Alkane/Water Partition Coefficient Calculation Based on the Modified AM1 Method and Internal Hydrogen Bonding Sampling Using COSMO-RS

Panagiotis C. Petris,* Paul Becherer, and Johannes G. E. M. Fraaije

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We introduce a physics-based model for calculating partition coefficients of solutes between water and alkanes, using a combination of a semi-empirical method for COSMO charge density calculation and statistical sampling of internal hydrogen bonds (IHBs). We validate the model on the experimental partition data (∼3500 molecules) of small organics, drug-like molecules, and statistical assessment of modeling of proteins and ligand drugs. The model combines two novel algorithms: a bond-correction method for improving the calculation of COSMO charge density from AM1 calculations and a sampling method to deal with IHBs. From a comparison of simulated and experimental partition coefficients, we find a root-mean-square deviation of roughly one log 10 unit. From IHB analysis, we know that IHBs can be present in two states: open (in water) and closed (in apolar solvent). The difference can lead to a shift of as much as two log 10 units per IHB; not taking this effect into account can lead to substantial errors. The method takes a few minutes of calculation time on a single core, per molecule. Although this is still much slower than quantitative structure−activity relationship, it is much faster than molecular simulations and can be readily incorporated into any screening method.
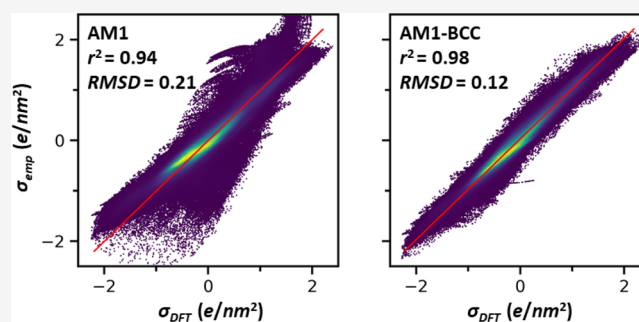
1. INTRODUCTION

In previous publications, we have discussed automated coarse-graining protocols to calculate partition coefficients[1] and diffusion coefficients.[2] In this work, we focus on providing a fast and reliable method for calculating the alkane/water partition coefficient of lead-like molecules with a molecular weight of up to 800. As far as modeling is concerned, predicting large-molecule partitioning in a practical industrial setting is extraordinarily difficult. One must be able to accurately calculate the COSMO charge density distribution of molecules with a molecular weight of up to 800 and then the free energy of partition (here by COSMO-RS[3]) in a reasonable amount of time (a few core minutes). Finally, internal hydrogen bonding (IHB) redistribution should be considered, taking proper account of statistical thermodynamics. In simple terms, when the molecule is dissolved in water, it is energetically favorable to expose the IHB to the solvent, while in an organic solvent, the molecule forms IHB to inhibit its interaction with the surrounding apolar environment.

In a recent statistical assessment of modeling of proteins and ligand (SAMPL5) competition to predict cyclohexane−water partition coefficients of lead-like molecules,[4−6] it was found that COSMO-RS beats the competition, including force-field methods, in both calculation time and accuracy. From a medicinal chemist point of view, one has only a few choices.

The increasing size of drugs/drug candidates and the need for fast estimation in medicinal chemistry screening require significant advancements in *ab initio*[7] and force-field-based[8,9] methods to compete against COSMO-RS,[4] UNIFAC,[10] and quantitative structure−activity relationship (QSAR) approaches.[11] QSAR approaches are the ones requiring the least computational resources, although they are not always precise, and most importantly, they do not provide physics-based insights to improve on them. A statistical thermodynamic model like COSMO-RS or UNIFAC can provide accurate results in relatively short time but cannot really be used in any system without careful consideration. More specifically, charged molecules cannot really be handled using COSMO-RS in its current implementation. Certain functional groups are not described adequately enough, regardless of the method used to extract COSMO charge densities. Theoretically, these issues are resolved by using *ab initio* and force-field-based models. Such approaches will always be the most accurate ones but very
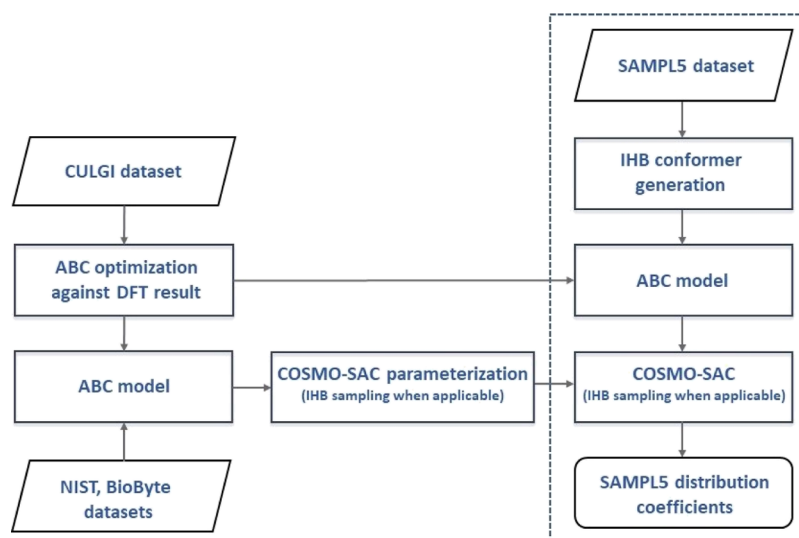
Figure 1. Protocol's flowchart is used to optimize the ABC model, generate IHB conformers, parameterize the COSMO-SAC coefficients, and predict the SAMPL5 distribution coefficients (denoted by a dashed rectangle).

expensive computationally. Depending on the system, the properties, and the application of interest, one should decide which method to use cautiously, as this is the most important choice in building a successful workflow.

Despite their shortcomings, COSMO-RS and UNIFAC that provide accurate estimations considerably faster than force-field-based methods are increasingly being used in the literature. In other domains, such as octanol−water partition coefficients, one could perhaps rely on massive amounts of data to build an ultra-fast machine learning correlation or a QSAR model. However, for the systems of current interest, drug-like molecules partitioning in water and alkane solvents, there are simply not enough data to calibrate correlations. Hence, a statistical approach cannot be adopted, and models must necessarily be physics-based.

A suitable candidate can be found in the COSMO-RS model because it requires only a handful of parameters, all with realistic physical meaning, while it is, at the same time, precise. Unfortunately, in standard COSMO-RS, one needs to calculate the charge envelope by performing a costly density functional theory (DFT) calculation, which could take many days for the molecular targets of our interest. Approximate methods exist that combine fragments extracted from a library (COSMO-frag[12,13]), but the library and extraction method are not publicly available nor described in sufficient detail to rebuild these methods.

The second significant problem is that many lead-like molecules exhibit IHBs as these amplify drug-design specificity. From analyzing small molecules such as salicylic acid and *o*-nitrophenol, it is clear that IHBs can have a dramatic effect on the partition coefficient. The difference between open (in water) and closed (in hydrocarbon solvent) state can account for roughly one to two log 10 unit per IHB.[14−16] When extended to large molecules (i.e., MW 800) that can easily have 3−5, if not more, IHBs, the effect is dramatic, and getting this right will make or break any prediction tool.

