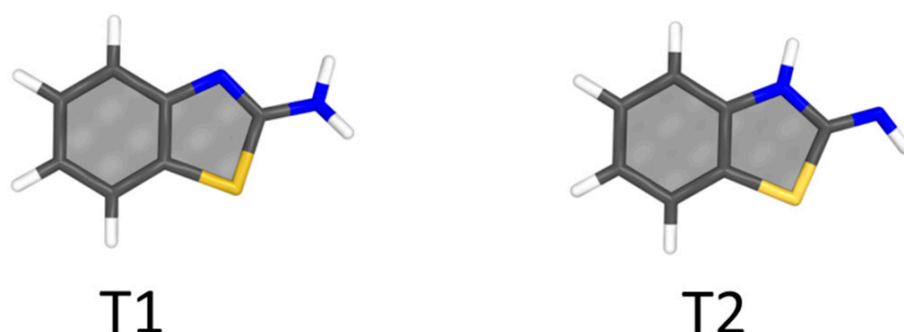


Figure 6. Quantum-chemically predicted conformations of T1 and T2. The relative electronic energy difference $\Delta E_{\text{rel}} = E(\text{T2}) - E(\text{T1})$ is 5.4 kcal/mol in the gas phase and 5.6 kcal/mol in the implicit solvent when calculated with B3LYP/def2-TZVP/D3.

Initial DFT structure optimizations in the implicit solvent confirmed the amino isomer (T1) to be more stable—by 5.4 kcal/mol in the gas phase and by 5.6 kcal/mol in the implicit solvent.

The results of our microsolvation analysis are presented in Figure 7, where the FEBISS bar charts for both tautomers are shown. For T1, we observe six consecutive water molecules with a solute–water distance below 3 Å, of which two water molecules have a ΔA value lower than -10.00 kcal/mol (Figure 7A). Regarding T2, there are also six water molecules ranked in a row, with a solute–water distance smaller than 3 Å (Figure 7B), while three water molecules have a ΔA value lower than -10.50 kcal/mol. If we inspect these water positions in more detail (Figure 7C), we find for T1 that the most tightly bound water is a distal water with a coordination to the amine nitrogen, whereas the next two water molecules are pre-aligned to form a network between the two nitrogen nuclei to facilitate proton transfer. In the case of the trans-imino isomer T2 (Figure 7D), the first two water molecules with $\Delta A = -10.60$ kcal/mol can be considered to be pre-aligned to form a water chain and to facilitate proton transfer. The third water also coordinates to the imine nitrogen, but is almost perpendicular to the aromatic ring.

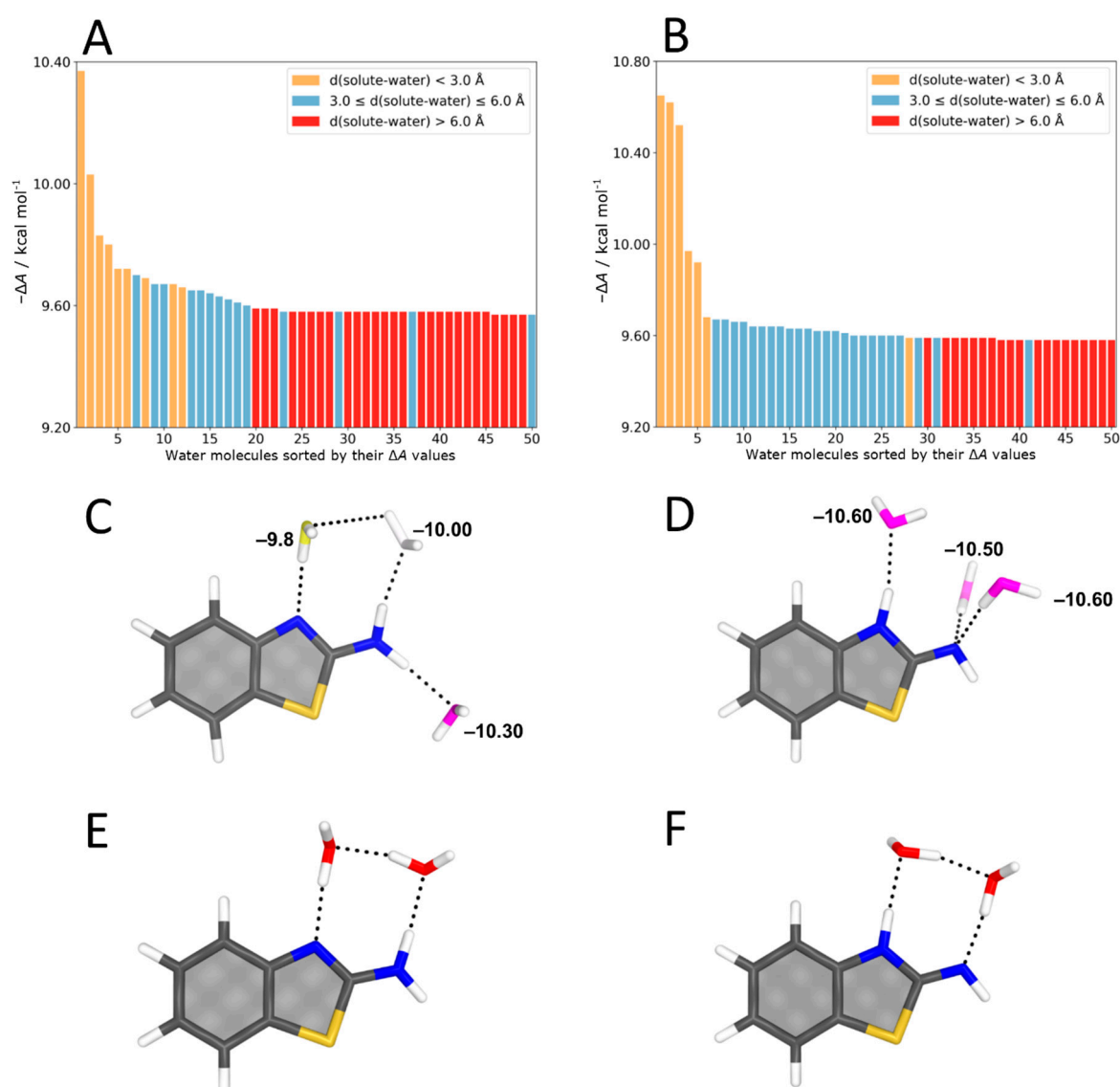


Figure 7. Ranking of the automatically placed water molecules around the T1 (A) and the T2 (B) conformer of 2-aminobenzothiazole. Representation of T1 (C) and T2 (D) water clusters with three explicit water molecules; water molecules are color-coded according to their free energy from pink (very favorable) to yellow (less favorable); B3LYP/def2-TZVP/D3 optimized conformers of T1-(H₂O)₂ (E) and T2-(H₂O)₂ (F).

Our FEBISS algorithm clearly predicts two water molecules to form an intermolecular hydrogen bonded chain to facilitate the proton transfer from the amino to the imino tautomer, and vice versa. Hence, we set up the microsolvated structures with two explicit water molecules to yield T1-(H₂O)₂ and T2-(H₂O)₂ and optimized them (Figure 7E,F). The microsolvation shifts the equilibrium by decreasing the energy difference between the two isomers $\Delta E_{\text{rel}} = E(\text{T2}-(\text{H}_2\text{O})_2) - E(\text{T1}-(\text{H}_2\text{O})_2)$ to 2.9 kcal/mol (gas phase) and 4.5 kcal/mol (implicit solvent), respectively—a finding in agreement with previously published results [49]. More strikingly, however, the predicted number of two water molecules forming a water chain and aiding the proton transfer process coincides with previous results found by trial and error [49]. Wazzan et al. reported that two water molecules minimize the barrier for proton transfer from T1 to T2, whereas chains consisting of one or three water molecules result in higher barriers.

