Published in final edited form as:
J Chem Theory Comput. 2012 May 8; 8(5): 1721–1736. doi:10.1021/ct300035u.

The Structure, Thermodynamics and Solubility of Organic Crystals from Simulation with a Polarizable Force Field

Michael J. Schnieders[1], **Jonas Baltrusaitis**[2], **Yue Shi**[1], **Gaurav Chattree**[1], **Lianqing Zheng**[3], **Wei Yang**[3,4,*], and **Pengyu Ren**[1,*]

[1]Department of Biomedical Engineering, University of Texas at Austin, Austin, TX 78712

[2]Departments of Chemistry and Chemical/Biochemical Engineering, University of Iowa, Iowa City, IA, 52242

[3]The Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306

[4]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306

Abstract

An important unsolved problem in materials science is prediction of the thermodynamic stability of organic crystals and their solubility from first principles. Solubility can be defined as the saturating concentration of a molecule within a liquid solvent, where the physical picture is of solvated molecules in equilibrium with their solid phase. Despite the importance of solubility in determining the oral bioavailability of pharmaceuticals, prediction tools are currently limited to quantitative structure–property relationships that are fit to experimental solubility measurements. For the first time, we describe a consistent procedure for the prediction of the structure, thermodynamic stability and solubility of organic crystals from molecular dynamics simulations using the polarizable multipole AMOEBA force field. Our approach is based on a thermodynamic cycle that decomposes standard state solubility into the sum of solid-vapor sublimation and vapor-liquid solvation free energies $\Delta G^{o}_{solubility}=\Delta G^{o}_{sub}+\Delta G^{o}_{solv}$, which are computed via the orthogonal space random walk (OSRW) sampling strategy. Application to the *n*-alkylamides series from aeetamide through octanamide was selected due to the dependence of their solubility on both amide hydrogen bonding and the hydrophobic effect, which are each fundamental to protein structure and solubility. On average, the calculated absolute standard state solubility free energies are accurate to within 1.1 kcal/mol. The experimental trend of decreasing solubility as a function of *n*-alkylamide chain length is recapitulated by the increasing stability of the crystalline state and to a lesser degree by decreasing favorability of solvation (i.e. the hydrophobic effect). Our results suggest that coupling the polarizable AMOEBA force field with an orthogonal space based free energy algorithm, as implemented in the program *Force Field X,* is a consistent procedure for predicting the structure, thermodynamic stability and solubility of organic crystals.

INTRODUCTION

Organic crystals are a critical class of materials for the pharmaceutical industry[1] and represent an emerging approach for the low-pressure storage of gases within crystalline metal-organic frameworks (MOFs).[2,3] Specific examples of organic crystals include the polymorphs of ice[4], pharmaceutical tablets such as paracetamol[5,6] (acetaminophen) and acetylsalicylic acid[7,8] (aspirin), and those responsible for diseases such as gout[9] (monosodium urate monohydrate) and kidney stones[10] (calcium oxalate). Key properties of organic crystals that are valuable to predict computationally include their atomic structure,

*Corresponding Authors: pren@mail.utexas.edu, yyang2@fsu.edu.

thermodynamic stability and their solubility (i.e. pharmaceutical tablets within the digestive tract). Although significant progress has been made toward *in silico* crystal structure prediction (CSP),[11–14] computational prediction of the thermodynamic stability and solubility of organic crystals from first principles has remained a challenging unsolved problem.[11,15]

For CSP, the underlying algorithms are characterized by application of a search protocol over space groups, unit cell dimensions and atomic coordinates using intermolecular potentials defined by fixed partial charge [16–20] or fixed atomic multipole force fields.[21–24] Although a set of fixed atomic multipoles can model the quantum mechanical electrostatic potential outside the molecular surface of a polar organic molecule in vacuum to high accuracy, changes in electron distribution due to alternate conformations or transfer into a condensed phase environment are neglected unless polarization is accounted for.[25,26] Therefore, candidate low energy crystal structures are usually rescored using electronic structure calculations based on density functional theory (DFT) with dispersion corrections.[27,28] When these algorithms were applied during the 4th blind test of CSP organized by the Cambridge Crystallographic Data Center (CCDC) in 2007, which featured 4 target organic crystals, 7 of 14 participating groups predicted at least one structure correctly and 4 groups had multiple successful predictions.[12] Moreover, in the 5th blind test of CSP in 2010, each of the 6 target organic crystals was predicted correctly by at least one of the 14 participating groups.[28]

However, despite its importance to the rational design of bioavailable pharmaceutical tablets, quantitative prediction of the thermodynamic stability and solubility of organic crystals from first principles has lagged behind both CSP and advancements in the computation of liquid state thermodynamics.[11,15,28] In the context of predicting solubility, the solvation free energy[15,29] and stability of amorphous solids[30] have been computed using Monte Carlo statistical mechanics simulations with the OPLS-AA force field,[18] however, a procedure to predict the thermodynamic stability of crystalline solutes has not been presented. An alternative to predicting solubility from first principles (i.e. from an atomic resolution model of molecular energetics) is the revised General Solubility Equation[31,32] (GSE), which predicts the molar aqueous solubility S (i.e. the saturating concentration of the solute within solvent in mol/L) based on knowledge of the Celsius melting point (MP) of the crystal and the octonal-water partition coefficient (K_{ow}) of the molecule

$$\log S = 0.5 - (MP - 25)/10 + \log K_{ow} \qquad \text{Equation 1}$$

The usefulness of the revised GSE in drug design settings where MP, K_{ow} or both are unknown is reduced. Therefore, practical tools for solubility prediction are currently limited to quantitative structure–property relationships (QSPR) and trained on experimental data[33–45] as discussed in recent reviews.[15,46–48]

To address the challenge of computing the structure, thermodynamics and solubility of organic crystals from first principles, this work brings together recent advances from three complementary fields including polarizable atomic multipole force field development, free energy simulation methods and algorithms for efficient treatment of long-range electrostatics inherent to lattice summation. First, the importance of a polarizable atomic multipole force field to accurate modeling of crystal energy landscapes[11] arises from aspherical features of the electron density inherent to lone pairs and π-electrons,[25] and from perturbations such as conformational flexibility (i.e. intra-molecular polarization)[26] or transfer between vapor, liquid and crystalline environments (i.e. inter-molecular polarization). These critical features of the crystal energy landscape are neglected by fixed

atomic charge force fields such as Amber[49], CHARMM[50] or OPLS-AA.[18] In this work, the AMOEBA force field captures aspherical atomic electron density using permanent atomic multipoles through quadrupole order and transferability between confirmations and phases using explicit treatment of electronic polarization via induced dipoles.[51–53] Second, large energy barriers[11] that necessitate advanced statistical mechanics theories to efficiently converge sampling of molecular configurations characterize crystal potential energy surfaces.[54–56] Here we overcome these large energy barriers using the orthogonal space random walk (OSRW) strategy.[56–58] On the other hand, traditional methods such as free energy perturbation[59] or thermodynamic integration[60] require knowledge of the crystal structure coordinates, while use of OSRW allows coordinate prediction. Finally, algorithms for long-range van der Waals and electrostatics forces such as particle-mesh Ewald (PME) summation[61–63] must leverage the space group symmetry of crystalline solids to achieve computational efficiently, which has recently been implemented within the program *Force Field X*[64] On the other hand, widely used molecular dynamics programs such as Amber[65], CHARMM[66] and TINKER[67] require explicit sampling of the entire unit cell, which both increases the cost per energy evaluation while also slowing the kinetics of crystal nucleation.

In the next section, background on the polarizable AMOEBA force field, solubility thermodynamic cycle and space group symmetry extension of PME are presented. This is followed by theory describing the OSRW free energy algorithm, including details of the derivatives of the AMOEBA potential required by OSRW and are a superset of the derivatives required for thermodynamic integration, which are presented here for the first time. The overall implementation of AMOEBA and OSRW in *Force Field X* is validated by comparison to monovalent ion solvation free energies computed using standard free energy perturbation. Prediction of the structure, thermodynamics and solubility of the *n*-alkylamides from acetamide to octanamide is then presented. The *n*-alkylamides series was chosen due to the dependence of its solubility trend on both amide hydrogen bonding and the hydrophobic effect, which are both of broad importance to biomolecular structure and function.

BACKGROUND

A. The Potential Energy Surface of Organic Crystals

The non-covalent interactions between molecules of an organic crystal can be decomposed into van der Waals and electrostatic forces. Until recently, the atomic resolution force fields used to study crystal energetics and liquid state properties were based on the approximation that the electrostatic charge of each atom is invariant, or fixed, with respect to its local environment.[68] In reality, the electron density deforms in response to the electric field of its environment. For example, negatively charged electron density moves toward the positive potential of a cation such as Na+ or K+ and away from the negative potential of an anion such as Cl−. The average electric field within a crystal is different from that experienced by a molecule when solvated or in vapor. In the past, this necessitated choosing between fixed atomic charges that were optimal for either the crystalline phase or the liquid phase, but not both. Fortunately, increasing computational power has facilitated explicit inclusion of electronic polarization into force fields for organic molecules and thereby improved their transferability between vacuum, liquid and crystalline phases.[52,69–75] In this work, the crystal potential energy landscape is defined by the polarizable Atomic Mulitpole Optimized Energetics for Biomolecular Applications (AMOEBA) force field[52] for organic molecules.[53,76] The AMOEBA potential is then sampled using the OSRW algorithm described below in order to compute the thermodynamic stability of crystals while also predicting their atomic coordinates.

B. Solubility Thermodynamic Cycle for Organic Crystals

Solubility defines the equilibrium concentration of a solute in contact with an excess of its solid state. The physical process reaches equilibrium when the rate of molecules joining the solid state (assumed to be crystalline in this work) from solution equals the rate of molecules leaving the solid to be solvated. Although the required system size and time scale for explicit atomic resolution simulation of the crystalline-solution phase equilibrium is prohibitive, the standard state solubility free energy ($\Delta G^{o}_{solubility}$) can be decomposed into the sum of standard state sublimation (ΔG^{o}_{sub}) and solvation (ΔG^{o}_{solv}) free energy steps

$$\Delta G^{o}_{solubility} = \Delta G^{o}_{sub} + \Delta G^{o}_{solv} \qquad \text{Equation 2}$$

as shown in the thermodynamic cycle of Figure 1. Although solvation free energies are routinely computed to high precision,[77,78] calculation of the sublimation free energy by annihilation of the entire crystalline solid is novel to this work and is depicted in Figure 2. A correction is applied to the simulation result to account for the change in volume between a crystalline solid containing 1 mole of molecules and the standard state concentration of 1 M chosen for both the vapor and solvated states.

C. Long-Range Forces in Organic Crystals

Evaluation of the polarizable AMOEBA electrostatic energy for a crystalline lattice builds on the theory for lattice summation proposed by Ewald.[79] The real space portion of the theory was described by Ren and Ponder in the context of the AMOEBA water model[80] and was based on earlier work by Smith.[81] The reciprocal space portion is accelerated by fast Fourier transforms (FFT) as described by Darden and colleagues for fixed charge[61,62] and multipolar[63] atomic electrostatics. Further acceleration of PME for crystals has been achieved via explicit support for space group symmetry,[64] which has been applied to biomolecular X-ray refinement.[64,82] Without explicit support for space group symmetry, the alchemical transformation shown in Figure 2 would need to consider n_{symm} molecules, where n_{symm} is the number of symmetry operators of the space group. For an acetamide crystal, which has 18 molecules in its unit cell (n_{symm}=18), the kinetics of nucleation for the alchemical transformation may be computationally prohibitive. In particular, the approximate 18x increase in the cost per energy evaluation is exacerbated by the need to convergence an 18x larger free energy change, which is accompanied by crystal nucleation kinetics.

METHODS

A. Orthogonal Space Random Walk

In order to effectively overcome large barriers in the crystalline free energy landscape, a second-order generalized ensemble method called the orthogonal-space random walk (OSRW) is applied.[56,57] The central idea of generalized ensemble (GE) free energy simulation methods is to employ a modified ensemble, which allows free energy barriers in the target ensemble to be crossed more frequently in order to efficiently sample the distributions required for free energy estimations. A first-order GE modification is based on the order parameter λ and the biasing energy term $f_m(\lambda)$ in the following target potential:

$$U_m = U_{AMOEBA}(\lambda) + f_m(\lambda) \qquad \text{Equation 3}$$

where λ=1 corresponds to the solute interacting with the condensed phase environment and λ=0 to the solute uncoupled from the condensed phase environment. In the first-order GE regime, the biasing term $f_m(\lambda)$ is commonly updated to approach $-G_{AMOEBA}(\lambda)$, which is

the negative of the λ-dependent free energy profile corresponding to the canonical ensemble with the AMOEBA potential energy function. In this way, an order parameter space random walk can be achieved to uniformly sample all the target states. To adaptively estimate $G_{AMOEBA}(\lambda)$, three major recursion schemes have been developed. The first is adaptive umbrella sampling,[83] in which free energy estimations are based on order parameter probability distributions. The second is the adaptive biasing force (ABF) method,[84–86] in which free energy estimations are based on the thermodynamic integration (TI) formula.[60,87] Finally, multiplicative approaches, such as the metadynamics method for molecular dynamics simulations[55] and the Wang-Landau method for Monte-Carlo simulations,[88] are realized through dynamic force-balancing.