The present work is structured specifically to resolve the two issues mentioned above (fast estimation of COSMO charge density and the IHB effect). In the Systems and Methods section, we describe the methodology that we have developed to predict COSMO charge density in a fast and accurate way, using the AM1 semi-empirical method[17−20] alongside a bond charge correction.[21,22] To cope with the IHB effect, a separate algorithm was developed to generate the conformations that include all the possible open (no IHB formed) and closed (IHB formed) states for a given molecule. We also discuss the way to extract the partition coefficient and the data sets used. In the Results section, we present the COSMO charge density as calculated from our newly developed method, AM1-bond charge correction-COSMO (ABC), and we compare it with the corresponding DFT outcome. We optimize COSMO-SAC coefficients against the thermodynamic and partitioning data. We apply the ABC model with the IHB algorithm and the optimized COSMO-SAC coefficients to the BioByte and SAMPL5 data set. In the Discussion section, we report a few improvements that could make the model more accurate. The Supporting Information contains the BCC parameters for ABC, the molecules extracted from the BioByte and SAMPL5 data set, and the Kuhn patterns used to identify IHB. All algorithms, including AM1, are implemented in CULGI software version 13.0,[23] installed on a Dell desktop PC equipped with an Intel Xeon E5-2670 dual processor.

2. SYSTEMS AND METHODS

This section is separated into four different subsections explaining the methods we have developed and the systems studied. We present the "AM1-bond-charge-correction-COSMO" (AM1-BCC-COSMO, ABC for short) algorithm we have developed for calculating the COSMO charge density of a molecule. We include a detailed explanation of the IHB algorithm we developed to generate all possible IHB conformers. Subsequently, we illustrate the methodology for parameterizing COSMO-SAC coefficients and the features and characteristics of the data sets we applied the protocol to (CULGI,[1] NIST standard reference database 103b as implemented in the NIST ThermoData Engine (TDE) version 7.0, BioByte,[24] and SAMPL5[4,25,26]). A schematic representation of the implemented protocol can be found in the form of a flowchart in Figure 1.

2.1. ABC Algorithm. We introduce a COSMO charge density calculation method based on Dewar's AM1,[17−20] a widely used and fast semi-empirical quantum method. Our

implementation of COSMO in AM1 is largely based on the method described by Klamt and Schüürmann,[24] specifically solving the linear system from eq 9 in that paper directly. However, for calculating the interaction of the electron density (orbitals) with the COSMO surface charges, we follow the implementation by Klamt in MOPAC 7.[27−29] There, the electron density is approximated by a collection of atom-centered multipoles (monopoles, dipoles, linear quadrupoles, and square quadrupoles) in the same way as this is used for calculating the two-center repulsion integrals in MNDO,[30,31] using the same parameters as MNDO. An additional approximation is then made by treating the multipoles as point dipoles instead of physical dipoles (taking the limit to zero distance between charges while keeping the multipole moments constant). This leads to relatively simple analytical expressions for the electrostatic potential arising from the electron density. The explicit expressions can be found in the Supporting Information. Additionally, our implementation of COSMO in AM1 generates smoother COSMO surfaces than MOPAC. The construction of the solvent-excluded surface is based substantially on the work of Sanner et al.[32]

AM1 by itself does not give accurate electrostatic moments and as a result gives unsatisfying COSMO charge density,[33] which we propose to repair by a QSAR type of approach. Our new ABC algorithm is a bond charge correction that works similar to the BCC method for force fields.[21,22] Compared to the force-field BCC algorithm, we do not optimize for atomic partial charges but for COSMO surface charge density. The BCC parameters are calculated by comparing DFT to AM1 results; no experimental input is needed, but instead, we can profit from matching vast amounts of purely computer-generated data. The algorithm itself can be described as follows:

- Perform an AM1 calculation with COSMO as usual, giving the COSMO surface charge density σ_M^{AM1} of the molecule.
- Apply the AM1-BCC method (with modified parameters to be determined) to obtain atom-centered correction charges q_M^{BCC}
- Calculate the COSMO surface charge density that results from these correction charges alone, $\Delta\sigma_M^{BCC}$ by solving separately the COSMO equations using only the correction charges.

Because the COSMO equations are linear, this COSMO correction charge density is added to the charge density from the original AM1-COSMO calculation to obtain an improved COSMO surface charge density $\sigma_M^{AM1-BCC} = \sigma_M^{AM1} + \Delta\sigma_M^{BCC}$. In this way, the improvement of the COSMO calculation is a straightforward post-processing step that can be performed after a regular AM1-COSMO calculation. The improved COSMO sigma profile can then be used in COSMO-RS-type calculations, after re-optimizing the COSMO-RS parameters for this source of sigma profiles.

The modified AM1-BCC parameters for this procedure are found by fitting the parameters to minimize the difference between the AM1-BCC-COSMO surface charge density $\sigma_M^{AM1-BCC} = \sigma_M^{AM1} + \Delta\sigma_M^{BCC}$ and the σ_M^{DFT} resulting from a full DFT COSMO calculation. This is in contrast with the force-field BCC method, where the parameters are fitted to reproduce atomic partial charges. For the optimization, we must construct an objective function to be minimized, which expresses the deviation of the AM1-BCC-COSMO results from the full DFT calculations. The surface charge density is calculated on a

discretized (triangulated) surface, leading to a representation as a collection of local surface charge densities $\sigma_{M,I,}$ one for each surface segment i in the molecule M. In the objective function, we multiply each $\sigma_{M,i}$ with the surface area $a_{M,i}$ of the segment to properly weigh the size of the surface segments. We, then, minimize numerically the following objective function

$$F = \sum_{M,i} [a_{M,i}(\sigma_{M,i}^{AM1\text{-}BCC} - \sigma_{M,i}^{DFT})]^2 \tag{1}$$

Minimizing this function (the total square deviation) is equivalent to minimizing the root-mean-square deviation. In order to calculate the difference between the surface charge density distributions segment by segment, we perform a COSMO calculation on the DFT BP86/def2-TZVP[34−39]-optimized geometry for both the full DFT calculation and the AM1-BCC-COSMO calculation, using the same COSMO surface tessellation for both cases.

The set of molecules that we used for this parameterization is a subset of an in-house database of over 11,000 molecules ("the Culgi molecular database") that all have pre-computed COSMO information, calculated with DFT as described above. To obtain a manageable subset of this database while making sure that we have enough occurrences of each BCC bond type, we followed this procedure:

- For each bond type in the list of BCC bond types, we determined the molecules that contain this type.
- If there were fewer than 10 molecules containing a bond type, we did not include that bond type in our list. This means, for example, that we cannot apply our AM1-BCC-COSMO on iodine-containing molecules because we only have a handful of these in our entire database, not enough for a meaningful parameterization of the iodine-containing bond types. The more than 300 bond types that were present in the original AM1-BCC scheme were thus reduced to 149 types. Of these 149 types, there are 17 that are bonds between atoms of identical type, where the correction charge is necessarily 0. That leaves 132 non-trivial bond types. These parameters suffice for the vast majority of molecules in the Culgi Molecules Database.
- For each bond type that has 10 or more molecules containing it, 10 molecules with the lowest molecular weight were selected.
- The overlap between these sets of 10 molecules left us with 965 molecules to parameterize 132 non-trivial bond charge-correction parameters.

