

Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods

Sebastian Spicher,[†] Christoph Plett,[†] Philipp Pracht, Andreas Hansen, and Stefan Grimme*



Cite This: *J. Chem. Theory Comput.* 2022, 18, 3174–3189



Read Online

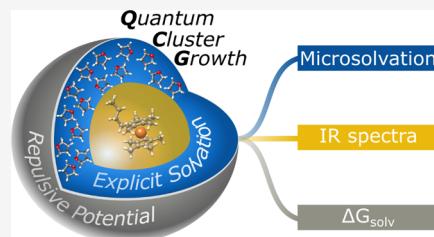
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: An automated and broadly applicable workflow for the description of solvation effects in an explicit manner is introduced. This method, termed quantum cluster growth (QCG), is based on the semiempirical GFN2-xTB/GFN-FF methods, enabling efficient geometry optimizations and MD simulations. Fast structure generation is provided using the intermolecular force field xTB-IFF. Additionally, the approach uses an efficient implicit solvation model for the electrostatic embedding of the growing clusters. The novel QCG procedure presents a robust cluster generation tool for subsequent application of higher-level (e.g., DFT) methods to study solvation effects on molecular geometries explicitly or to average spectroscopic properties over cluster ensembles. Furthermore, the computation of the solvation free energy with a supermolecular approach can be carried out with QCG. The underlying growing process is physically motivated by computing the leading-order solute–solvent interactions first and can account for conformational and chemical changes due to solvation for low-energy barrier processes. The conformational space is explored with the NCI–MTD algorithm as implemented in the CREST program, using a combination of metadynamics and MD simulations. QCG with GFN2-xTB yields realistic solution geometries and reasonable solvation free energies for various systems without introducing many empirical parameters. Computed IR spectra of some solutes with QCG show a better match to the experimental data compared to well-established implicit solvation models.



1. INTRODUCTION

Current chemical research, for example, on stereoselective catalysis in organic synthesis,¹ electrochemical capacitors for energy storage,² or protein folding in living organisms,^{3,4} is connected with the fundamental question of how solvent molecules interact with solutes and surfaces in the condensed phase. To answer this question, an adequate description of solvation effects is inevitable. Nowadays, theoretical chemistry is capable of providing highly accurate quantum mechanical (QM) calculations in the gas phase, whereas most experiments are carried out in solution. To compare experimental findings and theoretical simulations, a reliable solvation model has to be included in the calculation. Quantitative theoretical predictions of thermodynamic properties for molecules in solution require an accurate description of the interaction between solvent molecules themselves and their very specific interaction with the solute.⁵ Therefore, the calculation of mass densities, enthalpies of vaporization, heat capacities, surface tensions, dielectric constants, solvation free energies, and other properties of molecules in solution remains a challenging task for computational chemistry and is part of current theoretical research.^{6–8}

Methods for evaluating solvation effects can be roughly classified into two categories.⁹ Explicit solvation models^{10–12} describe the individual solvent molecules whereas implicit models^{13,14} treat the solvent as a continuous medium, mainly characterized by its dielectric constant ϵ .¹³ Combinations of explicit/implicit solvation models are conceivable, in which, for

example, the first solvation shell is built up from explicitly placed solvent molecules, while the remaining shells are treated implicitly by continuum embedding. Furthermore, each of these methods may be conducted at the classical molecular mechanical (MM) or QM level of theory or the combination of both in so-called QM/MM schemes.^{15,16} Recently, great advances were made in the context of polarizable molecular mechanics models that are commonly used to compute many molecular properties and that can reproduce solvation effects.^{17–19} Continuum solvation models consider the solvent as a continuous isotropic medium. The solvent is replaced with an electric “reaction field” that represents a statistical average of all solvent degrees of freedom (DOFs) at thermal equilibrium.¹⁴ The solute is placed in a suitably shaped hole in the medium, thus creating a cavity. In standard polarizable continuum models (PCMs),²⁰ the polar electrostatics can be decoupled from the nonpolar interactions, and the solvation free energy can be written as

$$\delta G_{\text{solv}} = \delta G_{\text{cavity}} + \delta G_{\text{disp}} + \delta G_{\text{elec}} \quad (1)$$

Received: March 10, 2022

Published: April 28, 2022



where the three terms are the free energy of cavitation, dispersion, and electrostatic energy.²¹ The latter contribution in PCM models is mostly approximated through solving the Poisson–Boltzmann (PB) equation^{22,23} or further simplified by using the generalized Born (GB)^{24–27} model. Nowadays, the most commonly applied version of the PCM model is a reformulation of dielectric PCM in terms of the integral equation formalism termed IEFPCM.²⁸ The conductor-like screening model/conductor-like screening model for real solvents (COSMO²⁹/COSMO-RS^{30,31}) is a variation of PB PCM. In COSMO, the dielectric permittivity is set to infinity ($\epsilon = \infty$), which defines the solvent as a conductor. COSMO-RS results from the combination of the COSMO approach with the statistical thermodynamics of interacting surfaces.³² The COSMO approach is also employed in a further variation of PCM, the conductor-like PCM (CPCM).³³ Other continuum solvation models, for example, SMD, also include the QM charge density of a solute molecule interacting with a continuum description of the solvent.³⁴ Implicit solvation models are computationally efficient and were successfully applied in many computational studies (see, e.g., refs 35–38). COSMO and COSMO-RS are the default models for computational chemistry work at the density functional theory (DFT) level in our group. Nevertheless, the main disadvantage of implicit models remains the inadequate description of very polar or charged species, mainly because strong directional and local interactions between solute and solvent molecules, that is, noncovalent interactions (NCIs), ionic, and hydrogen bonds, are not treated properly. By neglecting explicit solvent molecules, important interactions are missing or only described poorly.^{39–42} Explicit solvation models use molecular dynamics (MD)^{43,44} or Monte Carlo (MC) statistical mechanics^{45,46} simulations to generate molecular ensembles and corresponding energies. Differences in free energies are obtained by applying free energy perturbation theory,⁴⁷ thermodynamic integration (TI),⁴⁸ or Bennett's acceptor ratio method.⁴⁹ Free-energy methods that use data from MD or MC simulations require a large number of steps to converge and, thus, suffer from the issue of sufficient phase space sampling to estimate the ratio of partition functions. For a short overview of explicit solvation treatments and related approaches, see refs 50–53. At this point, quantum cluster equilibrium methods should also be mentioned, comprising the essential idea of applying statistical mechanics to quantum chemically calculated clusters to obtain thermodynamic properties of the liquid (condensed) and the vapor phase.^{54–56} In the SAMPL5 challenge of calculating host–guest binding free energies, it was found that methods involving explicit solvent molecules, in general, perform better than implicit solvation models.⁵⁷ Therefore, recent developments turned their sights on hybrid cluster–continuum model approaches, where explicit water molecules were added to the continuum model to describe alkane complexation in self-assembled capsules or calculate the solvation free energies of small molecules in aqueous solution.^{58–60} In a recent study by Bensberg et al., the electrostatic PCM energy was partially replaced with an explicit solvent molecule treatment in the context of sub-system DFT, which was already successfully applied to systems of different sizes containing water and cyclohexane as solvents.⁶¹

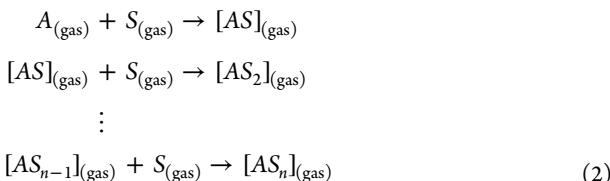
Here, the idea of combined cluster–continuum models is extended significantly in terms of molecular size and versatility. We propose a new QM and force-field (FF)-based hybrid solvation model, in which large molecular clusters of any solute

are generated fully automatically by successively adding explicit solvent molecules. With cluster sizes up to a few hundreds of atoms and the possibility to include all elements up to radon ($Z \leq 86$), a large part of the chemical compound space can be covered and any solute–solvent combination is in principle accessible, with no restriction in the charge or spin state. This newly developed procedure is denoted as quantum cluster growth (QCG) and can be applied to study the effect of explicit solvation on various properties at the QM level of theory. For computational efficiency reasons, semiempirical QM (SQM) methods in combination with even faster FFs are employed⁶² in the generation process. The resulting cluster ensembles may serve as input for subsequent high-level DFT or wave function theory calculations. In the context of microsolvation, the herein proposed cluster growing algorithm can automatize the detection of important interaction sites and, therefore, replace laborious approaches^{63–66} or can be an alternative to other workflows.^{67,68} An example for a recently developed, automated explicit solvation workflow is the AutoSolvate toolkit for generating clusters of organic solutes in several solvents.⁶⁹ Furthermore, the QCG algorithm is extended to compute solvation free energies using a novel supermolecular ansatz. In contrast to many existing proposals, this includes the explicit calculation of cluster entropies giving access to δG_{solv} and δH_{solv} values.

After a description of the theoretical background, technical details of the QCG algorithm are given and the cluster generation process is examined statistically. In a first application example, the quality of the generated clusters is assessed in comparison to other (micro)solvation tools. The effects of explicit solvation on molecular geometries are evaluated in the framework of MD simulations and infrared (IR) spectra calculation. Solvation free energies are computed for a test set of small organic molecules in comparison to established implicit solvation models. As an outlook for future applications, free association energies of supermolecular complexes in solution are calculated.

2. THEORETICAL BACKGROUND

QCG represents a fully automated approach to describe a molecule in solution in an explicit manner at a QM level of theory. Therefore, molecular clusters of the solute (A) with a given number n of solvents (S) are generated by adding one solvent molecule at a time to an energetically favorable position



The square brackets in eq 2 indicate a NCI complex. The formation of new covalent chemical bonds between solvent and solute is possible as long as the underlying QM or FF method can describe it and deserves no special attention. First, the fully automated cluster ensemble generation procedure is outlined in Section 2.1. In Section 2.2, an extension to the QCG algorithm is proposed that enables the computation of solvation free energies.

2.1. Cluster Ensemble Generation. The automated cluster generation part of the QCG algorithm may be

subdivided into two steps. Both are illustrated in the QCG workflow in Figure 1.

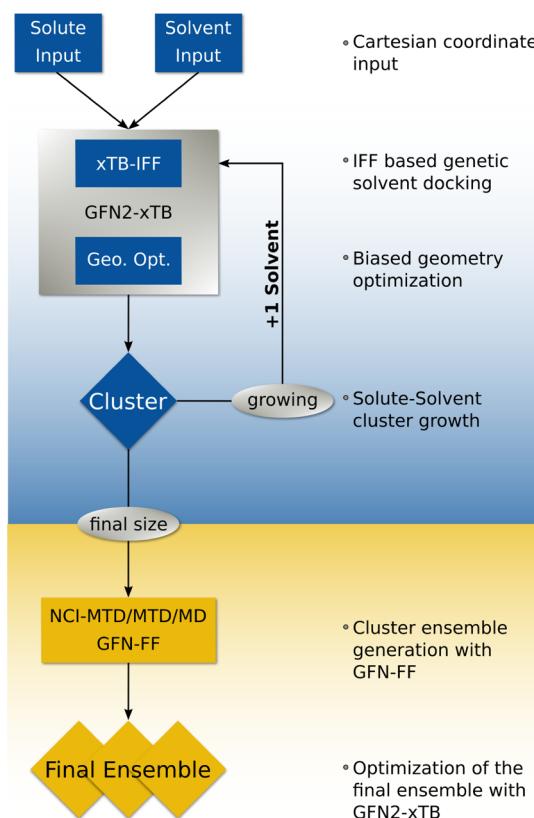


Figure 1. Schematic illustration of the QCG algorithm.

