PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Predictions of the Physicochemical Properties of Amino Acid Side Chain Analogs Using Molecular Simulation

Alauddin Ahmed[ab] and Stanley I Sandler*[a]

[a]*Center for Molecular and Engineering Thermodynamics, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, USA*

[b]*Current Address: Mechanical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109, USA*

**Corresponding Author: sandler@udel.edu*

(Version 11 December 2015)

ABSTRACT

A candidate drug compound is released for clinical trails (*in vivo* activity) only if its physicochemical properties meet desirable bioavailability and partitioning criteria. Amino acid side chain anlogs play vital role in the functionalities of protein and peptides and as such are important in drug discovery. We demonstrate here that the predictions of solvation free energies in water, in 1-octanol, and self-solvation free energies computed using force field–based expanded ensemble molecular dynamics simulation provide good accuracy compared to existing empirical and semi-empirical methods. These solvation free energies are then, as shown here, used for the prediction of a wide range of physicochemical properties important in the assessment of bioavailability and partitioning of compounds. In particular, we consider here the vapor pressure, the solubility, Henry's law constant and activity coefficients in both water and 1-octanol, and the air-water, air-octanol, and octanol-water partition coefficients of amino acid side chain analogs computed from the solvation free energies. The calculated solvation free energies

using different force fields are compared against each other and with available experimental data. The protocol here can also be used for a newly designed drug and other molecules where force field parameters and charges are obtained from density functional theory.

Keywords: Amino Acid; Side Chain Analogs; Solubility; Vapor Pressure; Octanol/Water Partition Coefficient; Solvation Free Energy; Self-Solvation Free Energy; Expanded Ensemble; Henry's Law Constant.

I. INTRODUCTION

The high cost ($500 million to $2 billion)[1,2] of developing a new drug is mainly due to very limited success rate (0.01%) of clinical tests. The fulfillment of the so-called ADME (absorption, distribution, metabolism, and excretion)[3,4] requirement of a potential drug candidate is important in a drug discovery project. Here we compute and use the solvation free energies of a compound in water, 1-octanol, and itself for the prediction of physicochemical properties related to ADME.

The bioavailability and partitioning of a chemical compound depend on its physicochemical properties including vapor pressure of the liquid (P_{vap}^{liq}) or solid (P_{vap}^{solid}), solubility in water (C^{wat}) and in an organic solvent (e.g., solubility in 1-octanol, C^{oct}), and the air-water (K_{AW}), air-octanol (K_{AO}) and octanol-water (K_{OW}) partition coefficients.[5,6] The determination of the physicochemical properties in the prescreening of real or virtual pharmaceuticals and chemicals may require expensive and time consuming synthesis and laboratory measurements. To avoid this, common practice is to calculate these properties using empirical or semi-empirical multi-parameter correlative models.[5,6] Such predictions can be satisfactory for the compounds used in the training set for the model, or for compounds with

similar structure and functional groups to those in the training set. But predictions from such models may not be accurate for other compounds.

A theoretically-based alternative is to use force field-based molecular simulation in which the van der Waals dispersion and Coulombic interactions are considered explicitly. An attractive feature of this approach is that it can be used to predict properties of chemicals knowing only their chemical structure. This is advantageous for a newly proposed chemical, and for chemicals with only limited data or data of limited accuracy, or in cases where extrapolation over a large temperature range leads to results of uncertain accuracy[7]. For example, the reported experimental vapor pressures for compounds of low volatility can disagree by an order of magnitude or more. Finding accurate experimental data for solubilities can also be difficult, with specific examples given by Jorgensen and Duffy[8], Kishi and Hashimoto[9] and Katritzky et al.,[10,11] all of whom pointed out that the experimental accuracy can be about 0.6 log unit.[12] Also, experimental octanol-air partition coefficient data are limited in the literature[13–15] and estimation methods are applicable to only certain classes of chemicals.[16–20]

One alternative to experimental measurement is the use of quantitative structure-property relationships (QSPR) as used in chemistry and biology[21] for the prediction of a variety of physicochemical properties. However, the most accurate QSPR modeling of vapor pressure and solubility requires at least some experimental vapor pressure[22] and solubility data for molecules with similar functionalities, and the training set in their development must contain data for a greater number molecules than there are fitting parameters in the regression. Some QSPR models also require the data on boiling points, critical pressures, and critical temperatures,[23–25] which may not be available. For compounds with low vapor pressures, the uncertainties and/or scatter in the experimental data results in poor QSPR predictions; in some cases there can be an order of magnitude difference between prediction and experiment. The UNIFAC (UNIversal Functional Activity Coefficient)[26] method can lead to predictions with large errors. Also, commercial

databases such as DIPPR (Design Institute of Physical Properties),[27] the TRC database,[28] and IUCLID (International Uniform Chemical Information Database)[29] contain inconsistencies (see Table 1 of Ref. 7) as identified by Olsen and Nielsen.[7]

Group contribution based methods[30–34] such as UNIFAC[35,36] and AQUAFAC (Aqueous functional group activity coefficients; application to hydrocarbons)[37–41] are commonly used for the prediction of solubilities, though, they require a melting point correction term for solid solutes. Hansch et al.[42] have correlated solubilities with octanol-water partition coefficients; however, this method is not applicable to solid solutes. Yalkowsky and Valvani[43] improved the Hansch method by incorporating the entropy of fusion and melting point as in the group contribution method of Irmann.[30]

The general solubility equation (GSE)[43] and its variants[44–47] require melting point and octanol/water partition coefficient data for the estimation of solubilities. Martinez and Gómez[48] used partition coefficients and entropies of fusion to predict aqueous solubilities for a class of compounds. Multiple linear regression models[32,33,49] are also used widely in solubility estimations. Also the mobile order theory[50–52] is used for solubility predictions; however, it is limited by the need for solute cohesion parameter information in addition to melting point corrections. Abraham and Le predicted solubility using solvation free energy relationships.[53]

The accuracy of the methods discussed above depend on structural information (such as size, shape, morphology, conformation, and polymorphs) and the nature of the self-interactions (e.g., hydrogen bonding, molecular orientation, polarizability) and solute-solvent interactions (e.g., solvent accessible area, preferential solvation of functional groups, hydrogen bonding, dipole moment, the hydrophobic or hydrophilic characteristics of the solutes in aqueous solution). Quite often the correlations and model calculations rely on other experimental data and the experimental uncertainties or scatter among different measurements accumulate in the final predictions.

As a result of the availability of faster computers and powerful algorithms, molecular simulation techniques are becoming an increasingly important tool for the prediction of thermophysical properties. The uncertainty in the use of molecular simulations is the adequacy and accuracy of the force fields used. This is an important consideration in the work reported here.