In first-order GE simulations the free energy surface along λ is flattened, however, "hidden" free energy barriers may exist in the space perpendicular to λ. These "hidden" free energy barriers create a sampling challenge due to slow environmental relaxation. To address this, the second-order GE simulation method, the orthogonal space random walk (OSRW) algorithm, was introduced based on the following modified energy function:

$$U_m = U_{AMOEBA}(\lambda) + f_m(\lambda) + g_m(\lambda, F_\lambda) \quad \text{Equation 4}$$

where $g_m(\lambda, F_\lambda)$ is targeted towards $-G_{AMOEBA}(\lambda, F_\lambda)$ the negative of the free energy profile along (λ, F_λ) in the ensemble corresponding to the energy function $U_{AMOEBA}(\lambda) - G_{AMOEBA}(\lambda)$, and $F_\lambda = \partial U_{AMOEBA}(\lambda)\, \partial\lambda$ In contrast to first-order GE methods, OSRW requires two recursion components to be adaptively updated. The recursion component responsible for $g_m(\lambda, F_\lambda)$ is called the "recursion kernel" and the recursion component responsible for $f_m(\lambda)$ is called the "recursion slave" because its target $-G_{AMOEBA}(\lambda)$ can be determined by the target of $g_m(\lambda, F_\lambda)$. In our implementation, the recursion slave is determined using the TI relationship and metadynamics is used for the recursion kernel. However, the recursion kernel can be based on any of the three recursion methods discussed above.

The free energy biasing potential $g_m(\lambda, F_\lambda)$ can be obtained by repetitively adding a relatively small Gaussian-shaped repulsive potential:

$$he^{\left(-\frac{|\lambda-\lambda(t_i)|^2}{2w_1^2}\right)} e^{\left(-\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right)} \quad \text{Equation 5}$$

that is centered at the state $[\lambda(t_i), F_\lambda(t_i)]$ being visited at time t_i and thereby discourages the system from often visited configurations. Repeating this procedure builds up the biasing potential

$$g_m(\lambda, F_\lambda) = \sum_{t_i} he^{\left(-\frac{|\lambda-\lambda(t_i)|^2}{2w_1^2}\right)} e^{\left(-\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right)} \quad \text{Equation 6}$$

and eventually flattens the underlying curvature of the free energy surface along the (λ, F_λ) space. Then, as in traditional metadynamics simulations, the free energy profile along the reaction coordinate (λ, F_λ) will eventually converge to $-G_{AMOEBA}(\lambda, F_\lambda)$ and is estimated as $-g_m(\lambda, F_\lambda)$. For any state λ', the generalized force distribution should be proportional to $e^{\{\beta g_m[\lambda F_\lambda(\lambda)]\}}$ and the free energy derivative at each state is

$$\left.\frac{dG}{d\lambda}\right|_{\lambda'} = \langle F_\lambda(\lambda')\rangle = \frac{\int_{F_\lambda} F_\lambda e^{[\beta g_m(\lambda,F_\lambda)]}\delta(\lambda - \lambda')}{\int_{F_\lambda} e^{[\beta g_m(\lambda,F_\lambda)]}\delta(\lambda - \lambda')} \quad \text{Equation 7}$$

Following the thermodynamic integration formula, the free energy change between the initial state λ_i (for example λ_i=0) and the target state λ (for example λ=1) is given by

$$G_{AMOEBA}(\lambda) = \int_{\lambda_i}^{\lambda} \left.\frac{dG}{d\lambda}\right|_{\lambda'} d\lambda' \quad \text{Equation 8}$$

In our OSRW implementation, the metadynamics strategy described in Eq. 6 serves as the recursion kernel and the TI based formula (Eqs. 7 and 8) serves as the recursion slave with $f_m(\lambda)$ set recursively as the instantaneous estimation of $-G_{AMOEBA}(\lambda)$

Our first goal was to derive, implement and parallelize derivatives with respect to λ of the polarizable AMOEBA potential energy function $U(\lambda, \mathbf{X})$ where λ defines a continuous, softcore transition between the asymmetric unit (solute) fully interacting with its symmetry mates (condensed phase environment) to decoupled in order to compute the sublimation free energy. Analogous decoupling schemes are used for computing solvation free energy (i.e. a solute decoupled from solvent) and binding free energy (i.e. ligand decoupled from the binding pocket of a solvated protein). The terms needed for the various λ-based sampling methods may include $\partial U(\lambda, \boldsymbol{X})/\partial\lambda$, $\partial^2 U(\lambda, \mathbf{X})/\partial\lambda^2$ and/or $\partial^2 U(\lambda, \mathbf{X})/\partial\mathbf{X}\,\partial\lambda$. We briefly present key aspects of these derivatives with special emphasis on unique advancements required for the AMOEBA buffered 14-7 van der Waals softcore term, the permanent multipole softcore term and the polarization energy. Parallelization depends on *Parallel Java*, which is a unified API for shared memory and cluster parallel programming.[89]

B. AMOEBA Intermolecular Interactions and Their λ Derivatives

1. Softcore Buffered 14-7 van der Waals—The AMOEBA model for water,[80,90] ions,[91] organic molecules[53] and proteins[52,92] uses a buffered 14-7 van der Waals term proposed by Halgren.[93] A softcore form that allows the potential to smoothly disappear as λ goes to zero has been described previously[92] and is given by

$$U_{vdW}(\lambda, r) = \lambda^{\beta}\varepsilon_{ij} t_1 t_2 \quad \text{Equation 9}$$

where for convenience the expression has been factored into the following two terms

$$t_1 = \frac{(1+\delta)^{n-m}}{\alpha(1-\lambda)^2 + (\rho+\delta)^{n-m}} \quad \text{Equation 10}$$

$$t_2 = \frac{1+\gamma}{\alpha(1-\lambda)^2 + \rho^m + \gamma} - 2 \quad \text{Equation 11}$$

where ε_{ij} is the well-depth, $\rho = r/r'$ is the atomic separation distance r normalized by the minimum energy distance r', the parameters (n=14, m=7, δ=0.07, γ=0.12) specify a buffered 14-7 form (a 6-12 Lennard-Jones potential is recovered by n=12, m=6, δ=0 and γ=0) and the parameters α and β tune the softcore transition.[92] In the original work, the buffered 14-7 potential was completely turned on before polarizable atomic multipole electrostatics was

considered,[92] however, it is more convenient to modify both the van der Waals and electrostatics potentials concomitantly over the λ path. Adoption of the latter approach necessitated different α and β parameters, which are discussed further in the Results section below.

The various derivatives of the softcore buffered 14-7 interaction required for OSRW are now described. First, the partial derivative of U_{vdw} with respect to an atomic coordinate is given by

$$\frac{\partial U_{vdW}(\lambda, r)}{\partial r_{i,\gamma}} = \varepsilon_{ij}\lambda^{\beta}\left(\frac{\partial t_1}{\partial r_{i,\gamma}}t_2 + t_1\frac{\partial t_2}{\partial r_{i,\gamma}}\right) \quad \text{Equation 12}$$

where

$$\frac{\partial t_1}{\partial r_{i,\gamma}} = \frac{(n-m)(1+\delta)^{n-m}(\rho+\delta)^{n-m-1}}{r\,[\alpha(1-\lambda)^2+(\rho+\delta)^{n-m}]^2}\frac{r_\gamma}{r} \quad \text{Equation 13}$$

and

$$\frac{\partial t_2}{\partial r_{i,\gamma}} = \frac{m(1+\gamma)\rho^{m-1}}{r\,[\alpha(1-\lambda)^2+\rho^m+\gamma]^2}\frac{r_\gamma}{r} \quad \text{Equation 14}$$

Next, the partial derivative of U_{vdw} with respect to λ is shown

$$\frac{\partial U_{vdW}(\lambda, r)}{\partial \lambda} = \varepsilon_{ij}\left[\beta\lambda^{\beta-1}t_1t_2 + \lambda^{\beta}\left(\frac{\partial t_1}{\partial \lambda}t_2 + t_1\frac{\partial t_2}{\partial \lambda}\right)\right] \quad \text{Equation 15}$$

where

$$\frac{\partial t_1}{\partial \lambda} = \frac{2\alpha(1-\lambda)(1+\delta)^{n-m}}{[\alpha(1-\lambda)^2+(\rho+\delta)^{n-m}]^2} \quad \text{Equation 16}$$

and

$$\frac{\partial t_2}{\partial \lambda} = \frac{2\alpha(1-\lambda)(1+\gamma)}{[\alpha(1-\lambda)^2+\rho^m+\gamma]^2} \quad \text{Equation 17}$$

For the 2nd derivative of the U_{vdw} with respect to λ it is convenient to separate the previous first derivative of Equation 15 into a sum of three terms as

$$\frac{\partial U_{vdW}(\lambda, r)}{\partial \lambda} = \varepsilon_{ij}(f_1+f_2+f_3) \quad \text{Equation 18}$$

such that $\partial^2 U_{vdW}(\lambda, r)/\partial\lambda^2$ is then given by

$$\frac{\partial^2 U_{vdW}(\lambda, r)}{\partial \lambda^2} = \varepsilon_{ij}\left(\frac{\partial f_1}{\partial \lambda} + \frac{\partial f_2}{\partial \lambda} + \frac{\partial f_3}{\partial \lambda}\right) \quad \text{Equation 19}$$

where

$$\frac{\partial f_1}{\partial \lambda} = \beta(\beta-1)\lambda^{\beta-2} t_1 t_2 + \beta\lambda^{\beta-1}\left(\frac{\partial t_1}{\partial \lambda} t_2 + t_1 \frac{\partial t_2}{\partial \lambda}\right) \quad \text{Equation 20}$$

$$\frac{\partial f_2}{\partial \lambda} = \beta\lambda^{\beta-1}\frac{\partial t_1}{\partial \lambda} t_2 + \lambda^{\beta}\left(\frac{\partial^2 t_1}{\partial \lambda^2} t_2 + \frac{\partial t_1}{\partial \lambda}\frac{\partial t_2}{\partial \lambda}\right) \quad \text{Equation 21}$$

and

$$\frac{\partial f_3}{\partial \lambda} = \beta\lambda^{\beta-1} t_1 \frac{\partial t_2}{\partial \lambda} + \lambda^{\beta}\left(\frac{\partial t_1}{\partial \lambda}\frac{\partial t_2}{\partial \lambda} + t_1 \frac{\partial^2 t_2}{\partial \lambda^2}\right) \quad \text{Equation 22}$$

In the above expressions, only $\partial^2 t_1/\partial\lambda^2$ and $\partial^2 t_2/\partial\lambda^2$ remain to be defined and are given by

$$\frac{\partial^2 t_1}{\partial \lambda^2} = \frac{8\alpha^2(1-\lambda)^2(1+\delta)^{n-m}}{[\alpha(1-\lambda)^2+(\rho+\delta)^{n-m}]^3} - \frac{2\alpha(1+\delta)^{n-m}}{[\alpha(1-\lambda)^2+(\rho+\delta)^{n-m}]^2} \quad \text{Equation 23}$$

and

$$\frac{\partial^2 t}{\partial \lambda^2} = \frac{8\alpha^2(1-\lambda)^2(1+\gamma)}{[\alpha(1-\lambda)^2+\rho^m+\gamma]^3} - \frac{2\alpha(1+\gamma)}{[\alpha(1-\lambda)^2+\rho^m+\gamma]^2} \quad \text{Equation 24}$$

Finally, the 2nd partial derivative with respect to both an atomic coordinate and λ is given by

$$\frac{\partial^2 U_{vdW}(\lambda, r)}{\partial \lambda \partial r_{i,\gamma}} = \varepsilon_{ij}\left(\frac{\partial f_1}{\partial r_{i,\gamma}} + \frac{\partial f_2}{\partial r_{i,\gamma}} + \frac{\partial f_3}{\partial r_{i,\gamma}}\right) \quad \text{Equation 25}$$

where

$$\frac{\partial f_1}{\partial r_{i,\gamma}} = \beta\lambda^{\beta-1}\left(\frac{\partial t_1}{\partial r_{i,\gamma}} t_2 + t_1 \frac{\partial t_2}{\partial r_{i,\gamma}}\right) \quad \text{Equation 26}$$

$$\frac{\partial f_2}{\partial r_{i,\gamma}} = \lambda^{\beta}\left(\frac{\partial^2 t_1}{\partial \lambda \partial r_{i,\gamma}} t_2 + \frac{\partial t_1}{\partial \lambda}\frac{t_2}{\partial r_{i,\gamma}}\right) \quad \text{Equation 27}$$

and

$$\frac{\partial f_3}{\partial r_{i,\gamma}} = \lambda^{\beta}\left(\frac{t_1}{\partial r_{i,\gamma}}\frac{\partial t_2}{\partial \lambda} + t_1\frac{\partial^2 t_2}{\partial \lambda \partial r_{i,\gamma}}\right) \quad \text{Equation 28}$$