The first part describes the cluster growth process, consisting of a repeating cycle, in which each turn increases the cluster size by one solvent molecule. As input, only the solute and solvent geometries are required. Optimal complexation (docking) positions for added solvents are determined with a genetic (global) optimization algorithm employing the intermolecular force field xTB-IFF.⁷⁰ The necessary QM information is generated on-the-fly with GFN2-xTB⁷¹ and consists of the Mulliken atomic charges, charge centers of localized molecular orbitals, frontier orbitals, and orbital energies. The interaction energy surface between the growing cluster and the added solvent molecule is screened, a few most

favorable positions at the xTB-IFF level are determined and re-optimized at the GFN2-xTB level of theory. Repulsive wall potentials (vide infra) are applied throughout the growing process to “shape” the solute–solvent cluster properly. Complete and consistent coverage of the solute with a minimum number of solvent molecules is the target. The great advantage of QCG in comparison to other explicit solvation tools is the general applicability as most of the required parameterization is inherent and already existing in the underlying SQM and FF methods. To highlight this characteristic, molecular clusters of various (exotic) solute–solvent combinations were generated with QCG and are depicted in Figure 2.

In QCG, the interaction energy of the solute with the surrounding solvent molecules is computed in each growing step. It could be taken as a convergence criterium for growing and is fulfilled when the change in energy by the addition of another solvent molecule becomes smaller than a certain threshold. We observed an irregular convergence behavior due to fluctuations in the solvent shell, which was also observed in other cluster–continuum approaches.^{59,68} Hence, the simple moving average of the interaction energy is chosen instead as an alternative convergence criterium, where the threshold is set to $10^{-4} E_h$.

The convergence behavior of the interaction energy E_{int} (and its moving average) with regard to the cluster size is shown Figure 3 for the previously introduced examples (cf. Figure 2). It is found that for systems (A) (acetophenone in CH_2Cl_2) and (C) (C_{60} in PCDA), the interaction energy converges when roughly the first solvation shell is filled. Interestingly, for (B) (butylferrocene in THF) and (D) (taxol in an eutectic solvent), filling the second solvation shell leads to rearrangements in the first shell, leading to an increase in the solute–solvent interaction energy. As the QCG algorithm optimizes the total cluster energy rather than the solute–solvent interaction, an increase in E_{int} can occur. Although the moving average would be applicable as a convergence threshold in general, the number of solvent molecules can vary in different cluster growth runs. Hence, in the following, the target number of added solvent molecules (user input) is chosen as an exit criterium in the algorithm due to practical reasons and to obtain better reproducibility.

The QCG algorithm is non-deterministic, and a single cluster geometry is of limited value. With increasing cluster size, the number of local minima on the PES grows drastically giving rise to millions of possible conformations already for

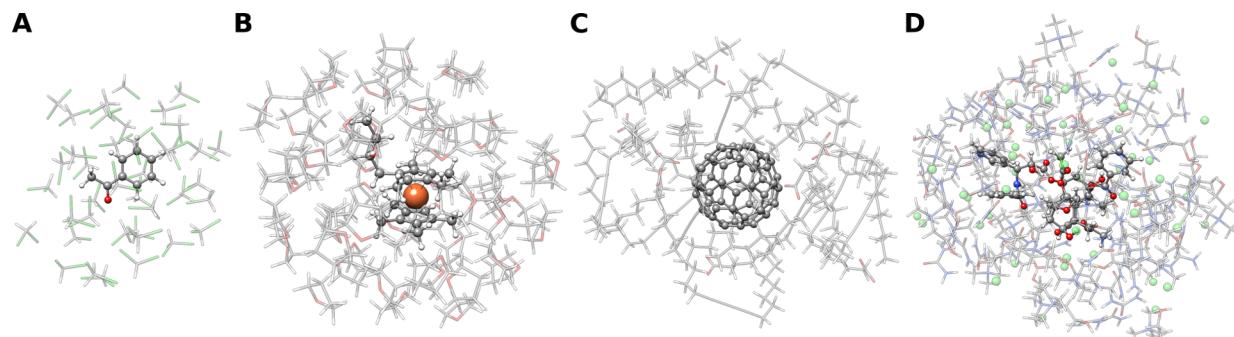


Figure 2. Examples of QCG-generated solute–solvent clusters: (A) acetophenone solvated by 40 explicit molecules of dichloromethane. (B) Butylferrocene (*n*-butylcyclopentadienyl(cyclopentadienyl)iron(II)) surrounded by 55 molecules of THF. (C) Fullerene C_{60} solvated by 10 PCDA (10,12-pentacosadiynoic acid) molecules. (D) Taxol within a eutectic solvent consisting of choline chloride and urea.

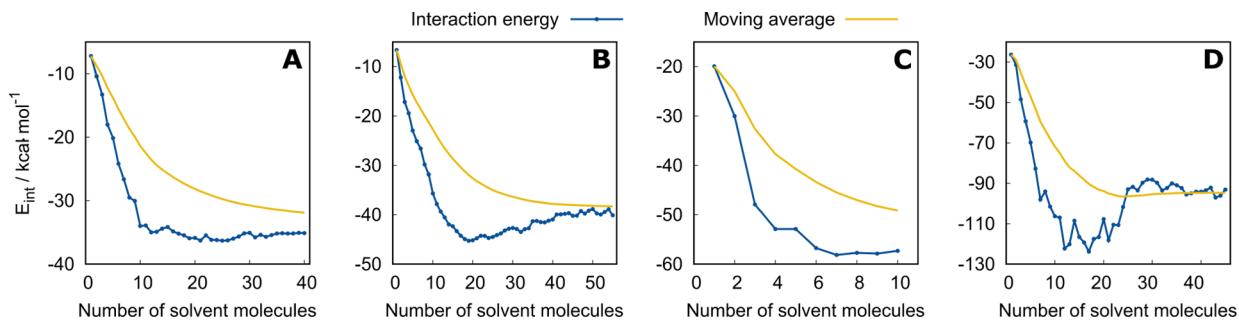


Figure 3. Solute–solvent interaction energy (blue) and moving average (yellow) as a function of cluster size for the examples in Figure 2.

medium-sized systems. Thus, statistical averages over many parallel-generated clusters must be computed to determine equilibrium properties and the entropic contribution of each cluster. Therefore, the second step of the cluster generation process in QCG is the sampling of phase space by a combination of MD and metadynamic (MTD) simulations. We employ the NCI–MTD algorithm as implemented in the CREST program to generate an ensemble of energetically low cluster structures.^{72,73} However, the computational demands of this algorithm for common computational resources can only be met with FFs. Therefore, the recently developed GFN-FF⁷⁴ method is used as the underlying level of theory. This combination has already been successfully applied in similar situations to determine protein conformations and to bind gases in metal–organic cages.^{75,76} As an alternative to NCI–MTD, similar algorithms are implemented in QCG, where just one MTD or MD simulation is performed instead of multiple ones. The cluster ensemble generated at the FF level is re-optimized at the GFN2-xTB level of theory. Thereupon, final single-point energy calculations are performed in the absence of any constraints (wall potentials) always at the same level to obtain the averaged ensemble energy \bar{E} . This can be carried out, with or without the implicit ALPB or GBSA solvation model,^{62,77} to minimize artificial surface effects.

2.1.1. Wall Potentials. An unbiased cluster growing process may lead to incomplete coverage of the solute and to an inconsistent description of the solvated system as it usually occurs in the bulk solvent. To enhance the coverage of the solute and to prevent irregularities in the growing process, a dynamic, repulsive outer wall potential is applied to shape the outermost solvent shell. The form of the repulsive potential is chosen as an ellipsoid. This choice allows to adopt the solute's geometry, varying from spherical to axial and resulting in a more uniform distribution of the solvent molecules around the solute.

Thus, the outer wall potential prevents accumulation of solvent molecules at a specific binding site. In addition, a second inner wall potential is applied to keep the solute fixed at the center of the growing cluster. This prevents, for example, the movement of hydrophobic molecules in polar solvents like water to the cluster surface. Both potentials are applied in the xTB-IFF docking steps, the GFN2-xTB geometry optimizations, and the conformer search (NCI–MTD). The arrangement of the potentials is illustrated for an exemplary system in Figure 4.

The reasonable choice of the potential is a challenging task. On the one hand, the potentials must ensure that any solute molecule is covered with any solvent, which is difficult for solute–solvent combinations with very different polarities and structures. On the other hand, the potentials should allow

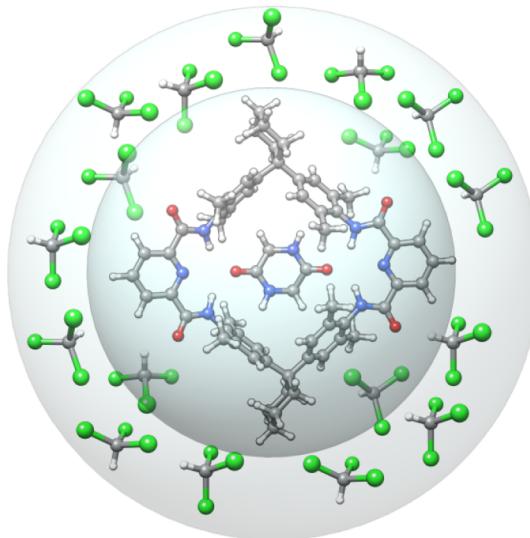


Figure 4. Inner and outer wall potential applied during the QCG procedure to shape the molecular clusters.

conformational reorganization during the growing process and hence should not be too restrictive. The QCG algorithm calculates the three principle axes of an ellipsoid according to different geometrical and solute/solvent specific criteria. The closer a molecule is to the surface of the ellipsoid cavity, the more repulsive is the applied wall potential. The energy contribution E_{pot} given by the ellipsoid potential is defined as a steep polynomial function

$$E_{\text{pot}} = \sum_i^N \left(\frac{\mathbf{R}_i - \mathbf{O}}{\mathbf{R}_{\text{pot}}} \right)^{10} \quad (3)$$

In eq 3, the summation runs over all atoms N , \mathbf{R}_i represents the Cartesian coordinates of atom i , \mathbf{O} is the center of the potential (i.e., the origin), and \mathbf{R}_{pot} represents the principle semi-axes of the ellipsoid potential parallel to $\mathbf{R}_i - \mathbf{O}$. The inner potential is only applied to the solute to fix its position during the cluster generation process. Therefore, it is rigid and does not change within the growing process. The outer potential is dynamic and increases as the number of solvent molecules within the cluster rises. For the solute and solvent molecules, diagonalization of the inertia tensor yields the principle moments and unit ellipsoid axes $a \geq b \geq c$, yielding important information about the molecular geometry. In addition, the excluded volume of overlapping spheres V is computed via analytic equations for solute and solvent molecules using the arvo package.⁷⁸ To further describe the geometry of a molecule, a structural factor F_a is introduced to account for further geometrical properties