Amino acids are the building blocks of proteins, nucleic acids, hormones, neurotransmitters, and an essential component of living organisms. The affinities of amino acid side chains for water and 1-octanol can determine the structure of proteins in solution.[54] The biological activities of peptides and proteins are also controlled by the solvation of the amino acid side chains, and are important in protein engineering and biochemistry. Here we calculate the physicochemical properties of amino acid side chain analogs (AASCAs) using molecular simulations.

The performance of small molecule force fields in reproducing experimental solvation free energies are commonly compared against one another with different water models and with either default partial atomic charges or those calculated from quantum or semi-classical methods. The accuracy of the force fields and the performance of the algorithms have been examined by others using amino acid side chain analogs as the test set.[55–61] One of the more complete studies was carried out by Shirts et al.[55] in which they compared the hydration free energies of amino acid side chain analogs using the AMBER, CHARMM, and OPLS-AA force fields, with TIP3P as a water model. Separately they compared the hydration free energies using the TIP3P, TIP4P, SPC, SPC/E, TIP3P-MOD, and TIP4P-Ew water models with the OPLS-AA (Optimized Potentials for Liquid Simulations-All Atom) force field for the amino acid side chain analogs.[56] As expected, they found that the hydration free energy results varied with the choice of force field and water model.

Later, more extensive simulations were carried out by Mobley et al.[62] and Shivakumar et al.[63,64] for several hundred molecules, including the amino acid side chain analogs. Also, the results of simulation using the GAFF(Generalized Amber Force Field)[62] have been tested against measured solvation free energies for 504 neutral compounds with the TIP3P water model. Shivakumar et al.[63,64] compared OPLS-2005/(SPC+default charge), GAFF/(SPC+AM1-BCC), CHARMm-MSI/(SPC+CHELPG) models for hydration free energy calculations. The general conclusion is that the resulting hydration free energies were found to vary with the choices for the water model, the partial atomic charges, and above all on the force fields. Recently, Caleman et al.[65] have, in addition to hydration free energies, used the OPLS-AA[66–69] and GAFF[70] force fields to compute densities, enthalpies of vaporization, heat capacities, surface tensions, isothermal compressibilities, volume expansion coefficients, and dielectric constants. Also, Vanommeslaeghe et al.[71] have reported densities and enthalpies of vaporization computed using CGenFF (CHARMM Generalized Force Field).[71] The TraPPE[72–74] force field has also been studied for reproducing experimental vapor pressures. Recently Garrido et al.[75] used different force fields for the calculation of octanol-water partition coefficients. However, what has not been studied is how well different force fields perform in the calculations of solubilities and air-solvent partition coefficients. Moreover, computations of self-solvation free energies, and from them vapor pressures, by molecular simulations are rarely found in the literature. Although important in drug design, vapor pressure and solubilities (in water and 1-octanol) of a solid solute in its subcooled state is rarely reported. As an alternative to force field-based simulations for computing solvation free energies, one can use quantum mechanics-based solvation models including the polarizable continuum model (PCM),76 Minnesota solvation models (MSM),77 and variants of conductor-like screening model (COSMO-RS,78 DCOSMO-RS,79 and COSMO-SAC80). A detailed discussion of the usage of these models is out of the scope of this work.

Here we demonstrate that use of a combined framework of molecular simulations and macroscopic thermodynamics to compute the physicochemical properties of drug-like small molecules. Also, the validity of the protocol is examined for subcooled properties of crystalline or amorphous compounds, as is the feasibility of using currently available small molecule force fields for the computation of physicochemical properties. Here, the water, 1-octanol and self-solvation free energies are computed using the four different force field models discussed above. From these solvation free energies, predictions for the thermophysical properties are made and compared with available experimental data.

II. METHODS

A. Solvation Free Energy Calculation: Expanded Ensemble Method

A number of methods are available for the calculation of solvation free energies,[81-84] and the choice of the method primarily depends on computational accuracy and efficiency. The availability of a phase space convergence check, a built-in error estimation mechanism, the ability to handle conformational barriers, and minimum post-processing are important criteria for choosing a suitable algorithm,[85,86] and this led us to use the expanded ensemble (EE) method[83] in the calculations here. In the EE algorithm the solvation free energy is calculated from a single simulation dividing the phase space Hamiltonian into subensembles for faster convergence, and connecting them using a random walk for the insertion of a solute into the solvent. The details of the algorithm can be found elsewhere.[83,87-89]

B. Simulation-Based Thermodynamics Protocol of Computing Physicochemical Properties

Molecular simulation uses information on the atomic-level interactions between a solute and a solvent from which ensemble averages of different thermodynamic quantities can be obtained; the solvation free energies are of interest here. The detailed protocol for computing physicochemical properties from solvation free energies can be found elsewhere[90] and shown

schematically in Figure 1. Following are the main equations for computing vapor pressure (P_{vap}),[91,92] solubility ($C_{solvent}$),[93] and the partition coefficients (octanol-water $\log_{10}K_{OW}$;[85,86,90] air-water $\log_{10}K_{AW}$;[91] and air-octanol: $\log_{10}K_{AO}$[91]) from solvation free energies

$$P_{vap} = P^0 M \exp\left(\frac{\Delta G_{self}}{RT}\right) \quad (1)$$

$$P_{vap}^{solid} = P_{vap}^{subco} exp\left\{\frac{\Delta S_{fus}}{R}\left(1 - \frac{T_m}{T}\right)\right\} \quad (2)$$

$$C_{solvent} \equiv M_{solvent} = \left(\frac{P_{vap}}{P^0}\right) \exp\left[-\frac{\Delta G_{solvent}}{RT}\right] \quad (3)$$

$$\log_{10}K_{OW} = \frac{(\Delta G_{wat} - \Delta G_{oct})}{2.303RT} \quad (4)$$

$$\left.\begin{aligned}\log_{10}K_{AW} &= -\frac{\Delta G_{wat}}{2.303RT}\\ \log_{10}K_{AO} &= -\frac{\Delta G_{oct}}{2.303RT}\end{aligned}\right\} \quad (5)$$

In eqn. (1), $P^0 = 24.45$ atm is the ideal gas pressure at 298 K at a density of 1 mol/L, M is the molarity of the solute in its pure liquid state, R is the gas constant, and T is the room temperature. Since *p*-cresol is a solid at room temperature and pressure, the subcooled liquid vapor pressure (P_{vap}^{subco}) was calculated using eqn. (1) and then corrected for the solid phase vapor pressure (P_{vap}^{solid}) using eqn. (2) where Δs_{fus} is the entropy of fusion at the melting temperature T_m (in degrees Kelvin). In eqn. (3), $\Delta G_{solvent}$ is the solvation free energy and $M_{solvent} = C_{solvent}$ is the solubility of the solute where "solvent" is either "water" or "octanol". The solubilities of solid *p*-cresol in water and 1-octanol were computed using the corrected vapor pressure (eqn. (2)) mentioned earlier.