The objective function was minimized using an L-BFGS-B minimizer as implemented in the SciPy library for the Python programming language.

The resulting set of bond charge corrections can be found in the Supporting Information. A more quantitative analysis of the results of applying this to the entire Culgi Molecules Database is presented in the Results section. However, an indicative demonstration of how well the new model works is provided in Figure 2. We show the COSMO charge density distribution of propanol calculated from AM1, ABC, and DFT, respectively. It is evident that ABC (red line) is far closer to DFT results (blue line) than plain AM1 (black line), even for a molecule without special and exotic functional groups, such as propanol.

DFT calculations were carried out with a modified version of NWChem 6.3,[40] with an improved COSMO mesh generation based substantially on the work of Sanner et al.,[32] and with outlying charge correction.[41] Molecules were geometry-
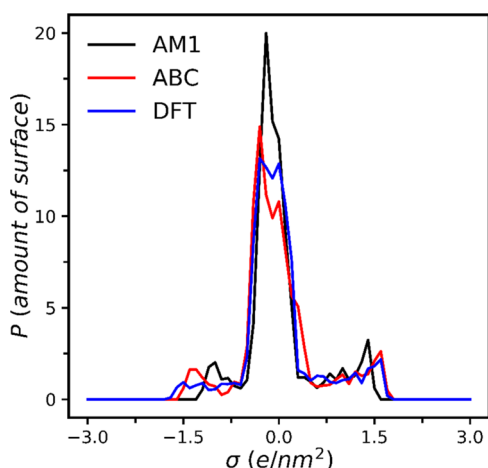
Figure 2. COSMO charge density distribution of propanol calculated using (a) AM1 semiempirical method, (b) ABC model, and (c) DFT with the modified NWChem software package.[40]

optimized in vacuum, using the def2-TZVP basis set[34,37] and BP86 exchange−correlation functional.[36−38] A COSMO calculation was then performed on that geometry using the def2-TZVPD basis set.[34,35,39] These are the recommended high-quality settings for COSMO-RS as used in COSMOtherm software[42] and also used in the entry of Klamt et al. in the SAMPL5 challenge.

2.2. COSMO-RS Partition Coefficients. The ABC algorithm is completed by reparameterization of some of the COSMO-SAC coefficients by fitting four sets of thermodynamic and partitioning data. Our implementation is based on COSMO-SAC,[43,44] but because COSMO-SAC and COSMO-RS differ only cosmetically, we will use the name COSMO-RS throughout the paper. With the ABC method, we can calculate the COSMO charge density of even a large molecule (MW 800, the upper range of SAMPL5 molecules) in just a few minutes on a single core. Afterward, one can easily derive the partition coefficient in less than a second.

The partition coefficient is defined as the ratio of concentrations in the two solvents as follows

$$K \equiv \frac{c_{\text{oil}}}{c_{\text{water}}} = \frac{V_{\text{water}}}{V_{\text{oil}}} e^{(-\Delta \mu^0/RT)}$$

(2)

where V_{water} and V_{oil} are the molar volume of water and oil, respectively, while the difference in chemical potentials $\Delta \mu^0$ is calculated from COSMO-RS through the activity coefficients γ, via

$$\Delta \mu^0 \equiv \ln \frac{\gamma_{\text{oil}}}{\gamma_{\text{water}}} = \sum_{\text{surfacepanel}} \Delta f(a_i, \sigma_i)$$

(3)

where a_i is the surface area and σ_i is the COSMO surface charge density on segment i. When two or more conformations of the same molecule are included in the partition coefficient calculation, the corresponding relation is transformed. For two conformations A and B, in oil/water solvents, the cumulative partition coefficient is provided by the following equation

$$K = \frac{e^{-E_{\text{oil}}(A)} + e^{-E_{\text{oil}}(B)}}{e^{-E_{\text{water}}(A)} + e^{-E_{\text{water}}(B)}}$$

(4)

where $E_{\text{oil}}(A)$ and $E_{\text{water}}(A)$ are the total energy required for the conformation A to move from the perfect conductor state into a

solvent like oil or water, respectively (same for conformation B). Such a quantity for a conformer i in a solvent S can be calculated as

$$E_S(i) = E_{\text{COSMO}}(i) + \mu_S(i)$$

(5)

where $E_{\text{COSMO}}(i)$ is the total quantum mechanical energy of the conformer i in the perfect conductor and $\mu_S(i)$ is the chemical potential of the conformer i in solvent S that accounts for the transition of this conformer from a perfect conductor into the solvent S.

COSMO-RS theory amounts to the integration over the surface, of a local free energy difference for a charged segment exposed to two different solvents $\Delta f = f_{\text{oil}} - f_{\text{water}}$. The local free energy function differs between various COSMO-RS flavors; in most cases, as we do here, one accommodates the polarization energy of a charged segment in a conductor and a hydrogen bonding term. In the limited sense of integration over the surface, the model is not too dissimilar from QSAR, where one integrates over different types of polar groups and hydrogen donors and acceptors. The essential difference is that in COSMO-RS, one does not employ a group description and group parameters but a generic interaction model based on electrostatics from quantum chemistry calculations, and therefore, one needs only a few physics-based parameters, instead of a whole zoo of groups and group parameters.

In this paper, we use the "SplitOH" COSMO-SAC model from Lin's group[45,46] (a recent approach by Chang et al.[47] suggests directional interactions for a more accurate representation of hydrogen bonding). The Staverman−Guggenheim combinatorial term is used. Instead of taking the published Lin parameters as is, we re-optimized the COSMO-SAC coefficients. This was carried out by minimizing the total root-mean-square error for four sets of thermodynamic and partitioning data:

- 400 pressure data points for binary vapor−liquid equilibria (VLEs)
- 400 excess Gibbs free energies for 50/50 liquid binary mixtures (derived from VLE data)
- 400 partition coefficients for octanol/water
- 212 partition coefficients for hexadecane/water

VLE data were obtained from the NIST standard reference database 103b as implemented in the NIST TDE version 7.0. Partition coefficients were obtained from the BioByte data set.[24] More information related to the parameterization scheme can be found in the Supporting Information. The optimized settings can be seen in the following table where c_{base} defines the interaction between any related segment, $c_{\text{OH−OH}}$ provides the (additional) interaction between segments with hydroxyl groups, $c_{\text{OH−OT}}$ gives the (additional) interaction between a segment containing a hydroxyl group and a segment with a different hydrogen bonding acceptor, and $c_{\text{OT−OT}}$ represents the interaction between two segments that do not contain any hydroxyl group. The segment surface area is the size of a thermodynamically independent segment. This is equivalent to the inverse of the so-called "coordination number" in chemical engineering models.