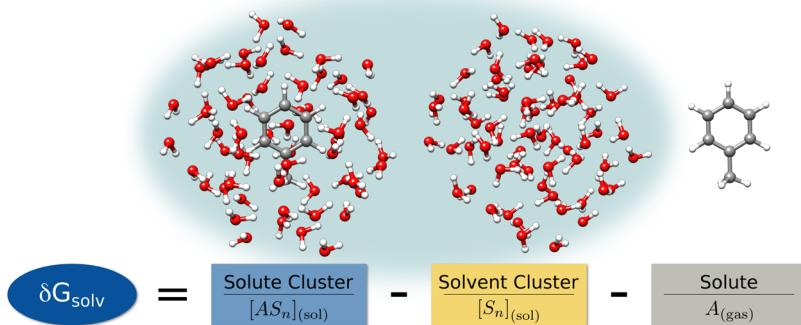


Figure 5. Example system of toluene in 60 molecules of water for the calculation of the solvation free energy δG_{solv} according to a supermolecular approach.

$$F_\alpha = \sqrt{1 + \frac{a - c}{\frac{1}{3}(a + b + c)}} \quad (4)$$

F_α is a measure of how strong a molecule's shape differs from an ideal sphere. For spherical molecules, F_α simply reduces to unity, and for arbitrarily shaped molecules, $F_\alpha \geq 1$ holds true. The volume and radius of a cavity hosting the solute and n solvent molecules need to be properly sized. For the outer cavity, all requirements are fulfilled in eq 5

$$R_{\text{out}} = \left[\frac{3}{4\pi} \left(\frac{F_\alpha}{2} \cdot n V_{\text{solvent}} + V_{\text{solute}} \right) \right]^{1/3} + \beta \cdot R_{\text{max}} + \gamma_1 \quad (5)$$

where the cavity radius is determined by calculating the third root of the added molecular volumes. The contribution of the solvent molecules to the overall volume is scaled by the geometrical correction factor F_α . R_{max} is the maximal internal distance within the solvent molecule scaled by the empirically determined factor β (usually = 0.5). γ_1 is an added constant to damp the long-range effects of the polynomial wall potential. Adding more solvent molecules to the cluster increases the radius of the cavity, leading to a dynamic outer wall potential. The static inner cavity is hosting the solute, and its radius is calculated according to eq 6. The radius is independent of the number of solvent molecules, where R_{max} is the maximal internal distance within the solute

$$R_{\text{in}} = \left(\frac{3F_\alpha}{4\pi} \cdot V_{\text{solute}} \right)^{1/3} + \beta \cdot R_{\text{max}} + \gamma_2 \quad (6)$$

To obtain the desired principle semi-axes R_{pot} (cf. eq 3) of an ellipsoid, the radii of the inner and outer cavity are projected on the unit axes of the solute. This results in an ellipsoid that is capable of hosting the cluster during the generation process.

2.2. Solvation Free Energies. One of the fundamental quantities to describe the interaction of a solute with the surrounding solvent is the free energy of solvation δG_{solv} that describes the change in free energy upon transferring a molecule from the (ideal) gas state to a solvent at a certain temperature and pressure.^{50,65,79,80} This quantity can be calculated with QCG in a supermolecular approach as the difference in total free energy of the isolated solute A (gas phase), and the “filled” $[AS_n]$ and “empty” $[S_n]$ clusters visualized in Figure 5.

This can be applied to any solute and requires the total free energies of the solute in the gas phase $G(A)_{(\text{gas})}$, the pure solvent cluster $G([S_n])_{(\text{sol})}$, and the solute cluster $G([AS_n])_{(\text{sol})}$ (eq 7), each consisting of the respective total electronic energy

and the sum of corrections from energy to free energy in the modified rigid rotor–harmonic oscillator approximation (mRRHO) including zero-point vibrational energy.⁸¹

$$\delta G_{\text{solv}}(A) = G([AS_n])_{(\text{sol})} - G([S_n])_{(\text{sol})} - G(A)_{(\text{gas})} \quad (7)$$

To keep the algorithm computationally efficient also for large molecular clusters, GFN2-xTB is employed to calculate the electronic energies and GFN-FF for the thermostatistical mRRHO contributions. In QCG, the solute is surrounded by a limited number of solvent molecules, which resembles only a cutout from the infinitely diluted solute in the condensed phase of the solvent. Hence, for finite cluster sizes, an additional embedding in a GBSA continuum model is employed.⁷⁷ This approach is common practice in cluster–continuum models and leads to faster convergence of δG_{solv} as a function of the cluster size.^{59,68,82} For $n \rightarrow \infty$, the effect of the continuum model vanishes. In the next step, it is important to distinguish between the free energy of an individual equilibrium structure and the ensemble value of the previously generated and optimized cluster structures (see Section 2.1). Assuming that all DOFs are separable, the free energy of the (cluster) structure ensemble (SE) is obtained as⁸³

$$G_{\text{SE}} = \bar{G} + G_{\text{conf}} \quad (8)$$

where G_{conf} is the conformational free energy part of N distinguishable conformers in the SE calculated from the Gibbs–Shannon entropy according to ref 84.

$$G_{\text{conf}} = -TS'_{\text{conf}} = RT \sum_i^N p_i \ln(p_i) \quad (9)$$

Hence, the conformational free energy G_{conf} is included if a complete ensemble of low-energy clusters is found. Furthermore, the average \bar{G} in eq 8 is given using the following equation

$$\bar{G} = \sum_i^N p_i G_i \quad (10)$$

with the Boltzmann weight p_i

$$p_i = \frac{e^{-G_i/k_B T}}{\sum_j^N e^{-G_j/k_B T}} \quad (11)$$

and the molecular free energy $G_i = E_i + G_{i,\text{mRRHO}}$ of the ensemble member i . Note that the molecular entropy of each species (solute cluster, solvent cluster, and solute) is explicitly calculated. Taking into account the SE of the solute clusters,

the solvent clusters, and the solute molecules, eq 7 must be re-written in terms of SEs

$$\delta G_{\text{solv}}(A) = G_{\text{SE}}([AS_n]_{(\text{sol})}) - G_{\text{SE}}([S_n]_{(\text{sol})}) - G_{\text{SE}}(A)_{(\text{gas})} \quad (12)$$

Hence, for a complete cluster ensemble, that is, $n \rightarrow \infty$, the solvation energy should approach the true solvation free energy for transferring 1 mol/L solute from the gas phase into solution with the same concentration. The required volume work $p\Delta V$ is added to eq 12. Considering the SE for the solute molecule in the gas phase is necessary for flexible molecules and mentioned here for the sake of completeness. The application examples (vide infra) are restricted to rather rigid molecules, and hence, the ensemble average for $[A]_{(\text{gas})}$ can be neglected. If all entropic contributions are discarded and only the zero-point vibrational and $H(T)$ terms are taken from the mRRHO calculation, the solvation enthalpy $\delta H_{\text{solv}}(A)$ is obtained directly, without additional computational effort.

$$\delta H_{\text{solv}}(A) = H_{\text{SE}}([AS_n]_{(\text{sol})}) - H_{\text{SE}}([S_n]_{(\text{sol})}) - H_{\text{SE}}(A)_{(\text{gas})} \quad (13)$$

Hence, the extended QCG approach also enables the calculation of solvation enthalpies and entropies separately, which is, to the best of the authors' knowledge, a unique feature among existing hybrid continuum models. The subtracting scheme given in eq 12 automatically includes the energy required to create a cavity inside the solvent G_{cavity} including entropy contributions and the loss of solvent–solvent van der Waals interactions. Hence, in contrast to implicit solvation models, the estimation of G_{cavity} is not needed in QCG theory. Moreover, the electronic energies of the ensembles are computed by single-point computations without the wall potentials to ensure that the biasing potential does not enter the solvation free energy directly.

2.2.1. Generating Reference Clusters. Reference (“empty”) clusters of pure solvent molecules are generated by employing the newly developed cut-fix-fill (CFF) algorithm. A schematic representation is given in Figure 6.

The computational effort invested into the cluster ensemble generation is recycled in the process of generating the reference cluster ensemble by replacing the solute molecule in the generated clusters with a number of solvent molecules.

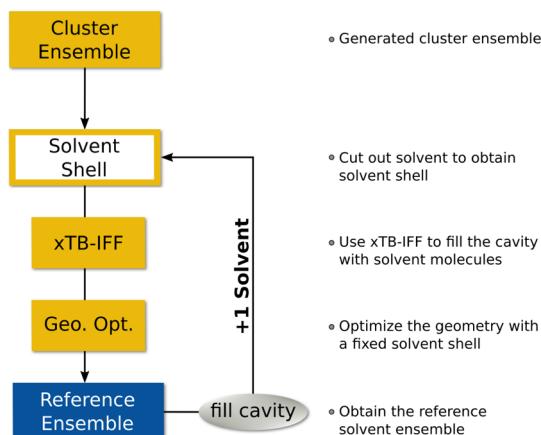


Figure 6. Workflow of the CFF algorithm. The solute is cut out from the cluster and replaced with solvent molecules. The energy of the desired reference cluster is interpolated from the filled cluster.

In a first step, the solute is cut out of the cluster and the remaining solvent shell is kept fixed in the following steps by setting the forces on the respective solvent molecules to zero. Second, analogous to the growing procedure, xTB-IFF docking and GFN2-xTB optimization steps are applied to fill the created cavity in a stepwise fashion. The cluster is considered to be filled when no further solvent molecule can be placed inside the cavity, which is supervised using energetic (positive interaction energy) and geometric (volume of inserted solvent molecules exceeds the solute volume) criteria. The CFF algorithm unavoidably increases the number of molecules. Thus, the energy of the reference cluster with the initial count of solvents is calculated from the energy of the filled cluster by multiplication with a factor $n/(n+x)$, where n is the initial number of solvents and x the number of solvents placed inside the cavity. The reason for using the CFF approach compared to an additional MD-based ensemble generation is that a close structural similarity of the solvent shell in both, the solute and the reference cluster, is ensured, leading to beneficial error compensation effects. This minimizes statistical noise and is important because slight changes in the solvation shell may cause significant changes in energy, especially for highly polar and protic solvents.

3. TECHNICAL DETAILS

The QCG algorithm is implemented in the open-source CREST⁸⁵ program, and a detailed manual with examples can be found online.⁸⁶ Here, we will limit the explanation only to the most important features. The QCG algorithm is invoked by crest <solute> -qcg <solvent>. The user has to provide only the solute and solvent input coordinates. As discussed in the previous sections, the QCG algorithm consists of (i) the cluster growth, (ii) the ensemble generation, and (iii) the reference cluster construction for the computation of solvation free energies, including the respective calculation of thermochemical contributions. The cluster growth is invoked with the addition of -grow to the command line input and can be carried out with all GFN methods. By default, GFN2-xTB is employed as the xTB-IFF docking is usually the time-limiting factor. The added number of solvent molecules can either be set manually (-nsolv) or automatically determined from a moving average threshold of the solute–solvent interaction energy (*cf.* Section 2.1). In the case of water as solvent, the outer wall potential is scaled per default by 0.7. This factor is increased by 5% each time the interaction energy is positive or if the default of 1.0 is reached. It can be adjusted by the user for any solvent.