Only the terms $\partial^2 t_1/\partial\lambda\partial r_{i,\gamma}$ and $\partial^2 t_2/\partial\lambda\partial r_{i,\gamma}$ remain to be defined and are given by

$$\frac{\partial^2 t_1}{\partial\lambda\partial r_{i,\gamma}} = \frac{4\alpha(1-\lambda)(n-m)(1+\delta)^{n-m}(\rho+\delta)^{n-m-1}}{r\,[\alpha(1-\lambda)^2+(\rho+\delta)^{n-m}]^3}\frac{r_\gamma}{r} \quad \text{Equation 29}$$

and

$$\frac{\partial^2 t_2}{\partial\lambda\partial r_{i,\gamma}} = \frac{4\alpha(1-\lambda)m(1+\gamma)^{m-1}}{r^{m}\,[\alpha(1-\lambda)^2+\rho^m+\gamma]^3}\frac{r_\gamma}{r} \quad \text{Equation 30}$$

2. Softcore for Multipolar Real Space Ewald Summation—For real space Ewald electrostatic interactions between permanent multipoles, it is essential to formulate inclusion of softcore behavior in a manner that facilitates concise higher order tensors and partial derivatives. Our approach replaces the atomic separation distance $r = |\mathbf{r}_j - \mathbf{r}_i|$ by $f = \sqrt{r^2+\alpha(1-\lambda)^2}$ and the parameter α tunes the softcore transition. The softcore electrostatic interaction energy between permanent multipoles through quadrupole order at sites i and j is given by

$$U_{\text{real}}(\lambda, r) = \lambda^{\beta}\sum_{l=0}^{4} G_{ij}^{l}(r)B_l(f) \quad \text{Equation 31}$$

where the coefficients $G_{ij}^{l}(r)$ for each order l depend on the multipole coefficients of each site and were given by Smith.[81] The function $B_l(u)$ is defined by

$$B_0(u) = \text{erfc}(\xi u)/u \quad \text{Equation 32}$$

for the 0th order term based on the complementary error function

$$\text{erfc}(\xi u) = \frac{2}{\sqrt{\pi}}\int_{\xi u}^{\infty}\exp(-s^2)\,ds \quad \text{Equation 33}$$

and higher order $B_l(u)$ terms are obtained from the recursion relationship

$$B_l(u) = \frac{1}{u^2}\left[(2l-1)B_{l-1}(u) + \frac{(2\xi^2)^l}{\xi\sqrt{\pi}}\exp(-\xi^2u^2)\right] \quad \text{Equation 34}$$

The derivative of the permanent real space electrostatic energy with respect to an atomic coordinate $r_{i,\gamma}$ of atom i is given by

$$\frac{\partial \mathrm{U}_{\mathrm{real}}(\lambda, r)}{\partial r_{i,\gamma}} = \lambda^{\beta} \sum_{l=0}^{4} \frac{\partial G_{ij}^{l}(r)}{\partial r_{i,\gamma}} \mathrm{B}_{l}(f) + \mathrm{G}_{ij}^{l}(r) \frac{\partial \mathrm{B}_{l}(f)}{\partial r_{i,\gamma}} \quad \text{Equation 35}$$

where the partial derivatives for $\partial G_{ij}^{l}(r)/\partial r_{i,\gamma}$ were given by Smith[81] and

$$\frac{\partial \mathrm{B}_{l}(f)}{\partial r_{i,\gamma}} = r_{\gamma} \mathrm{B}_{l+1}(f) \quad \text{Equation 36}$$

The 1st derivative with respect to λ is

$$\frac{\partial \mathrm{U}_{\mathrm{real}}(\lambda, r)}{\partial \lambda} = \sum_{l=0}^{4} \mathrm{G}(r)_{ij}^{l} \left[\beta \lambda^{\beta-1} \mathrm{B}_{l}(f) + \lambda^{\beta} \frac{\partial \mathrm{B}_{l}(f)}{\partial \lambda} \right] \quad \text{Equation 37}$$

where it can be shown that

$$\frac{\partial \mathrm{B}_{l}(f)}{\partial \lambda} = \alpha(1-\lambda) \mathrm{B}_{l+1}(f) \quad \text{Equation 38}$$

and the 2nd derivative is given by

$$\partial^{2} \mathrm{U}_{\mathrm{real}}(\lambda, r)/\partial \lambda^{2} = \sum_{l=0}^{4} G_{ij}^{l}(r) \left[\beta(\beta-1) \lambda^{\beta-2} \mathrm{B}_{l}(f) + 2\beta \lambda^{\beta-1} \frac{\partial \mathrm{B}_{l}(f)}{\partial \lambda} + \lambda^{\beta} \frac{\partial^{2} \mathrm{B}_{l}(f)}{\partial \lambda^{2}} \right] \quad \text{Equation 39}$$

where

$$\frac{\partial^{2} \mathrm{B}_{l}(f)}{\partial \lambda^{2}} = \alpha^{2}(1-\lambda)^{2} \mathrm{B}_{l+2}(f) - \alpha \mathrm{B}_{l+1}(f) \quad \text{Equation 40}$$

Finally, the 2nd derivative of the softcore permanent multipole real space electrostatic energy with respect to an atomic coordinate $r_{i,\gamma}$ of atom i and λ is given by

$$\frac{\partial^{2} \mathrm{U}_{\mathrm{real}}(\lambda, r)}{\partial \lambda \partial r_{i,\gamma}} = \beta \lambda^{\beta-1} \sum_{l=0}^{4} \frac{\partial \mathrm{G}_{ij}^{l}(r)}{\partial r_{i,\gamma}} \mathrm{B}_{l}(f) + \mathrm{G}_{ij}^{l}(r) \frac{\partial \mathrm{B}_{l}(f)}{\partial r_{i,\gamma}} + \lambda^{\beta} \sum_{l=0}^{4} \frac{\partial \mathrm{G}_{ij}^{l}(r)}{\partial r_{i,\gamma}} \frac{\partial \mathrm{B}_{l}(f)}{\partial \lambda} + \mathrm{G}_{ij}^{l}(r) \frac{\partial^{2} \mathrm{B}_{l}(f)}{\partial \lambda \partial r_{i,\gamma}} \quad \text{Equation 41}$$

where

$$\frac{\partial^{2} \mathrm{B}_{l}(f)}{\partial \lambda \partial r_{i,\gamma}} = \alpha(1-\lambda) r_{\gamma} \mathrm{B}_{l+2}(f) \quad \text{Equation 42}$$

3. Permanent Reciprocal Space Ewald Summation—The reciprocal space portion of particle-mesh Ewald summation due to permanent multipoles is not pairwise and is therefore calculated by interpolating between the potential due to the total system $\varphi_{\mathrm{recip}}^{\mathrm{total}}$ (i.e. infinite crystal, solute plus solvent, or ligand plus protein/solvent) and that due to only the

solute $\varphi_{\text{recip}}^{\text{solute}}$ (i.e. asymmetric unit, solute or ligand). The latter calculation is significantly faster than the computation of the reciprocal potential due to the entire condensed phase system minus the solute $\varphi_{\text{recip}}^{\text{condensed}}$ (i.e. symmetry mates without the asymmetric unit, solvent without the solute or protein/solvent without the ligand), which can be determined from the relationship

$$\varphi_{\text{recip}}^{\text{condensed}}=\varphi_{\text{recip}}^{\text{total}}-\varphi_{\text{recip}}^{\text{solute}} \quad \text{Equation 43}$$

The permanent reciprocal space potential is then given by

$$\varphi_{\text{recip}}(\lambda)=(1-\lambda)\varphi_{\text{recip}}^{\text{condensed}}+\lambda\varphi_{\text{recip}}^{\text{total}}=\varphi_{\text{recip}}^{\text{total}}+(\lambda-1)\varphi_{\text{recip}}^{\text{solute}} \quad \text{Equation 44}$$

and its 1st derivative with respect to λ is

$$\frac{\partial\varphi_{\text{recip}}(\lambda)}{\partial\lambda}=\varphi_{\text{recip}}^{\text{total}}-\varphi_{\text{recip}}^{\text{condensed}}=\varphi_{\text{recip}}^{\text{solute}} \quad \text{Equation 45}$$

while the 2nd derivative vanishes. The dependence of the permanent multipole reciprocal space energy $U_{\text{recip}}(\lambda)$ on λ can be formulated as a function of the reciprocal space electrostatic potential $\varphi_{\text{recip}}(\lambda)$ to give $U_{\text{recip}}(\varphi_{\text{recip}}(\lambda))$, which makes it clear that the energy and coordinate gradient can be computed as usual. The first derivative of the permanent multipole reciprocal space energy with respect to λ is given by $\partial U_{\text{recip}}\left(\varphi_{\text{recip}}(\lambda)\right)/\partial\lambda=U_{\text{recip}}\left(\varphi_{\text{recip}}^{\text{solute}}\right)$, the 2nd derivative with respect to λ is zero and terms of the form $\partial^2 U_{\text{recip}}(\varphi_{\text{recip}}(\lambda))/\partial\lambda\partial r_{i,\gamma}$ are analogous to the atomic coordinate gradient.

4. Polarization Energy—The AMOEBA polarization model is based on induced dipoles at each atomic center, which are polarized by the field of the permanent multipoles as well as the field of the induced dipoles themselves.[53] Pauli repulsion balances favorable deformation of the electron cloud, which is captured by smearing the induced dipole as described by Thole[94] via

$$\rho=\frac{3a}{4\pi}e^{-au^3} \quad \text{Equation 46}$$

where a is a dimensionless constant that controls the degree of smearing and $u = r_{ij}/(\alpha_i\alpha_j)^{(1/6)}$ is an effective distance that depends on the polarizability α of each atom.

Calculation of the induction energy along the alchemical path is similar to the calculation of the reciprocal space potential due to permanent multipoles because it is not pairwise and interpolation is used. For small values of λ, solute (i.e. asymmetric unit) atoms frequently overlap with solvent (i.e. symmetry mate) atoms because of the softcore modification to solute-solvent van der Waals interactions smoothly and completely eliminates Pauli repulsion by $\lambda = 0$. Although the Thole model damps polarization at short range, the self-consistent field (SCF) calculation can require a larger number of iterations than is usual at unphysical, small atomic separations. Therefore, the contribution of the solute to condensed phase polarization is not considered until λ achieves a value greater than $\lambda_{\text{pol}}^{\text{start}}=\sim 0.7$. An additional benefit of this approach is that approximately ¾ of the alchemical path requires a single condensed phase SCF calculation, rather than two. As the SCF is the most

computationally expensive aspect of evaluating the AMOEBA potential energy and its derivatives, this saves approximately a factor of 2 with regards to the overall efficiency of the OSRW algorithm.

For the interval $\lambda=0\ldots\lambda_{pol}^{start}$, the vapor polarization energy for the solute $U_{pol\ (v)}^{solute}$ is calculated and summed with the condensed phase energy in the absence of the solute $U_{pol}^{condensed}$. Over the interval $\lambda_{pol}^{start}\leq\lambda\leq 1$, this sum is smoothly turned off based on shifting and scaling λ to give a state variable for polarization $\lambda_{pol}=\frac{\lambda-\lambda_{pol}^{start}}{1-\lambda_{pol}^{start}}$. Concomitantly, the total polarization energy of the condensed phase system is turned on

$$U_{pol}(\lambda)=\begin{cases} U_{pol\ (v)}^{solute}+U_{pol}^{condensed} & 0\leq\lambda\leq\lambda_{pol}^{start} \\ \lambda_{pol}^{3}U_{pol}^{Total}+(1-\lambda_{pol}^{3})(U_{pol\ (v)}^{solute}+U_{pol}^{condensed}) & \lambda_{pol}^{start}\leq\lambda\leq 1 \end{cases} \quad \text{Equation 47}$$

The cubic dependence on λ ensures that the 1st derivative of $U_{pol}(\lambda)$ with respect to λ is zero at λ_{pol}^{start}

$$\frac{\partial U_{pol}(\lambda)}{\partial\lambda}=\begin{cases} 0 & 0\leq\lambda\leq\lambda_{pol}^{start} \\ 3\lambda_{pol}^{2}(U_{pol}^{total}-U_{pol\ (v)}^{solute}-U_{pol}^{condensed}) & \lambda_{pol}^{start}\leq\lambda\leq 1 \end{cases} \quad \text{Equation 48}$$

as are the 2nd derivatives

$$\frac{\partial^{2}U_{pol}(\lambda)}{\partial\lambda^{2}}=\begin{cases} 0 & 0\leq\lambda\leq\lambda_{pol}^{start} \\ 6\lambda_{pol}(U_{pol}^{total}-U_{pol\ (v)}^{solute}-U_{pol}^{condensed}) & \lambda_{pol}^{start}\leq\lambda\leq 1 \end{cases} \quad \text{Equation 49}$$

and

$$\frac{\partial^{2}U_{pol}(\lambda)}{\partial\lambda\partial r_{i,\gamma}}=\begin{cases} 0 & 0\leq\lambda\leq\lambda_{pol}^{start} \\ 3\lambda_{pol}^{2}\left(\frac{\partial U_{pol}^{total}}{\partial r_{i,\gamma}}-\frac{\partial U_{pol\ (v)}^{solute}}{\partial r_{i,\gamma}}-\frac{\partial U_{pol}^{condensed}}{\partial r_{i,\gamma}}\right) & \lambda_{pol}^{start}\leq\lambda\leq 1 \end{cases} \quad \text{Equation 50}$$

C. Parameterization of Amide Functional Groups for AMOEBA

A 1981 study of acetamide structures from a neutron diffraction experiment at 23 K and *ab initio* molecular orbital studies in vacuum demonstrated that there are significant equilibrium bond length differences for both C=O and C-N as a function of environment due to hydrogen bonding.[95] One consequence is that the total dipole moment *in vacuum* of the equilibrium vacuum and crystalline structures differ by approximately 10%, as shown in Table 1. This difference is large enough that the solvation free energy for members of the *n*-alkylamide series is sensitive to the bond lengths used during parameterization of AMOEBA electrostatics. The sensitivity can be rationalized from the quadratic dependence of the solvation free energy on the total molecular dipole moment. For AMOEBA parameterization, the C-N and C=O bond lengths of the *n*-alkylamides were fixed at 1.34 Å and 1.23 Å, respectively. This protocol yielded a dipole moment for acetamide of 3.98 Debye, which is intermediate between that of unrestrained vacuum optimization (3.69

Debye) and that from fixing the C=O and C-N bond lengths at condensed phase values (4.07 to 4.17 Debye).