C. Simulation Details

(i) Modeling Solvents and Amino Acid Side Chain Analogs

The amino acid side chain analogs (see Figure 2 for molecular structure) were modeled using the parameters of GAFF,[70] CGenFF,[71] and OPLS-AA[66–69] force fields. The initial geometries were generated using ArgusLab 4.0.1[94] and then optimized in the gas phase using

Gaussian 09[95] to obtain the minimum energy structure. The pdb (protein data bank) files created from the optimized geometries were used in AmberTools 1.5[96] to obtain the GAFF parameters and AM1-BCC charges. The pdb files were converted into mol2 format using OpenBabel 2.3.1[97] and then uploaded to the ParamChem[98-100] webpage to generate the parameters and charges for CGenFF, and manually validated for accuracy. The default charges and parameters for OPLS-AA force fields were collected from literature.

(ii) Simulation Conditions

All *NPT*-EE[83,87-89] simulations were carried out using MDynaMix package102,102 version 5.2.4 at room temperature and 1 atm pressure controlled by a Nose-Hoover[103,104] thermostat and barostat, respectively. The relaxation times for the thermostat and barostat were 30 fs and 700 fs, respectively. The initial configurations of all simulations were generated placing the optimized molecular geometries in a face centered cubic lattice and replicated using the minimum image convention and periodic boundary conditions.[105] The van der Waals interactions were truncated at 12 Å and conventional long range corrections[105] were applied. The Ewald summation method was used to account for the electrostatic interaction.[105-107] A long time step of 2 fs was used for integrating the equations of motion for the slower moving parts of the molecule and a short time step of 0.5 fs was used for integrating the equations of motion for the faster moving components, following the double time step algorithm of Tuckerman et al.[108]

In all solvation free energy calculations using the EE algorithm, including that for the self-solvation free energies, a single solute molecule was gradually inserted into the fluid of solvent molecules; the numbers of solvent molecules used for water, 1-octanol, and the pure liquid of amino acid side chain analogs were 500, 200, and 200, respectively. Twenty to thirty subensembles were used and the extent of the insertion of the solute molecule in a solvent was started from the fully disengaged state ($\lambda = 0$) and ended at the fully inserted state ($\lambda = 1$). The values of the insertion parameter, λ, were distributed uniformly[109] among the subensembles. The

balancing factors were optimized automatically using the Wang-Landau (WL) algorithm[110,111] with 20 iterations used for optimization. The initial increment in the WL optimization was 0.1 and later scaled by a factor of 0.5 after each iteration, and at each iteration the system passed at least twice between the first (*m=1*) and the last (*m=30*) subensembles.

To calculate the self-solvation free energy of subcooled tyr (*p*-cresol) at room temperature the initial configuration was first generated at very low density and then gradually compressed at a steady rate until the solid density was reached without crystallization or solidification. [An alternative approach would be starting from higher temperature, lower than the boiling temperature but higher than the freezing temperature, and quenching the system at a controlled rate.] Then the system was equilibrated in the *NPT* ensemble for 2 ns before initiating EE simulations at constant temperature and pressure. In a 20 ns simulation run several intermediate configurations, including the starting and ending configurations, were visually examined for further confirmation of the subcooled liquid phase.

D. Compilation of Experimental Data for Comparison with Simulation Results

The reliability of the experimental data effects on the assessment of the accuracy of computational methods[55,56]. For the analysis here the experimental hydration free energies were collected from the experimental database of Katrizky et al.[10], though in most of the simulation literature the comparison is with the measured data of Radzicka and Wolfenden.[112] These measured data sets for hydration free energy were found to be consistent with each other for the set of amino acid side chain analogs (AASCAs) studied here, other than for thr (ethanol) for which they differ 0.15 kcal/mol. The rationale for comparing with the data from Katrizky et al.[10] is that we examined the measured hydration free energies and the solvation free energies in 1-octanol from different sources, and examined the values of $\log_{10}K_{OW}$ calculated from these using eqn. 4. Although the data of Katrizky et al.[10] and Radzicka and Wolfenden[112] were found

to be consistent with each other for the hydration free energies of the AASCAs studied here, the solvation free energies in 1-octanol and the calculated $\log_{10}K_{OW}$ were significantly different between the two data sets. To understand the inconsistencies we compared the calculated $\log_{10}K_{OW}$ data with the recommended data of Sangster (Supplementary Figure 1 and Supplementary Table 1) and found that solvation free energies of Katrizky et al. agreed with the recommended $\log_{10}K_{OW}$ values better than those of Radzicka and Wolfenden.[112] This is the reason for choosing Katrizky et al. data for our comparisons. Also, the solvation free energies in octanol were compared with those calculated from the measured hydration free energies and recommended $\log_{10}K_{OW}$ data of Sangster et al. [113,114] The self-solvation free energies of ser (methanol), thr (ethanol), and 1-octanol were obtained from the experimental data of Katrizky et al.[10], and the self-solvation free energies of phe (toluene) and tyr (p-cresol) were calculated from their measured vapor pressure data reported in the EPIWEB 4.1[115] database and the measured molarity in the liquid phase using eqn. (1); the self-solvation free energy of solid *p*-cresol was calculated from its experimental sub-cooled vapor pressure and the molarity in the sub-cooled phase. The experimental data for the pure liquid or solid phase vapor pressures of the amino acid side chain analogs, the subcooled liquid vapor pressure of tyr (*p*-cresol) and the experimental aqueous solubility data of the amino acid side chain analogs were also obtained from EPIWEB 4.1.[115]

The solubilities of the amino acid side chain analogs in 1-octanol were calculated from their measured vapor pressures and their solvation free energies in 1-octanol. The experimental values for the logarithms of the air-water partition coefficients ($\log_{10}K_{AW}$) were obtained from Abraham et al.[116] and also independently calculated from the experimental solvation free energies using the eqn. (5). Both of these methods agree, on average, to within 0.02 log unit. Similarly, the logarithm of octanol-water partition coefficients ($\log_{10}K_{OA}$) were obtained from

Meylan and Howard[14], and also independently calculated from experimental solvation free energies using eqn. (5) for *p*-cresol. Both of these agree within 0.01 log unit.