To conduct the ensemble generation of the grown cluster, the -ensemble flag is employed (instead of -grow). As a starting point, a cluster after the growth algorithm is used that is either generated during the same or a previous run. Therefore, a single MD/MTD simulation or the NCI–MTD run type is available. By default, the latter is performed at the GFN-FF level of theory with MTD and MD simulations of 10 ps length, respectively. To obtain the qualitatively best cluster ensemble at reasonable computational costs, final geometry optimizations are conducted with GFN2-xTB. The MD/MTD length can also be varied by the command line input. Choosing the single MD/MTD simulation instead of the NCI–MTD run type, a GFN2-xTB MD/MTD simulation of 10 ps length is performed by default at 298 K. The computation of the solvation free energy is invoked by -gsolv and requires a solute–solvent cluster ensemble. δG_{solv} values are computed

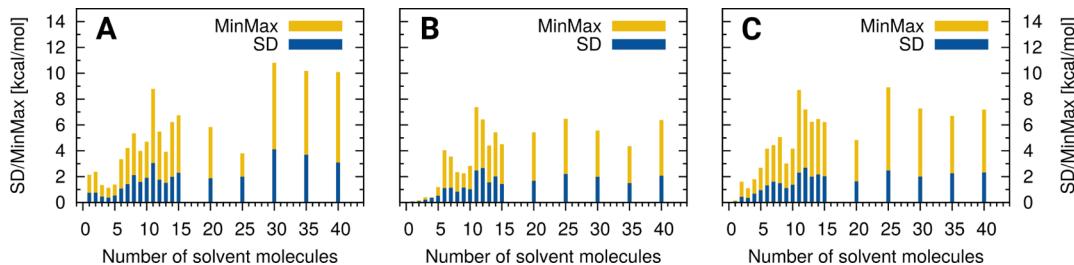


Figure 7. SD and spread between the energetically lowest and highest cluster (MinMax) given for (A) electronic energies after the cluster growth, (B) conformer sampling (ensemble generation), and (C) reference ensemble generation of ethanol in acetonitrile averaged over 10 runs each.

according to eq 12, where the free energies of the solute and reference clusters are obtained as the Boltzmann-weighted average of the ensemble (G_{SE} cf. eq 10). By default, every solute cluster populated by more than 10% is taken into account. Starting from these structures, the reference cluster ensemble is constructed using the CCF algorithm (cf. Section 2.2). Every geometry optimization during the CFF algorithm is performed with GFN2-xTB, similar to the ensemble generation step. For the final single-point calculation, the solute and reference clusters are additionally embedded in a continuum solvation model. Either the GBSA or ALPB solvation model can be chosen.^{62,77} The total free energy of a cluster is calculated as the sum of the single-point energy and thermostatistical contribution G_{mRRHO} . Therefore, QCG computes the harmonic vibrational frequencies of all solute and reference clusters populated more than 10% to ensure computational efficiency by considering the most relevant structures. For this step, GFN-FF is the default, but all GFN methods can be applied. Entering the liquid phase limits the translational and rotational DOFs of the solute molecule. To mimic this effect, the corresponding entropic contributions to G_{mRRHO} are reduced by 25% (cf. eq 14)

$$G_{mRRHO} = (H_{\text{trans}} + H_{\text{rot}} + H_{\text{vib}}) - T[0.75(S_{\text{trans}} + S_{\text{rot}}) + S_{\text{vib}}] \quad (14)$$

whereas the vibrational contribution remains unchanged. The scaling factor of 0.75 was empirically determined and can be adjusted for each solvent individually. Nevertheless, this imposes no restriction to the general applicability of the QCG approach. Finally, the conformational free energy part G_{conf} of the cluster ensembles and the volume work to transfer a solute molecule from an ideal gas to an ideal solution at molar concentration (1 mol L⁻¹) are included in QCG.

4. COMPUTATIONAL DETAILS

The QCG algorithm was applied as implemented in the CREST program.^{72,85} Unless stated otherwise, the default settings were employed throughout this work. Final single-point energies were calculated at the GFN2-xTB level of theory with the implicit GBSA solvation model for the solute and reference clusters.^{62,77} All GFN n -xTB and GFN-FF calculations were performed with the xtb 6.4.0 program package.⁸⁷ To decrease the statistical error, each solvation free energy computation with QCG was performed 10 times and averaged. The number of solvent molecules was determined to complete at least the first solvation shell. For comparison, solvation contributions to the free energy were calculated with COSMO-RS,³⁰ also including the volume work required for changing from an ideal gas at 1 bar to 1 mol L⁻¹ to solution. For the

COSMO-RS free energy, the BP_TZVP_C30_1601 parameterization was used. Two single-point calculations with BP86^{88,89}/def-TZVP⁹⁰ (one in the gas phase and one in an ideal conductor) were performed, and the output of these calculations was then processed using the COSMOtherm program package.^{31,91} The efficient B3LYP-3c⁹² and r²SCAN-3c⁹³ DFT composite methods were employed as implemented in the TURBOMOLE program package (version 7.5.1)^{94–96} together with the COSMO model. Harmonic vibrational frequencies were calculated analytically using the aforce implementation in TURBOMOLE and scaled by a factor of 0.97 in the case of B3LYP-3c.⁹² Visualization of molecules was performed with the UCSF Chimera (version 1.15)⁹⁷ program and gnuplot (version 5.0)⁹⁸ was employed for plotting.

5. RESULTS AND DISCUSSION

5.1. Reproducibility. The QCG algorithm consists of multiple steps (cf. section 2.1), with each of it containing non-deterministic components (i.e., docking or MD/MTD). Regarding the complexity of the phase space for a large explicitly solvated cluster, prohibitively long simulation times would be required to always converge to the same solution. In practice, the finite simulation time introduces a statistical error, which is investigated here. First, we determine the standard deviation (SD) resulting from the docking procedure in the growth step, the subsequent finite simulation times for the solute ensemble, and the reference ensemble generation. Exemplarily, a system of acetonitrile in ethanol (Figure 7) is elaborated. Different-sized clusters (1–40 solvent molecules) were generated 10 times each using the same settings and the SD and maximal spread (MinMax) respective to the averaged electronic energies were analyzed and their average over all number of solvents (\overline{SD} and $\overline{\text{MinMax}}$).

It is found that already the cluster growth algorithm consisting of xTB-IFF docking and GFN2-xTB optimizations (Figure 7A) yields (slightly) different structures for the repeated computations with the same setup and introduces scattering of the cluster energies. Averaging over all cluster sizes, the \overline{SD} and $\overline{\text{MinMax}}$ amount to 1.8 and 5.1 kcal mol⁻¹ with respect to the average energy. Different structures within the growth process result from the docking algorithm in xTB-IFF. Typically, the first few solvent molecules are always added at the same position, explaining the small SD and MinMax for small clusters. However, larger clusters exhibit energetically very similar docking positions and, hence, different binding sites are occupied. QCG employs the NCI–MTD algorithm to explore the low-energy conformational space after the growing process. In Figure 7B, the ensemble energy is shown as the Boltzmann-weighted sum of all clusters within a generated ensemble. Again, the statistical errors are evaluated over 10

equivalent runs. In general, the conformer sampling compensates partially for the scattering introduced by the growth process, as similar energetically low clusters are found by repeated conformational samplings. The average \overline{SD} reduces to 1.3 kcal mol⁻¹ and the $\overline{\text{MinMax}}$ value to 3.6 kcal mol⁻¹ upon the ensemble generation. For a small number of added solvents, the SD and MinMax values are close to zero. As the conformational space becomes larger with increasing cluster size, the SD and MinMax values also increase.

A further source of error is introduced using the CFF algorithm (Figure 7C). Differently shaped cavities within the frozen cluster shells can be filled with a variable number of solvent molecules. This inconsistency increases the average \overline{SD} and $\overline{\text{MinMax}}$ to 1.5 and 4.9 kcal mol⁻¹, respectively. Overall, the error of the CFF algorithm is larger than the error after ensemble search and comparable to the error of the growing process because the added solvent molecules introduce an additional conformational error. Additionally, problems arise if spatially different conformations of the solute molecule are possible. For example, a bent conformation of *n*-octanol can be placed in cavities, which is otherwise too small for the extended conformation.

Increased simulation times in the NCI-MTD run were investigated. Therefore, again 10 calculations were repeated per number of solvent molecules (ranging from 1 to 15) for different MTD lengths. The resulting SDs and MinMax values for each number of solvent molecules were averaged for each MTD length. These averages (\overline{SD} and $\overline{\text{MinMax}}$) are shown in Table 1.

Table 1. \overline{SD} and $\overline{\text{MinMax}}$ Values of the Ensemble Energy (in kcal mol⁻¹) for Ethanol in Acetonitrile with Different MTD Lengths During the NCI-MTD Step^a

	1.2 ps	10 ps	50 ps	100 ps	single MD (10 ps)
\overline{SD}	1.05	0.97	0.83	0.63	1.08
$\overline{\text{MinMax}}$	3.19	3.02	2.54	1.87	3.25

^aSD and MinMax are averaged over 15 different cluster sizes with up to 15 solvent molecules.

The comparison of different simulation times shows that longer MTDs systematically reduce the scattering, for example, increasing the MTD simulation time from 1.2 to 100 ps reduces the \overline{SD} by 40%. One also notices that the NCI-MTD algorithm employing 1.2 ps of simulation time has the same effect on the SD as a normal MD simulation of 10 ps. Increased simulation times further yield energetically lower clusters. Throughout this work, we chose the energy gain upon cluster formation as a measure of its quality. Figure 8 shows the energy difference between the clusters after the growth process and the most populated clusters (MPCs) after the conformational sampling. Except for some fluctuation, longer MTD times yield in general larger energy gains ($E_{\text{MPC}} - E_{\text{grow}}$). Thus, the MTDs in the NCI-MTD run should be chosen as long as possible to yield better ensembles and to reduce the statistical error. However, for larger cluster sizes, long MTD simulations become prohibitively expensive.

5.2. Cluster Quality. In this section, we want to assess the quality of the generated clusters with QCG, that is, the physical meaningfulness of the found binding motifs including solute–solvent and solvent–solvent interactions. For quantification, the energy of cluster formation E_{form} is computed according to

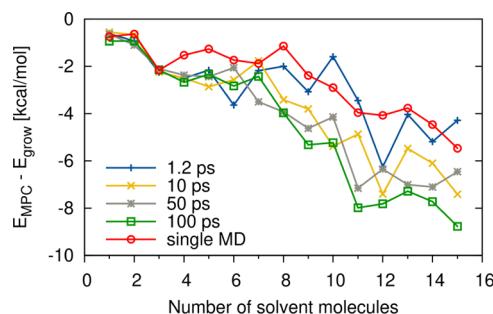


Figure 8. Difference in energy between the most populated cluster (MPC) found using the NCI-MTD algorithm and the input cluster from the growth step for different-sized clusters of ethanol in acetonitrile. Each value is averaged over 10 individual runs.

$$E_{\text{form}} = E_{\text{cluster}} - E_{\text{solute}} - n \cdot E_{\text{solvent}} \quad (15)$$

where the energy of the solute E_{solute} and n times the energy of the solvent molecule E_{solvent} are subtracted from the energy of the cluster E_{cluster} . The lower (more negative) the energy of formation, the higher is the quality of the assessed cluster.