One can use the suggested coefficients in Table 1 alongside a software package supporting COSMO-SAC calculations. An open source implementation of COSMO-SAC has recently become available.[48]

2.3. IHB Sampling. The special treatment of IHB in a molecule is related to the different conformations that the

Table 1. Optimized COSMO-SAC Coefficients

c_{base} (kcal/mol)· (Å/e²)	c_{OH-OH} (kcal/mol)· (Å/e²)	c_{OH-OT} (kcal/mol)· (Å/e²)	c_{OT-OT} (kcal/mol)· (Å/e²)	segment surface area (Å²)
6032.98	3175.29	2209.34	940.67	7.58

molecule can form when dissolved in a polar or apolar solvent. More specifically, if a molecule that exhibits an IHB is dissolved in a polar solvent, the acceptor (hydroxyl, amine, etc.) and the donor would be exposed to the solvent, the so-called "open" conformation. If dissolved in an apolar solvent (i.e., alkane), the molecule tends to form IHB so that the acceptor and donor are hidden from the solvent; this is the "closed" conformation. Therefore, we have conducted conformation analysis for molecules exhibiting IHB, so that all possible combinations of open and closed IHB structures will be included in the final analysis. Such analysis ensures that the effect of the solvent on the molecular conformation is explored and accounted for.

Following the work of Kuhn et al.[49] we have used the Kuhn rings to identify the presence of IHB in a molecule. The patterns do not include the trans-core IHBs that can be found in macrocycles, and we therefore omitted all such molecules from the present study. We have included an additional pattern for the IHBs between carboxylic groups and aromatic nitrogen. Finally, a few additional rings are included to make sure that we capture the IHBs present in smaller molecules like salicylic acid and *o*-nitrophenol.

The second challenge in IHB sampling is to find an automatic way of reproducing the conformations that would include all the combinations of open and closed IHBs. We start by assigning zero charges to all atoms, apart from the donor and acceptor atoms that take part in the formation of the IHB. To the donor/acceptor couple, we assign opposite charges if the closed conformation should be formed (attractive electrostatic interaction) or equal charges if the open conformation should be formed (repulsive electrostatic interaction). Using these charges, we run successive molecular dynamics runs followed by energy minimization, until a hydrogen bond distance criterion is met (distance below a certain value if an IHB is to be formed and distance above a certain value if the atoms should not form an IHB). Each molecular dynamics run consists of 1000 steps. We have selected an *NVT* ensemble with the Andersen thermostat, at a temperature of 298 K, with a 0.5 fs timestep coupled with the Dreiding force field.

As the second step, we re-assign the correct charges and geometrically optimize the conformation, so that we make sure that the structure is energetically favorable. This way, after the optimization step, some of the conformations might end up without the desirable IHB, although the vast majority of them will maintain the desired structure. Then, we apply AM1 geometry optimization to each conformation to ensure that we have a physically meaningful representation of the molecule. The COSMO charge density is then calculated using ABC for every conformation. Finally, we calculate the partitioning of the molecule between water and alkane using eq 4.

A characteristic example can be found in salicylic acid. Figure 3 illustrates the two different conformations that should be taken into account when calculating the partition coefficient of this molecule in water/alkane solvents. When the molecule is dissolved in water, the "open" conformation, depicted in Figure 3a, is practically the only one observed. In contrast, when dissolved in alkane solution, the "closed" conformation, represented in Figure 3b where an IHB is formed, is
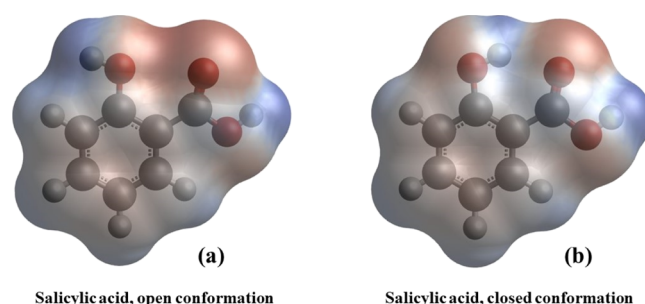

Figure 3. Salicylic acid's (a) open conformation, favorable in an aqueous environment, and (b) closed conformation, favorable in alkane solution. The COSMO surface is depicted in the form of a color map.

predominant. The COSMO surface is visualized in the form of a color map. The red surface areas correspond to a positive COSMO charge density, while the blue areas carry a negative COSMO charge density. The gray COSMO surface indicates neutral COSMO charge density.

The severe effect of leaving out the closed conformation is evident when estimating the salicylic acid partition coefficient in water/cyclohexane. Using eq 2, the open conformation partition coefficient is equal to −5.43, while the experimental one is −1.97. If we use eq 4 to include both conformations, we get to the value of −2.54, which is in good agreement with the experiment.

2.4. Data sets. We used the BioByte database[24] as the primary source for partition coefficients between water and alkane solvents (hexane, cyclohexane, heptane, isooctane, octane, decane, undecane, dodecane, and hexadecane). We have curated the data set manually: for the solutes with data in more than one solvent, we only accepted those entries where log 10 deviated less than 1.5 unit between the lowest and highest. We accepted all entries for which only one solvent is present, for lack of a good curation criterion. We have also omitted

- all tautomer-forming molecules (as they need to be separately investigated for different solvents, while we have not developed a special model for their treatment),
- all molecules with more than five rotatable bonds (because we focus mostly on the residual contribution, so we limit surface flexibility and folding effects), and
- all crown ethers, where water binding should be taken into account due to the extreme hydrogen bonding behavior.

The OpenBabel package is utilized to generate the molecular three-dimensional conformations from their corresponding SMILES strings. Subsequently, we stretch the molecule open to prevent energetically unfavorable starting conformations; this also avoids unphysical steric and intramolecular effects. To do so, we apply molecular dynamics without charges to open the molecule, followed by energy minimization with the corresponding force-field charges to get a molecule with realistic geometry. When the same molecule is encountered again for different solvents, we keep the generated conformation from the earlier occurrences to ensure that statistical mechanics from the molecular dynamics steps will not sample different conformations for different alkane solutes, hence providing different results.

As discussed in the previous section, a further distinction is made between the molecules with IHB (a few hundred) and those without IHB (a few thousand). The molecules are listed in the Supporting Information. A second set is the SAMPL5 data set of drugs and drug-like candidates.[4,25,26] In Figure 4, we
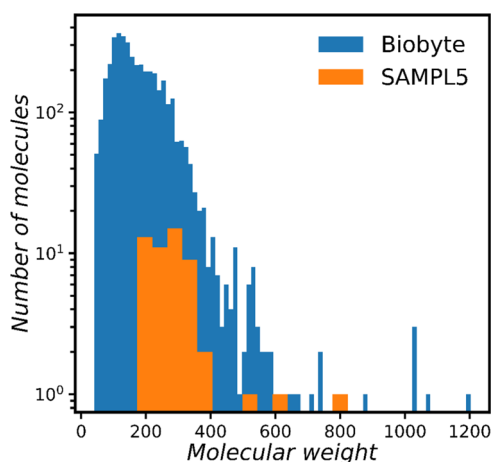
Figure 4. Molecular weight distribution for BioByte and SAMPL5 data sets.

present the molecular weight distribution for the BioByte and SAMPL5 data set. The average of the molecular weight distribution for the BioByte data set is equal to 180 and for SAMPL5 is equal to 290. This difference is expected because BioByte contains many smaller solutes, while SAMPL5 includes larger lead-like molecules.