III. RESULTS

A. Solvation free energies

The average unsigned errors (AUE) compared to the experimental data for the physicochemical properties of the 6 AASCAs computed here using the CGenFF/ParamChem (charge), GAFF/AM1-BCC (charge), and OPLS-AA force field parameters are summarized in Table 1, and detailed predictions for each of the compounds appear in the Supplementary Information (Supplementary Tables 2-10). The self-solvation free energies for the AASCAs are collected in Supplementary Tables 2a and 2b, and the average errors are summarized in Table 1.

The GAFF/AM1-BCC (charge) force fields reproduced the self-solvation free energy of 1-octanol within its standard deviations and the OPLS-AA force fields with their default charges led to self-solvation energy predictions to within twice of their standard deviations (Supplementary Table 2a). The predictions using the CGenFF/ParamChem (charge) force field were noticeably less accurate. The self-solvation free energies for ala (methane) and val (propane) were not calculated since they are gases at room temperature. For tyr (*p*-cresol), which is solid at room temperature, its subcooled self-solvation free energies were calculated as mentioned earlier.

Supplementary Table 3 displays the hydration free energies for the AASCAs computed here using CGenFF/ParamChem (charge) force field and the TIP3P water model. However, in computing physicochemical properties using the GAFF/AM1-BCC (charge) and OPLS-AA force fields, the hydration free energies were collected from the simulation data obtained from literature[55,62] where the water model was also TIP3P. The average errors of the predictions are

summarized in Table 1, where we see that the OPLS-AA and CGenFF force fields, in that order, lead to the most accurate predictions.

The average errors in the predicted solvation free energies for the AASCAs in 1-octanol modeled by GAFF/AM1-BCC (charge), CGenFF/ParamChem (charge) and OPLS-AA force fields are presented in Table 1, and detailed predictions for each of the compounds appear in the Supplementary Table 4. Overall, the CGenFF force field with ParamChem[98-100] assigned charges led to better predictions than the other force fields; predictions using the OPLS-AA force field being the second most accurate. Note that all methods led to predicted solvation free energies that were slightly less negative than the experimental data, so that a lower solubility is predicted.

B. Vapor Pressures from Self-Solvation Free Energies

The predicted (logarithms of) the vapor pressures ($\log_{10}P_{vap}^{liq/sol}$) for each of 1-octanol and the amino acid side chain analogs calculated from the self-solvation free energies obtained from EE simulations using eqn. (1) are given in Supplementary Tables 5a and 5b, respectively. A summary of the statistical averages is given in Table 1, but does not include 1-octanol and *p*-cresol. For 1-octanol the GAFF/AM1-BCC (charges) FF led to a predicted logarithm vapor pressure with unsigned error of only 0.07 (in Pa units) compared to the experimental datum. The predictions based on the OPLS-AA and CGenFF led to decreasing in accuracy in that order. The use of CGenFF led to underestimates of the 1-octanol vapor pressure, while the use of the OPLS-AA FF led to an over estimate.

For the amino acid side chain analogs (excluding ala and tyr), use of the OPLS-AA force field led to the best predictions, though it overestimated the vapor pressure in all cases. Predictions using the CGenFF and GAFF force fields were, on average, comparable in

reproducing experimental vapor pressures, though both led to overestimates the vapor pressures, and were less accurate than the OPLS-AA force field.

C. Solubilities

Supplementary Table 6 gives the predicted logarithms of the aqueous solubilities (mol/l) of the amino acid side chain analogs (excluding tyr) compared with experiment, and Table 1 contains the summary of relevant statistics. The OPLS-AA and CGenFF force fields led to predictions of comparable accuracy in reproducing the experimental aqueous solubilities; followed by the GAFF with AM1-BCC charges.

Supplementary Table 7 gives the predicted logarithms of the solubilities (in mol/L) of the amino acid side chain analogs (excluding tyr) in 1-octanol together with the experimental data, and Table 1 contains the summary of relevant statistics. In this case predictions based on the CGenFF force field are the most accurate followed by the GAFF and OPLS-AA force fields in that order.

D. Sub-Cooled Versus Solid Phase Vapor Pressures and Solubilities of p-Cresol

The sub-cooled vapor pressure obtained from the self-solvation free energy of *p*-cresol and the solubility predictions are summarized in Table 2. The difference in vapor pressures of the subcooled and solid phases is within 0.1 log unit, and the difference in solubilities is within 0.07 log unit. These differences are due to the small free energy changes on the solid-to-subcooled liquid phase transitions since the melting point of *p*-cresol (35.5^{o}C) is only slightly above room temperature.

E. Octanol-Water, Air-Water and Air-Octanol Partition Coefficients

Supplementary Table 8 displays the logarithms of the octanol-water partition coefficients ($\log_{10}K_{ow}$) for each of the compounds calculated from the solvation free energies in Supplementary Tables 3 and 4 using eqn. (4) together with the experimental data for comparison. The average errors are summarized in Table 1. CGenFF/ParamChem (charge) force field leads to the most accurate predictions, and OPLS-AA is the next most accurate.

The logarithms of the air-water partition coefficients of the amino acid side chain analogs calculated from the aqueous solvation free energies obtained from simulation together with the experimental data are given in Supplementary Table 9, and the average errors are given in Table 1. Predictions for this property based on the OPLS-AA FF best reproduced the experimental air-water partition, followed by the predictions based on the CGenFF force field.

Supplementary Table 10 contains the air-octanol partition coefficients calculated from the solvation free energies in 1-octanol obtained from the simulations here. For the air-octanol partition coefficients, the property predictions of the CGenFF were most accurate followed by the OPLS-AA, and GAFF force fields, respectively.

IV. DISCUSSION

The self-solvation free energy predictions from molecular simulation for the molecules of interest here have not been reported previously. However, Winget et al.[91] reported vapor pressures calculated the using the SM5.4, SM5.2R, and SM5.0R solvation models,[117] from which self-solvation free energies can be calculated. Their results were found to vary with the level of quantum theory and basis sets used in obtaining the optimized minimum energy geometry. We have compared our self-solvation predictions for the liquid solutes with those obtained from the calculated vapor pressures of Winget et al.[91] (Supplementary Table 11a), and the self-solvation free energy of solid p-cresol at room temperature with the data of Thompson et al.[93]

(Supplementary Table 11b). For the liquid amino acid side chain analogs the results from using the OPLS-AA/(default charge) force field are of comparable accuracy to that obtained with the best of the solvation models, and better than the predictions of many of them, though the SM5.4/AM1 and SM5.4/PM3 predictions are the most accurate. For solid tyr (*p*-cresol) the results from simulation with the OPLS-AA force field are more accurate than those obtained from any of the solvation models mentioned above, and from using other force fields. A possible cause of error in the solvation models is that the macroscopic surface tension of liquid *m*-cresol is used for *p*-cresol, which affects the surface tension-based semi-empirical component of the self-solvation free energy.

