

---

Theses and Dissertations

---

Spring 2015

## Computational prediction of organic crystal thermodynamics using molecular dynamics

Jooyeon Park  
*University of Iowa*

Follow this and additional works at: <https://ir.uiowa.edu/etd>

 Part of the [Biomedical Engineering and Bioengineering Commons](#)

Copyright © 2015 Jooyeon Park

This thesis is available at Iowa Research Online: <https://ir.uiowa.edu/etd/1721>


---

### Recommended Citation

Park, Jooyeon. "Computational prediction of organic crystal thermodynamics using molecular dynamics." MS (Master of Science) thesis, University of Iowa, 2015.  
<https://doi.org/10.17077/etd.bxnjyflx>

---

Follow this and additional works at: <https://ir.uiowa.edu/etd>

 Part of the [Biomedical Engineering and Bioengineering Commons](#)

COMPUTATIONAL PREDICTION OF ORGANIC CRYSTAL THERMODYNAMICS  
USING MOLECULAR DYNAMICS

by

Jooyeon Park

A thesis submitted in partial fulfillment  
of the requirements for the Master of  
Science and Engineering degree in Biomedical Engineering  
in the Graduate College of  
The University of Iowa

May 2015

Thesis Supervisor: Assistant Professor Michael J. Schnieders

Copyright by  
JOOYEON PARK  
2015  
All Rights Reserved

Graduate College  
The University of Iowa  
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

MASTER'S THESIS

---

This is to certify that the Master's thesis of

Jooyeon Park

has been approved by the Examining Committee  
for the thesis requirement for the Master of Science and Engineering  
degree in Biomedical Engineering at the May 2015 graduation.

Thesis Committee: \_\_\_\_\_  
Michael J. Schnieders, Thesis Supervisor

\_\_\_\_\_  
Michael Mackey

\_\_\_\_\_  
Leonard R. MacGillivray

\_\_\_\_\_  
Lewis L. Stevens

To the ones I love

## ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Michael J. Schnieders. He has been a wonderful advisor and has given me a lot of wonderful opportunities to advance my academic and professional career forward. He has opened my eyes to the computational approaches to study biomaterials I have never known before. I would also like to thank Drs. Michael Mackey, Leonard MacGillivray and Lewis Stevens for their guidance. Thanks to Drs. Brian McClain and Dainius Macikenas and many people in Vertex Pharmaceuticals' Material Discovery and Characterization group, experimental validation set and practical applications were explored. Dr. Jonas Baltrusaitis helped with the quantum calculations to compare our potential energy values. In addition, some of the beautiful figures in this thesis and other publications could not have happened without Ian Nessler.

The absolute organic crystal thermodynamics work featured in this thesis is also published in the Journal of Chemical Theory and Computation.

## ABSTRACT

Computational prediction of organic crystal structure and thermodynamics is essential for material design, crystal engineering and drug development. However, accurate computational tools for organic crystal thermodynamics calculations are lacking, and experimental data sets for validation of computational methods is limited. Most crystal structure predictions and stability calculations depend solely on potential energy, which is often an insufficient approximation to thermodynamic stability. This thesis proposes and validates both absolute and relative free energy calculations for small organic compounds, thus presenting an accurate computational tool that overcomes the shortcomings of potential-energy-based models.

The solubility of organic molecules can be computed from a thermodynamic cycle that decomposes standard state solubility into the sum of solid-vapor sublimation (i.e. the thermodynamic stability of the crystal) and vapor-liquid solvation free energies  $\Delta G_{\text{solubility}}^{\circ} = \Delta G_{\text{sub}}^{\circ} + \Delta G_{\text{solv}}^{\circ}$ . Crystal polymorphs have different values of  $\Delta G_{\text{sub}}^{\circ}$ , and thus different solubilities. Although this is of critical importance to the pharmaceutical industry, robust computational methods to predict this quantity from first principles are lacking. Over the past few decades, alchemical simulation methods to compute solvation free energy using classical force fields have become widely used. However, analogous methods for determining the free energy of the sublimation/deposition phase transition are currently limited. This thesis describes an absolute thermodynamic approach based on growth of the asymmetric unit into a crystal via alchemy (GAUCHE). GAUCHE computes deposition free energy  $\Delta G_{\text{dep}}^{\circ} = -\Delta G_{\text{sub}}^{\circ} = \Delta G_{\text{Vol}} + \Delta G_{\text{AU}} + \Delta G_{\text{AU} \rightarrow \text{UC}}$  as the sum of an entropic term to account for compressing a 1 M vapor into the molar volume of