The lattice energy (E_{lattice}) for the AMOEBA acetamide model will be compared to periodic solid-state *ab initio* calculations based on the program suite CRYSTAL'09,[96,97] which uses functions localized at atoms as the basis for expansion of the crystalline orbitals via a linear combination of atomic orbitals (LCAO) technique. All-electron Gaussian type basis sets and the hybrid B3LYP[103,104] Hamiltonian was used. The DFT exchange–correlation contribution is evaluated by numerical integration over the unit cell volume. Radial and angular points of the grid were generated through Gauss–Legendre radial quadrature and Lebedev two-dimensional angular point distributions with a pruned grid of 75 radial and 974 angular points. The level of accuracy in evaluating the Coulomb and Hartree–Fock exchange series was controlled by five parameters,[97] and values of 7, 7, 7, 7, 16 were used. The reciprocal space integration was performed by sampling the Brillouin zone with the 6 × 6 × 6 Pack-Monkhorst net.[105] Structure optimizations were performed using analytical energy gradients with respect to atomic coordinates with cell parameters fixed,[106–108] within a quasi-Newton scheme combined with the Broyden–Fletcher–Goldfarb–Shanno scheme for Hessian updating.[109–112] Convergence was checked on both gradient components and nuclear displacements and was signaled when the RMS gradient was 0.00015 Hartree/Bohr and RMS displacement was 0.0006 Bohr. The 6-31G*[113,114] and 6-311G**[115] basis sets were used and condensed phase energies were corrected for BSSE via the counterpoise method.[116]

The lattice energy was computed as

$$E_{\text{lattice}} = \frac{E_{\text{unit}}}{Z} - E_{\text{mol}} \qquad \text{Equation 51}$$

where E_{unit} is the total energy of the unit cell, Z is the number of the molecules in the unit cell and E_{mol} is the total energy of acetamide optimized in vacuum. Observing the results in Table 2, the AMOEBA lattice energy for acetamide compares well to the *ab initio* values before the zero point energy (ZPE) was applied. After the ZPE is taken into account, AMOEBA overestimates the lattice energy by a little more than 2 kcal/mol at 0°K compared to DFT. The effect of this limitation on crystal thermodynamics near room temperature is difficult to project.

D. Previous AMOEBA Free Energy Perturbation Protocol

The first applications of the AMOEBA force field to the computation of free energy were limited to procedures that did not require the derivative of the potential energy with respect to λ, such as thermodynamic integration, which was not been completed until the present work. The original protocol, based on free energy perturbation (FEP), is described here for ion solvation and will be compared to OSRW in the Results section that follows.[91] For FEP, particle growth was completed with the charge and polarizability of each ion set to zero. During particle growth, the van der Waals radius (r) and well depth (ε) parameters were defined by

$$\mathrm{r}(\lambda) = 1 + \lambda(r_{\text{final}} - 1) \qquad \text{Equation 52}$$

and

$$\varepsilon(\lambda)=\lambda\varepsilon_{\text{final}} \quad \text{Equation 53}$$

respectively, for λ = (0.01,0.05,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0). During the subsequent charging stage, the charge (q) and polarizability (α) of each ion were defined by

$$\mathrm{q}(\lambda)=\lambda q_{\text{final}} \quad \text{Equation 54}$$

and

$$\alpha(\lambda)=\lambda\alpha_{\text{final}} \quad \text{Equation 55}$$

respectively, for λ = (0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0) where the first charging trajectory is equivalent to the last growth trajectory. For each λ value, 200 psec of constant volume molecular dynamics was performed using an 18.643 Å cubic box containing 216 AMOEBA water molecules. The Helmholtz free energy between adjacent λ values $\Delta\mathrm{A}(\lambda_i \rightarrow \lambda_{i+1})$ was calculated via half-steps as

$$\Delta\mathrm{A}\left(i \rightarrow i+\frac{1}{2}\right)=-\mathrm{k_B}T\ln\langle\exp\left(\frac{E_{\lambda_i}-E_{\lambda_{i+1/2}}}{\mathrm{k_B}T}\right)\rangle_{\lambda_i} \quad \text{Equation 56}$$

plus

$$\Delta\mathrm{A}\left(i+\frac{1}{2} \rightarrow i+1\right)=\mathrm{k_B}T\ln\langle\exp\left(\frac{E_{\lambda_{i+1}}-E_{\lambda_{i+1/2}}}{\mathrm{k_B}T}\right)\rangle_{\lambda_{i+1}} \quad \text{Equation 57}$$

using the last 150 psec of each trajectory.[91]

The standard state solvation free energy $\Delta G^{\mathrm{o}}_{\mathrm{solv}}$ (a relative free energy due to transferring the solute from vapor to solution phase) is independent of the solution concentration (C_s) as long as the standard state of the vapor phase is chosen to be the same as that of solution (e.g. 1 M), and the solution is dilute enough that solute molecules do not interact (no excess free energy of solvation). The standard state concentration $C°$ then cancels as shown below

$$\Delta G^{\mathrm{o}}_{\mathrm{solv}}=-RT\ln\left(\frac{C_s/C^{\mathrm{o}}}{C_v/C^{\mathrm{o}}}\right)=-RT\ln\left(\frac{C_s}{C_v}\right) \quad \text{Equation 58}$$

In this case, solvation free energy simulation results are essentially independent of the number of water molecules beyond a lower threshold. For example, the volume per molecule at 1M is 1660 $\mathrm{Å}^3$, while for the ion solvation simulations the volume per molecule was approximately 6500 $\mathrm{Å}^3$ and for the *n*-alkylamides discussed below it was 15000 $\mathrm{Å}^3$. In both cases, there are negligible interactions between solute periodic images.

RESULTS

A. Validation on Ion Solvation Free Energy

The first free energy calculations using the AMOEBA water model[80,90] focused on the solvation free energy of K+, Na+ and CI− ions and demonstrated very good agreement with experimental results for the solvation energy of KCl and NaCl salts.[91] Here we compare the

free energy methodology used in the initial ion solvation work, described above, with our OSRW approach in terms of accuracy and efficiency. The current OSRW protocol follows Grossfield *et al.* in terms of simulation box size (18.643 Å), number of water molecules (216), temperature (300 K), use of a Berendsen weak coupling thermostat (0.1 psec time constant),[117] time step (1.0 fsec), SCF convergence criteria (0.01 RMS Debye), van der Waals cutoff (12 Å) and ensemble (NVT), but differs in a few notable ways that are now described.[91] First, the current work uses of a softcore version of the buffered 14-7 van der Waals potential[92] as described above, which avoids the difficulties encountered in the original particle growth protocol as $\lambda \rightarrow 0$. Secondly, the original work used conventional Ewald summation[80] whereas the more efficient PME approach is used for the current AMOEBA/OSRW protocol as implemented in *Force Field X*[63,64] The PME parameters used throughout this work include a real space cutoff of 7.0 Å, an Ewald convergence parameter of 0.545, 5th order B-spline interpolation and a reciprocal grid spacing of 1.2 per Å. Finally, whereas the original protocol divided particle growth and charging into separate steps, our approach simultaneously performs 1.) growth of the van der Waals potential, 2.) charging of the condensed phase permanent electrostatic interactions and 3.) interpolation between vapor and condensed phase ligand polarization. Coupling ligand van der Waals growth with scaling up of electrostatic interactions between the ligand and condensed phase environment may be more efficient than separation into two stages. In addition, the latter approach does not require a vapor-discharging step for polyatomic ligands because intramolecular solute electrostatics are used throughout the protocol.

The ion solvation free energies under the AMOEBA/OSRW protocol are converged after only 0.5 nsec as shown in Figure 3. Furthermore, a plot of the cumulative free energy as a function of λ demonstrates smooth behavior, although small fluctuations in the ensemble average of $\partial U/\partial\lambda$ remain as shown in Figure 4. The original results of Grossfield *et al* were referenced to a vapor standard state of 1 atm. Therefore, the AMOEBA/OSRW single ion solvation results were corrected to account for the cost to compress an ideal gas to 1M, which is given by

$$T\Delta S = -k_{\mathrm{B}}T\ln\left(\frac{V_{1\mathrm{M}}}{V_{1\mathrm{atm}}}\right) = -k_B T\ln\left(\frac{1}{24.46}\right) = 3.197\cdot k_B T \qquad \text{Equation 59}$$

where k_{B} is the Boltzmann constant, T is absolute temperature and the resulting correction is 1.9 kcal/mol.

In Table 3, we compare the AMOEBA/OSRW ion solvation energies for K+, Na+ and Cl− to the original values for AMOEBA, OPLS-AA and CHARMM27 that were computed using FEP by Grossfield *et al.*[91] In each case, the AMOEBA/OSRW result is approximately 1 kcal/mol more positive than the original calculation using FEP. This is partially explained by the lack of a softcore definition of ion-solvent van der Waals interactions in the original work, which is consistent with underestimating the cost to form a cavity. This effect cancels for the relative solvation free energy between K+ and Na+ such that OSRW and FEP agree to within 0.1 kcal/mol. For the whole salts, the original FEP solvation energies were too negative by approximately 0.5 kcal/mol, while our results with OSRW are too positive by 1.2–1.5 kcal/mol. However, the original conclusions of Grossfield *et al* are unchanged.

B. Structure, Thermodynamics and Solubility of *n*-Alkylamides

1. Experimental *n*-Alkylamide Unit Cell Parameters, Solvation Free Energy and Standard State Solubility—

In this work we limit ourselves to canonical ensemble crystal thermodynamics based on the experimentally determined space group and unit cell parameters for the *n*-alkylamides as shown in Table 4.[101,12–24] The space group for each *n*-

alkylamide is $P2_1/c$,[121] except for the rhombohedral structure of acetamide.[101] For propanamide through octanamide, the length of *a* increases monotonically with chain length, while *b* and *c* are nearly constant.[121] Atomic coordinates for acetamide,[101] propanamide,[122] butanamide,[124] and pentanamide[123] have been solved and deposited in the Cambridge Crystallography Database permitting comparisons with the lowest energy structures found during this work. Although experimental structures for hexanamide, heptanamide and octanamide are currently unknown, we present predictions below as an illustration that the current approach yields structural information in addition to thermodynamics.

In addition to structural information, experimental values for the melting point[121] and solubility[125] for *n*-alkylamides have been reported and are summarized in Table 5. Furthermore, the solvation free energy of acetamide (−9.68 kcal/mol) and propanamide (−9.38 kcal/mol) have been determined in the context of studying amino acid side-chain analogues.[126] We note that the melting points of acetamide and propanamide are lower by 20°K than those for the longer chain length *n*-alkylamides. Experimental solubility values for acetamide, butanamide, pentanamide and hexanamide are available from the *Handbook of Aqueous Solubility Data,*[125] although uncertainty estimates were not provided. Discrepancies between the saturating concentrations of acetamide from three sources suggest errors on the order of ~0.5 kcal/mol.[125]

2. *n*-Alkylamide Solubility from Sublimation and Solvation Free Energies—The sublimation free energy for each *n*-alkylamide was calculated based on sampling the AMOEBA potential for organic molecules[53] using OSRW[56–58] by alchemically transforming a single molecule in vapor ($\lambda = 0$) into an infinite, perfectly symmetric crystal ($\lambda = 1$) via strict enforcement of space group symmetry[64] within the *Force Field X* program.[64,82,129] Each *n*-alkylamide molecule repeatedly sampled the minimum crystalline free energy basin ($\lambda = 1$) over the course of the simulation, as well as alternative conformations relatively higher in free energy. Each initial *n*-alkylamide conformation was a vacuum minima used for parameterization of the AMOEBA model in combination with an initial λ value of 0. Any reasonable starting confirmation produces identical thermodynamics once convergence has been reached. The sublimation simulations employed the parameters described above for ion solvation, with one notable difference. Unlike the ion and *n*-alkylamide solvation free energy calculations that used molecular dynamics, sublimation simulations used stochastic dynamics due to the small number of atoms being simulated.