3. RESULTS

In this section, we demonstrate how we acquired the correction charges for the ABC model for some of the most representative functional groups. Subsequently, we have applied the protocol (ABC alongside IHB sampling) to predict the partition and distribution coefficients in the BioByte and SAMPL5 data sets, respectively.

3.1. ABC Optimization. As explained in the Systems and Methods section, we optimized the ABC model against DFT results for a wide range of functional groups found in the CULGI database (extensively analyzed in previous publications[1]), such that the surface charge density disparity between ABC and DFT is minimized. The quadratic error in the surface charge density between AM1 and DFT and between ABC and DFT (for a few of the most frequently appearing functional groups) is represented using boxplots in Figure 5.

By comparing either the median value or the relative position of a boxplot between AM1 and DFT (Figure 5a) and between ABC and DFT (Figure 5b), one can see that for all functional groups, apart from sulfides, there is a moderate to substantial improvement using ABC instead of AM1. We can see the molecules with hypervalent sulfur and phosphorus in the upper regions of Figure 5a: "sulfonamide" and "phosphate"—the latter including thiophosphates. In Figure 5b, these groups improve significantly, with errors ending up in the same range as more "regular" molecules. It is known that AM1 does not deal well with these atom types, and this is also clear from looking at the COSMO surface of these molecules when comparing the DFT COSMO surface charge density with the AM1 COSMO surface charge density, as we show graphically in the Supporting Information for the pyridine-3-sulfonamide molecule. The nitriles show a moderately large improvement. In general, for large molecules, the dominant source of error is often formed by large alkyl tails.

The values of the optimized parameters also bear this out. For the bonds involving hypervalent S and P, the charge shift is often as large as 0.4 to even 0.9 electron charges. For nitriles and similar groups, this may still be up to 0.1 to 0.2, but in the vast majority of other bonds, the charge displacement is below that value. To quantify our metric in familiar COSMO terms, if there is a difference in σ_{error} of one unit between the estimation of AM1 and ABC, this corresponds to approximately 0.85 $k_B T$ difference in the total surface segment energy. Because the introduced bond charge correction significantly improves AM1 results, we are justified to apply it in calculating the COSMO charge density and, subsequently, the partition coefficient.

3.2. Protocol Application to the BioByte Data Set. We have applied the ABC model to the BioByte data set to predict the alkane/water partition coefficient. For molecules that exhibit IHBs, we applied the IHB sampling algorithm as described in the Systems and Methods section.

For molecules containing carboxyl groups, we use only the syn conformation (with the hydrogen pointing in roughly the same direction as the carbonyl oxygen) and not the anti-conformation (with the hydrogen pointing away) because the former is the most stable and energetically favorable one as far too polar behavior is observed when we sample energetically unfavorable carboxyl anti-conformations (donor and acceptor exposed to the
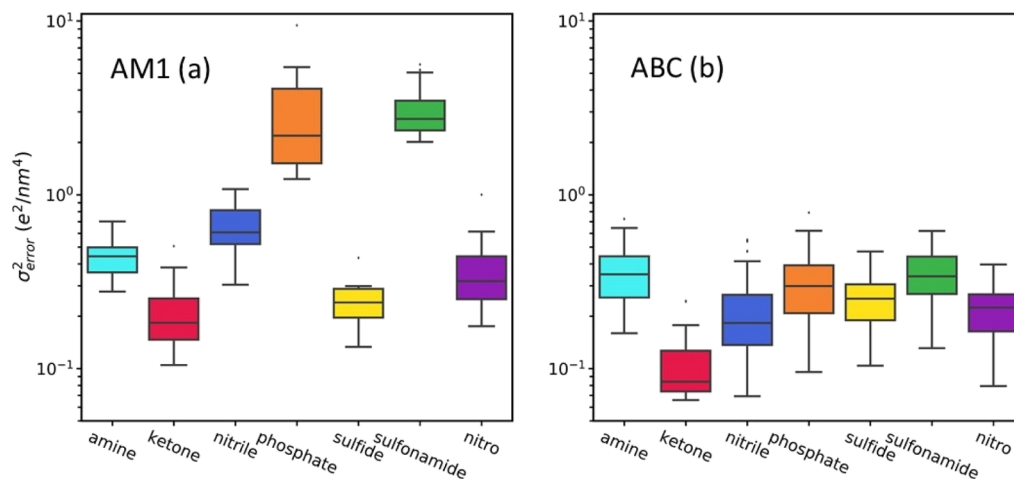

Figure 5. Quadratic surface charge density error between DFT and (a) AM1 and (b) ABC for a selection of different functional groups represented in a boxplot fashion.
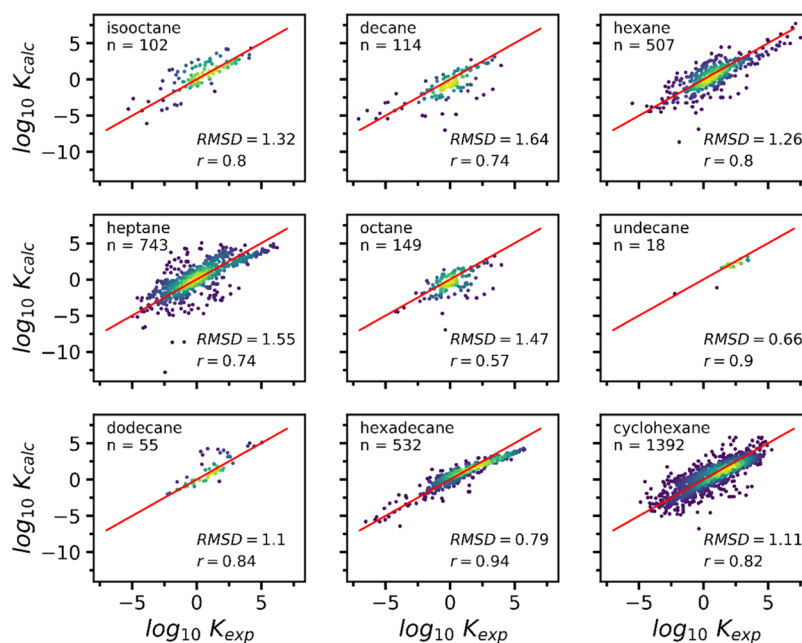
Figure 6. Computed partition coefficient, from ABC and IHB sampling, plotted against the experimental partition coefficient of the BioByte data set for different solvents. The red line corresponds to $y = x$. A color density map is used for regions with overlapping data points.

solvent). The necessity of special treatment for carboxyl groups has already been demonstrated in force-field-based works in the literature.[50,51]

Figure 6 compares the experimental K_{exp} and calculated K_{calc} partition coefficients for each hydrocarbon solvent (namely, isooctane, decane, hexane, heptane, octane, undecane, dodecane, hexadecane, and cyclohexane). A color density map is used for regions with overlapping data points. Prior to optimization, using the SplitOH coefficients as suggested by Lin's group,[45,46] we obtained RMSD = 1.92 with a correlation coefficient of 0.80. Using the optimized COSMO-SAC coefficients (shown in Table 1), we calculated RMSD = 1.24 with a correlation coefficient of 0.81. This is a significant improvement, considering how diverse the data set is.