The performance of QCG in terms of cluster generation (-grow) is compared to different algorithms. This intention turned out to be rather difficult as, to the best of the authors' knowledge, hardly any other algorithm exists that automatically generates solute–solvent clusters. One program package is AMBERtools.⁹⁹ Therein, different solvent models can be chosen, but they are restricted to a few solvents. One of these is water, for example, employing the TIP3P model.¹⁰⁰ During the cluster generation with AMBERtools, a sphere with a user-defined radius is cut out of a pre-equilibrated box of 216 water molecules. Subsequently, the solute is placed inside this cavity, and solvent molecules that collide are removed. We also implemented a second competitor that relies only on geometrical criteria. Therein, solvent molecules are randomly placed around the solute within a sphere of a given radius until the entire volume is filled. To ensure a reliable comparison, the structures resulting from AMBERtools and the space-filling algorithm were post-optimized with GFN2-xTB as in the QCG algorithm.

As an example, phenylalanine was solvated with 60 water molecules. Figure 9 depicts the structures resulting from the QCG algorithm and AMBERtools after geometry optimization with GFN2-xTB. Table 2 shows the corresponding formation energies for the three methods.

The space-filling algorithm shows the highest formation energies and hence performs the worst. Tentatively this can be assigned to the neglect of intermolecular interactions during cluster generation. Even though subsequent GFN2-xTB geometry optimizations improve the structure, the starting point is still too far off from a global minimum to be repaired in just a single geometry optimization. The AMBERtools procedure yields a lower (better) formation energy than that of the space-filling model. The priorly equilibrated water box ensures a more physical solvent structure, and for bulk water, the TIP3P water model delivers a good description. For a molecular cluster, this description may not be optimal as explicit solute–solvent interactions during the cluster generation are neglected. This may lead to an unreasonable solute surrounding, which can only partially be compensated by the GFN2-xTB geometry optimization. The QCG algorithm optimizes the solute–solvent NCIs during each step of the

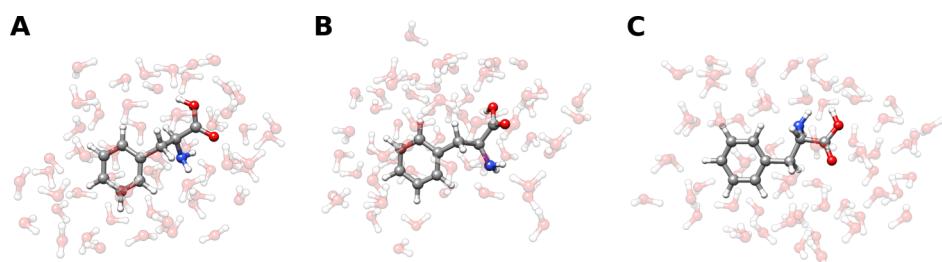


Figure 9. Phenylalanine surrounded by 60 water molecules generated using QCG (A), the TIP3P water model (B), and the space-filling algorithm (C) with subsequent GFN2-xTB geometry optimization.

Table 2. Energy of Formation (in kcal mol^{-1}) for the Cluster out of Phenylalanine and 60 Water Molecules Computed With GFN2-xTB and $r^2\text{SCAN-3c}$ single-point DFT Computations^a

	QCG (-grow)	AMBERtools (TIP3P)	space-filling
GFN2-xTB	-712.4	-684.1	-665.3
$r^2\text{SCAN-3c}/\text{GFN2-xTB}$	-656.7	-632.1	-622.8

^aThe QCG algorithm is compared to AMBERtools and the space-filling algorithm, both with subsequent GFN2-xTB geometry optimizations.

iterative cluster growth and, hence, yields the best cluster formation energies. At the $r^2\text{SCAN-3c}/\text{GFN2-xTB}$ level, the energy gain is $24.6 \text{ kcal mol}^{-1}$ ($28.3 \text{ kcal mol}^{-1}$ for GFN2-xTB) larger than with the TIP3P model and $33.9 \text{ kcal mol}^{-1}$ ($47.1 \text{ kcal mol}^{-1}$) larger than with the space-filling approach. The generation of an initial energetically low cluster is mandatory for the subsequent ensemble generation (conformer sampling) in QCG. Due to the enormous complexity of the phase space and the limitation to finite simulation lengths, the outcome of the NCI-MTD algorithm strongly depends on the quality of the input cluster. Hence, the elaborated growing routine is indispensable for QCG. The computational cost of the cluster growth (16 min on four cores of an Intel Xeon CPU E3-1270 v5@3.60 GHz) is much lower than that of the subsequent ensemble generation (2 h 4 min with the same

CPU). In passing, we note the good agreement between the total cluster formation energies of the very reliable $r^2\text{SCAN-3c}$ DFT composite and semi-empirical GFN2-xTB methods (deviation of about 10%).

5.3. Microsolvation. For small cluster sizes, the QCG algorithm shows small statistical errors (cf. Section 5.1). Hence, it seems to be a promising and easy-to-use tool for applications in the context of microsolvation. For demonstration, ensembles of benzoic acid and 2-amino-benzothiazol were generated with QCG (-ensemble) at the GFN2-xTB level of theory adding three explicit water molecules to the solutes. Distinct conformations of the found clusters were energetically sorted by $r^2\text{SCAN-3c}$ single-point calculations. Selected structures are shown in Figure 10. Relative $r^2\text{SCAN-3c}$ (in black) and GFN2-xTB (in blue) energies are given, respectively.

In the case of benzoic acid (Figure 10A), the most favorable structures at the $r^2\text{SCAN-3c}$ level are cyclic conformations, including three water molecules and the carboxylic acid functional group. The energetic order of the conformers for (A1, A2) changes at the GFN2-xTB level. Structures (A3, A4) form a cyclic arrangement between two water molecules and the carboxylic acid group. The energy difference between these structures of $1.7 \text{ kcal mol}^{-1}$ ($1.5 \text{ kcal mol}^{-1}$) occurs due to the different position of the third water molecule, forming either an intermolecular HB to the C=O or the O-H group of the benzoic acid molecule. In the least favorite conformation (A5),

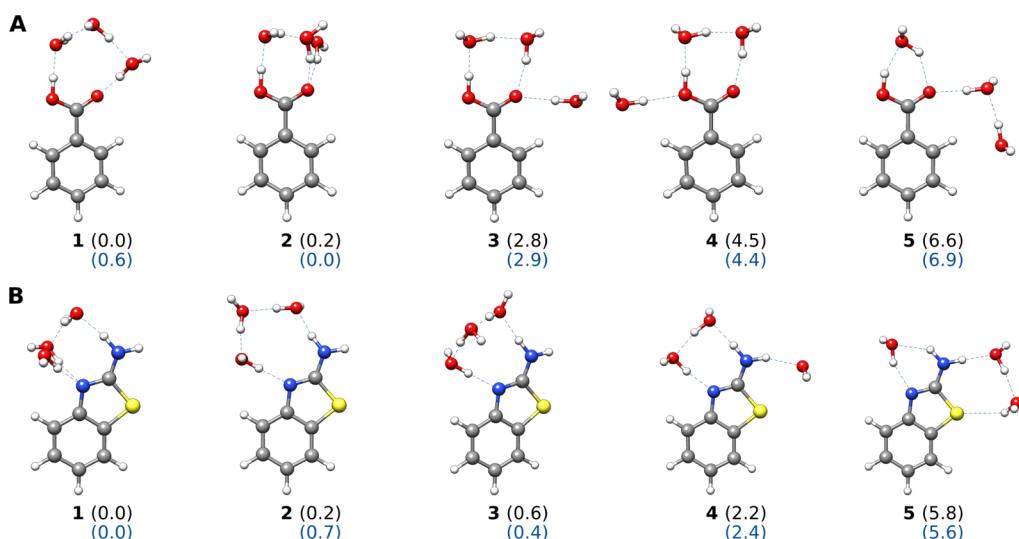


Figure 10. Microsolvated SE of benzoic acid (A) and aminobenzothiazole (B) with three explicit water molecules, respectively. Shown are five different conformations each. Relative conformational energies (in kcal mol^{-1}) were calculated with $r^2\text{SCAN-3c}$ (in black) and GFN2-xTB (in blue) and are given in parenthesis.

only two water molecules interact with the COOH group. For the structures (A3–A5), the energetic order of GFN2-xTB and r^2 SCAN-3c agree. For 2-amino-benzothiazol (Figure 10 B), the lowest energies are observed for the structures with intramolecular HBs between the water molecules and the amine fragments. Again, cyclic arrangements of the amine group with three water molecules are favored over those with two or one water molecules, respectively. The energetically higher clusters also contain intramolecular HBs to the sulfide group, which are less strong. For small energy differences between the conformers ($\leq 1 \text{ kcal mol}^{-1}$), a partially different ordering is observed for GFN2-xTB compared to r^2 SCAN-3c. Thus, in the context of microsolvation, we recommend a re-ranking by DFT as already suggested in ref 83 for non-rigid molecules because noncovalently bound clusters may also be regarded as highly flexible. To further validate the QCG ensembles, they were compared to results from a different microsolvation approach proposed in ref 67. Here, the ensembles for the same systems were obtained from MD simulations with subsequent grid inhomogeneous solvation theory (GIST) analysis.¹⁰¹ The resulting structures after B3LYP^{102,103}-D3¹⁰⁴/def2-TZVP¹⁰⁵ optimization are similar to the QCG ones and have the same interaction motifs. For example, the GIST analysis found a structure similar to (A3), and also a cluster with the same water docking positions as (B4) was observed. Even though the energy ranking depends on the chosen density functional, the inclusion of an implicit solvation model, and on the respective methods employed for the thermostatistical contributions, both approaches (GIST analysis and QCG) yield similar orderings. In terms of computation time, QCG, with subsequent r^2 SCAN-3c single-point calculations takes only a few minutes on a regular desktop computer (Figure 11).

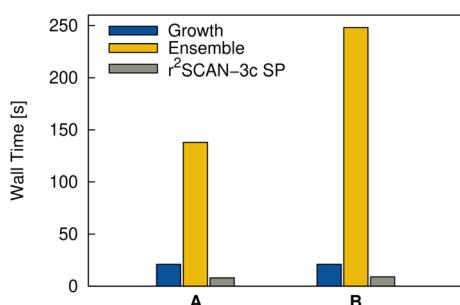


Figure 11. Computational timings of benzoic acid (A) and aminobenzothiazole (B) for the cluster growth, the ensemble generation, and the r^2 SCAN-3c single-point calculation with three explicit water molecules. The calculations were performed on four cores of an Intel Xeon CPU E3-1270 v5@3.60 GHz.