4.3. Benzotriborneol–Water Complexes

Our largest test compound was (+)-*syn*-benzotriborneol, which was synthesized by Fabris et al. [48]. The structure can be characterized as a C₃-symmetric triterpene, with a central aromatic ring and three hydroxyl groups (Scheme 1). By means of X-ray crystal structure and neutron diffraction analysis, it was proven that the molecule acts as an efficient water receptor. In total, three water molecules were found in the X-ray crystal structure of (+)-*syn*-benzotriborneol. Out of those three water molecules, one molecule is bound in the center of the molecule's cavity, forming hydrogen bonds with the surrounding hydroxyl groups. We aimed at reproducing those experimental outcomes and subjected (+)-*syn*-benzotriborneol to our FEBISS calculation.

The resulting bar chart is shown in Figure 8A, where the first 50 water molecules—ranked according to their ΔA values—are shown. Out of those 50 water molecules, six consecutive water molecules have a solute–water distance below 3 Å. Most strikingly, we can observe a single water molecule with a ΔA value of around −13.00 kcal/mol, which clearly stands out from the rest of the water molecules.

To compare the X-ray crystal structure with the FEBISS water position, we plotted both in Figure 8B. Most remarkable, our microsolvation protocol exactly predicted the position of the most strongly bound central water molecule ($\Delta A = -13.00$ kcal/mol), where the water molecule and the hydroxyl groups of (+)-*syn*-benzotriborneol both act as hydrogen bond donor and acceptor. As visible from the overlay of the FEBISS and the crystal structure (Figure 8B (right)), the two water molecules are in almost perfect alignment with a root mean square deviation (RMSD) of 0.215 Å. Indeed, DFT reoptimization hardly changed the structure (Figure 8C (right)), with averaged distances of the water and the three hydroxyl groups being 1.87 Å, 1.82 Å (crystal), and 1.88 Å (FEBISS), respectively. Considering the next five predicted water positions, Figure 8B (middle) reveals that they form hydrogen bonds with the three hydroxy groups. They show also good agreement with the additional two crystal waters. The RMSD between the second most favorable FEBISS water and the crystal water is 0.481 Å, while it is 1.421 Å between the third crystal water and the nearest FEBISS water.

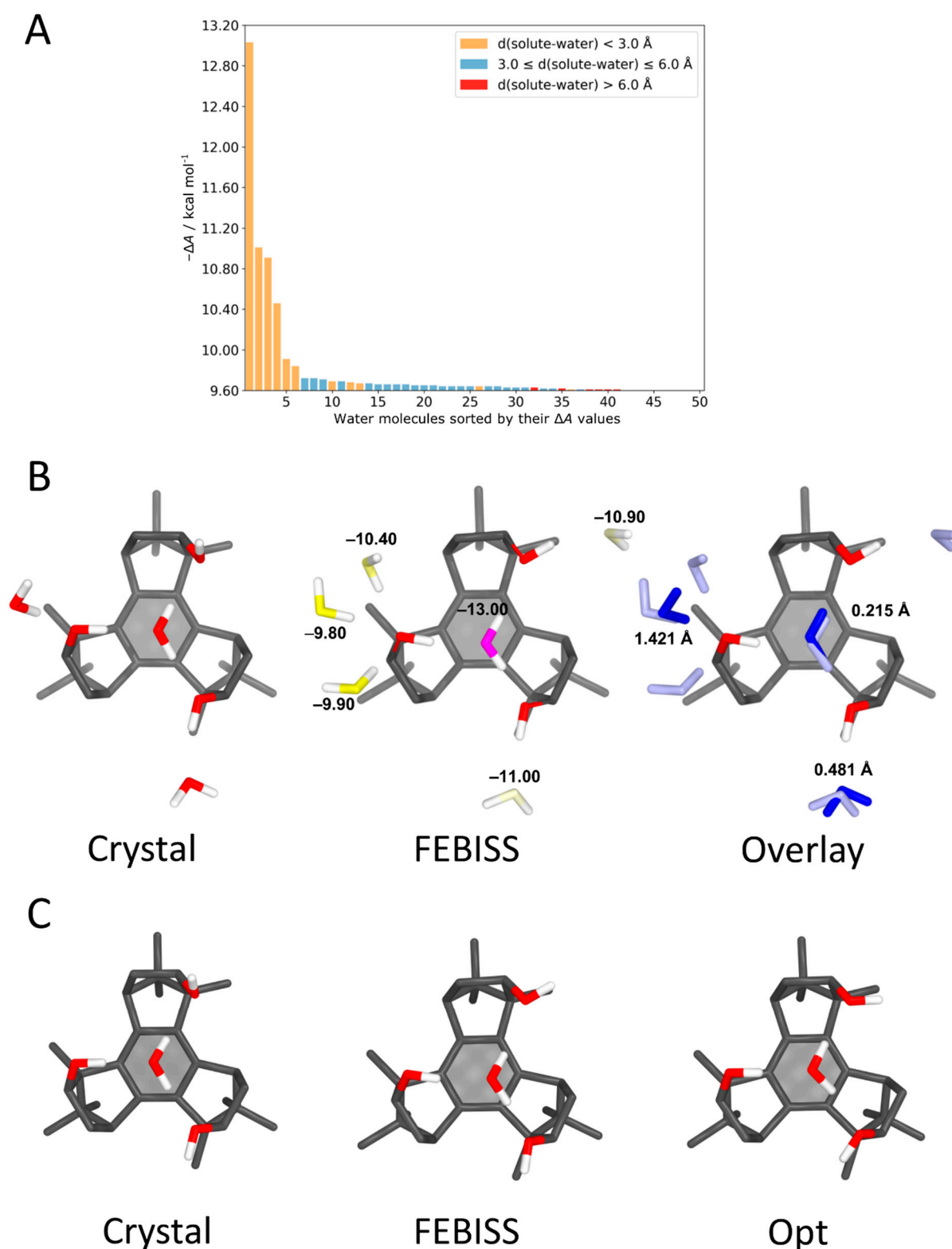


Figure 8. (A) Automated water placement of (+)-*syn*-benzotriborneol. (B) Comparison of the crystal structure with crystal waters (left) and our FEBISS water positions (middle) as well as an overlay of the two (right). FEBISS water positions are depicted in light blue, and crystal waters in dark blue; the depicted values indicate free energies of solvation (middle) and root mean square deviation between the crystal and the nearest FEBISS water (right). (C) Comparison of the crystal structure with one explicit water (left), the FEBISS water position (middle) and the B3LYP/def2-TZVP/D3 optimized (+)-*syn*-benzotriborneol-(H₂O) cluster structure (right).

4.4. Benzoic Acid–Water Complexes

Due to the amphiphilic nature of benzoic acid (BA), we included the molecule in our study to investigate whether our algorithm can distinguish between the hydrophilic and hydrophobic site. To assess the algorithm's performance when working with charged molecules, we also studied the deprotonated benzoate.

The hydroxyl group of BA exists in two distinct orientations, which are denoted as *syn* and *anti*, whereas the *syn* conformation is by 6.0 kcal/mol (gas phase) and 3.2 kcal/mol (implicit solvent) lower in energy (compare Figure S4 of Supplementary Materials). This is in agreement with previous experimental and computational investigations, [53–56,58,59] where values range between 6.0 kcal/mol and 6.5 kcal/mol in the gas phase [55,56,59]. The FEBISS bar charts for *syn*-BA and *anti*-BA as well as benzoate are depicted in Figure 9A–C. All three graphs show the ranking of the first 50 automatically placed water molecules. Regarding *syn*-BA and *anti*-BA, we observe three water molecules, with a solute–water distance below 3 Å (Figure 9A). Out of those three water molecules, two show similar ΔA values, while one water molecule stands out with a ΔA value of -11.20 kcal/mol and -10.60 kcal/mol, respectively. On the contrary, the bar chart of the BA anion shows numerous water molecules with a solute–water distance lower than 3 Å (Figure 9C). In total, this includes 15 consecutive water molecules, which have ΔA values between -10.00 and -14.00 kcal/mol. These solvent molecules are almost exclusively located around the charged carboxylate group as visible from the initial microsolvated structures with $n = 1, 2, 3, 6, 9, 15$ displayed in Figure S5 in Supplementary Materials. Only for $n = 15$, we see water molecules being located on top of the phenyl ring.