The difference between the self-solvation free energies of *p*-cresol calculated from the experimental vapor pressures in the solid phase and the subcooled phase is only 0.15 kcal/mol, which is smaller than the uncertainties in most of the solvation free energies calculated here, and of the variation in the reported experimental values. This suggests that eqn. (3) can be used for computing solubilities rather than using the solid vapor pressure correction of eqn. (2) provided that, as here, the melting temperature of the solid is not very different from room temperature. If there is a large temperature difference, the use of eqn. (3) directly without the vapor pressure correction may lead to substantial errors.

Supplementary Table 12a shows the comparison of vapor pressures predicted from the self-solvation free energies computed using simulations here and the Minnesota solvation models.[117] Different combinations of quantum theory level and basis sets were used in the solvation models for the calculation of self-solvation free energies and then the vapor pressures. The comparison was made for the amino acid side chain analogs common to this study and ref. 91, namely toluene, ethanol, and methanol. Though the best results are obtained with the SM5.4/AM1 and SM5.4/PM3 solvation models, the predictions using the OPLS-AA force field are quite good,

and is better than many of the other solvation model calculations. Supplementary Table 12b compares the predictions of the vapor pressure of *p*-cresol, a solid solute at room temperature, from the different methods, and there we see that the prediction from simulation using any of the three force fields considered here is somewhat more accurate than the solvation model predictions.

Supplementary Table 13 shows the comparison of the predicted solubilities for phe (toluene) and tyr (*p*-cresol) from our calculations using different force fields and the results obtained by Thompson et al.[93] using the Minnesota solvation models.[117] The logarithms of aqueous solubilities (in mol/L) and the unsigned errors for phe (toluene), liquid at room temperature, and *p*-cresol, solid at room temperature, are shown in that table. For phe (toluene) the predictions using OPLS-AA force field is the most accurate among the different force fields and solvation models, while the best results for *p*-cresol are obtained with the CGenFF that leads to more accurate results than any of the solvation models. Considering both compounds together, using the OPLS-AA force field in EE simulation leads to the best predictions of aqueous solubilities, and better than those of the solvation models.

Egan[4] discussed the limited accuracy of solubility measurements and modeling[118-121] due to polymorphism, pH-dependent ionization, crystal melting point, salt forms, and formulation solvents (organic solvents, water, methylcellulose, polyethylene glycol, water and the others). Nonetheless, the approach presented here can be to compute the subcooled solubility from first principles atomistic simulations without any supporting measured data.

V. CONCLUSIONS

In this communication the solvation free energies of amino acid side chain analogs in water, 1-octanol and their own pure liquids (for species liquid at room temperature) were calculated using the expanded ensemble method with three different force fields. These results

were then used to compute the vapor pressures at room temperature, the octanol-water, air-water and air-octanol partition coefficients, and the solubilities in water and 1-octanol. By comparing with measured values of these properties, we find that over all the use of the OPLS-AA force field leads to the most accurate predictions. The CGenFF force field also led to good, but generally slightly less accurate, predictions. We also find the using that using expanded ensemble simulations and the OPLS-AA or GCenFF force fields lead to predictions that are, in general, as good as those obtained from the semi-empirical SM5.X class of solvation models. Although the method we have described is computationally expensive, it has the advantage of not relying on somewhat arbitrarily defined functional groups or the availability of descriptors and their parameters as is the case in semi-empirical methods.

The protocol presented here can be used at the prescreening stage of drug development or for environmentally important properties of compounds for which only their geometrical structure are known.

ACKNOWLEDGEMENTS

This work has been supported by NSF Grant GOALI-0853685. This work partially used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

REFERENCES

1 C. Adams and V. Brantner, *Health Aff.*, 2006, **25**, 420.

2 J. A. DiMasi, R. W. Hansen and H. G. Grabowski, *J. Heal. Econ.*, 210AD, **22**, 151–85.

3 W. L. Jorgensen, in *Drug Design: Structure- and ligand-based approaches*, ed. J. K. M. Merz, D. Ringe and C. H. Reynolds, Cambridge University Press, New York, 2010, pp. 1-24.

4 W. J. Egan, in *Drug Design: Structure- and ligand-based approaches*, ed. J. K. M. Merz, D. Ringe and C. H. Reynolds, Cambridge University Press, New York, 2010, pp.165-180.

5 B. R. White, C. R. Wagner, D. G. Truhlar and E. A. Amin, 2008, **24**, 1718–1732.

6 K. M. Merz, D. Ringe and C. H. Reynolds, Ed., *Drug Design: Structure- and Ligand-Based Approaches*, Cambridge University Pres, New York, 2010.

7 E. Olsen and F. Nielsen, *Molecules*, 2001, **6**, 370–389.

8 W. L. Jorgensen and E. M. Duffy, *Adv. Drug. Deliv. Rev.*, 2002, **54**, 355–66.

9 H. Kishi and Y. Hashimoto, *Chemosphere*, 1989, **18**, 1749–1759.

10 A. R. Katritzky, A. A. Oliferenko, P. V. Oliferenko, R. Petrukhin and D. B. Tatham, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1794–1805.

11 A. R. Katritzky, A. A. Oliferenko, P. V. Oliferenko, R. Petrukhin and D. B. Tatham, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1806–1814.

12 W. L. Duffy and E. M. Jorgensen, *J. Am. Chem. Soc.*, 2000, **2**, 2878–2888.

13 M. H. Abraham, J. Le, W. E. Acree Jr., P. W. Carr and A. J. Dallas, *Chemosphere*, 2001, **44**, 855–863.

14 W. M. Meylan and P. H. Howard, *Chemosphere*, 2005, **61**, 640–644.

15 K. Sepassi and S. H. Yalkowsky, *Ind. Eng. Chem. Res.*, 2007, **46**, 2220–2223.

16 J. W. Chen, X. Quan, Y. Z. Zhao, F. L. Yang, K.-W. Schramm and A. Kettrup, *Bull. Environ. Contam. Toxicol.*, 2001, **66**, 755–761.

17 J. Chen, X. Xue, K.-W. Schramm, X. Quan, F. Yang and A. Kettrup, *Chemosphere*, 2002, **48**, 535–544.

18 J. W. Chen, T. Harner, P. Yang, X. Quan, S. Chen, K.-W. Schramm and A. Kettrup, *Chemosphere*, 2003, **51**, 577–584.

19 J. Chen, X. Xue, K.-W. Schramm, X. Quan, F. Yang and A. Kettrup, *Comput. Biol. Chem.*, 2003, **27**, 165–171.

20 T. Puzyn and J. Falandysz, *Atmos. Environ.*, 2005, **39**, 1439–1446.

21 C. Hansch and A. Leo, *Exploring QSAR: Volume 1: Fundamentals and applications in chemistry and biology*, American Chemical Society, 1995.