Each *n*-alkylamide was fully flexible at all points in the simulation such that conformations or orientations that were not appropriate for the crystalline phase create a large force on the metadynamics particle favoring a return to vapor. Once in vapor, the molecule can easily change conformation and/or orientation. When a conformation is sampled in vapor that is consistent with the crystalline phase, a term in the force applied to the metadynamics particle ($-\partial U/\partial \lambda$) favors changing into condensed phase. Once in an energetically favorable crystalline state, the time dependent biasing potential builds up to allow the simulation to return to the vapor state. Before convergence is reached, each *n*-alkylamide sampled the correct crystalline conformation tens of times, in addition to exhaustively sampling torsional and rigid body degrees of freedom in vapor.

Based on the results shown in Table 6, the standard deviation for the sublimation free energy of each *n*-alkylamide averages only 0.25 kcal/mol after 100 nsec simulations for acetamide through hexanamide and 200 nsec simulations for heptanamide and octanamide. The sublimation free energies increase monotonically for propanamide through octanamide, which each share the P $2_1/c$ space group. The rhombohedral acetamide crystal is not

expected to follow the trend of the larger *n*-alkylamides, since its crystal packing is substantially different from the others.

The standard state solvation free energies for each *n*-alkylamides are given in Table 7 and were calculated based on 5 nsec simulations with the AMOEBA model for water[80,90] and organic molecules.[53] The OSRW[56–58] sampling method was used to alchemically solvate a single molecule in vapor ($\lambda = 0$) into a cubic box of water ($\lambda = 1$) using the *Force Field X* program. The simulation system was equilibrated with 500 waters within a cubic box of 24.662 Å. All other simulation parameters were changed from those reported for ion solvation. As for the sublimation free energy simulations, the initial conformation of each *n*-alkylamides was a vacuum energy minimum used for parameterization in tandem with an initial λ value of 0. The calculated standard state solvation free energies of acetamide and propanamide of −10.3 and −8.3 kcal/mol, respectively, agree with experiment to within 0.8 kcal/mol on average. We note that mean standard deviation for the solvation free energies of 0.3 kcal/mol is comparable to that for the sublimation free energies.

From the thermodynamic cycle depicted in Figure 1, summing the sublimation free energies given in Table 6 with the solvation free energies given in Table 7 yields the standard state solubility free energies as reported in Table 8. For acetamide through butanamide, the standard state sublimation and solvation free energy values nearly cancel, which yields absolute standard state solubility free energy values near zero. For the *n*-alkylamides from pentanamide through octanamide, the combined trends of increasingly stable crystals coupled with decreasingly favorable solvation due to the hydrophobic effect leads to increasingly lower solubility, as shown in Figure 5. The reason for the relatively lower solubility predicted for hexanamide by our method remains unclear. Until further studies can be completed, possible explanations include experimental uncertainty, force field accuracy or undetected sampling difficulties for the larger molecules.

3. *n*-Alkylamide Crystal Structures—As the OSRW procedure initially approaches the crystalline state (i.e. as λ approaches 1), conformations that are sampled will likely be in the free energy minimum region. However, there may exist hidden free energy barriers that promote sampling of conformations whose $\partial U/\partial\lambda$ values are inconsistent with the free energy minimum. In either case, the OSRW procedure will build up repulsive Gaussian potentials that drive $\partial U/\partial\lambda$ toward a random walk and thereby accelerate convergence of crystalline conformations. As a validation of the OSRW sampling procedure and the AMOEBA force field, samples were saved every 0.1 nsec that were subsequently completely optimized. The lowest potential energy structures for acetamide through pentanamide are shown in Figures 6–9, respectively. Except in the case of butanamide, the structure of the AMOEBA potential energy minimum agrees with the experimental X-ray structure. However, the AMOEBA models are more accurate with regards to the placement of hydrogen atoms, which results from limitations in the isolated atom scattering model (IAM) used in many X-ray structure refinement programs.[130]

DISCUSSION

A. AMOEBA Transferability Between Vacuum, Liquid and Solid Phases

An important motivation for developing polarizable force fields is to improve transferability between vacuum, liquid and solid phases and between environments with vastly different dielectric constants. [52,68,72] The inherent dependence of solubility on the equilibrium between a molecule in a solid crystalline phase and solvated in a liquid phase provides a test of the transferability of the AMOEBA force field.[53,80] The mean signed error between the experimental and calculated absolute standard state solubility values for the *n*-alkylamides

of 1.1 kcal/mol (Table 8) is a promising first step toward the prediction of solubility from first principles to chemical accuracy.

B. Efficient Computation of Free Energies via OSRW

Advancements in potential energy functions require further work to integrate them with existing alchemical free energy methods such as thermodynamic integration, meta-dynamics[55,131] and OSRW.[56–58] In this work, we have presented derivatives of the AMOEBA potential that are required for OSRW, which include $\partial U(\lambda, \mathbf{X})/\partial\lambda$, $\partial U^2(\lambda, \mathbf{X})/\partial\lambda^2$ and $\partial U^2(\lambda, \mathbf{X})/\partial\mathbf{X}\partial\lambda$. In particular, our approach includes softcore modifications to the buffered 14-7 van der Waals term and to the real space portion of Ewald summation in order to facilitate a smooth transition between full strength intermolecular interactions and non-interacting. As demonstrated, an approach that allows simultaneous sampling of the λ and $\partial U(\lambda, \mathbf{X})/\partial\lambda$ spaces is very effective for sampling the rough free energy landscape of the deposition/sublimation phase transition.

C. Deposition/Sublimation Phase Transitions in *Force Field X*

To the best of our knowledge, *Force Field X*[64,82,129] is the first and only implementation of an algorithm that permits calculation of the free energy of the deposition/sublimation phase transition from first principles. This requires two layers of complexity beyond the implementation of AMOEBA in other widely used computer codes such as TINKER,[67] OpenMM[132] and Amber.[65] First, explicit support for space group symmetry and replicates of the unit cell is necessary,[64] otherwise the complete annihilation of the crystalline phase becomes untenable. Consider the case of acetamide, which includes 18 copies of the molecule (i.e. the asymmetric unit) within the unit cell and 144 copies within a replicated super cell large enough to allow application of the minimum image convention. The 1 or 2 orders of magnitude increase in the degrees of freedom not only increases the cost of a single energy and gradient evaluation, but also exponentially increases the complexity of the crystalline free energy surface. Second, the crystalline annihilation algorithm in *Force Field X* depends on a meta-dynamics sampling algorithm in order to escape the large crystalline free energy barriers. Therefore, derivatives of the polarizable multipole AMOEBA potential energy function such as $\partial U(\lambda, \mathbf{X})/\partial\lambda$, $\partial^2 U(\lambda, \mathbf{X})/\partial\lambda^2$ and $\partial^2 U(\lambda, \mathbf{X})/\partial\mathbf{X}\partial\lambda$ have been implemented not only for P1 unit cells, but also for the asymmetric unit of each of the 230 space groups.

Acknowledgments

We thank Drs. Jeff Blaney and JW Feng for helpful discussions regarding the importance of crystal thermodynamics and solubility to the pharmaceutical industry, which provided the original motivation for this work. Assistance from Dr. Johnny Wu to parameterize the *n*-alkylamides using the *Poltype* automated parameterization tool and from Dr. Bartolomeo Civalleri to use the CRYSTAL09 software is gratefully acknowledged. We appreciate a Burroughs Wellcome Fund 2012 Collaborative Research Travel Grant to M.S. to support the collaboration between the Ren and Yang labs. The National Institute of General Medical Sciences (R01GM079686) and the Robert A. Welch Foundation (F-1691) for support (P.R.). This contribution was also supported by grants MCB 0919983 (W.Y.) and AGS-0927944 (J.B.) from the National Science Foundation (NSF) and by grant UL1RR024979 (J.B.) from the National Center for Research Resources, a part of the National Institutes of Health. TeraGrid and the Texas Advanced Computing Center provided high performance computing resources supported by NSF grant number TG-MCB100057.

References

1. Blagden N, de Matas M, Gavan PT, York P. Advanced Drug Delivery Reviews. 2007; 59:617. [PubMed: 17597252]
2. James SL. Chemical Society Reviews. 2003; 32:276. [PubMed: 14518181]

3. Furukawa H, Ko N, Go YB, Aratani N, Choi SB, Choi E, Yazaydin AO, Snurr RQ, O'Keeffe M, Kim J, Yaghi OM. Science. 2010; 329:424. [PubMed: 20595583]
4. Matsumoto M, Saito S, Ohmine I. Nature. 2002; 416:409. [PubMed: 11919626]
5. Haisa M, Kashino S, Maeda H. Acta Crystallographica Section B. 1974; 30:2510.
6. Haisa M, Kashino S, Kawai R, Maeda H. Acta Crystallographica Section B. 1976; 32:1283.
7. Wheatley PJ. Journal of the Chemical Society. 1964:6036.
8. Vishweshwar P, McMahon JA, Oliveira M, Peterson ML, Zaworotko MJ. Journal of the American Chemical Society. 2005; 127:16802. [PubMed: 16316223]
9. Terkeltaub R. Nature Reviews Rheumatology. 2010; 6:30.
10. Orson WM. The Lancet. 2006; 367:333.
11. Price SL. Physical Chemistry Chemical Physics. 2008; 10:1996. [PubMed: 18688351]
12. Day GM, Cooper TG, Cruz-Cabeza AJ, Hejczyk KE, Ammon HL, Boerrigter SXM, Tan JS, Della Valle RG, Venuti E, Jose J, Gadre SR, Desiraju GR, Thakur TS, van Eijck BP, Facelli JC, Bazterra VE, Ferraro MB, Hofmann DWM, Neumann MA, Leusen FJJ, Kendrick J, Price SL, Misquitta AJ, Karamertzanis PG, Welch GWA, Scheraga HA, Arnautova YA, Schmidt MU, van de Streek J, Wolf AK, Schweizer B. Acta Crystallographica Section B. 2009; 65:107.
13. Karamertzanis PG, Kazantsev AV, Issa N, Welch GWA, Adjiman CS, Pantelides CC, Price SL. Journal of Chemical Theory and Computation. 2009; 5:1432.
14. Price, SL.; Price, LS. Solid State Characterization of Pharmaceuticals. John Wiley & Sons, Ltd; 2011. Computational Polymorph Prediction; p. 427
15. Jorgensen WL, Duffy EM. Advanced Drug Delivery Reviews. 2002; 54:355. [PubMed: 11922952]
16. Lii JH, Allinger NL. Journal of the American Chemical Society. 1989; 111:8576.
17. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. Journal of the American Chemical Society. 1996; 118:2309.
18. Jorgensen WL, Maxwell DS, Tirado-Rives J. Journal of the American Chemical Society. 1996; 118:11225.
19. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. The Journal of Physical Chemistry B. 1998; 102:3586.
20. Jorgensen WL, Tirado-Rives J. Proceedings of the National Academy of Sciences. 2005; 102:6665.
21. Stone AJ, Alderton M. Molecular Physics. 1985; 56:1047.
22. Stone, AJ. The Theory of Intermoleular Forces. Vol. 32. Clarendon Press; Oxford: 1996.
23. Stone AJ. Journal of Chemical Theory and Computation. 2005; 1:1128.
24. Stone AJ. Science. 2008; 321:787. [PubMed: 18687950]
25. Price SL. Journal of the Chemical Society, Faraday Transactions. 1996; 92:2997.
26. Ouvrard C, Price SL. Crystal Growth & Design. 2004; 4:1119.
27. Civalleri B, Zicovich-Wilson CM, Valenzano L, Ugliengo P. CrystEngComm. 2008; 10:405.
28. Bardwell DA, Adjiman CS, Arnautova YA, Bartashevich E, Boerrigter SXM, Braun DE, Cruz-Cabeza AJ, Day GM, Della Valle RG, Desiraju GR, van Eijck BP, Facelli JC, Ferraro MB, Grillo D, Habgood M, Hofmann DWM, Hofmann F, Jose KVJ, Karamertzanis PG, Kazantsev AV, Kendrick J, Kuleshova LN, Leusen FJJ, Maleev AV, Misquitta AJ, Mohamed S, Needs RJ, Neumann MA, Nikylov D, Orendt AM, Pal R, Pantelides CC, Pickard CJ, Price LS, Price SL, Scheraga HA, van de Streek J, Thakur TS, Tiwari S, Venuti E, Zhitkov IK. Acta Crystallographica Section B. 2011; 67:535.
29. Jorgensen WL, Duffy EM. Bioorganic & Medicinal Chemistry Letters. 2000; 10:1155. [PubMed: 10866370]
30. Lüder K, Lindfors L, Westergren J, Nordholm S, Persson R, Pedersen M. Journal of Computational Chemistry. 2009; 30:1859. [PubMed: 19115279]
31. Yalkowsky SH, Valvani SC. Journal of Pharmaceutical Sciences. 1980; 69:912. [PubMed: 7400936]