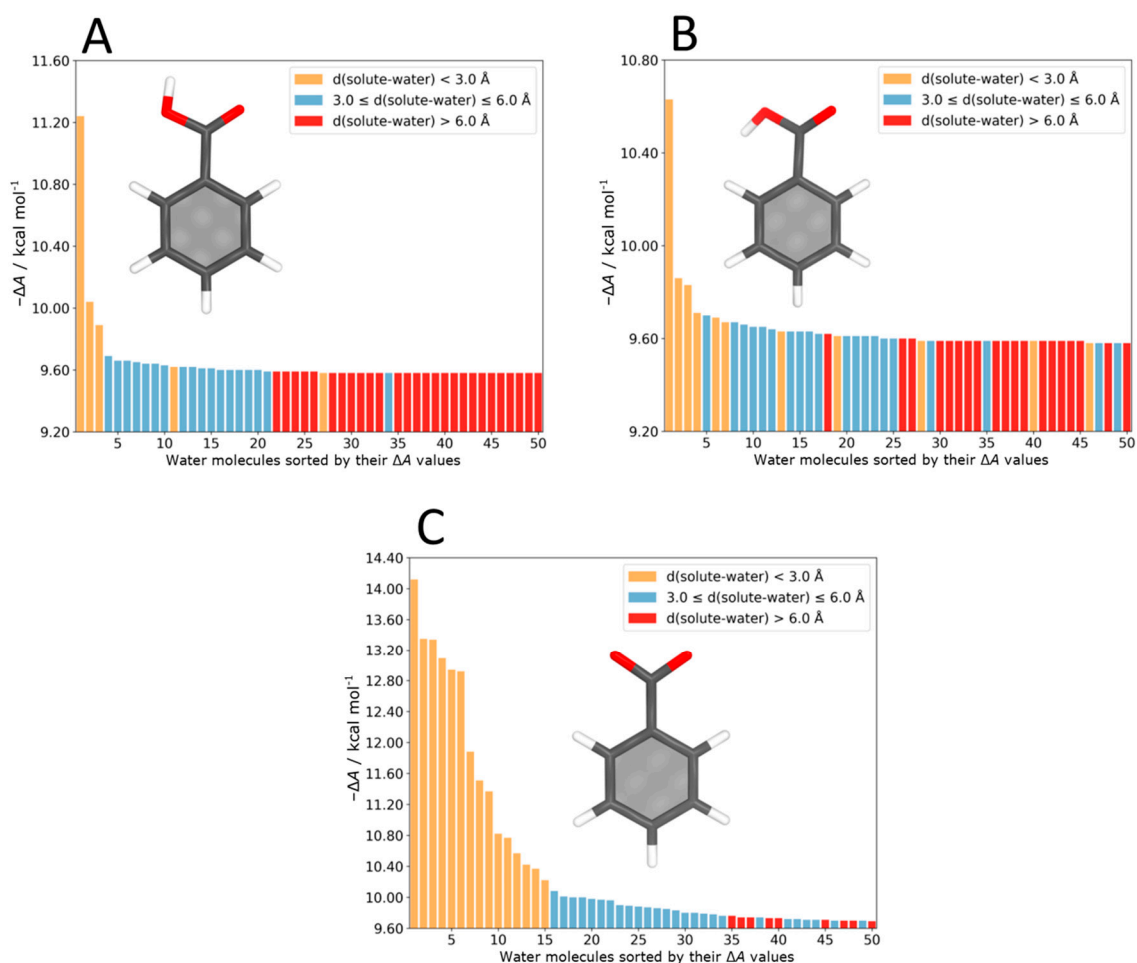


Figure 9. Bar chart of the FEBISS calculation for *syn*-BA (A) and *anti*-BA (B) as well as of benzoate (C).

For the two conformers of BA, we investigated the energy difference between the *syn*- and the *anti*-structure in the gas phase and in implicit solvent for various cluster sizes $\text{BA}-(\text{H}_2\text{O})_n$ ($n = 0-3$). When looking at the listed energy differences of the complexes in the gas phase and in implicit solvent (Table 1), distinct differences in the energies can be found—lacking a systematic pattern: for example, for $n = 1$, a relative energy difference of -9.0 kcal/mol in favor of *syn* is determined for the gas phase in agreement with previous findings of -8.9 kcal/mol, [54] but this value increases to -13.4 kcal/mol for $n = 2$ before collapsing to -2.3 kcal/mol for $n = 3$. The energy spread is less pronounced for implicit solvent calculations (2nd row, Table 1), where energy differences are consistent with the exception of $n = 2$.

The microsolvated clusters optimized with an additional implicit solvent model are depicted in Figure 10. The microsolvated clusters in gas phase are shown in Figure S6 in Supplementary Materials. Almost all gas phase and implicit solvent structures are remarkably similar the only exception being *syn*- $\text{BA}-(\text{H}_2\text{O})$, where the water molecule forms two H-bonds to the solute in the gas phase (Figure S6) and one H-bond in the implicit solvent structure (Figure 10). However, when comparing the *syn*- and *anti*-species with $n = 1$ and $n = 2$ (Figure 10), it can be seen that the locations of the explicit solvent molecules are very different. In case of the $\text{BA}-(\text{H}_2\text{O})_2$ even the number of hydrogen bonds differs between the *syn*- and the *anti*-conformer, which skews the energy difference. For $\text{BA}-(\text{H}_2\text{O})_3$, however, the two structures are reasonably similar. Consequently, the *syn-anti* energy differences in the gas phase and the implicit solvent are also similar with -2.3 kcal/mol and -2.2 kcal/mol, respectively.

Table 1. Relative stability differences between the microsolvated *syn*- and *anti*-conformer of BA in the gas phase (ΔE_{gas}) and in the implicit solvent (ΔE_{solv}). Structures were optimized with B3LYP/def2-TZVP/D3. Relative electronic energy differences are given in kcal/mol.

Structure	BA	BA-(H ₂ O)	BA-(H ₂ O) ₂	BA-(H ₂ O) ₃
$\Delta E_{\text{gas}}(\textit{syn-anti})$	-6.0	-9.0	-13.4	-2.3
$\Delta E_{\text{solv}}(\textit{syn-anti})$	-3.2	-2.7	-6.2	-2.2

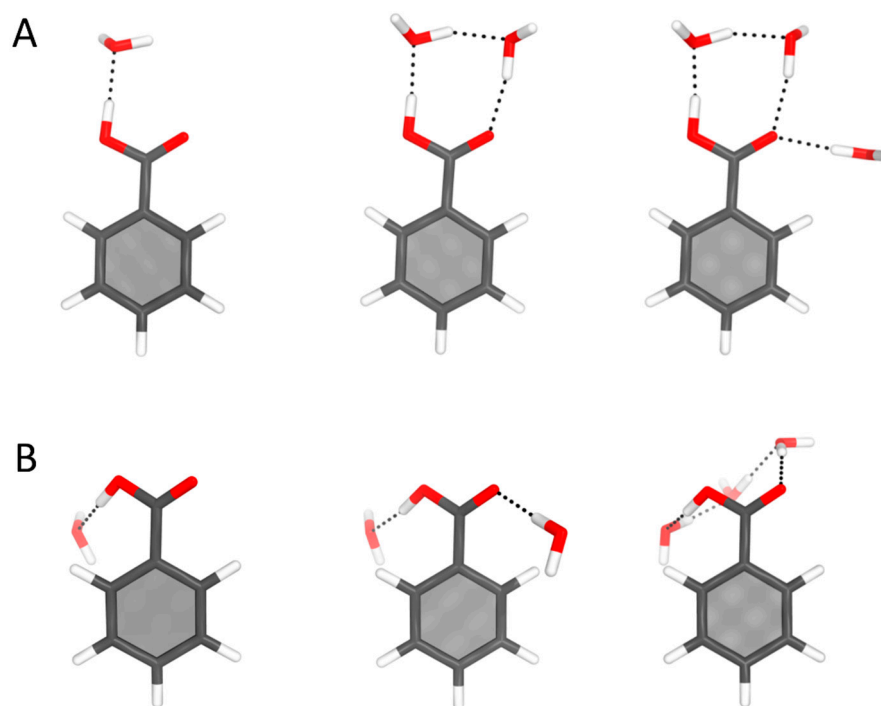


Figure 10. Microsolvated structures of *syn*-BA (A) and *anti*-BA (B) after B3LYP/def2-TZVP/D3 optimization in the implicit solvent.