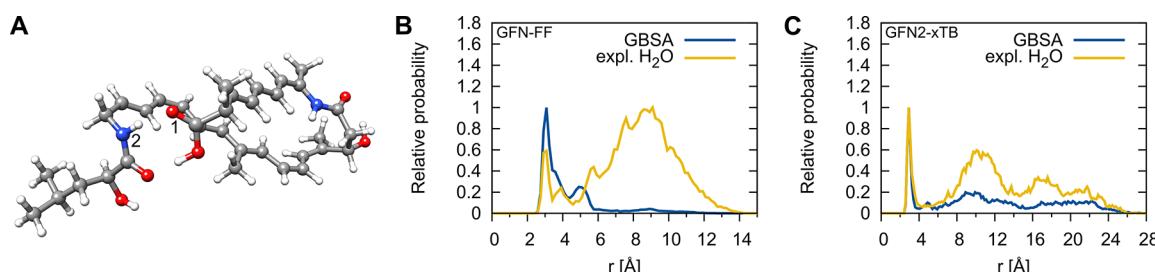


Figure 12. (A) Gas phase-optimized structure of the energetically lowest bacillaene conformer found at the GFN2-xTB level of theory. Distance distribution functions of atoms O1 and N2 obtained from GFN-FF (B) and GFN2-xTB (C) MD simulations employing the implicit GBSA solvation model and from a GFN-FF and GFN2-xTB MD simulation of the QCG cluster containing 100 water molecules.

The QCG approach combined with r^2 SCAN-3c single-point calculation (3 min 7 s and 4 min 39 s) outperforms other approaches solely based on MD simulations^{67,106} in terms of computation time. This computational efficiency coupled with high accuracy and robustness makes QCG a promising candidate for future applications in calculating free energies of reactions based on microsolvated structures.

5.4. Molecular Dynamics. The broad field of QCG applications also includes the impact of explicit solvation effects on geometries during MD simulations. Depending on the solvation model, qualitatively different geometries might result in solution and change the course and outcome of MD simulations. This is illustrated here in MD simulations of the natural antibiotic bacillaene¹⁰⁷ applying two different solvent models. First, 100 explicit water molecules were added with QCG to the gas phase-optimized geometry, and 1 ns GFN-FF and GFN2-xTB MD simulations at 298.15 K were performed under the application of an external wall potential. For comparison, MD simulations with equivalent settings were run using the implicit GBSA model. As a measure for the here relevant folding process, radial distribution functions (RDFs) of the O1–N2 interatomic distance (Figure 12A) were calculated from GFN-FF and GFN2-xTB trajectories. These RDFs show the frequency of occurrence for the intramolecular HBs that are mainly responsible for the ring-shaped geometry in the gas phase.

The RDFs computed with GFN-FF (Figure 12B) and GFN2-xTB (Figure 12C) show similar trends. Distances below 3 Å indicate that two intramolecular HBs are present, meaning the structure is similar to the gas phase. The second peak shortly above 3 Å shows a small elongation of the O–N distance, which is attributed to the cleavage of the HB between O1 and N2 under preservation of the second intramolecular HB. Other peaks occurring up to 6 Å show the formation of intramolecular HBs that were not present at the beginning, indicating a significant change of the structure. Distances of more than 6 Å indicate the dissociation and the separation of the two ends of the bacillaene chain. In general, GFN2-xTB yields a more elongated conformation compared to GFN-FF, reflected in a non-vanishing RDF beyond 14 Å. The MD results with GBSA reveal that in both simulations, the intramolecular HBs are predominantly formed yielding a closed structure. In the GFN2-xTB/GBSA MD simulation, a noteworthy occurrence of structures with no intramolecular HBs is obtained, which is strongly amplified by explicit water molecules. The same picture holds for GFN-FF. Hence, we conclude that the explicitly modeled solvent molecules lead mainly to an open structure of bacillaene and favor intermolecular over intramolecular HBs. Similar observations

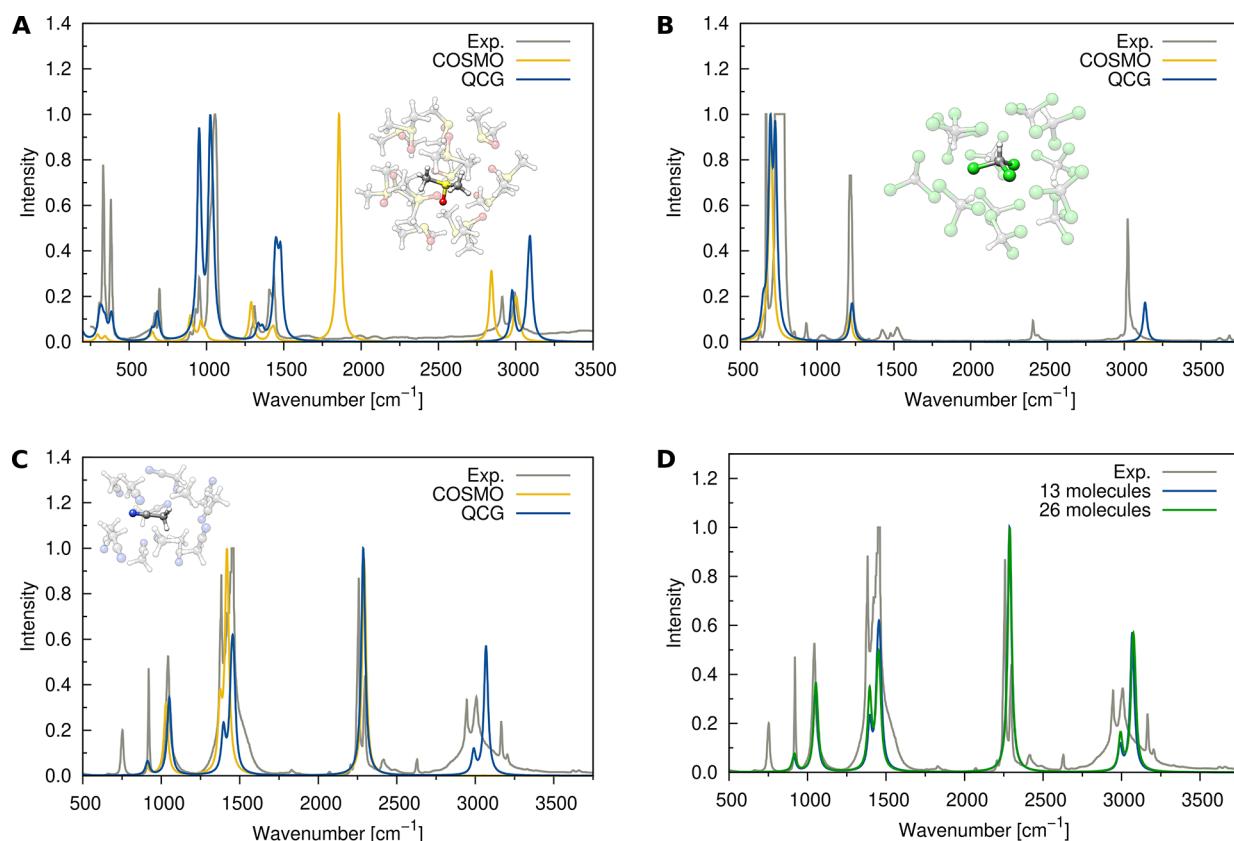


Figure 13. IR-spectra of liquid DMSO (A), CHCl₃ (B), and CH₃CN (C) computed at the B3LYP-3c level of theory from explicit QCG clusters and with the COSMO model in comparison to experimental data. Different QCG cluster sizes are investigated in (D) for CH₃CN.

were reported in a recent study of methyl lactate in explicit solvent clusters.¹⁰⁸

5.5. IR Spectra. The QCG algorithm was already successfully applied for IR spectra calculations in solution.^{108,109} Therein, it was demonstrated that solvation effects play an important role in liquid-phase IR spectra. Continuing on this work, the QCG approach was applied for IR spectra calculations of organic liquids in comparison to experimental data. The COSMO model was additionally tested as an alternative. For the explicit description with QCG, clusters consisting of one solvent shell were generated and the energetically lowest structure from the ensemble search was then re-optimized at the B3LYP-3c level of theory. To minimize finite-cluster size effects, the vibrations of the central solute molecule were separated from the vibrations of the surrounding molecules by increasing their atomic mass to shift the vibrational frequencies to the low-frequency region. For the COSMO approach, input geometries for the calculation of liquid-phase IR spectra were generated by optimizing gas-phase geometries at the B3LYP-3c level of theory employing the COSMO model. The results for different organic liquids are shown in Figure 13. In agreement with refs 108 and 110, only the most populated cluster was taken into account for IR spectra calculation as the effect of considering the entire SE is rather small.

The IR spectra computed with the COSMO model show peaks that are not present in the experimental or the QCG cluster spectrum, for example, in the case of liquid DMSO (Figure 13A), a peak of highest intensity occurs at ~1850 cm⁻¹. Moreover, the C–Cl (Figure 13B) and C–H (Figure 13C) stretching vibration signals are incorrectly almost not

visible in the case of chloroform and acetone. On the contrary, the QCG cluster ansatz yields peaks close to the experimental ones regarding position and intensity. Furthermore, changes in symmetry due to the surrounding solvent cannot be seen with COSMO but are captured by QCG. For example, the asymmetric C–Cl stretching vibrations at around 750 cm⁻¹ (Figure 13B) are different. For COSMO, the system has high C_{3v} symmetry leading to fewer signals than in the QCG cluster calculation where each chlorine atom is slightly different. This leads to a loss of symmetry and to multiple peaks for the C–Cl stretching and bending vibrations in better agreement with the experimental IR spectrum. Adding more explicit solvent molecules than one solvent shell has a minor influence, which is shown exemplarily for liquid acetonitrile (Figure 13D). The frequencies resulting from the differently sized clusters are mostly identical and the intensities differ only slightly, which is consistent with previous findings in ref 109. The overall good performance of the QCG approach is quantified using the respective spectral match score (for definition, see ref 92) reported in Table 3. For DMSO and CHCl₃, the IR spectra computed for QCG clusters yield an improvement of 12.0 and 24.8%, respectively, over the COSMO approach. For CH₃CN, the two different solvation models show similar match scores because COSMO performs better at around ~1450 cm⁻¹ while QCG provides the more realistic spectrum for the C–H stretching vibrations at ~3000 cm⁻¹.

Similar differences in IR spectra between implicit and explicit solvent models were also found in another study,¹¹¹ where the placement of solvent molecules was carried out

Table 3. Spectral Match score (in %, Identical Spectra Yield 100%) between the Computed IR Spectra at the B3LYP-3c Level Employing the QCG and COSMO Approaches and the Experimental Spectra for Liquid DMSO, CHCl₃, and CH₃CN

COSMO	molecule	QCG
56.7	DMSO	68.7
42.2	CHCl ₃	67.0
61.2	CH ₃ CN	60.8

manually. The automated procedure of QCG allows for the efficient study of different clusters of various sizes.