32. Jain N, Yalkowsky SH. Journal of Pharmaceutical Sciences. 2001; 90:234. [PubMed: 11169540]
33. Kühne R, Ebert RU, Kleint F, Schmidt G, Schüürmann G. Chemosphere. 1995; 30:2061.
34. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Advanced Drug Delivery Reviews. 1997; 23:3.
35. Mitchell BE, Jurs PC. Journal of Chemical Information and Computer Sciences. 1998; 38:489.
36. Abraham MH, Le J. Journal of Pharmaceutical Sciences. 1999; 88:868. [PubMed: 10479348]
37. Huuskonen J. Journal of Chemical Information and Computer Sciences. 2000; 40:773. [PubMed: 10850781]
38. Meylan W, Howard P. Perspectives in Drug Discovery and Design. 2000; 19:67.
39. Klopman G, Zhu H. Journal of Chemical Information and Computer Sciences. 2001; 41:439. [PubMed: 11277734]
40. Liu R, So SS. Journal of Chemical Information and Computer Sciences. 2001; 41:1633. [PubMed: 11749590]
41. McElroy NR, Jurs PC. Journal of Chemical Information and Computer Sciences. 2001; 41:1237. [PubMed: 11604023]
42. Tetko IV, Tanchuk VY, Kasheva TN, Villa AEP. Journal of Chemical Information and Computer Sciences. 2001; 41:1488. [PubMed: 11749573]
43. Yan A, Gasteiger J. Journal of Chemical Information and Computer Sciences. 2002; 43:429. [PubMed: 12653505]
44. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA. Bioorganic & Medicinal Chemistry. 2008; 16:7944. [PubMed: 18701302]
45. Huuskonen J, Livingstone DJ, Manallack DT. SAR and QSAR in Environmental Research. 2008; 19:191. [PubMed: 18484495]
46. John SD. Drug Discovery Today. 2005; 10:289. [PubMed: 15708748]
47. Wang J, Hou T. Combinatorial Chemistry & High Throughput Screening. 2011; 14:328. [PubMed: 21470182]
48. Mitchell JBO. Future Medicinal Chemistry. 2011; 3:451. [PubMed: 21452981]
49. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. Journal of the American Chemical Society. 1996; 118:2309.
50. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. Journal of Physical Chemistry B. 1998; 102:3586.
51. Ren P, Ponder JW. Journal of Computational Chemistry. 2002; 23:1497. [PubMed: 12395419]
52. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. The Journal of Physical Chemistry B. 2010; 114:2549. [PubMed: 20136072]
53. Ren P, Wu C, Ponder JW. Journal of Chemical Theory and Computation. 2011; 7:3143. [PubMed: 22022236]
54. Kong X, Brooks CL III. The Journal of Chemical Physics. 1996; 105:2414.
55. Laio A, Parrinello M. Proceedings of the National Academy of Sciences. 2002; 99:12562.
56. Zheng L, Chen M, Yang W. Proceedings of the National Academy of Sciences. 2008; 105:20227.
57. Zheng L, Chen M, Yang W. The Journal of Chemical Physics. 2009; 130:234105. [PubMed: 19548709]
58. Lee S, Chen M, Yang W, Richards NGJ. Journal of the American Chemical Society. 2010; 132:7252. [PubMed: 20446682]
59. Zwanzig RW. The Journal of Chemical Physics. 1954; 22:1420.
60. Kirkwood JG. The Journal of Chemical Physics. 1935; 3:300.
61. Darden T, York D, Pedersen L. The Journal of Chemical Physics. 1993; 98:10089.
62. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. The Journal of Chemical Physics. 1995; 103:8577.

63. Sagui C, Pedersen LG, Darden TA. The Journal of Chemical Physics. 2004; 120:73. [PubMed: 15267263]
64. Schnieders MJ, Fenn TD, Pande VS. Journal of Chemical Theory and Computation. 2011; 7:1141.
65. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. Journal of Computational Chemistry. 2005; 26:1668. [PubMed: 16200636]
66. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. Journal of Computational Chemistry. 2009; 30:1545. [PubMed: 19444816]
67. Ponder, JW. TINKER: Software Tools for Molecular Design. 5.0. Saint Louis, MO: p. 2009
68. Ponder, JW.; Case, DA. Advances in Protein Chemistry;. Vol. 66. Academic Press; 2003. Force fields for protein simulations; p. 27
69. Maple JR, Cao YX, Damm WG, Halgren TA, Kaminski GA, Zhang LY, Friesner RA. Journal of Chemical Theory and Computation. 2005; 1:694.
70. Patel S, Brooks CL. Molecular Simulation. 2006; 32:231.
71. Piquemal JP, Chevreau H, Gresh N. Journal of Chemical Theory and Computation. 2007; 3:824.
72. Lopes PEM, Roux B, MacKerell AD. Theoretical Chemistry Accounts. 2009; 124:11. [PubMed: 20577578]
73. Holt A, Karlström G. International Journal of Quantum Chemistry. 2009; 109:1255.
74. Wang J, Cieplak P, Li J, Hou T, Luo R, Duan Y. The Journal of Physical Chemistry B. 2011; 115:3091. [PubMed: 21391553]
75. Wang J, Cieplak P, Li J, Wang J, Cai Q, Hsieh M, Lei H, Luo R, Duan Y. The Journal of Physical Chemistry B. 2011; 115:3100. [PubMed: 21391583]
76. Wu JC, Chattre G, Ren P. Theoretical Chemistry Accounts. 2012 (inpress).
77. Shirts MR, Pitera JW, Swope WC, Pande VS. The Journal of Chemical Physics. 2003; 119:5740.
78. Klimovich P, Mobley D. Journal of Computer-Aided Molecular Design. 2010; 24:307. [PubMed: 20372973]
79. Ewald PP. Annalen der Physik. 1921; 369:253.
80. Ren P, Ponder JW. The Journal of Physical Chemistry B. 2003; 107:5933.
81. Smith W. CCP5 Information Quaterly. 1982; 4
82. Fenn TD, Schnieders MJ. Acta Crystallographica Section D. 2011; 67:957.
83. Bartels C, Karplus M. Journal of Computational Chemistry. 1997; 18:1450.
84. Darve E, Pohorille A. The Journal of Chemical Physics. 2001; 115:9169.
85. Bitetti-Putzer R, Yang W, Karplus M. Chemical Physics Letters. 2003; 377:633.
86. Swendsen RH, Fasnacht M, Rosenberg JM. Computer Physics Communications. 2005; 169:274.
87. Carter EA, Ciccotti G, Hynes JT, Kapral R. Chemical Physics Letters. 1989; 156:472.
88. Wang F, Landau DP. Physical Review Letters. 2001; 86:2050. [PubMed: 11289852]
89. Kaminsky, A. Parallel Java: A unified API for shared memory and cluster parallel programming in 100% Java “. Parallel and Distributed Processing Symposium; 2007; Long Beach, CA.
90. Ren P, Ponder JW. The Journal of Physical Chemistry B. 2004; 108:13427.
91. Grossfield A, Ren PY, Ponder JW. Journal of the American Chemical Society. 2003; 125:15671. [PubMed: 14664617]
92. Jiao D, Golubkov PA, Darden TA, Ren P. Proceedings of the National Academy of Sciences. 2008; 105:6290.
93. Halgren TA. Journal of the American Chemical Society. 1992; 114:7827.
94. Thole BT. Chemical Physics. 1981; 59:341.
95. Jeffrey GA, Ruble JR, McMullan RK, DeFrees DJ, Pople JA. Acta Crystallographica Section B. 1981; 37:1885.
96. Dovesi R, Orlando R, Civalleri B, Roetti C, Saunders VR, Zicovich-Wilson CM. Zeitschrift für Kristallographie. 2005; 220:571.

97. Dovesi, R.; Saunders, VR.; Roetti, R.; Orlando, R.; Zicovich-Wilson, CM.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, NM.; Bush, IJ.; D'Arco, P.; Llunell, M. CRYSTAL09 User's Manual. University of Torino; Torino: 2009. CRYSTAL09 User's Manual
98. Kitano M, Kuchitsu K. Bulletin of the Chemical Society of Japan. 1973; 46:3048.
99. Ottersen T. Acta Chemica Scandinavica. 1975; A29:939.
100. Zobel D, Luger P, Dreissig W, Koritsanszky T. Acta Crystallographica Section B. 1992; 48:837.
101. Denne WA, Small RWH. Acta Crystallographica Section B. 1971; 27:1094.
102. Hamilton W. Acta Crystallographica. 1965; 18:866.
103. Lee C, Yang W, Parr RG. Physical Review B. 1988; 37:785.
104. Becke AD. The Journal of Chemical Physics. 1993; 98:5648.
105. Monkhorst HJ, Pack JD. Physical Review B. 1976; 13:5188.
106. Civalleri B, D'Arco P, Orlando R, Saunders VR, Dovesi R. Chemical Physics Letters. 2001; 348:131.
107. Doll K. Computer Physics Communications. 2001; 137:74.
108. Doll K, Saunders VR, Harrison NM. International Journal of Quantum Chemistry. 2001; 82:1.
109. Shanno DF. Mathematics of Computation. 1970; 24:647.
110. Goldfarb D. Mathematics of Computation. 1970; 24:23.
111. Fletcher R. The Computer Journal. 1970; 13:317.
112. Broyden CG. IMA Journal of Applied Mathematics. 1970; 6:76.
113. Holst M, Saied F. Journal of Computational Chemistry. 1993; 14:105.
114. Gatti C, Saunders VR, Roetti C. The Journal of Chemical Physics. 1994; 101:10686.
115. Krishnan R, Binkley JS, Seeger R, Pople JA. The Journal of Chemical Physics. 1980; 72:650.
116. Boys SF, Bernardi F. Molecular Physics. 1970; 19:553.
117. Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, Haak JR. Journal of Chemical Physics. 1984; 81:3684.
118. Schmid R, Miah AM, Sapunov VN. Physical Chemistry Chemical Physics. 2000; 2:97.
119. Friedman, HL.; Krishnan, CV. Thermodynamics of Ion Hydration. In: Franks, F., editor. Water: A Comprehensive Treatise;. Vol. 3. Plenum Press; New York: 1973.
120. Tissandier MD, Cowen KA, Feng WY, Gundlach E, Cohen MH, Earhart AD, Coe JV, Turtle TR. The Journal of Physical Chemistry A. 1998; 102:7787.
121. Turner JD, Lingafelter EC. Acta Crystallographica. 1955; 8:549.
122. Usanmaz A, Adler G. Acta Crystallographica Section B. 1982; 38:660.
123. Kopylovich MN, Kukushkin VY, Haukka M, Fraústo da Silva JJR, Pombeiro AJL. Inorganic Chemistry. 2002; 41:4798. [PubMed: 12206707]
124. Lewis TC, Tocher DA. Acta Crystallographica Section E. 2005; 61:1985.
125. Yalkowsky, SH.; He, Y. Handbook of Aqueous Solubility Data;. CRC Press; 2003.
126. Wolfenden R, Andersson L, Cullis PM, Southgate CCB. Biochemistry. 1981; 20:849. [PubMed: 7213619]
127. Romero C, González M. Journal of Thermal Analysis and Calorimetry. 2008; 92:705.
128. Romero CM, Gonzalez ME. Journal of Chemical & Engineering Data. 2009; 55:2326.
129. Fenn TD, Schnieders MJ, Brunger AT. Acta Crystallographica Section D. 2010; 66:1024.
130. Schnieders MJ, Fenn TD, Pande VS, Brunger AT. Acta Crystallographica Section D. 2009; 65:952.
131. Barducci A, Bonomi M, Parrinello M. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2011; 1:826.
132. Eastman P, Pande VS. Journal of Computational Chemistry. 2010; 31:1268. [PubMed: 19847780]

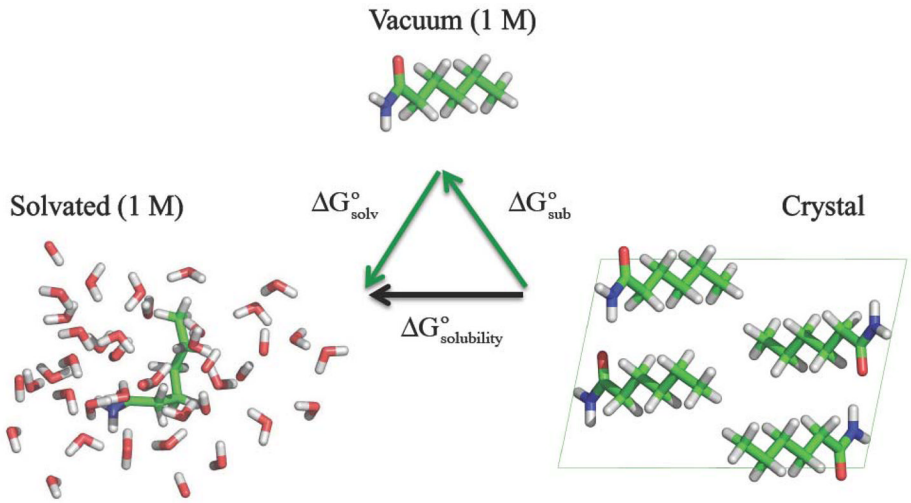

Figure 1.

This hexanamide thermodynamic cycle demonstrates that standard state solubility free energy ($\Delta G^{o}_{solubility}$) can be decomposed into the sum of standard state sublimation (ΔG^{o}_{sub}) and solvation (ΔG^{o}_{solv}) free energy steps.