4.5. Helicene–Water Complexes

The most apolar molecule of our study is [4]-helicene, which belongs to the family of polycyclic aromatic hydrocarbons (PAHs) [90]. Such molecules and their interaction with water are of interest for studying ice-grain formation [51]. [4]-Helicene–water clusters have been previously investigated by Domingos et al., who provided experimental data on the arrangement of a single water molecule around [4]-helicene in the gas phase, using high-resolution rotational spectroscopy. The results of their study served as further benchmark in evaluating the performance of our FEBISS algorithm.

Our automated solvent placement protocol yielded the free energy bar chart depicted in Figure 11A. We observe that none of the water molecules exceeds a ΔA value of -9.70 kcal/mol and that all solute–water distances correspond to values larger than 3 \AA . An exception is the water ranked at position #12 with a solute–water distance of 2.9 \AA , barely below 3 \AA . Due to the small differences in free energy, a representation of [4]-helicene and the first 14 consecutive water molecules from our FEBISS calculation is shown in Figure 11B. It is evident that the algorithm hardly places any water molecules in close proximity to the aromatic rings, but rather locates them in the surroundings of the solute's hydrogen atoms.

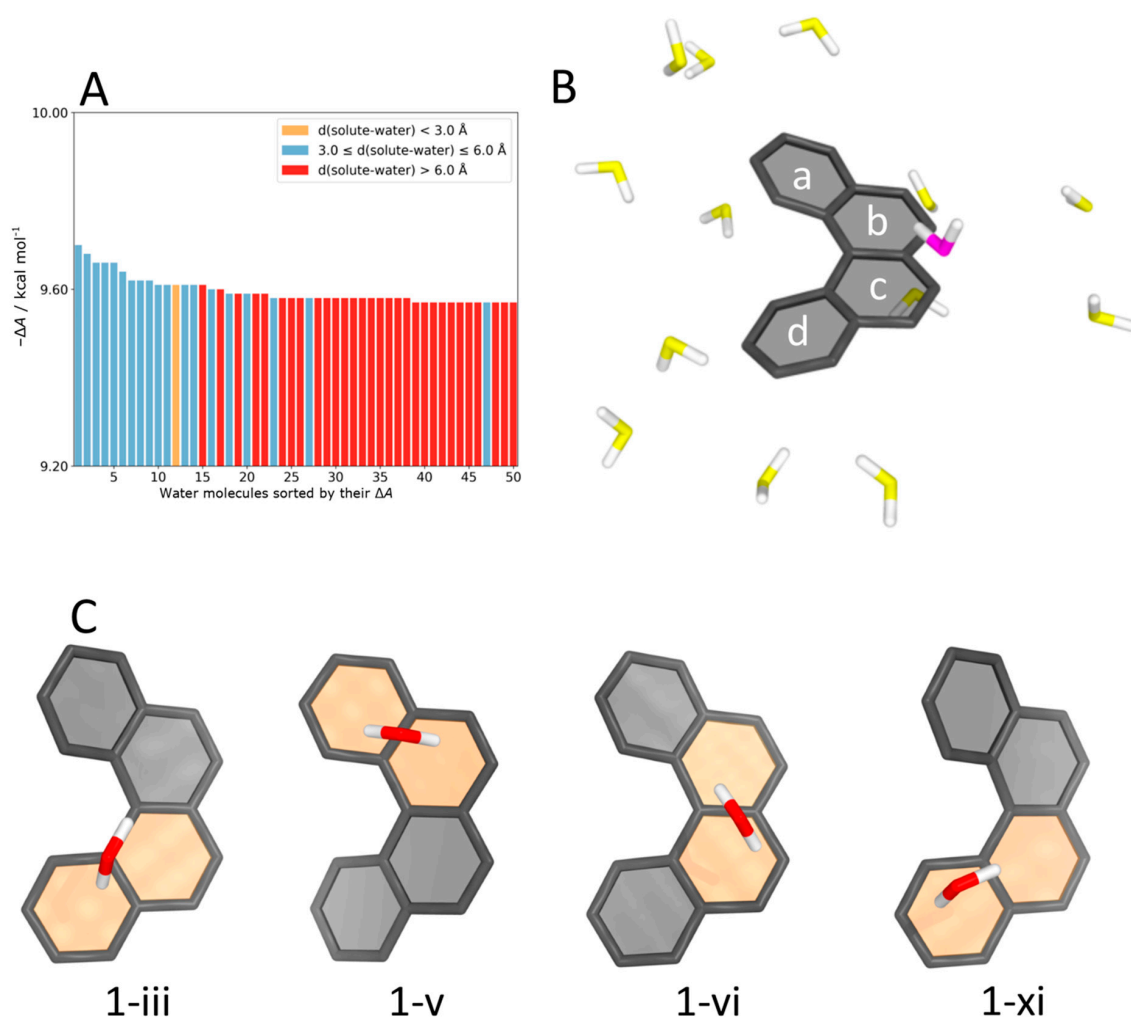


Figure 11. (A) FEBISS graph of [4]-helicene. (B) [4]-helicene with the 14 highest ranked water molecules as obtained from FEBISS, molecules are color-coded according to their ΔA from favorable (magenta) to less favorable (yellow). (C) Representative low energy conformation of the [4]-helicene–(H_2O) cluster after quantum chemical optimization (B3LYP/def-TZVP/D3). The orange area highlights the aromatic ring, which interacts with the explicit water molecule. Roman numbers indicate the rank of the water position in the FEBISS bar chart.

From the FEBISS bar chart and from inspection of the positioned water molecules, we could not infer which water positions would yield low-energy monohydrated [4]-helicene clusters (denoted as [4]-helicene-(H₂O)). Hence, we generated 14 such clusters with the water molecules ranked at position #1 to #14 in the FEBISS graph and subjected them to DFT optimizations. Several low-energy clusters were found (Table S4, Supplementary Materials) and selected low-energy conformers are depicted in Figure 11C. They are denoted as 1-iii ($\Delta E = 0.17$ kcal/mol), 1-v ($\Delta E = 0.15$ kcal/mol), 1-vi ($\Delta E = 0.21$ kcal/mol), and 1-xi ($\Delta E = 0.00$ kcal/mol), whereby the number 1 is a shorthand notation for [4]-helicene-(H₂O) and the small Roman number specifies the rank of the water molecule in the FEBISS bar chart. All other conformers were either identical to these or enantiomers.

The relative electronic energies are very similar for all four complexes and all clusters show a similar water arrangement, with the water hydrogen atoms directed toward one of the aromatic rings. In cluster 1-iii and 1-xi, the hydrogen atoms interact with ring number c and d, respectively. In 1-v, the water is located over rings a and b, whereas in cluster 1-vi, it is located over rings b and c. These results are in agreement with reported experimental and theoretical investigations with the exception of cluster 1-iii [51].

5. Discussion

General Performance. We devised a computational protocol to quantum chemical microsolvation, where the location, orientation, and interaction of individual solvent molecules with the solute are automatically calculated based on MD and GIST. Our methodology provides a rational approach to generate microsolvated clusters based on physical properties. The fully automated workflow is implemented in our Python program FEBISS. The results are presented in a bar chart, where the solvent molecules are ranked according to their free energy. By simultaneously displaying the distances to the solute, the user has two criteria to choose the number of solvent molecules for a following quantum chemical calculation. The algorithm to obtain the solvent orientations is limited to water at the moment but will be extended to arbitrary solvents in the future. Consequently, as of now, our FEBISS program can only generate solute–water clusters.