Looking in more detail, taking each solvent separately into account, we notice that the model appears better for some solvents and not that adequate for others. For instance, heptane exhibits RMSD = 1.55, while hexadecane shows RMSD = 0.79. Heptane has more data points, but it is also the noisiest of the two. This noise does not appear to be a feature of the solvent; we have verified this for the same molecule, partitioning is almost identical for both solvents. This is to be expected because for all alkanes, the COSMO profile is essentially a delta-function, that is, all these hydrocarbon molecules have an almost zero COSMO charge density everywhere on the entire molecular surface.

However, we have identified several molecules of rather exotic structures and functional groups being experimentally studied only in the case of heptane (like morin and xanthine). ABC is less accurate in predicting the partitioning of such molecules, which are more common in the heptane data set, making this particular solvent appear to be much more problematic and prone to error. This is an indication of how important it would be to improve upon functional groups that are not described with high accuracy in the current model.

The scatter plots of hexadecane and heptane show a downward curvature for larger values of the partition coefficient, which could be attributed to a combination of conformational changes and insufficient alkyl–water parameterization. In the current ABC implementation, all conformations are obtained by AM1 minimization; no sampling is applied (apart from cases where IHBs occur and we apply a special sampling as explained in the IHB sampling section). We have checked that almost all compounds with large partition coefficients are hydrocarbons. Overall, the fit is entirely satisfactory. The attained accuracy is typical for what one can expect for oil–water partition coefficients from COSMO-RS. Still, here we have achieved this result by applying semi-empirical calculations (a few minutes per molecule) rather than full and much more expensive quantum DFT calculations (hours per molecule).[4]

3.3. Protocol Application to the SAMPL5 Data Set. We can, also, apply the developed protocol in predicting the distribution coefficient for water/cyclohexane of the SAMPL5 data set containing drug-like molecules. The distribution coefficient is given by the following relation

$$\log_{10}D = \log_{10}K + \log_{10}\alpha \tag{6}$$

where K is the partition coefficient of the neutral molecule and α is the neutral fraction of molecules at pH = 7.4; pK values are estimated with the program cxcalc from ChemAxon.

As we observe in Figure 7 our model gives RMSD = 1.88, slightly outperforming Klamt's result of RMSD = 2.07, while our model correlation coefficient is equal to 0.81, comparable to Klamt's 0.85. The overall comparison indicates that we can obtain similar results, if not improved, using ABC to calculate COSMO charge density and IHB sampling in the case of IHB. At this point, we need to stress that on comparing the RMSD with respect to the experimental data and uncertainty in measurements, the difference between our model and Klamt's COSMO-RS is insignificant. Therefore, in the SAMPL5 challenge, our protocol would have performed similar to the winner.

What deserves further investigation is that both ABC and Klamt's model deviate from the experimental values in two ways: for polar molecules, partition coefficients are too low (log too
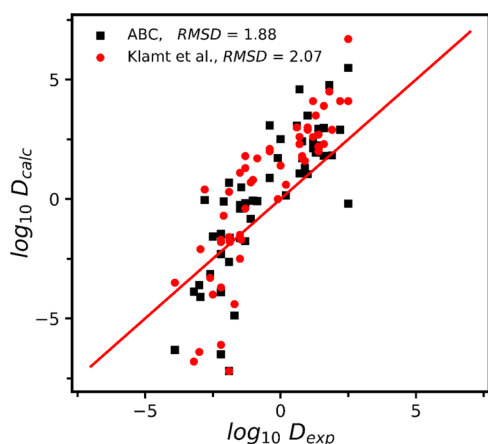
Figure 7. Distribution coefficient calculated from ABC and IHB sampling, alongside the results from Klamt et al.[4] against the experimental data of the SAMPL5 challenge. The red line corresponds to $y = x$.

negative), and for less polar molecules, the partition coefficients are systematically too high (log too positive). In fact, it seems as if the entire curve is shifted a few units in the positive direction. These two effects have been discussed in extenso by Klamt[4] and the SAMPL5 organizers,[5] without any clear conclusion. Other prediction methods of very different backgrounds (force fields) seem to suffer from this same phenomenon, so it could be concluded that this is partly an experimental effect. It could also be due to limited conformational sampling. Apart from molecules exhibiting IHBs, we did not use any conformational sampling. We simply took the conformation of the lowest AM1 energy in a vacuum. In contrast, force-field methods do sample conformations while Klamt also used conformational sampling. Considering that ABC performs similar to or better than all of them at a fraction of the computation time, the sampling issue remains a puzzle.

4. DISCUSSION

To our knowledge, there is no similar large-scale comparison between the calculated and experimental partition coefficients. As far as we know, in works using molecular dynamics, only a handful or at the most a few tens (SAMPL5) of molecules are investigated.[52,53] These data sets are too small to draw statistically valid conclusions, and anyway, a molecular dynamics simulation on the scale of the BioByte database is not possible in a reasonable amount of time. On the QSAR level, we also found only a few studies: a study on ~100 molecules using Abraham descriptors (a fine RMSD 0.2−0.3 but a rather poor correlation r^2 0.5−0.7)[54] and an older QSAR study on ~100 molecules by the group of Leo (the source of some of the experimental data we use here).[55] These studies have been limited to a few tens or a few hundred molecules at most, and apart from this, it is always challenging to extract physical insights from QSAR modeling.

For the ubiquitous log P (octanol/water) predictions (COSMO-RS was recently used to predict octanol/water partitioning in small drug molecules[56]), the situation is very different and large-scale comparisons abound (see, e.g., the review[57] that compares methods based on a data set of more than 96,000 molecules). For hydrocarbon predictions, the situation is much less satisfactory. This is worrisome because on theoretical and experimental grounds,[58−63] one can expect that hydrocarbon partitioning is a much better predictor for bilayer

permeability than log P: in the bilayer core, no hydrogen bonding with the surrounding lipid is possible and one is better off with a proxy solvent that cannot form hydrogen bonds.