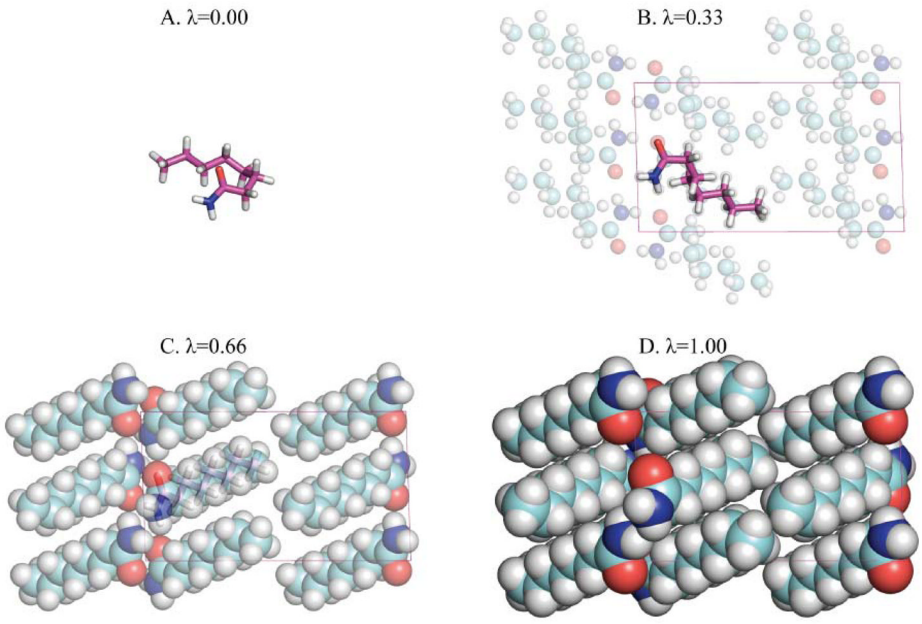

Figure 2.

An alchemical path for the calculation of the sublimation free energy change of octanamide is depicted. The infinite crystalline material (λ=1) is annihilated to leave a single molecule of octanamide in an infinite vacuum (λ=0). As the λ state variable is reduced, softcore descriptions of both van der Waals and multipolar electrostatics smoothly eliminate interactions between the asymmetric unit copy of octanamide (magenta) and its symmetry mates (cyan). The molecule is fully flexible for all λ values during the length of a simulation. No constraint on either molecular conformation or orientation is applied.

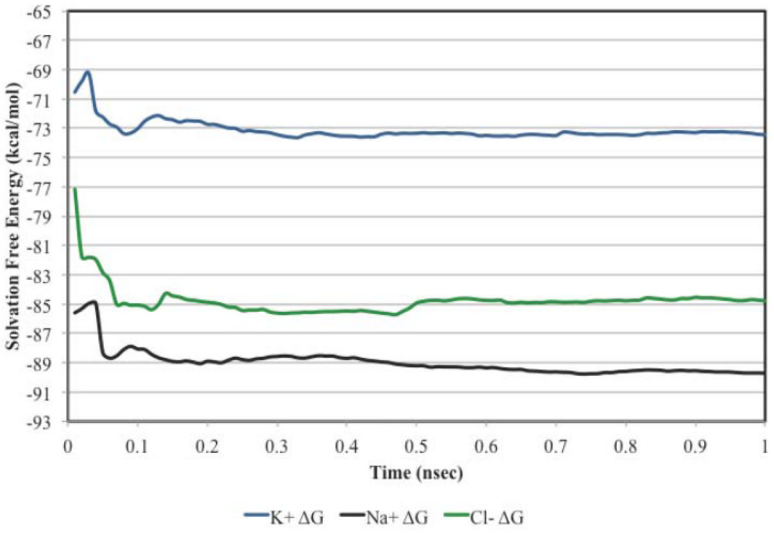

Figure 3.
Convergence of the AMOEBA/OSRW ion solvation free energy simulations as a function of time.

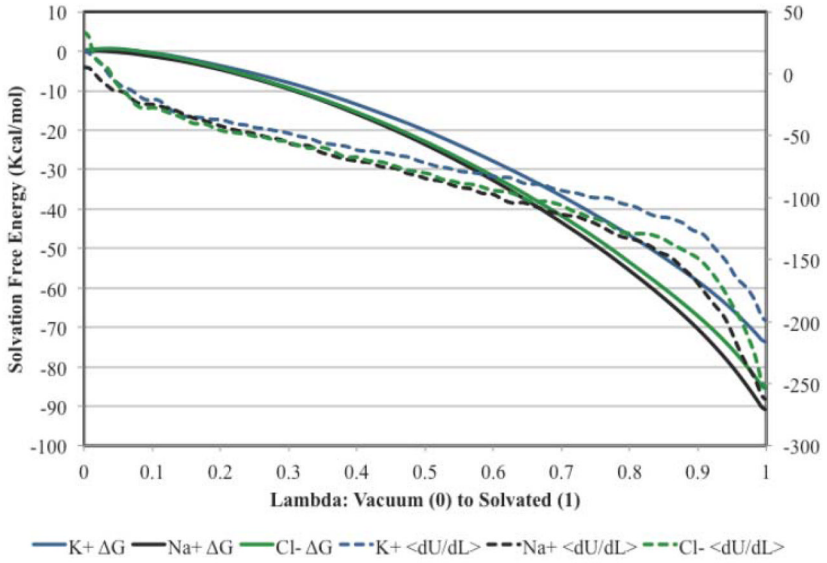

Figure 4.
Ion solvation free energy as a function of the state variable.

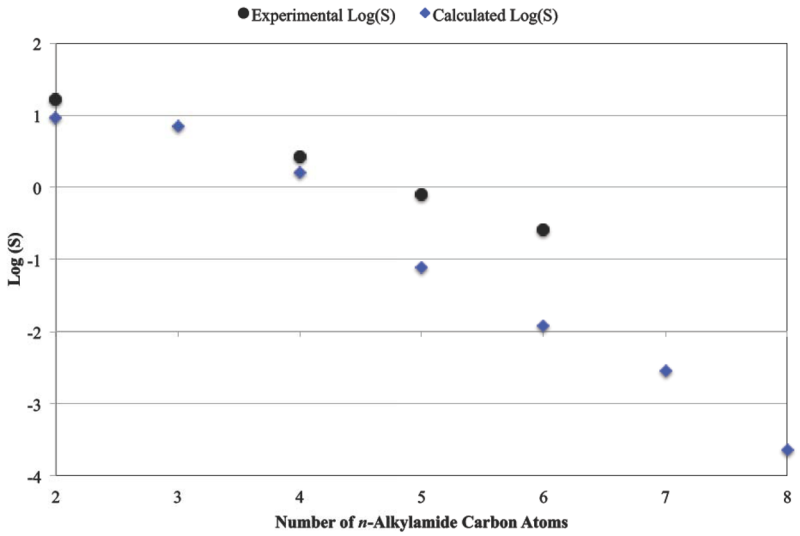

Figure 5.

Shown are experimental and calculated Log(*S*) values for the *n*-alkylamides (*S* has units of mol/L) from acetamide to octanamide. There is a monotonic trend in both the experimental and calculated values toward lower solubility with each additional CH_2 group due to increasingly favorable deposition and to a lesser extent from unfavorable solvation.

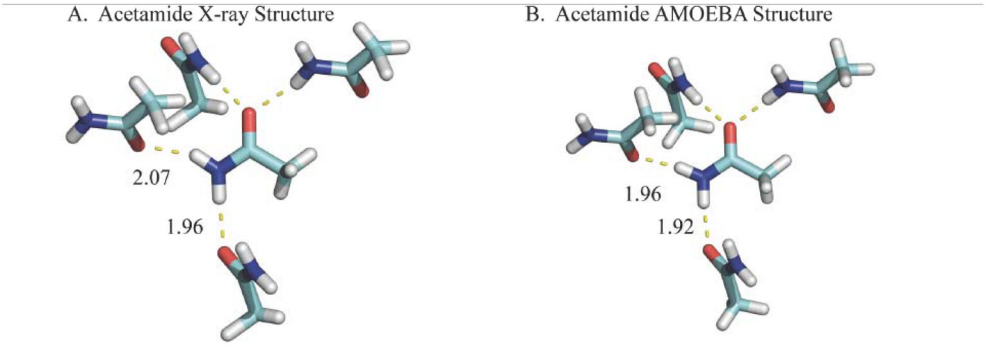

Figure 6.
The conformations of the acetamide X-ray structure (ACEMID01[101]) and lowest potential energy structure sampled by AMOEBA/OSRW agree.

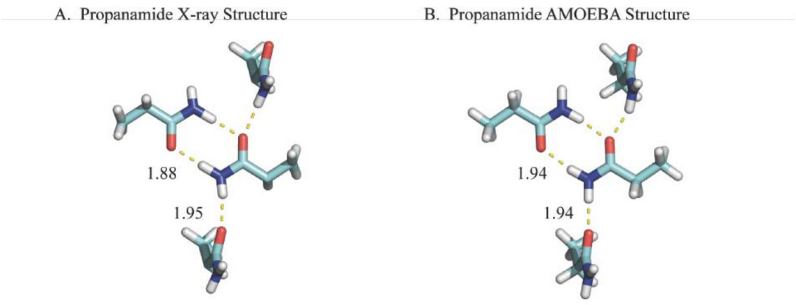

Figure 7.
The heavy atom positions of the propanamide X-ray structure (ZZZKAY01[122]) and lowest potential energy structure sampled by AMOEBA/OSRW agree, although a hydrogen from the central carbon atom was not included in the experimental model.

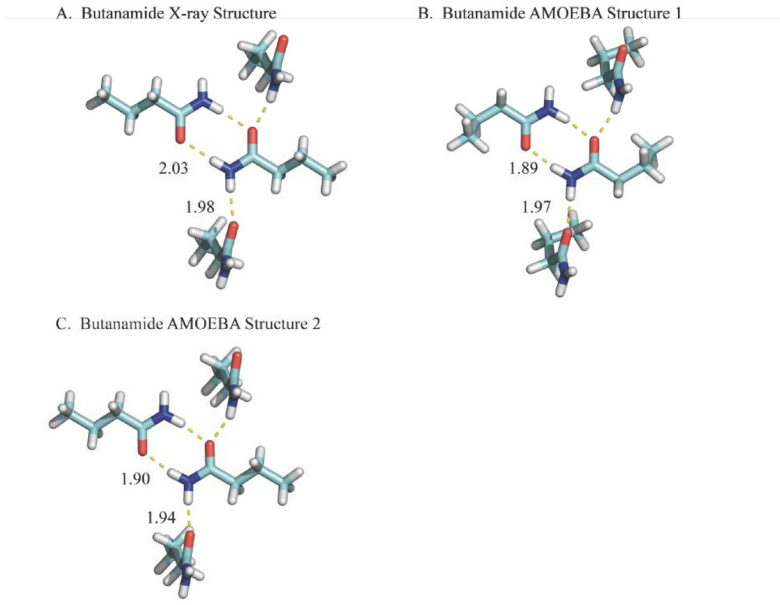

Figure 8.

The conformations of the butanamide X-ray structure (ZZZKDQ01[124]) and the 2nd lowest potential energy structure sampled by AMOEBA/OSRW agree (Structure 2). The lowest AMOEBA potential energy structure (Structure 1) exhibits a gauche conformation across the penultimate C-C bond and is lower in energy than the all trans conformation by ~0.4 kcal/mol.

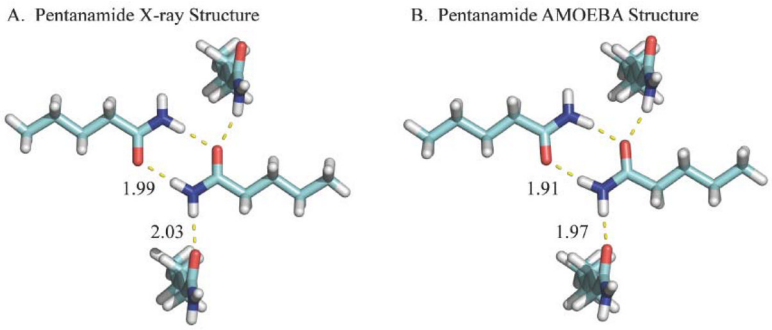

Figure 9.
The conformations of the pentanamide X-ray structure (ZZZKJQ01[123]) and the lowest potential energy structure sampled by AMOEBA/OSRW agree.