Of course, the performance of this protocol intrinsically depends on the applied force field and water model used for the initial MD calculations as well as the simulation time. These factors need to be evaluated for each investigated molecule. While the generalized Amber force field (GAFF) in combination with the TIP3P water model—which we used—was shown to yield good results for small organic compounds, [91–93] individual molecules may profit from a more tailored force field [94]. As pointed out, the force field can easily be modified as long as the used software package generates a trajectory and topology file convertible to be readable by Cpptraj.

To obtain accurate results, it is also necessary to assign reasonable partial charges. As previously reported, RESP charges yield better results than AM1-BCC, [93] a finding we could confirm for urea (*vide infra*).

Overall, our top-down protocol to quantum chemical microsolvation yields very good agreement with available experimental and theoretical data; specific findings and limitations will be discussed for each investigated species.

Urea-Water Clusters. Urea's strong interaction with solvent molecules in its close proximity has been recognized by spectroscopic and theoretical investigations, even though the extent of this interaction is still under debate [95]. Theoretical investigations based on MD simulations tend to find higher hydration numbers than experimental studies with up to eight hydrogen bonds [50,96–98]. In contrast, by dielectric electron relaxation spectroscopy Agieienko et al. found a hydration number of only 2 [92]. Our determined hydration pattern of urea suggests up to six strongly interacting water molecules, which is in agreement with previous theoretical studies [50,52,97,98]. In particular, the positions of our water molecules agree with 3D-RISM results [50] and with the areas of hydration determined by the *ab initio* MD study of Weiss and Hofer [52]. The latter study also finds a bifurcating water molecule localized between the two amide groups, which is

consistent with our results. The three oxygen coordinating water molecules found in the here presented microsolvated urea complexes are in alignment with the average number of hydrogen bonds to the urea oxygen (2.07 ± 0.75) predicted by Weiss and Hofer [52]. Still, the number and position of the strongly interacting water molecules is sensitive to the parametrization scheme, in particular, to the assigned partial charges. The water placement obtained by employing AM1-BCC charges (Figure S3) predicted four water molecules interacting with the urea oxygen and one with the two amide groups. This is clearly a poorer agreement with available reference data [50,52] concerning the number and positions of the water molecules.

Considering the dipole moment of urea in water, experimental studies found a value of $\mu_{\text{eff}} = 10.0 \pm 0.1$ D in infinite solution [92]. While the calculated dipole moment of urea in implicit solvent is only 6.10 D, positioning one water molecule in between the amide groups (urea-(H₂O)) already increased μ to > 9 D. The addition of five more waters to yield urea-(H₂O)₆ had little effect on the dipole and for the large cluster, urea-(H₂O)₁₃, μ decreased to 2.98 D. This finding shows that a parallel alignment of solute and solvent dipole moments is beneficial [92]. It may also indicate that out of the 13 water molecules in close proximity to the solute, some might have more bulk-like properties or show high dynamics. Hence, calculating molecular properties from a single urea-(H₂O)₁₃ cluster does not yield accurate results. Explicitly describing these “bulk-like” solvent molecules by one conformation is not adequate.

Aminobenzothiazole–Water Clusters. Concerning the amine–imine (T1 to T2) tautomerization process in ABT, our FEBISS algorithm predicted exactly two waters forming a hydrogen bonded chain and aiding the proton transfer reaction. These two solvent molecules were previously only identified by trial and error [49]. While these two bridging water molecules were correctly identified as strongly interacting solvent molecules by FEBISS, a distal amine-coordinated water was determined to have the most favorable free energy for T1 (see Figure 7). Obviously, to compare stabilities between two configurations, the microsolvated clusters T1 and T2 have to be structurally similar to obtain reliable results. Hence, the user has to select meaningful water positions based on the available data, such as solute–solvent interaction strength and location of solvent molecules.

Benzotriboorneol–Water Clusters. Our FEBISS algorithm predicted the position of the most strongly bound central water molecule of the water receptor (+)-*syn*-benzotriboorneol in quantitative agreement with experimental X-ray and neutron diffraction data [48]. Remarkably, not only the position of the oxygen atom but also the orientations of the hydrogens are in agreement with experiment—a finding that highlights the accuracy and reliability of our microsolvation protocol.

The positions of the next five water molecules are all in close proximity to the three hydroxyl groups as expected based on chemical intuition. Two of the solvent molecules are also in good agreement with X-ray crystal water positions (Figure 8B (right)). However, differences are expected as packing effects and symmetry mates influence the position of crystal waters and the environment in the crystal structure clearly differs from that in solution.

Benzoic Acid–Water Clusters. Concerning the microsolvation of benzoic acid, the FEBISS bar charts (Figure 9) clearly show that our algorithm is able to distinguish between charged and neutral species as indicated by the distinct water placement. The benzoate anion attracts more water molecules to aid charge compensation than the protonated BA species, which is in alignment with chemical intuition. The comparison of the microsolvated *syn*- and *anti*-conformer of BA further illustrates that in any explicit-implicit solvation model, the microsolvated structures have to be similar to be comparable in energy. In particular, the number of hydrogen bonds must be carefully examined. If it differs as, for example, in the BA-(H₂O)₂ cluster, it has to be evaluated whether this is inherent to the specific geometry or whether this is due to additional solvent–solvent interaction as found in our case. Still, energy differences between the *syn*- and *anti*-cluster calculated in the gas phase and in implicit solvation became very similar for three explicitly considered water

molecules. As these are the only water molecules placed in close proximity to the solute ($d < 3 \text{ \AA}$, see Figure 9) by our FEBISS algorithm, it can be assumed that solvation effects of the COOH group are sufficiently described.

Comparing the cluster geometries, the predicted position of the most favorable water molecule in the gas phase optimized *syn*-BA-(H₂O) complex (Figure S6)—bridging the hydroxyl and the carbonyl group—agrees with spectroscopic microwave data [57] and theoretical studies [33,96]. For the implicit solvent optimization of *syn*-BA-(H₂O), no such bridging was found but a single hydrogen bond was formed between BA and H₂O (Figure 10). Such a structure further suggests that a single water molecule is not sufficient to describe the microsolvation environment present in the bulk phase. Larger clusters with $n > 2$ reported in a previous study [99] showed some differences with our structures. However, these deviations can be attributed to the different methodologies to retrieve the microsolvated molecules: Krishnakumar [99] et al. used MD snapshots, whereas we use averaged highly populated positions.

Helicene–Water Clusters. Helicene was the most challenging system for our FEBISS algorithm as the interaction with water was expected to be weak due to its apolar nature. Indeed, all predicted water positions had a distance of $>3 \text{ \AA}$ and their interaction with the solute was barely larger than the value for a typical bulk water (see red bars in Figure 11A). In a pragmatic approach, we generated 14 [4]-helicene-(H₂O) clusters and obtained four low-energy conformers. When comparing our computed results with the rotational spectroscopy data from Domingos et al., [51] we can see striking similarities regarding the binding mode and position of a single water molecule toward [4]-helicene. In agreement with their data, we found the most tightly bound water molecule (cluster 1-xi) to be aligned on top of ring d and c, and also found two more conformers with the water located on top of rings a and b (1-v) and rings b and c (1-vi), respectively. Our fourth cluster, 1-iii, with the water placement similar to 1-xi was not reported by Domingos, but it remains unclear why.

Although the ranking of the water molecules according to their free energy provided little insight into which solvent molecules to choose for subsequent quantum chemical studies, the FEBISS algorithm yielded rational and reasonable starting structures for optimization. Nonetheless, we would like to point out here that the poorer performance for helicene is not a weakness of the methodology per se but can be attributed to the apolar nature of the molecule, its low solubility in water, and potentially lower accuracy of the force field for such species.