22 C. Liang and D. A. Gallagher, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 321–324.

23 D. Mackay, A. Bobra, D. W. Chan and W. Y. Shlu, *Environ. Sci. Technol.*, 1982, **16**, 645–649.

24 J. McGarry, *Ind. Eng. Chem. Process Des. Dev*, 1983, **22**, 313–322.

25 D. S. Mishra and S. H. Yalkowsky, *Ind. Eng. Chem. Res.*, 1991, **30**, 1609–1612.

26 T. Jensen, A. Fredenslund and P. Rasmussen, *Ind. Eng. Chem. Fundam.*, 1981, **20**, 239–246.

27. DIPPR. Data Compilation of Pure Compound Properties. U.S. Department of Commerce. National Institute of Standard and Technology: Gaithersburg. Now avaialable from Technical Database Services Inc. 135 West 50th Street. New York, 1999. www.tdstds.com.

28 Texas A & M University. Thermodynamics Research Center., 1998.

29 IUCLID. IUCLID. International Uniform Chemical Information Database on CD-ROM. European Commission EUR 17283. European Chemicals Bureau Existing Chemicals. Joint Research Centre: Ispra (VA) Italy, 2000; http://ecb.ei.jrc.it.

30 F. Irmann, *Chem. Eng. Tech.*, 1965, **37**, 789–798.

31 G. Klopman and H. Zhu, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 439–445.

32 G. Klopman, S. Wang and D. M. Balthasar, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 474–482.

33 R. Kuhne, R.-U. Ebert, F. Kleint, G. Schmidt and G. Schuurmann, *Chemosphere*, 1995, **30**, 2061–2077.

34 N. N. Nirmalakhandan and R. E. Speece, *Environ. Sci. Technol.*, 1989, **23**, 708–713.

35 A. T. Kan and M. B. Tomson, *Environ. Sci. Technol.*, 1996, **30**, 1369–1376.

36 S. Banerjee, *Environ. Sci. Technol.*, 1985, **19**, 456–457.

37 P. Myrdal, G. H. Ward, R.-M. Dannenfelser, D. Mishra and S. H. Yalkowsky, *Chemosphere*, 1992, **24**, 1047–1061.

38 P. Myrdal, G. H. Ward, P. Simamora and S. H. Yalkowsky, *SAR QSAR Environ. Res.*, 1993, **1**, 53–61.

39 S. Pinsuwan, P. B. Myrdal, Y.-C. Lee and S. H. Yalkowsky, *Chemosphere*, 1997, **35**, 2503–2513.

40 Y.-C. Lee, P. B. Myrdal and S. H. Yalkowsky, *Chemosphere*, 1996, **33**, 2129–2144.

41 P. B. Myrdal, A. M. Manka and H. Yalkowsky, *Chemosphere*, 1995, **30**, 1619–1637.

42 C. Hansch, J. E. Quinlan and G. L. Lawrence, *J. Org. Chem.*, 1968, **33**, 347–350.

43 S. H. Yalkowsky and S. C. Valvani, *J. Pharm. Sci.*, 1980, **69**, 912–922.

44 Y. Ran, N. Jain and S. H. Yalkowsky, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1208–1217.

45 Y. Ran and S. H. Yalkowsky, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 354–357.

46 N. Jain and S. H. Yalkowsky, *J. Pharm. Sci.*, 2001, **90**, 234–252.

47 Y. Ran, Y. He, G. Yang, J. L. H. Johnson and S. H. Yalkowsky, *Chemosphere*, 2002, **48**, 487–509.

48 F. Martinez and A. Gomez, *Phys. Chem. Liq.*, 2002, **40**, 411–420.

49 J. Huuskonen, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 773–777.

50 P. Ruelle and U. W. Kesselring, *J. Pharm. Sci.*, 1998, **87**, 987–997.

51 P. Ruelle and U. W. Kesselring, *J. Pharm. Sci.*, 1998, **87**, 998–1014.

52 P. Ruelle and U. W. Kesselring, *J. Pharm. Sci.*, 1998, **87**, 1015–1024.

53 M. H. Abraham and J. Le, *J. Pharm. Sci.*, 1999, **88**, 868–880.

54 J. T. Edsall and H. A. McKenzie, *Adv. Biophys.*, 1983, **16**, 53–183.

55 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *J. Chem. Phys.*, 2003, **119**, 5740–5761.

56 M. R. Shirts and V. S. Pande, *J. Chem. Phys.*, 2005, **122**, 134508.

57 B. Lin and B. M. Pettitt, *J. Comput. Chem.*, 2010, **32**, 878–885.

58 B. Hess and N. F. A. van der Vegt, *J. Chem. Phys. B*, 2006, **110**, 17616–17626.

59 J. L. MacCallum and D. P. Tieleman, *J. Comput. Chem.*, 2003, **24**, 1930–1935.

60 A. Villa and A. E. Mark, *J. Comput. Chem.*, 2002, **23**, 548–553.

61 J. Chang, A. M. Lenhoff and S. I. Sandler, *J. Phys. Chem. B*, 2007, **111**, 2098–2106.

62 D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts and K. A. Dill, *J. Chem. Theory. Comput.*, 2009, **5**, 350–358.

63 D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley and W. Sherman, *J. Chem. Theory. Comput.*, 2010, **6**, 1509–1519.

64 D. Shivakumar, Y. Deng and B. Roux, *J. Chem. Theory Comput.*, 2009, **5**, 919–930.

65 C. Caleman, P. J. Van Maaren, M. Hong, J. S. Hub, L. T. Costa and D. van der Spoel, *J. Chem. Theory. Comput.*, 2012, **8**, 61–74.

66 W. L. Jorgensen and J. Tirado-Rives, *Proc. Natl. Acad. Sci . U. S. A.*, 2005, **102**, 6665–6670.

67 M. L. P. Price, D. Ostrovsky and W. L. Jorgensen, *J. Comput. Chem.*, 2001, **22**, 1340–1352.

68 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

69 W. Damm, A. Frontera, J. Tirado-Rives and W. L. Jorgensen, *J. Comput. Chem.*, 1997, **18**, 1955–1970.

70 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

71 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell Jr., *J. Comput. Chem.*, 2010, **31**, 671–690.

72 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

73 C. D. Wick, M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 2000, **104**, 8008–8016.

74 C. D. Wick, J. I. Siepman, W. L. Klotz and M. R. Schure, *J. Chromatogr. A*, 2002, **954**, 181–90.

75 N. M. Garrido, A. J. Queimada, M. Jorge, E. A. Macedo and I. G. Economou, *J. Chem. Theory Comput.*, 2009, **5**, 2436–2446.