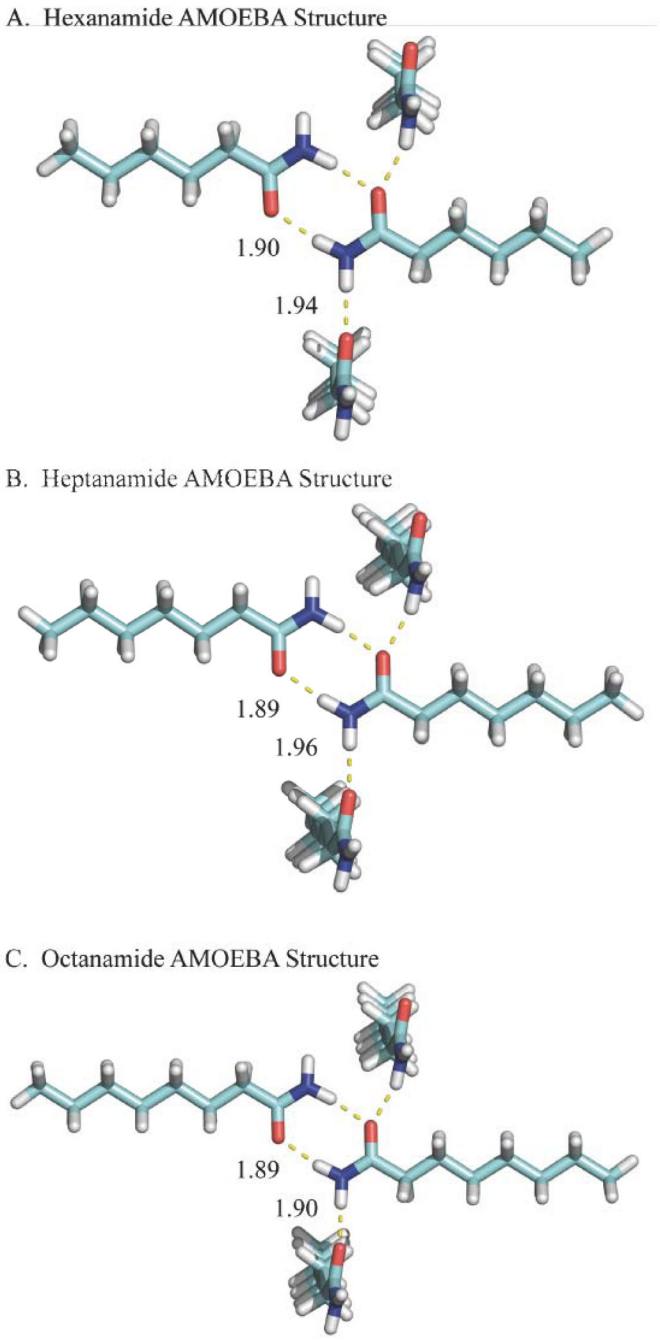

Figure 10.

Shown are the lowest potential energy structures for hexanamide, heptanamide and octanamide that were determined via AMOEBA/OSRW sampling, although experimental structures are not yet available for comparison. The two hydrogen-bonds between a single pair of amides is consistent with the shorter chain experimental structures. In addition, all C-C bonds are in a trans conformation.

Table 1

Shown are C-N and C=O bond lengths for acetamide from various sources including vacuum and crystalline experiments, as well as vacuum and crystalline QM optimizations. The experimental vacuum C-N bond length of 1.38 Å is reduced to 1.33–1.34 Å in condensed phase based on experimental crystal structures. Conversely, the experimental vacuum C=O bond length of 1.22 Å increases to 1.24–1.26 Å in condensed phase. These modest changes in equilibrium geometry have a significant effect on the molecular electron distribution, as demonstrated by the increase in the total dipole moment from a vacuum value of ~3.7 Debye to the condensed phase geometry value of ~4.1 Debye. Optimization of all atomic coordinates (ALL) or only carbon and hydrogen coordinates (C, H) was performed at the 6–31G* level in the environment listed, except for two cases optimized at the B3LYP/6311++G** and B3LYP/6311G** levels. The B3LYP/6311G** geometry optimization was performed using periodic boundary conditions[96,97] with cell parameters fixed and with inclusion of a dispersion correction.[27] The total dipole moment of each conformation was then calculated in vacuum at the MP2/6-311++G(2d, 2p) level.

Source	Environment	Opt.	Temp. (K)	C-N	C=O	Dipole
6-31G*	Vacuum	All		1.359	1.197	3.69
Electron Diffraction[98]	Vacuum	C, H		1.380	1.220	3.71
B3LYP/6311++G**	Vacuum	All		1.369	1.215	3.78
AMOEBA Parameterization	Vacuum	C, H		1.340	1.230	3.98
X-ray (ACEMID02)[99]	R3c	C, H	108	1.336	1.243	4.07
Neutron (ACEMID03)[95]	R3c	C, H	23	1.337	1.250	4.10
X-ray (ACEMID05)[100]	R3c	C, H	23	1.335	1.252	4.12
X-ray (ACEMID01)[101]	R3c	C, H	283–303	1.338	1.258	4.14
X-ray (ACEMID)[102]	Pccn	C, H	283–303	1.334	1.260	4.17
B3LYP/6311G**	R3c	ALL		1.336	1.249	4.17

Table 2

Shown is the lattice energy for acetamide using the AMOEBA potential in comparison to *ab initio* results with and without ZPE correction. Energy values correspond to structural minima of their respective potentials.

Potential	$E_{lattice}$	$E_{lattice}$ with ZPE
AMOEBA	−20.78	
B3LYP/6-31G*	−20.54	−18.28
B3LYP/6-311G**	−20.06	−17.84

Table 3

A comparison of AMOEBA ion solvation free energies using both FEP and OSRW protocols is presented. The AMOEBA/FEP, OPLS-AA and CHARMM27 results are taken from Grossfield *et al.*[91] All values are in units of kcal/mol.

Method	Potassium	Sodium	K+ -> Na+	Chloride	KCl Salt	NaCl Salt
AMOEBA/OSRW	−71.7	−88.9	−17.2	−83.5	−155.3	−172.4
AMOEBA/FEP	−72.6	−89.9	−17.3	−84.6	−157.2	−174.5
OPLS-AA	−65.7	−83.9	−18.2	−85.3	−151.0	−169.2
CHARMM27	−70.3	−91.0	−20.7	−91.2	−161.5	−182.2
Schmid[118]	−69.3	−86.8	−17.5	−87.2	−156.5	−174.0
Friedman[119]	−80.8	−98.3	−17.5	−75.8	−156.6	−174.1
Tissandier[120]	−84.1	−101.3	−17.2	−72.7	−156.8	−174.0

Table 4

Shown are the temperature (K), space group (SG), lattice lengths (Å) and angles (degrees), volume (Å^3), number of molecules per unit cell (Z) and volume per molecule (Å^3) for the *n*-alkylamide series. Although Turner and Lingafelter reported lattice lengths and angles for propanamide through octanamide at room temperature[121], atomic coordinates were not determined. Later work by separate groups led to experimental structures for propanamide[122], butanamide[124] and pentanamide.[123]

n-alkylamide	Temp.	SG	*a*	*b*	*C*	*α*	*β*	*γ*	Vol.	Z	Vol./Z
Acetamide	298[a]	R3c	11.53	11.53	13.59	90	90.0	120	1563.4	18	86.9
Propanamide	298[b]	$P2_1/c$	8.97	5.78	9.76	90	113.5	90	464.1	4	116.0
	203[c]	$P2_1/c$	8.67	5.76	9.81	90	113.8	90	448.2	4	112.1
Butanamide	298[b]	$P2_1/c$	10.02	5.79	9.94	90	100.9	90	566.3	4	141.6
	150[d]	$P2_1/c$	9.81	5.92	9.70	90	112.1	90	522.6	4	130.7
Pentanamide	298[b]	$P2_1/c$	11.08	6.13	9.84	90	100.7	90	656.7	4	164.2
	150[e]	$P2_1/c$	11.06	5.83	9.79	90	103.5	90	613.8	4	153.5
Hexanamide	298[b]	$P2_1/c$	14.11	5.65	9.78	90	101.6	90	763.8	4	190.9
Heptanamide	298[b]	$P2_1/c$	15.29	5.65	9.90	90	105.0	90	826.1	4	206.5
Octanamide	298[b]	$P2_1/c$	17.39	5.69	9.78	90	91.9	90	967.2	4	241.8

[a] Denne and Small[101]

[b] Turner and Lingafelter[121]

[c] Usanmaz and Alder[122]

[d] Lewis and Tocher[124]

[e] Pombeiro *et al.*[123]

Table 5

The experimental melting point,[121] solubility,[125] volume entropy (ΔG_{vol}) for moving from a saturated concentration to a 1 M standard state concentration, experimental estimate for the excess free energy lost (ΔG_{excess}) moving from a saturated concentration to the 1 M standard state[127,128] and finally, the absolute standard state solubility free energy ($\Delta G^{o}_{solubility}$) for the *n*-alkylamides through octanamide are given. The absolute standard state solubility is the sum of ΔG_{vol} and ΔG_{excess}. All free energy values are in units of kcal/mol. The experimental solubilities for propanamide, heptanamide and octanamide are currently not available.

***n*-Alkylamide**	**Melting Point (K)**	**Solubility (mg/ml)**	**ΔG_{vol}**	**ΔG_{excess}*a***	$\Delta G^{o}_{solubility}$
Acetamide	352–354	705, 2000	−1.46, −2.08	1.9	0.44, −0.18
Propanamide	350–351				
Butanamide	386–388	230	−0.57	0.3	−0.27
Pentanamide	378–379	80	0.14		0.14
Hexanamide	372–374	30	0.80		0.80
Heptanamide	368–369				
Octanamide	378–379				

[a]Excess free energy values were estimated via multiplication of experimental molal saturating concentrations[128] by experimental pairwise free energy interaction coefficients (g_{xx}).[127] Higher order coefficients g_{xxx} and g_{xxxx} were neglected due to their uncertainty.[127]

Table 6

Shown are the *n*-alkylamide standard state sublimation free energy values ΔG^{o}_{sub} (kcal/mol) based on the mean of 5 independent simulations along with their standard deviations (σ). A standard state volume correction ΔG_{vol} is added to the simulation result ΔG_{sim} to account for the increase in entropy due to moving from the volume occupied by the crystalline solid (see Table 4) into 1 M vapor concentration.

	Simulation								
***n*-Alkylamide**	**1**	**2**	**3**	**4**	**5**	**ΔG_{sim}**	**ΔG_{vol}**	**ΔG^{o}_{sub}**	**σ**
Acetamide	12.7	12.8	12.7	12.6	12.5	12.7	−1.8	10.9	0.1
Propanamide	10.1	10.1	10.0	9.8	10.2	10.1	−1.6	8.5	0.2
Butanamide	11.8	12.2	11.8	12.1	11.9	12.0	−1.5	10.5	0.2
Pentanamide	12.5	12.6	12.8	12.3	12.8	12.6	−1.4	11.2	0.2
Hexanamide	13.5	13.1	13.1	13.4	13.4	13.3	−1.3	12.0	0.2
Heptanamide	14.9	13.6	14.6	14.5	13.6	14.2	−1.3	13.0	0.6
Octanamide	15.3	15.5	16.1	15.4	15.2	15.5	−1.2	14.4	0.4

Table 7

Shown are the *n*-alkylamide standard state solvation free energies ΔG^{o}_{solv} (kcal/mol) based on the mean of 5 independent simulations, along with their standard deviations (σ).

	Simulation						
***n*-Alkylamide**	**1**	**2**	**3**	**4**	**5**	ΔG^{o}_{solv}	**σ**
Acetamide	−10.2	−10.2	−10.5	−10.4	−10.3	−10.3	0.1
Propanamide	−8.6	−8.4	−8.2	−8.0	−8.4	−8.3	0.2
Butanamide	−10.8	−10.9	−10.1	−10.3	−10.4	−10.5	0.3
Pentanamide	−9.8	−9.6	−9.3	−9.9	−9.9	−9.7	0.2
Hexanamide	−9.4	−9.1	−9.4	−9.6	−9.5	−9.4	0.2
Heptanamide	−9.6	−9.3	−9.6	−10.2	−8.9	−9.5	0.5
Octanamide	−10.1	−8.7	−9.4	−9.2	−9.6	−9.4	0.5

Table 8

Shown are the calculated and experimental absolute standard state solubility free energies $\Delta G^{o}_{solubility}$ for the *n*-alkylamides from acetamide to octanamide (kcal/mol) based on summing ΔG^{o}_{sub} and ΔG^{o}_{solv}. The

			$\Delta G^{o}_{solubility}$			Log (*S*)[a]		
***n*-Alkylamide**	ΔG^{o}_{sub}	ΔG^{o}_{solv}	**Calc.**	**Expt.**	**Error**	**Calc.**	**Expt.**	**Error**
Acetamide	10.9	−10.3	0.6	−0.18	0.8	1.0	1.2	0.3
Propanamide	8.5	−8.3	0.1			0.9		
Butanamide	10.5	−10.5	0.0	−0.27	0.3	0.2	0.4	0.2
Pentanamide	11.2	−9.7	1.5	0.14	1.4	−1.1	−0.1	1.0
Hexanamide	12.0	−9.4	2.6	0.80	1.8	−1.9	−0.6	1.3
Heptanamide	13.0	−9.5	3.5			−2.6		
Octanamide	14.4	−9.4	5.0			−3.6		
Mean					1.1			0.7

[a] Solubility (*S*) in units of mol/L is calculated from $S = \exp[(-\Delta G^{o}_{solubility} + \Delta G_{excess})/k_B T]$ where ΔG_{excess} was given in Table 5, k_B is the Boltzmann constant and *T* is temperature (298°K).