Applicability. The FEBISS protocol provides a physically sound basis to obtain microsolvated clusters. However, the assigned water locations are averaged positions. If the solute–solvent cluster is dominated by one conformation, this is a good approximation. However, depending on the investigated system and property of interest, it might be necessary to consider multiple solute conformations. In addition, it might also be necessary to subject the obtained solute–solvent cluster to additional sampling of the water positions—this time at a quantum chemical or semi-empirical level to generate a structural ensemble that may be used for Boltzmann averaging.

Still, even when only a single conformer of a microsolvated cluster is considered, DFT was shown to be well suited to describe solute–solvent interactions, [100] supporting the added value of low-cost microsolvation approaches to approximately model (bulk) solvent effects.

6. Conclusions

In this paper, we presented a quantitative approach to quantum chemical microsolvation based on MD simulations and GIST analysis, which assigns water position and orientations based on physical quantities. Our protocol provides a rigorous basis as to where to locate the explicit solvent molecules, how to orient them, and how many to include, thereby allowing a rational setup of microsolvated clusters.

We demonstrated that our computational protocol is applicable to a number of different molecules ranging from biologically relevant urea and benzoic acid to apolar helicene.

We retrieved solute–solvent complexes that are in very good agreement with experimental and theoretical reference data. Although currently limited to water as a solvent, we will extend our algorithm to arbitrary solvents in the future.

Our methodology can be used to set up Quantum Mechanics/Molecular Mechanics models, where the number of solvent molecules in the QM region may be determined by our FEBISS algorithm. In addition, our protocol paves the way to study explicit solute–solvent interactions in large complexes that cannot be treated by *ab initio* MD.

Supplementary Materials: The following are available online. Figure S1: Radial distribution functions $g(r)$ from the 300 ns simulation trajectory of urea. Table S1: Radial distribution functions (RDFs) for certain urea-water atom pairs, calculated from our MD simulation, in comparison to previously published results. Figure S2: Urea-(H₂O)_{*n*} complexes for $n = 1$ and $n = 6$ as obtained by the FEBISS algorithm. The free energy of the solvent molecules is given in kcal/mol and they are color-coded according to this value from strongly favored (magenta) to less favored (yellow). Figure S3: FEBISS bar chart (A) and initial urea-(H₂O)₅ cluster when AM1-BCC charges were used in the MD simulation. Figure S4: Quantum chemically optimized structures of *syn*-BA and *anti*-BA. The relative energies were calculated with B3LYP/def2-TZVP/D3 in water modelled as implicit. Table S2: Relative energies (kcal/mol) and dipole moments (Debye) of the *syn*- and *anti*-conformer of benzoic acid calculated with B3LYP/def2-TZVP/D3 in the gas phase and in implicit solvent. Figure S5: Benzoate-(H₂O)_{*n*} clusters as obtained from the FEBISS algorithm for $n = 1–3$ water molecules (top) and $n = 6, 9, 15$ water molecules (bottom). The water molecules are color-coded according to their interaction strength from magenta (most favorable) to yellow (less favorable). Figure S6: Gas phase optimized microsolvated cluster of benzoic acid in the *syn*- and *anti*-conformation. Table S3: Relative energies (kcal/mol) and dipole moments (Debye) of tautomer T1 and T2 calculated with B3LYP-D3(BJ)/def2-TZVP for in the gas phase and in implicit solvent. Figure S7a: Structural overview of the B3LYP/def2-TZVP/D3 optimized microsolvated clusters of T1-(H₂O)_{*n*} and T2(H₂O)_{*n*} with $n = 1–3$ water molecules. The initial water placement was done with our FEBISS algorithm. Figure S7b: Structural overview of the B3LYP/def2-TZVP/D3 optimized microsolvated clusters of T1-(H₂O)_{*n*} and T2-(H₂O)_{*n*} with $n = 4–6$ water molecules. The initial water placement was done with our FEBISS algorithm. Table S4: Relative energies in kcal/mol for the first 14 [4]-helicene-(H₂O) clusters.

Author Contributions: M.P. developed the concept of the project, M.S. (Miguel Steiner) wrote the software with support of M.S. (Michael Schauerl), M.S. (Miguel Steiner) and T.H. performed testing and data acquisition. M.S. (Miguel Steiner) and T.H. visualized the results and wrote the original draft. M.S. (Michael Schauerl) and M.P. reviewed, revised and edited the manuscript. M.P. provided funding acquisition, project administration and resources. All authors agree to the final version of this manuscript.

Funding: M.P. would like to thank the Austrian Science Fund (FWF) for financial support (M-2005 and P 33528).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Software available at <https://github.com/PodewitzLab/FEBISS> (accessed on 25 February 2021). Raw data available upon request from the authors.

Acknowledgments: The authors would like to thank Johannes Kraml for helpful discussions about GIST. The computational results presented in this work were achieved using the HPC infrastructure of the University of Innsbruck (leo3e) and the Vienna Scientific Cluster VSC3. Open Access Funding by the Austrian Science Fund (FWF).

Conflicts of Interest: The authors declare that there are no conflicts of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Sherwood, J.; Clark, J.H.; Fairlamb, I.J.S.; Slattery, J.M. Solvent effects in palladium catalysed cross-coupling reactions. *Green Chem.* **2019**, *21*, 2164–2213. [[CrossRef](#)]
2. Pirrung, M.C. Acceleration of Organic Reactions through Aqueous Solvent Effects. *Chem. Eur. J.* **2006**, *12*, 1312–1317. [[CrossRef](#)] [[PubMed](#)]
3. Reichardt, C. Solvents and Solvent Effects: An Introduction. *Org. Process. Res. Dev.* **2007**, *11*, 105–113. [[CrossRef](#)]
4. Reichardt, C.; Welton, T. Solute-Solvent Interactions. In *Solvents and Solvent Effects in Organic Chemistry*; Wiley: Hoboken, NJ, USA, 2010; pp. 7–64.
5. Elmi, A.; Cockroft, S.L. Quantifying Interactions and Solvent Effects Using Molecular Balances and Model Complexes. *Acc. Chem. Res.* **2021**, *54*, 92–103. [[CrossRef](#)]
6. Varghese, J.J.; Mushrif, S.H. Origins of complex solvent effects on chemical reactivity and computational tools to investigate them: A review. *React. Chem. Eng.* **2019**, *4*, 165–206. [[CrossRef](#)]
7. Sunoj, R.B.; Anand, M. Microsolvated transition state models for improved insight into chemical properties and reaction mechanisms. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12715–12736. [[CrossRef](#)]
8. Pliego, J.R.; Riveros, J.M. Hybrid discrete-continuum solvation methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, 25. [[CrossRef](#)]
9. Cramer, C.J.; Truhlar, D.G. A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008**, *41*, 760–768. [[CrossRef](#)] [[PubMed](#)]
10. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094. [[CrossRef](#)] [[PubMed](#)]
11. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396. [[CrossRef](#)] [[PubMed](#)]
12. Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 699–709. [[CrossRef](#)]
13. Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, UK, 2009.
14. Hutter, J. Car-Parrinello molecular dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *2*, 604–612. [[CrossRef](#)]
15. Kühne, T.D. Second generation Car-Parrinello molecular dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 391–406. [[CrossRef](#)]
16. Kelly, C.P.; Cramer, C.J.; Truhlar, D.G. Adding Explicit Solvent Molecules to Continuum Solvent Calculations for the Calculation of Aqueous Acid Dissociation Constants. *J. Phys. Chem. A* **2006**, *110*, 2493–2499. [[CrossRef](#)] [[PubMed](#)]
17. Thar, J.; Zahn, S.; Kirchner, B. When Is a Molecule Properly Solvated by a Continuum Model or in a Cluster Ansatz? A First-Principles Simulation of Alanine Hydration. *J. Phys. Chem. B* **2008**, *112*, 1456–1464. [[CrossRef](#)]
18. Pliego, J.A.J.R.; Riveros, J.M. The Cluster–Continuum Model for the Calculation of the Solvation Free Energy of Ionic Species. *J. Phys. Chem. A* **2001**, *105*, 7241–7247. [[CrossRef](#)]
19. Bryantsev, V.S.; Diallo, M.S.; Iii, W.A.G. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B* **2008**, *112*, 9709–9719. [[CrossRef](#)] [[PubMed](#)]
20. Marenich, A.V.; Ding, W.; Cramer, C.J.; Truhlar, N.G. Resolution of a Challenge for Solvation Modeling: Calculation of Dicarboxylic Acid Dissociation Constants Using Mixed Discrete–Continuum Solvation Models. *J. Phys. Chem. Lett.* **2012**, *3*, 1437–1442. [[CrossRef](#)]
21. Zins, E.-L. Microhydration of a Carbonyl Group: How does the Molecular Electrostatic Potential (MESP) Impact the Formation of $(\text{H}_2\text{O})_n:(\text{R}_2\text{C}=\text{O})$ Complexes? *J. Phys. Chem. A* **2020**, *124*, 1720–1734. [[CrossRef](#)]
22. Gadre, S.R.; Babu, K.; Rendell, A.P. Electrostatics for Exploring Hydration Patterns of Molecules. 3. Uracil. *J. Phys. Chem. A* **2000**, *104*, 8976–8982. [[CrossRef](#)]
23. Kalai, C.; Alikhani, M.E.; Zins, E.L. The molecular electrostatic potential analysis of solutes and water clusters: A straightforward tool to predict the geometry of the most stable micro-hydrated complexes of beta-propiolactone and formamide. *Theor. Chem. Acc.* **2018**, *137*, 20. [[CrossRef](#)]
24. Sure, R.; El Mahdali, M.; Plajer, A.; Deglmann, P. Towards a converged strategy for including microsolvation in reaction mechanism calculations. *J. Comput. Mol. Des.* **2021**, *20*, 1–20. [[CrossRef](#)]
25. Lanza, G.; Chiacchio, M.A. The water molecule arrangement over the side chain of some aliphatic amino acids: A quantum chemical and bottom-up investigation. *Int. J. Quantum Chem.* **2020**, *120*, 17. [[CrossRef](#)]
26. Simm, G.N.; Türtcher, P.L.; Reiher, M. Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144–1155. [[CrossRef](#)] [[PubMed](#)]
27. Basdogan, Y.; Groenenboom, M.C.; Henderson, E.; De, S.; Rempe, S.B.; Keith, J.A. Machine Learning-Guided Approach for Studying Solvation Environments. *J. Chem. Theory Comput.* **2019**, *16*, 633–642. [[CrossRef](#)] [[PubMed](#)]
28. Jesus, W.S.; Prudente, F.V.; Marques, J.M.C.; Pereira, F.B. Modeling microsolvation clusters with electronic-structure calculations guided by analytical potentials and predictive machine learning techniques. *Phys. Chem. Chem. Phys.* **2021**, *23*, 1738–1749. [[CrossRef](#)]
29. Kongsted, J.; Mennucci, B.; Coutinho, K.; Canuto, S. Solvent effects on the electronic absorption spectrum of camphor using continuum, discrete or explicit approaches. *Chem. Phys. Lett.* **2010**, *484*, 185–191. [[CrossRef](#)]