76 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999-3094.

77 A. V Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database version 2007*, University of Minnesota, Minneapolis, MN, 2009.

78 A. Klamt, J. Phys. Chem., 1995, **99**, 2224-2235.

79 A. Klamt and M. Diedenhofen, *J. Phys. Chem. A,* 2015**, 119**, 5439-5445.

80 S. T. Lin and S. I. Sandler, *Ind. Eng. Chem. Res.* 2002, **41**, 899–913.

81 C. Chipot and A. Pohorille, *Free energy calculations: theory and applications in chemistry and biology. Springer: Berlin, 2007.*, Springer, Berlin, 2007.

82 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.

83 A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.*, 1992, **96**, 1776–1783.

84 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.

85 A. Ahmed and S. I. Sandler, *J. Chem. Phys.*, 2012, **136**, 154505.

86 L. Yang, A. Ahmed and S. I. Sandler, *J. Comput. Chem.*, 2013, **34**, 284-293.

87 A. P. Lyubartsev, A. Laaksonen and P. N. Vorontsov-Velyaminov, *Mol. Phys.*, 1994, **82**, 455–471.

88 A. P. Lyubartsev, O. K. Førrisdahl and A. Laaksonen, in *2nd International Conference on Natural Gas Hydrates. Toulouse (France), June 2-6, 1996*, 1996, pp. 311–318.

89 A. P. Lyubartsev, S. P. Jacobsson, G. Sundholm and A. Laaksonen, *J. Chem. Phys. B*, 2001, **105**, 7775–7782.

90 A. Ahmed and S. I. Sandler, *J. Chem. Theory Comput.*, 2013, **9**, 2389–2397.

91 P. Winget, G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys. B*, 2000, **104**, 4726–4734.

92 C. F. Grain, in *Handbook of Chemical Property Estimation Methods: Environemntal Behavior of Organic Compounds, (ed.)*, ed. W. J. Lyman, W. F. Reehl, D. H. Rosenblatt, McGraw-Hill Book Co., New York, NY, 1982.

93 J. D. Thompson, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys.*, 2003, **119**, 1661–1670.

94 M. A. Thompson, ArgusLab, http://www.arguslab.com/arguslab.com/ArgusLab.html.

95 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. M. Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 user's reference*, 2010.

96 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

97 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminformatics*, 2011, **3**, 33.

98 ParamChem, https://www.paramchem.org/AtomTyping/, (accessed May 05, 2012).

99 K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell Jr., *J. Chem. Inf. Model.,* 2012, **52**, 3155−3168.

100 K. Vanommeslaeghe and A. D. MacKerell Jr., *J. Chem. Inf. Model.,* 2012, **52**, 3144−3154.

101 A. P. Lyubartsev and A. Laaksonen, *Comput. Phys. Commun.*, 2000, **128**, 565-589.

102 J. P. M. Jämbeck, F. Mocci, A. P. Lyubartsev and A.Laaksonen, *J. Comput. Chem.*, 2013, **34**, 187-197.

103 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

104 W. G. Hoover, *Physi. Rev. A*, 1985, **31**, 1695–1697.

105 R. J. Sadus, *Molecular simulation of fluids: theory, algorithms, and object-orientation.*, Elsevier, Amsterdam, 1999.

106 P. P. Ewald, *Ann. der Phys.*, 1921, **369**, 253–287.

107 M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, USA, 1989.

108 M. Tuckerman, B. J. Berne and G. J. Martyna, *J. Chem. Phys.*, 1992, **97**, 1990.

109 A. Lyubartsev and A. Laaksonen, *MDynaMix Package Version 5.2.4 User Manual*, Stockholm University, Stockholm, Sweden, 2011.

110 F. Wang and D. P. Landau, *Phys. Rev. Lett.*, 2001, **86**, 2050–2053.

111 K. M. Aberg, A. P. Lyubartsev, S. P. Jacobsson and A. Laaksonen, *J. Chem. Phys.*, 2004, **120**, 3770–3776.

112 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.

113 Sangster Research Laboratories, http://logkow.cisti.nrc.ca/logkow/index.jsp, (accessed May 05, 2012).

114 J. Sangster, *Octanol-water partition coefficients: fundamentals and physical chemistry*, John Wiley & Son Ltd, 1997.

115 EPIWEB 4.1 (US EPA, 2011). Estimation Programs Interface SuiteTM for Microsoft® Windows, v 4.10 2011.

116 M. H. Abraham, G. S. Whiting, R. Fuchs and E. J. Chambers, *J. Chem. Soc., Perkin Trans. 2*, 1990, 291–300.

117 A. V Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database version 2007*, University of Minnesota, Minneapolis, MN, 2009.

118 B. Faller, J. Wang, A. Zimmerlin, L. Bell, K. Hamon, Jacques Whitebread, Steven Azzaoui, D. Bojanic and L. Urban, *Expert Opini. Drug Metab. Tox.*, 2006, **2**, 823.