Several improvements could be introduced to enhance the accuracy of our model. A step in this direction could be to correct for specific functional groups whose behavior is not well captured using ABC. As illustrated in the Results, amines and sulfides are not significantly improved by the method. However, even for more frequently appearing and well-described groups like nitros, a serious improvement could be obtained. A representative molecule containing a nitro group is *o*-nitro-phenol, which also happens to exhibit an IHB. After our IHB algorithm, we calculate the partition coefficient from the three different conformations, and finally, we estimate the total partitioning. As we can see in Table 2, the *ab initio* calculation is

Table 2. Partition Coefficient of *O*-Nitrophenol Calculated Using ABC and DFT, Compared to the Experimental Data

	ABC + SplitOH (optimized)	DFT + SplitOH	experiment
Log K	−1.34	0.35	1.48

far closer to the experimental result than our semiempirical ABC method. This is a serious indication that a more detailed quantum description of the molecule could increase the accuracy of our model.

5. CONCLUSIONS

We have developed a fast and precise method to calculate COSMO charge density and, from that, the partition coefficient. At the core of the model, a bond charge correction is applied to the AM1 semi-empirical method, while special treatment for IHB is introduced.

The charge shifts for the bond charge correction were optimized by an iterative process so that the COSMO charge density difference between the new model, ABC, and DFT results was minimized. For molecules that exhibit IHBs, we built an algorithm that generates the majority of the possible combinations of open and closed IHBs to avoid severe underestimation in the partitioning calculation.

Subsequently, we re-parameterized some of the COSMO-SAC coefficients, namely, c_{base}, $c_{OH−OH}$, $c_{OH−OT}$, $c_{OT−OT}$, and the segment surface area, against the experimental data of several thermodynamic and partitioning quantities. Having obtained the optimized COSMO-SAC coefficients, we applied the whole methodology to predict partition coefficients in the BioByte data set and distribution coefficients of drug-like molecules participating in the SAMPL5 challenge. The results were entirely satisfactory and in the case of SAMPL5, comparable to (and perhaps slightly better than) the competition winner[4] at a fraction of the computational cost, as ABC bypasses the expensive DFT computations.

The fact that the model is so fast makes it a suitable candidate for large-scale screening studies of lead-like molecules as it provides fast and reliable results. The accuracy could be further improved, considering that ABC results are still inferior to DFT results for specific functional groups. Hence, in future studies, we will focus on finding quantum mechanical descriptors that can correct for the missing information related to the electronic structure of the interacting molecule.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c01478.

Description and detailed explanation of the tables and force-field files available for download, comparison of AM1, ABC, and DFT for the pyridine-3-sulfonamide molecule, elaborate explanation of orbital−COSMO interaction, and additional information related to COSMO-SAC parameterization (PDF)

Parameter file for assigning charge deltas for AM1-BCC-COSMO, molecules extracted from the BioByte and SAMPL5 data set, and Kuhn patterns used to identify IHB (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Panagiotis C. Petris − *Leiden Institute of Chemistry, Leiden University, 2300 RA Leiden, The Netherlands; Siemens Industry Software Netherlands B.V., 2333 BD Leiden, The Netherlands;* Ⓞ orcid.org/0000-0002-8257-4539; Email: panos.petris@siemens.com

Authors

Paul Becherer − *Siemens Industry Software Netherlands B.V., 2333 BD Leiden, The Netherlands*

Johannes G. E. M. Fraaije − *Leiden Institute of Chemistry, Leiden University, 2300 RA Leiden, The Netherlands; Siemens Industry Software Netherlands B.V., 2333 BD Leiden, The Netherlands*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.0c01478

Author Contributions

The article was written through equal contributions of all authors. All authors have given approval to the final version of the article.

Notes

The authors declare the following competing financial interest(s): The authors are employees of Siemens Industry Software Netherlands B.V.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Fraaije, J. G. E. M.; Van Male, J.; Becherer, P.; Serral Gracià, R. Coarse-Grained Models for Automated Fragmentation and Parametrization of Molecular Databases. *J. Chem. Inf. Model.* **2016**, *56*, 2361−2377.

(2) Fraaije, J. G. E. M.; van Male, J.; Becherer, P.; Serral Gracià, R. Calculation of Diffusion Coefficients through Coarse-Grained Simulations Using the Automated-Fragmentation-Parametrization Method and the Recovery of Wilke-Chang Statistical Correlation. *J. Chem. Theory Comput.* **2018**, *14*, 479−485.

(3) Klamt, A. *COSMO-R S: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier, 2005.

(4) Klamt, A.; Eckert, F.; Reinisch, J.; Wichmann, K. Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set. *J. Comput. Aided Mol. Des.* **2016**, *30*, 959−967.

(5) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **2016**, *30*, 927−944.

(6) Rustenburg, A. S.; Dancer, J.; Lin, B.; Feng, J. A.; Ortwine, D. F.; Mobley, D. L.; Chodera, J. D. Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **2016**, *30*, 945−958.

(7) Chuman, H.; Mori, A.; Tanaka, H.; Yamagami, C.; Fujita, T. Analyses of the partition coefficient, log P, using *ab initio* MO parameter and accessible surface area of solute molecules. *J. Pharm. Sci.* **2004**, *93*, 2681−2697.

(8) Bhatnagar, N.; Kamath, G.; Chelst, I.; Potoff, J. J. Direct calculation of 1-octanol-water partition coefficients from adaptive biasing force molecular dynamics simulations. *J. Chem. Phys.* **2012**, *137*, 014502.

(9) Ogata, K.; Hatakeyama, M.; Nakamura, S. Effect of atomic charges on octanol-water partition coefficient using alchemical free energy calculation. *Molecules* **2018**, *23*, 425.

(10) Kuramochi, H.; Noritomi, H.; Hoshino, D.; Kato, S.; Nagahama, K. Application of UNIFAC models to partition coefficients of biochemicals between water and n-octanol or n-butanol. *Fluid Phase Equilib.* **1998**, *144*, 87−95.

(11) Zissimos, A. M.; Abraham, M. H.; Barker, M. C.; Box, K. J.; Tam, K. Y. Calculation of Abraham descriptors from solvent-water partition coefficients in four different systems; evaluation of different methods of calculation. *J. Chem. Soc., Perkin Trans. 2* **2002**, 470−477.

(12) Hornig, M.; Klamt, A. COSMOfrag: A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. *J. Chem. Inf. Model.* **2005**, *45*, 1169−1177.

(13) Wittekindt, C.; Goss, K.-U. Screening the partition behavior of a large number of chemicals with a quantum-chemical software. *Chemosphere* **2009**, *76*, 460−464.

(14) Berthelot, M.; Laurence, C.; Lucon, M.; Rossignol, C.; Taft, R. W. Partition coefficients and intramolecular hydrogen bonding. 2. The influence of partition solvents on the intramolecular hydrogen-bond stability of salicyclic acid derivatives. *J. Phys. Org. Chem.* **1996**, *9*, 626−630.

(15) Klamt, A.; Reinisch, J.; Eckert, F.; Graton, J.; Le Questel, J.-Y. Interpretation of experimental hydrogen-bond enthalpies and entropies from COSMO polarisation charge densities. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7147−7154.