30. Mensch, C.; Bultinck, P.; Johannessen, C. The effect of protein backbone hydration on the amide vibrations in Raman and Raman optical activity spectra. *Phys. Chem. Chem. Phys.* **2019**, *21*, 1988–2005. [[CrossRef](#)] [[PubMed](#)]
31. Da Silva, E.F.; Svendsen, H.F.; Merz, K.M. Explicitly Representing the Solvation Shell in Continuum Solvent Calculations. *J. Phys. Chem. A* **2009**, *113*, 6404–6409. [[CrossRef](#)]
32. Nocito, D.; Beran, G.J.O. Averaged Condensed Phase Model for Simulating Molecules in Complex Environments. *J. Chem. Theory Comput.* **2017**, *13*, 1117–1129. [[CrossRef](#)]
33. Karimova, N.V.; Luo, M.; Grassian, V.H.; Gerber, R.B. Absorption spectra of benzoic acid in water at different pH and in the presence of salts: Insights from the integration of experimental data and theoretical cluster models. *Phys. Chem. Chem. Phys.* **2020**, *22*, 5046–5056. [[CrossRef](#)]
34. Basdogan, Y.; Maldonado, A.M.; Keith, J.A. Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, 19. [[CrossRef](#)]
35. Bodnarchuk, M.S.; Viner, R.; Michel, J.; Essex, J.W. Strategies to Calculate Water Binding Free Energies in Protein–Ligand Complexes. *J. Chem. Inf. Model.* **2014**, *54*, 1623–1633. [[CrossRef](#)] [[PubMed](#)]
36. Spyraakis, F.; Ahmed, M.H.; Bayden, A.S.; Cozzini, P.; Mozzarelli, A.; Kellogg, G.E. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J. Med. Chem.* **2017**, *60*, 6781–6827. [[CrossRef](#)]
37. Bucher, D.; Stouten, P.F.; Triballeau, N. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J. Chem. Inf. Model.* **2018**, *58*, 692–699. [[CrossRef](#)]
38. Rudling, A.; Orro, A.; Carlsson, J. Prediction of Ordered Water Molecules in Protein Binding Sites from Molecular Dynamics Simulations: The Impact of Ligand Binding on Hydration Networks. *J. Chem. Inf. Model.* **2018**, *58*, 350–361. [[CrossRef](#)] [[PubMed](#)]
39. Jukic, M.; Konc, J.; Janezic, D.; Bren, U. ProBiS H₂O MD Approach for Identification of Conserved Water Sites in Protein Structures for Drug Design. *ACS Med. Chem. Lett.* **2020**, *11*, 877–882. [[CrossRef](#)]
40. Kearney, B.M.; Schwabe, M.; Marcus, K.C.; Roberts, D.M.; DeChene, M.; Swartz, P.; Mattos, C. DRoP: Automated detection of conserved solvent-binding sites on proteins. *Proteins Struct. Funct. Bioinform.* **2020**, *88*, 152–165. [[CrossRef](#)]
41. Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R.C.; Gilson, M.K.; Kurtzman, T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37*, 2029–2037. [[CrossRef](#)]
42. Nguyen, C.N.; Young, T.K.; Gilson, M.K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101. [[CrossRef](#)]
43. Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10*, 2769–2780. [[CrossRef](#)]
44. Loeffler, J.R.; Schauerl, M.; Liedl, K.R. Hydration of Aromatic Heterocycles as an Adversary of pi-Stacking. *J. Chem. Inf. Model.* **2019**, *59*, 4209–4219. [[CrossRef](#)]
45. Loeffler, J.R.; Fernández-Quintero, M.L.; Schauerl, M.; Liedl, K.R. STACKED – Solvation Theory of Aromatic Complexes as Key for Estimating Drug Binding. *J. Chem. Inf. Model.* **2020**, *60*, 2304–2313. [[CrossRef](#)]
46. Schauerl, M.; Czodrowski, P.; Fuchs, J.E.; Huber, R.G.; Waldner, B.J.; Podewitz, M.; Kramer, C.; Liedl, K.R. Binding Pose Flip Explained via Enthalpic and Entropic Contributions. *J. Chem. Inf. Model.* **2017**, *57*, 345–354. [[CrossRef](#)] [[PubMed](#)]
47. Schauerl, M.; Podewitz, M.; Ortner, T.S.; Waibl, F.; Thoeny, A.; Loerting, T.; Liedl, K.R. Balance between hydration enthalpy and entropy is important for ice binding surfaces in Antifreeze Proteins. *Sci. Rep.* **2017**, *7*, 11901. [[CrossRef](#)]
48. Fabris, F.; De Lucchi, O.; Nardini, I.; Crisma, M.; Mazzanti, A.; Mason, S.A.; Lemée-Cailleau, M.-H.; Scaramuzzo, F.A.; Zonta, C. (+)-*syn*-Benzotriboorneol an enantiopure C₃-symmetric receptor for water. *Org. Biomol. Chem.* **2012**, *10*, 2464–2469. [[CrossRef](#)]
49. Wazzan, N.; Safi, Z.; Al-Barakati, R.; Al-Qurashi, O.; Al-Khateeb, L. DFT investigation on the intramolecular and intermolecular proton transfer processes in 2-aminobenzothiazole (ABT) in the gas phase and in different solvents. *Struct. Chem.* **2019**, *31*, 243–252. [[CrossRef](#)]
50. Ishida, T.; Rossky, P.J.; Castner, E.W. A Theoretical Investigation of the Shape and Hydration Properties of Aqueous Urea: Evidence for Nonplanar Urea Geometry. *J. Phys. Chem. B* **2004**, *108*, 17583–17590. [[CrossRef](#)]
51. Domingos, S.R.; Martin, K.; Avarvari, N.; Schnell, M. Water Docking Bias in 4 Helicene. *Angew. Chem. Int. Ed.* **2019**, *58*, 11257–11261. [[CrossRef](#)] [[PubMed](#)]
52. Weiss, A.K.H.; Hofer, T.S. Urea in aqueous solution studied by quantum mechanical charge field-molecular dynamics (QMCF-MD). *Mol. Biosyst.* **2013**, *9*, 1864–1876. [[CrossRef](#)] [[PubMed](#)]
53. Bruno, G.; Randaccio, L. A refinement of the benzoic acid structure at room temperature. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 1711–1712. [[CrossRef](#)]
54. Nagy, P.I.; Smith, D.A.; Alagona, G.; Ghio, C. *Ab initio* studies of free and monohydrated carboxylic acids in the gas phase. *J. Phys. Chem.* **1994**, *98*, 486–493. [[CrossRef](#)]
55. Stepanian, S.; Reva, I.; Radchenko, E.; Sheina, G. Infrared spectra of benzoic acid monomers and dimers in argon matrix. *Vib. Spectrosc.* **1996**, *11*, 123–133. [[CrossRef](#)]
56. Godfrey, P.D.; McNaughton, D. Structural studies of aromatic carboxylic acids via computational chemistry and microwave spectroscopy. *J. Chem. Phys.* **2013**, *138*, 24303. [[CrossRef](#)] [[PubMed](#)]
57. Schnitzler, E.G.; Jäger, W. The benzoic acid–water complex: A potential atmospheric nucleation precursor studied using microwave spectroscopy and *ab initio* calculations. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2305–2314. [[CrossRef](#)] [[PubMed](#)]