119 S. R. Johnson and W. Zheng, *AAPS J*, 2006, **8**, E27–40.

120 J. Wang and T. Hou, *Comb. Chem. High Throughput Screen.*, 2006, **14**, 328–38.

121 K. V. Balakin, N. P. Savchuk and I. V. Tetko, *Curr. Med. Chem.*, 2006, **13**, 223–241.

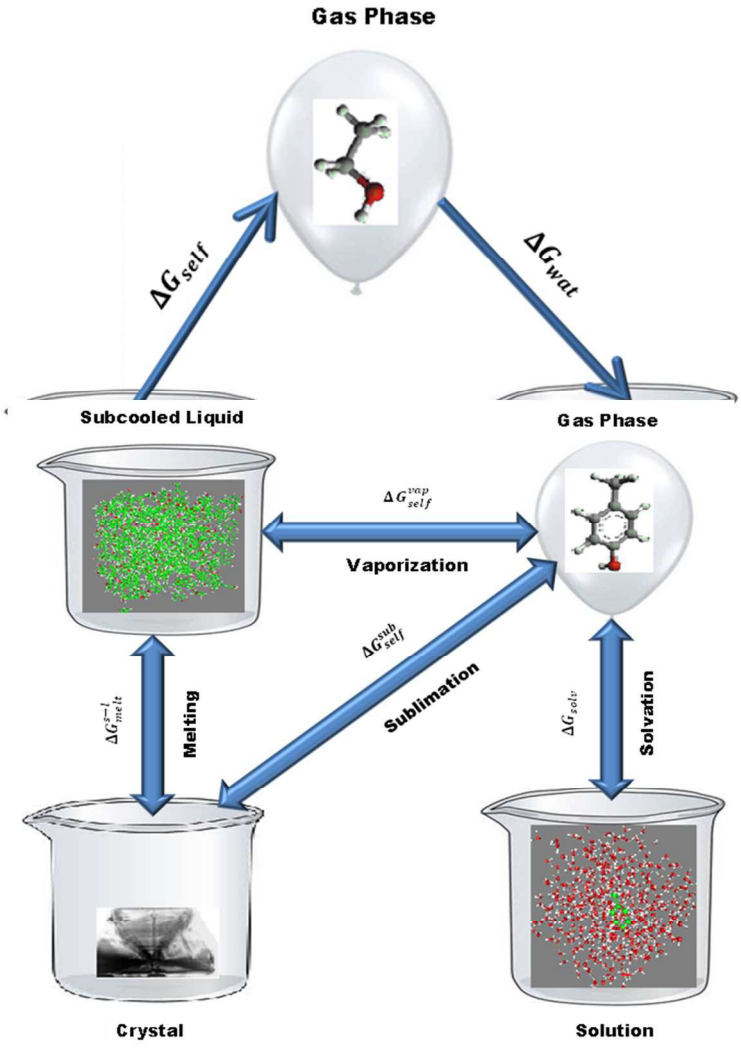

Figure 1 Schematics of solvation mechanism of liquid solute (top) and solid or subcooled liquid (bottom) .

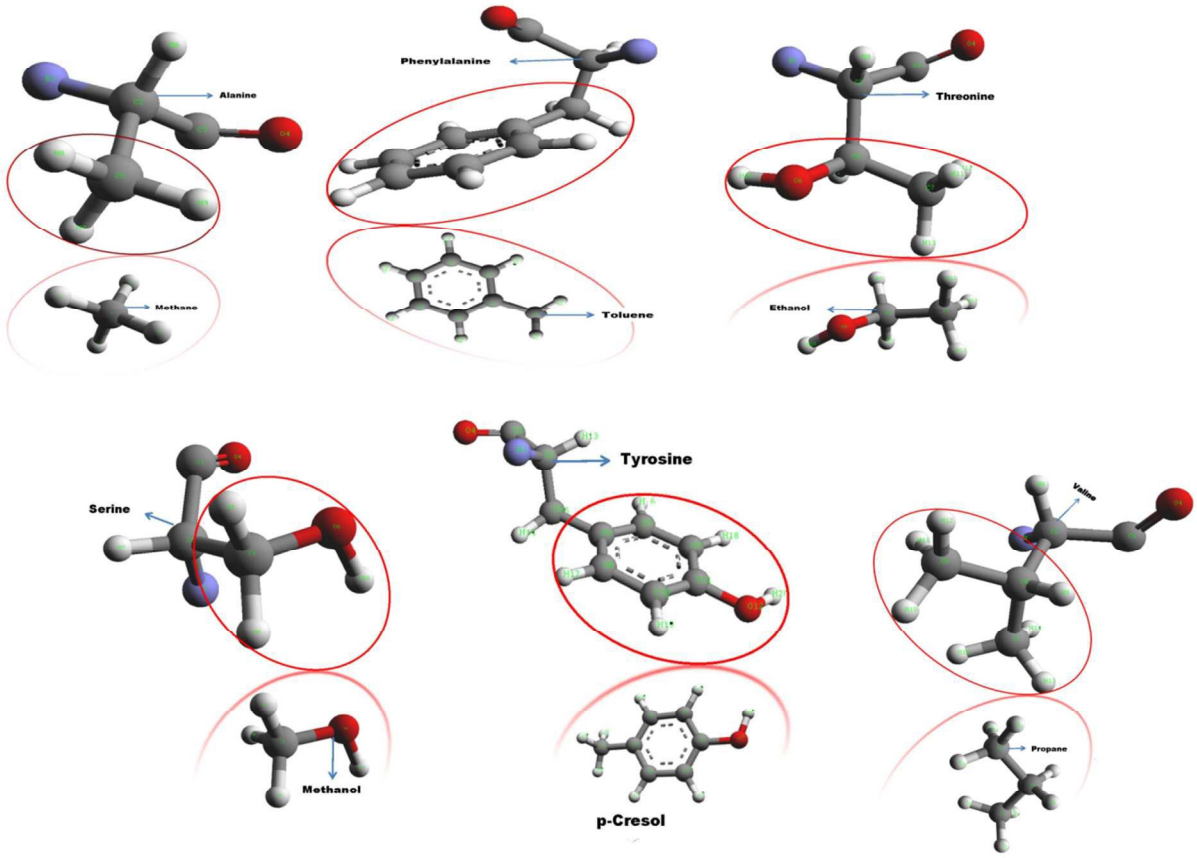

Figure 2 Amino acid side chain analogs.

Table 1: Average unsigned errors of the physicochemical properties of amino acid side chain analogs.

Physicochemical Property	GAFF/AMI-BCC (charge)	CGenFF/ParamChem (charge)	OPLS-AA
ΔG_{oct} (kcal/mol)	0.90	0.54	0.71
ΔG_{wat} (kcal/mol)	0.85	0.83	0.61
ΔG_{self} (kcal/mol)	0.90	0.45	0.33
$log_{10}P_{vap}$ (P_{vap} in Pa)	0.43	0.27	0.24
$log_{10}C_{wat}$ (S_{wat} in mol/L)	0.97	0.28	0.28
$log_{10}C_{oct}$ (S_{oct} in mol/L)	0.30	0.20	0.40
$log_{10}K_{OW}$ (log unit)	0.37	0.33	0.27
$log_{10}K_{AW}$ (log unit)	0.63	0.62	0.46
$log_{10}K_{AO}$ (log unit)	0.67	0.41	0.45

Table 2: Logarithm of vapor pressures in Pa unit ($log_{10}P_{vap}^{liq/sol}$) and logarithms of solubilities in mol/L units ($log_{10}S_{wat}$) of Tyr (*p*-cresol) in water and 1-octanol.

Force Fields	Sub-cooled vapor pressure and solubilities						Solid phase vapor pressure and solubilities					
	$log_{10}P_{vap}^{liq}$		$log_{10}C_{wat}$		$log_{10}C_{oct}$		$log_{10}P_{vap}^{sol}$		$log_{10}C_{wat}$		$log_{10}C_{oct}$	
	Expt.	Sim.	Expt.	Sim.	Expt.	Sim.	Expt.	Sim.	Expt.	Sim.	Expt.	Sim.
CGenFF	1.27	1.60	---	-1.85	---	1.12	1.17	1.52	-0.70	-1.92	1.26	1.05
GAFF		1.73		-0.74		0.79		1.65		-0.81		0.71
OPLS-AA		1.59		-0.96		1.19		1.51		-1.03		1.11