5.6. Solvation Free Energies. With QCG, it is possible to compute solvation free energies in a supermolecular fashion as described in Section 2.2. With the here discussed first proof-of-principle examples, we want to evaluate the performance of QCG in comparison to the (in our opinion) most accurate implicit model available (COSMO-RS¹¹²) and experimental values. Therefore, δG_{solv} were computed for 45 different systems at 298 K, combining polar and apolar organic molecules of different sizes. These are part of the SMD fit set³⁴ and are assumed to be also part of other fit sets for implicit solvation models. Each QCG calculation was performed 10 times, and the arithmetic mean was taken. The ensembles were generated with the NCI-MTD run type employing the default settings (e.g., an MTD time of 10 ps). For every solute, 25 solvent molecules were modeled explicitly, ensuring that at least the first solvation shell is completely covered. For water as a solvent, the number was increased to 40 molecules to ensure a complete first solvent shell. Figure 14 shows the correlation between QCG/COSMO-RS δG_{solv} values and the respective experimental results.³⁴ The corresponding statistical errors are given in Table 5.

COSMO-RS performs very well with an MD of 0.2 kcal mol⁻¹ and an MAD of 0.5 kcal mol⁻¹, respectively. The solvation free energies computed with QCG have an MD of -0.1 kcal mol⁻¹ and an MAD of 2.3 kcal mol⁻¹. Referring to the analysis of the statistical errors (cf. Section 5.1), it is clear

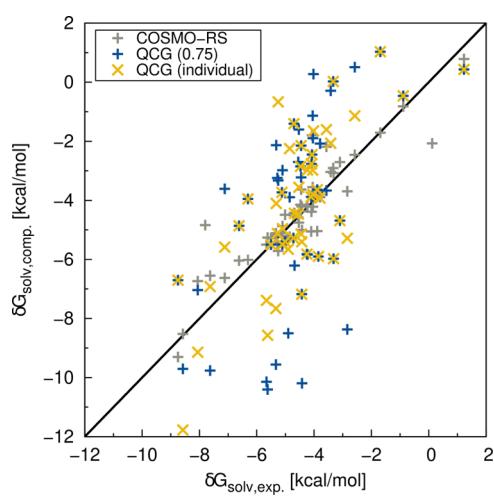


Figure 14. Correlation plot of δG_{solv} values of 33 small organic molecules computed with QCG and COSMO-RS in comparison to experimental values. QCG values are averaged over 10 individual runs. They are given for a global scaling factor of the translational and rotational entropy of 0.75 and for an empirically adjusted solvent-specific one (Table 4).

that observed scattering from the growth and ensemble steps is also inherent in the computed δG_{solv} values (see Supporting Information for further details). Taking this into account and regarding the fact that QCG subtracts huge energies in a "brute force" approach, the reasonable agreement with the experiment seems encouraging. Even though the accuracy of COSMO-RS cannot be reached, the already reasonable δG_{solv} values computed with QCG represent a promising starting point for further improvements, especially because of the universal applicability to arbitrary solute–solvent combinations. Currently, the following sources of error can be addressed. First, in contrast to other hybrid cluster continuum models, QCG generates two different sets of solute and solvent clusters, and thus, the errors within the electronic energies and thermochemical contributions occur twice. Also, the GFN2-xTB error for the interaction energies does not cancel out completely between the solute–solvent and the reference–solvent ensemble. Second, the ensemble might not be of similar quality for every system because an MTD time of 10 ps was chosen as a compromise between computational costs and accuracy. Longer simulation times should be tested comprehensively in the future. Third, using only solute–solvent clusters that are at least populated by 10% for the calculation of the solvation free energy leads to an error of the conformational entropy and thus the conformational free energy G_{conf} . For example, a small cluster of three chloroform molecules has approximately a G_{conf} value of -5.2 kcal mol⁻¹ at 298.15 K, while taking only clusters that are populated by at least 10% yield a lower G_{conf} of -1.3 kcal mol⁻¹. However, partial error compensation of solute–solvent and solvent–solvent ensembles is expected. Including more solute–solvent clusters would lead to a much higher computational expense. Fourth, the rather simple GBSA model applied for electrostatic bulk screening is computationally efficient but might introduce further errors. A future implicit solvent model that describes the electrostatic response of a solvent and accounts for polarization should improve the description of the bulk. Finally, the scaling factor for the translational and rotational entropy contributions was empirically determined to be 0.75 as average over many different solvents. An improvement can be obtained by adjusting it for different solvents (Table 4), leading to a 0.8 kcal mol⁻¹ lower MAD (Table 5).

Table 4. Adjusted Scaling Factors for the Rotational and Translational Entropy Contributions of Different Solvents

solvent	scaling factors
benzene	0.65
water	0.75
CH ₃ CN	0.85
DMSO	0.90

Various technical aspects can be improved, thereby increasing the computational cost, for example, increasing the MTD times or including also less populated solute–solvent clusters. Additionally, $\Delta \delta G_{\text{solv}}$ for reactions may be easier to compute than absolute δG_{solv} due to error cancellation.

6. CONCLUSIONS

We developed and tested an automated and broadly applicable model for a QM description of explicit solvation. This procedure termed QCG is based on the GFN-FF and GFN2-xTB methods in combination with xTB-IFF, giving

Table 5. Statistical Measures (in kcal mol⁻¹) for the δG_{solv} Values of Small Organic Molecules Computed with QCG/NCI-MTD and COSMO-RS in Comparison to Experimental Values^a

	QCG(0.75)	QCG (individual)	COSMO-RS
MD	-0.07	-0.02	0.21
MAD	2.34	1.57	0.49
Rmsd	2.86	2.12	0.75
SD	2.89	2.14	0.73

^aQCG values are given for a scaling factor of the translational and rotational entropy of 0.75 and for each solvent individually adjusted one.

access to fast geometry optimizations, MD simulations, and docking steps. The conformational space exploration of the generated molecular clusters is conducted using the NCI-MTD algorithm of the CREST program. This enables systematic improbability of the QCG approach by extending the simulation time within the conformer search for a smaller statistical error and structures with lower energy in the ensemble. The QCG algorithm includes only very few empirical parameters (wall potential and translational/rotational scaling), as most of the required parameterization for efficient treatment is inherent in the underlying QM/FF methods. The presented approach is unique in regard to the fully automated cluster growth and ensemble generation of arbitrary solute–solvent combinations. We tested the QCG approach on a large variety of chemical systems, reaching from small organic molecules to large anti-cancer drugs in eutectic solvents. We individually analyzed the different underlying steps (-grow, -ensemble, and -gsolv) in the workflow and found that the reproducibility in terms of molecular structures and energies is good for small cluster sizes. Increasing their size leads to notable deviations in terms of structures and energies between the same calculations performed multiple times. The incomplete sampling of a cluster's phase space was determined to be the main source of errors. Nevertheless, the increase in simulation times during the conformer sampling (ensemble generation) reduced the statistical energy error significantly. We showed that QCG can be straightforwardly applied in microsolvation studies. Moreover, many computed properties may benefit from including explicit solvent molecules and QCG offers a simple and automated way to generate these structures that can be used, for example, for simulating geometries and IR spectra in solution. Here, significant improvements were observed compared to the COSMO model. QCG further represents a physically reasonable procedure for the calculation of solvation free energies by including all terms in δG_{solv} explicitly. Even though the accuracy of the established, highly-parameterized implicit COSMO-RS model was not reached, reasonable results coupled with universal applicability are promising for future improvements.

In conclusion, the new, freely available QCG tool can help to investigate and understand solvation effects at a molecular level. Due to its computational demand, QCG is not meant to replace existing, efficient continuum models. QCG establishes an alternative solvation tool that is capable to obtain reasonably accurate results for complex molecular systems where implicit methods reach their limits. The universal applicability to arbitrary solute–solvent combinations is a unique feature that is yet missing in the portfolio of solvation

tools, and we hope that QCG will be useful for computational chemistry.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00239>.

Computational results, availability, and statistical error measures ([PDF](#))

IR spectra, molecular dynamics, cluster quality, and microsolvation ([ZIP](#))

AUTHOR INFORMATION

Corresponding Author

Stefan Grimme — Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany; orcid.org/0000-0002-5844-4371; Email: grimme@thch.uni-bonn.de

Authors

Sebastian Spicher — Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany; orcid.org/0000-0001-6541-3680

Christoph Plett — Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany; orcid.org/0000-0002-3911-2478

Philipp Pracht — Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany; orcid.org/0000-0002-8495-9504

Andreas Hansen — Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany; orcid.org/0000-0003-1659-8206

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.2c00239>

Author Contributions

[†]S.S. and C.P. contributed equally

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the DFG in the framework of the “Gottfried-Wilhelm-Leibniz” prize to S.G. S.S. thanks the “Fond der chemischen Industrie (FCI)” for financial support. The authors thank F. Bohle, M. Bursch, S. Ehler, D. Menche, and M. Reuter-Schniete for helpful input and discussions.

REFERENCES

- (1) Lindström, U. M. Stereoselective organic reactions in water. *Chem. Rev.* **2002**, *102*, 2751–2772.
- (2) Toupin, M.; Brousse, T.; Bélanger, D. Charge storage mechanism of MnO₂ electrode used in aqueous electrochemical capacitor. *Chem. Mater.* **2004**, *16*, 3184–3190.
- (3) Nicholls, A.; Sharp, K. A.; Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **1991**, *11*, 281–296.
- (4) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884.