58. Onda, M.; Asai, M.; Takise, K.; Kuwae, K.; Hayami, K.; Kuroe, A.; Mori, M.; Miyazaki, H.; Suzuki, N.; Yamaguchi, I. Microwave spectrum of benzoic acid. *J. Mol. Struct.* **1999**, *482–483*, 301–303. [CrossRef]
59. Aarset, K.; Page, E.M.; Rice, D.A. Molecular Structures of Benzoic Acid and 2-Hydroxybenzoic Acid, Obtained by Gas-Phase Electron Diffraction and Theoretical Calculations. *J. Phys. Chem. A* **2006**, *110*, 9014–9019. [CrossRef]
60. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531–3541. [CrossRef]
61. Roe, D.R.; Cheatham, T.E. [PTRAJ and CPPTRAJ]: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [CrossRef]
62. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
63. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210. [CrossRef]
64. Case, D.A.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, I.T.E.; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; et al. *AMBER 2018*; University of California: San Francisco, CA, USA, 2018.
65. The PyMOL Molecular Graphics System, Version 1.8. Available online: <https://sourceforge.net/projects/pymol/files/pymol/1.8/> (accessed on 25 February 2021).
66. Available online: <https://github.com/PodewitzLab/FEBISS.git> (accessed on 25 February 2021).
67. Available online: <https://github.com/liedllab/gigist> (accessed on 25 February 2021).
68. Kraml, J.; Kamenik, A.S.; Waibl, F.; Schauerl, M.; Liedl, K.R. Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces. *J. Chem. Theory Comput.* **2019**, *15*, 5872–5882. [CrossRef] [PubMed]
69. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Rev. C.01*; Gaussian, Inc.: Wallingford, CT, USA, 2016.
70. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280. [CrossRef]
71. Jakalian, A.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641. [CrossRef]
72. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef]
73. Wallnoefer, H.G.; Liedl, K.R.; Fox, T. A challenging system: Free energy prediction for factor Xa. *J. Comput. Chem.* **2011**, *32*, 1743–1752. [CrossRef] [PubMed]
74. Salomon-Ferrer, R.; Götz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888. [CrossRef] [PubMed]
75. Adelman, S.A.; Doll, J.D. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering: General Formulation for Classical Scattering off Harmonic Solids. *J. Chem. Phys.* **1976**, *64*, 2375–2388. [CrossRef]
76. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
77. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]
78. Kovalenko, A.; Hirata, F. Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, *110*, 10095–10112. [CrossRef]
79. Group, C.C. *Molecular Operating Environment*; 2018.
80. Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol.* **2014**, *4*, 91–100. [CrossRef]
81. TURBOMOLE V7.3 2018, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, Since 2007. Available online: <http://www.turbomole.com/> (accessed on 25 February 2021).
82. Becke, A.D. Density-Functional Thermochemistry III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [CrossRef]
83. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789. [CrossRef] [PubMed]
84. Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100. [CrossRef] [PubMed]
85. Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality For H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. [CrossRef] [PubMed]
86. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104. [CrossRef]
87. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. [CrossRef] [PubMed]
88. Klamt, A.; Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* **1993**, *5*, 799–805. [CrossRef]

89. Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J.C.W.; Eckert, F. COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193. [[CrossRef](#)]
90. Hirschfeld, F.L.; Schmidt, G.M.J.; Sandler, S. 398. The Structure of Overcrowded Aromatic Compounds. Part VI. The Crystal Structure of Benzo[C]phenanthrene and of 1,12-Dimethylbenzo[C]phenanthrene. *J. Chem. Soc.* **1963**, 2108–2125. [[CrossRef](#)]
91. Mobley, D.L.; Dumont, É.; Chodera, J.D.; Dill, K.A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254. [[CrossRef](#)]
92. Mobley, D.L.; Bayly, C.I.; Cooper, M.D.; Shirts, M.R.; Dill, K.A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350–358. [[CrossRef](#)]
93. Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009**, *5*, 919–930. [[CrossRef](#)]
94. Bunken, C.; Reiher, M. Self-Parametrizing System-Focused Atomistic Models. *J. Chem. Theory Comput.* **2020**, *16*, 1646–1665. [[CrossRef](#)] [[PubMed](#)]
95. Agieienko, V.; Buchner, R. Urea hydration from dielectric relaxation spectroscopy: Old findings confirmed, new insights gained. *Phys. Chem. Chem. Phys.* **2015**, *18*, 2597–2607. [[CrossRef](#)] [[PubMed](#)]
96. Åstrand, P.-O.; Wallqvist, A.; Karlström, G.; Linse, P. Properties of Urea–Water Solvation Calculated from a New *ab initio* Polarizable Intermolecular Potential. *J. Chem. Phys.* **1991**, *95*, 8419–8429. [[CrossRef](#)]
97. Kallies, B. Coupling of solvent and solute dynamics—Molecular dynamics simulations of aqueous urea solutions with different intramolecular potentials. *Phys. Chem. Chem. Phys.* **2001**, *4*, 86–95. [[CrossRef](#)]
98. Stumpe, M.C.; Grubmüller, H. Aqueous Urea Solutions: Structure, Energetics, and Urea Aggregation. *J. Phys. Chem. B* **2007**, *111*, 6220–6228. [[CrossRef](#)] [[PubMed](#)]
99. Krishnakumar, P.; Maity, D.K. Microhydration of a benzoic acid molecule and its dissociation. *New J. Chem.* **2017**, *41*, 7195–7202. [[CrossRef](#)]
100. Chen, J.; Chan, B.; Shao, Y.; Ho, J. How accurate are approximate quantum chemical methods at modelling solute–solvent interactions in solvated clusters? *Phys. Chem. Chem. Phys.* **2020**, *22*, 3855–3866. [[CrossRef](#)] [[PubMed](#)]