(16) Chen, D.; Zhao, M.; Tan, W.; Li, Y.; Li, X.; Li, Y.; Fan, X. Effects of intramolecular hydrogen bonds on lipophilicity. *Eur. J. Pharm. Sci.* **2019**, *130*, 100−106.

(17) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(18) Dewar, M. J. S.; Zoebisch, E. G. Extension of AM1 to the halogens. *J. Mol. Struct.: THEOCHEM* **1988**, *180*, 1−21.

(19) Dewar, M. J. S.; Healy, E. F.; Holder, A. J.; Yuan, Y.-C. Comments on a comparison of AM1 with the recently developed PM3 method. *J. Comput. Chem.* **1990**, *11*, 541−542.

(20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1—A New General-Purpose Quantum-Mechanical Molecular-Model. *J. Am. Chem. Soc.* **1993**, *115*, 5348.

(21) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(22) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(23) Culgi. *Chemistry Unified Language Interface (Culgi), 13.0*; Culgi BV: Leiden, 2019.

(24) *BioByte. Bioloom*, 1993−2019.

(25) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aided Mol. Des.* **2017**, *31*, 147−161.

(26) 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.

(27) 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.

(28) Stewart, J. J. P. MOPAC 7. 1993, http://openmopac.net/Downloads/Downloads.html.

(29) Stewart, J. J. P. MOPAC 7.1. 2006, http://openmopac.net/Downloads/Downloads.html.

(30) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.

(31) Dewar, M. J. S.; Thiel, W. A semiempirical model for the two-center repulsion integrals in the NDDO approximation. *Theor. Chim. Acta* **1977**, *46*, 89−104.

(32) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305−320.

(33) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(34) Peterson, K. A. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16−18 elements. *J. Chem. Phys.* **2003**, *119*, 11099.

(35) Florian, W.; Reinhart, A. 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*, 3279−3305.

(36) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(37) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(38) Rappoport, D.; Furche, F. Property-optimized gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(39) Perdew, J. P. Erratum: Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *34*, 7406.

(40) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477−1489.

(41) Klamt, A.; Jonas, V. Treatment of the outlying charge in continuum solvation models. *J. Chem. Phys.* **1996**, *105*, 9972−9981.

(42) Reinisch, J.; Diedenhofen, M.; Wilcken, R.; Udvarhelyi, A.; Glöß, A. Benchmarking Different QM Levels for Usage with COSMO-RS. *J. Chem. Inf. Model.* **2019**, *59*, 4806−4813.

(43) Lin, S.-T.; Chang, J.; Wang, S.; Goddard, W. A.; Sandler, S. I. Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model. *J. Phys. Chem. A* **2004**, *108*, 7429−7439.

(44) Mullins, E.; Oldland, R.; Liu, Y. A.; Wang, S.; Sandler, S. I.; Chen, C.-C.; Zwolak, M.; Seavey, K. C. Sigma-profile database for using COSMO-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2006**, *45*, 4389−4415.

(45) Hsieh, C.-M.; Sandler, S. I.; Lin, S.-T. Improvements of COSMO-SAC for vapor-liquid and liquid-liquid equilibrium predictions. *Fluid Phase Equilib.* **2010**, *297*, 90−97.

(46) Chen, W.-L.; Hsieh, C.-M.; Yang, L.; Hsu, C.-C.; Lin, S.-T. A Critical Evaluation on the Performance of COSMO-SAC Models for Vapor-Liquid and Liquid-Liquid Equilibrium Predictions Based on Different Quantum Chemical Calculations. *Ind. Eng. Chem. Res.* **2016**, *55*, 9312−9322.

(47) Chang, C.-K.; Chen, W.-L.; Wu, D. T.; Lin, S.-T. Improved Directional Hydrogen Bonding Interactions for the Prediction of Activity Coefficients with COSMO-SAC. *Ind. Eng. Chem. Res.* **2018**, *57*, 11229−11238.

(48) Bell, I. H.; Mickoleit, E.; Hsieh, C.-M.; Lin, S.-T.; Vrabec, J.; Breitkopf, C.; Jäger, A. A Benchmark Open-Source Implementation of COSMO-SAC. *J. Chem. Theor. Comput.* **2020**, *16*, 2635−2646.

(49) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601−2611.

(50) Klimovich, P. V.; Mobley, D. L. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J. Comput. Aided Mol. Des.* **2010**, *24*, 307−316.

(51) Paluch, A. S.; Mobley, D. L.; Maginn, E. J. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7*, 2910−2918.

(52) Bannan, C. C.; Calabró, G.; Kyu, D. Y.; Mobley, D. L. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.* **2016**, *12*, 4015−4024.

(53) Lara, A.; Riquelme, M.; Vöhringer-Martinez, E. Partition coefficients of methylated DNA bases obtained from free energy calculations with molecular electron density derived atomic charges. *J. Comput. Chem.* **2018**, *39*, 1728−1737.

(54) Abraham, M. H.; Acree, W. E., Jr.; Leo, A. J.; Hoekman, D.; Cavanaugh, J. E. Water-Solvent Partition Coefficients and Δ Log P Values as Predictors for Blood-Brain Distribution; Application of the Akaike Information Criterion. *J. Pharm. Sci.* **2010**, *99*, 2492−2501.

(55) Tayar, N. E.; Tsai, R.-S.; Testa, B.; Carrupt, P.-A.; Leo, A. Partitioning of solutes in different solvent systems: the contribution of hydrogen-bonding capacity and polarity. *J. Pharm. Sci.* **1991**, *80*, 590−598.

(56) Loschen, C.; Reinisch, J.; Klamt, A. COSMO-RS based predictions for the SAMPL6 logP challenge. *J. Comput. Aided Mol. Des.* **2020**, *34*, 385−392.

(57) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of LogP Methods on more than 96,000 Compounds. *J. Pharm. Sci.* **2009**, *98*, 861−893.

(58) Walter, A.; Gutknecht, J. Permeability of small nonelectrolytes through lipid bilayer membranes. *J. Membr. Biol.* **1986**, *90*, 207−217.

(59) Xiang, T.-X.; Anderson, B. D. Influence of chain ordering on the selectivity of dipalmitoylphosphatidylcholine bilayer membranes for permeant size and shape. *Biophys. J.* **1998**, *75*, 2658−2671.

(60) Liu, X.; Testa, B.; Fahr, A. Lipophilicity and Its Relationship with Passive Drug Permeation. *Pharm. Res.* **2011**, *28*, 962−977.

(61) Xiang, T.-X.; Anderson, B. D. Liposomal drug transport: A molecular perspective from molecular dynamics simulations in lipid bilayers. *Adv. Drug Delivery Rev.* **2006**, *58*, 1357−78.

(62) Xiang, T. X.; Anderson, B. D. Molecular distributions in interphases: statistical mechanical theory combined with molecular dynamics simulation of a model lipid bilayer. *Biophys. J.* **1994**, *66*, 561−572.

(63) Xiang, T. X.; Anderson, B. D. The relationship between permeant size and permeability in lipid bilayer membranes. *J. Membr. Biol.* **1994**, *140*, 111−22.