- (5) van der Spoel, D.; Zhang, J.; Zhang, H. Quantitative predictions from molecular simulations using explicit or implicit interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *12*, No. e1560.
- (6) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
- (7) Hub, J. S.; Caleman, C.; van der Spoel, D. Organic molecules on the surface of water droplets—an energetic perspective. *Phys. Chem. Chem. Phys.* **2012**, *14*, 9537–9545.
- (8) Daver, H.; Algarra, A. G.; Rebek, J.; Harvey, J. N.; Himo, F. Mixed Explicit–Implicit Solvation Approach for Modeling of Alkane Complexation in Water-Soluble Self-Assembled Capsules. *J. Am. Chem. Soc.* **2018**, *140*, 12527–12537.
- (9) Smith, P. E.; Pettitt, B. M. Modeling solvent in biomolecular systems. *J. Phys. Chem.* **1994**, *98*, 9700–9711.
- (10) Levy, R. M.; Gallicchio, E. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531–567.
- (11) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J. Chem. Inf. Model.* **2015**, *55*, 1192–1201.
- (12) Skynner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191.
- (13) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (14) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (15) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (16) König, G.; Pickard, F. C.; Mei, Y.; Brooks, B. R. Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4. *J. Comput. Aided Mol. Des.* **2014**, *28*, 245–257.
- (17) Schwörer, M.; Wichmann, C.; Tavan, P. A polarizable QM/MM approach to the molecular dynamics of amide groups solvated in water. *J. Chem. Phys.* **2016**, *144*, 114504.
- (18) Wang, L.-P.; Van Voorhis, T. A Polarizable QM/MM Explicit Solvent Model for Computational Electrochemistry in Water. *J. Chem. Theory Comput.* **2012**, *8*, 610–617.
- (19) Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable embedding QM/MM: the future gold standard for complex (bio)systems? *Phys. Chem. Chem. Phys.* **2020**, *22*, 14433–14448.
- (20) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the revision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.
- (21) Cammi, R.; Tomasi, J. Remarks on the use of the apparent surface charges (ASC) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *J. Comput. Chem.* **1995**, *16*, 1449–1458.
- (22) Davis, M. E.; McCammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **1990**, *90*, 509–521.
- (23) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (24) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (25) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- (26) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- (27) Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.
- (28) Mennucci, B.; Cammi, R.; Tomasi, J. Excited states and solvatochromic shifts within a nonequilibrium solvation approach: A new formulation of the integral equation formalism method at the self-consistent field, configuration interaction, and multiconfiguration self-consistent field level. *J. Chem. Phys.* **1998**, *109*, 2798–2807.
- (29) 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**, 799–805.
- (30) Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (31) Eckert, F.; Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* **2002**, *48*, 369–385.
- (32) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 699–709.
- (33) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (34) 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.
- (35) Eckert, F.; Leito, I.; Kaljurand, I.; Kütt, A.; Klamt, A.; Diedenhofen, M. Prediction of acidity in acetonitrile solution with COSMO-RS. *J. Comput. Chem.* **2009**, *30*, 799–810.
- (36) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Prediction of SAMPL2 aqueous solvation free energies and tautomeric ratios using the SM8, SM8AD, and SMD solvation models. *J. Comput. Aided Mol. Des.* **2010**, *24*, 317–333.
- (37) Assaf, K. I.; Florea, M.; Antony, J.; Henriksen, N. M.; Yin, J.; Hansen, A.; Qu, Z.-w.; Sure, R.; Klapstein, D.; Gilson, M. K.; Grimme, S.; Nau, W. M. HYDROPHOBES Challenge: A Joint Experimental and Computational Study on the Host–Guest Binding of Hydrocarbons to Cucurbiturils, Allowing Explicit Evaluation of Guest Hydration Free-Energy Contributions. *J. Phys. Chem. B* **2017**, *121*, 11144–11162.
- (38) Fleck, N.; Heubach, C.; Hett, T.; Spicher, S.; Grimme, S.; Schiemann, O. Ox-SLIM: Synthesis of and Site-Specific Labelling with a Highly Hydrophilic Tritly Spin Label. *Chem.—Eur. J.* **2021**, *27*, 5292–5297.
- (39) Larsson, D. S. D.; van der Spoel, D. Screening for the Location of RNA using the Chloride Ion Distribution in Simulations of Virus Capsids. *J. Chem. Theory Comput.* **2012**, *8*, 2474–2483.
- (40) Larsson, D.; Lindahl, E. A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling. *J. Chem. Theory Comput.* **2010**, *31*, 2593–2600.
- (41) Zhang, H.; Tan, T.; van der Spoel, D. Generalized Born and Explicit Solvent Models for Free Energy Calculations in Organic Solvents: Cyclodextrin Dimerization. *J. Chem. Theory Comput.* **2015**, *11*, 5103–5113.
- (42) Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017**, *13*, 1034–1043.
- (43) Alder, B. J.; Wainwright, T. E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459–466.
- (44) Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys. Rev.* **1964**, *136*, A405.
- (45) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

- (46) Metropolis, N.; Ulam, S. The monte carlo method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
- (47) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (48) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (49) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (50) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **2010**, *6*, 1509–1519.
- (51) Martins, S. A.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Prediction of solvation free energies with thermodynamic integration using the general amber force field. *J. Chem. Theory Comput.* **2014**, *10*, 3570–3577.
- (52) Brancato, G.; Rega, N.; Barone, V. A hybrid explicit/implicit solvation method for first-principle molecular dynamics simulations. *J. Chem. Phys.* **2008**, *128*, 144501.
- (53) Åqvist, J. Ion–water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- (54) Weinhold, F. Quantum cluster equilibrium theory of liquids: General theory and computer implementation. *J. Chem. Phys.* **1998**, *109*, 367–372.
- (55) Kirchner, B.; Spickermann, C.; Lehmann, S. B. C.; Perl, E.; Langner, J.; von Domaros, M.; Reuther, P.; Uhlig, F.; Kohagen, M.; Brüssel, M. What can clusters tell us about the bulk?: Peacemaker: Extended quantum cluster equilibrium calculations. *Comput. Phys. Commun.* **2011**, *182*, 1428–1446.
- (56) Zaby, P.; Ingenmey, J.; Kirchner, B.; Grimm, S.; Ehrlert, S. Calculation of improved enthalpy and entropy of vaporization by a modified partition function in quantum cluster equilibrium theory. *J. Chem. Phys.* **2021**, *155*, 104101.
- (57) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput. Aided Mol. Des.* **2017**, *31*, 1–19.
- (58) Daver, H.; Algarra, A. G.; Rebek, J., Jr; Harvey, J. N.; Himo, F. Mixed Explicit–Implicit Solvation Approach for Modeling of Alkane Complexation in Water-Soluble Self-Assembled Capsules. *J. Am. Chem. Soc.* **2018**, *140*, 12527–12537.
- (59) Wu, W.; Kieffer, J. New Hybrid Method for the Calculation of the Solvation Free Energy of Small Molecules in Aqueous Solutions. *J. Chem. Theory Comput.* **2018**, *15*, 371–381.
- (60) Kumar, G.; Qu, Z.-W.; Ghosh, S.; Grimm, S.; Chatterjee, I. Boron Lewis Acid-Catalyzed Regioselective Hydrothiolation of Conjugated Dienes with Thiols. *ACS Catal.* **2019**, *9*, 11627–11633.
- (61) Bensberg, M.; Türtscher, P. L.; Unsleber, J. P.; Reiher, M.; Neugebauer, J. Solvation Free Energies in Subsystem Density Functional Theory. *J. Chem. Theory Comput.* **2022**, *18*, 723.
- (62) Bannwarth, C.; Caldeweyher, E.; Ehrlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimm, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1493.
- (63) Sure, R.; El Mahdali, M.; Plajer, A.; Deglmann, P. Towards a converged strategy for including microsolvation in reaction mechanism calculations. *J. Comput. Aided Mol. Des.* **2021**, *35*, 473–492.
- (64) Pliego, J. R., Jr; Riveros, J. M. Hybrid discrete-continuum solvation methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, No. e1440.
- (65) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous solvation free energies of ions and ion–water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J. Phys. Chem. B* **2006**, *110*, 16066–16081.
- (66) 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.
- (67) Steiner, M.; Holzknecht, T.; Schauperl, M.; Podewitz, M. Quantum Chemical Microsolvation by Automated Water Placement. *Molecules* **2021**, *26*, 1793.
- (68) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *J. Comput. Chem.* **2020**, *41*, 1144–1155.
- (69) Hruska, E.; Gale, A.; Huang, X.; Liu, F. AutoSolvate: A toolkit for automating quantum chemistry design and discovery of solvated molecules. *J. Chem. Phys.* **2022**, *156*, 124801.
- (70) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *J. Chem. Phys.* **2017**, *147*, 161708.
- (71) Bannwarth, C.; Ehrlert, S.; Grimm, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (72) Pracht, P.; Bohle, F.; Grimm, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.
- (73) Grimm, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
- (74) Spicher, S.; Grimm, S. Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem., Int. Ed.* **2020**, *59*, 15665–15673.
- (75) Spicher, S.; Abdullin, D.; Grimm, S.; Schiemann, O. Modeling of spin–spin distance distributions for nitroxide labeled biomacromolecules. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24282–24290.
- (76) Spicher, S.; Bursch, M.; Grimm, S. Efficient Calculation of Small Molecule Binding in Metal–Organic Frameworks and Porous Organic Cages. *J. Phys. Chem. C* **2020**, *124*, 27529–27541.
- (77) Ehrlert, S.; Stahn, M.; Spicher, S.; Grimm, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4250–4261.
- (78) Buša, J.; Džurina, J.; Hayryan, E.; Hayryan, S.; Hu, C. K.; Plavka, J.; Pokorný, I.; Skivánek, J.; Wu, M. C. ARVO: A Fortran package for computing the solvent accessible surface area and the excluded volume of overlapping spheres via analytic equations. *Comput. Phys. Commun.* **2005**, *165*, 59–96.
- (79) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. The proton’s absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
- (80) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (81) Grimm, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem.—Eur. J.* **2012**, *18*, 9955–9964.
- (82) Tomaník, L.; Muchová, E.; Slavíček, P. Solvation energies of ions with ensemble cluster-continuum approach. *Phys. Chem. Chem. Phys.* **2020**, *22*, 22357–22368.
- (83) Grimm, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **2021**, *125*, 4039–4054.
- (84) Pracht, P.; Grimm, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551–6568.
- (85) *Conformer-Rotamer Ensemble Sampling Tool Based on the xtb Semiempirical Extended Tight-Binding Program Package Crest*, 2021, <https://github.com/grimme-lab/crest>.
- (86) *Documentation for xtb and Related Software*, 2021, <https://xtb-docs.readthedocs.io/>.
- (87) *Semiempirical Extended Tight-Binding Program Package xtb*, version 6.4.0., 2020, <https://github.com/grimme-lab/xtb>.
- (88) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

- (89) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822.
- (90) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (91) COSMOtherm, C3.0, Release 1601; COSMOlogic GmbH & Co KG, 2020, <http://www.cosmologic.de>.
- (92) Pracht, P.; Grant, D. F.; Grimme, S. Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra. *J. Chem. Theory Comput.* **2020**, *16*, 7044–7060.
- (93) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A “Swiss army knife” composite electronic-structure method. *J. Chem. Phys.* **2021**, *154*, 064103.
- (94) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91–100.
- (95) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (96) TURBOMOLE V7.5.1 2020, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007; TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- (97) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (98) Williams, T.; Kelley, C.; Gnuplot 5.0: An Interactive Plotting Program. <http://gnuplot.sourceforge.net/>, 2018.
- (99) Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., et al. Amber 2021; University of California: San Francisco, 2021.
- (100) 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.
- (101) 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.
- (102) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (103) 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: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (104) 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.
- (105) 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.
- (106) 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 β -propiolactone and formamide. *Theor. Chem. Acc.* **2018**, *137*, 144.
- (107) Patel, P. S.; Huang, S.; Fisher, S.; Pirnik, D.; Aklonis, C.; Dean, L.; Meyers, E.; Fernandes, P.; Mayerl, F. Bacillaene, a Novel Inhibitor of Prokaryotic Protein Synthesis Produced by *Bacillus subtilis*. *J. Antibiot.* **1995**, *48*, 997–1003.
- (108) Katsyuba, S. A.; Spicher, S.; Gerasimova, T. P.; Grimme, S. Revisiting conformations of methyl lactate in water and methanol. *J. Chem. Phys.* **2021**, *155*, 024507.
- (109) Katsyuba, S. A.; Spicher, S.; Gerasimova, T. P.; Grimme, S. Fast and Accurate Quantum Chemical Modeling of Infrared Spectra of Condensed-Phase Systems. *J. Phys. Chem. B* **2020**, *124*, 6664–6670.
- (110) Katsyuba, S. A.; Gerasimova, T. P.; Spicher, S.; Bohle, F.; Grimme, S. Computer-aided simulation of infrared spectra of ethanol conformations in gas, liquid and in CCl₄ solution. *J. Comput. Chem.* **2022**, *43*, 279–288.
- (111) Bünnemann, K.; Merten, C. Solvation of a chiral carboxylic acid: effects of hydrogen bonding on the IR and VCD spectra of α -methoxyphenylacetic acid. *Phys. Chem. Chem. Phys.* **2017**, *19*, 18948–18956.
- (112) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1338.