the crystal asymmetric unit ( $V_{AU}$ ) plus two simulation steps. In the first simulation step, the deposition free energy  $\Delta G_{AU}$  for a system composed of only  $N_{AU}$  asymmetric unit (AU) molecule(s) is computed beginning from an arbitrary conformation in vacuum. In the second simulation step, the change in free energy  $\Delta G_{AU \rightarrow UC}$  to expand the asymmetric unit degrees of freedom into a unit cell (UC) composed of  $N_{UC}$  independent molecules is computed. This latter step accounts for the favorable free energy of removing the constraint that every symmetry mate of the asymmetric unit has an identical conformation and intermolecular interactions. The current work is based on NVT simulations, which requires knowledge of the crystal space group and unit cell parameters from experiment, but not *a priori* knowledge of crystalline atomic coordinates. GAUCHE was applied to 5 organic molecules whose sublimation free energy has been measured experimentally, based on the polarizable AMOEBA force field and more than a microsecond of sampling per compound. The mean unsigned and root-mean-square errors were only 1.6 and 1.7 kcal/mol, respectively, which indicates that GAUCHE is capable of accurately predicting sublimation thermodynamics.

For polymorphic systems, we propose a relative thermodynamic approach, that is similar to the second simulation step of GAUCHE, where  $\Delta G_{P1 \rightarrow P2}$  is calculated instead of  $\Delta G_{AU \rightarrow UC}$ . A relative approach reduces statistical uncertainty compared to taking the difference between two absolute calculations; thus, it is more appropriate for determining the fairly small thermodynamic stability differences between two polymorphs. For example, the experimental free energy difference for our paracetamol test system is only 0.93 kcal/mol. Although both quantum and AMOEBA potential calculations predict that form II of paracetamol is more stable crystal than form I, our relative free energy



calculation predicts the opposite stability ranking therefore agrees with the experiment. Decomposition of free energy into entropy and enthalpy indicates that the favorable entropy change contributes to the greater thermodynamic stability of form I over form II. Although the exact magnitude of entropy and enthalpy changes differs across literature data as well as our data, the favorable entropic contribution is consistent. Further calculations over the temperature range from 100 to 308 K show the temperature dependence of free energy, which follows the parabolic trend observed in experiments. Our results show that relative polymorph stability methods can accurately capture the temperature dependence of stability and overcome the inherent limitations of potential energy based stability rankings. Thus, it is now possible for crystal structure prediction to be based on free energy differences rather than potential energy differences.

## PUBLIC ABSTRACT

Pharmaceutical development takes 12-15 years and billions of dollars in cost. Rigorous computational structure and properties prediction methods are needed to find optimal formulation solutions quickly and to reduce the costs associated with pharmaceutical development. The focus of this thesis is on the study of organic crystal structure and stability, which is crucial to the formulation process.

The existing approximations to the free energy are widely used but desperately inaccurate representation of the true thermodynamic stability, which highlights the need for this research. To overcome the limitations of existing methods, we propose and validate two accurate computational methods to calculate thermodynamic stability of organic crystals.

One is an absolute method, where the stability of each crystalline system is calculated. The other is a relative method, where the stability of one crystalline system is analyzed relative to another crystalline system. Our results quantitatively demonstrated the error in estimating thermodynamic stability of organic crystals with potential energy rather than free energy by comparing the results of existing methods, our methods and experiment.

This research shows promise in accurately predicting thermodynamics of organic compounds computationally for both absolute and relative approaches. For future directions, incorporation of our free energy calculation scheme into crystal structure prediction procedure will lead to a more accurate and reliable structure prediction than potential energy based methods.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	III
ABSTRACT .....	IV
PUBLIC ABSTRACT .....	VII
TABLE OF CONTENTS .....	VIII
LIST OF TABLES.....	X
LIST OF FIGURES.....	XII
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: BACKGROUND.....	4
2.1: Force Fields.....	4
2.1.1: Atomic Multipole Optimized Energetics for Biomolecular Applications.....	4
2.1.2: Dispersion Corrected Density Functional Theory .....	8
2.2: Statistical Mechanics .....	9
2.2.1: Molecular Dynamics .....	9
2.2.2: Metadynamics & Orthogonal Space Random Walk.....	9
2.2.3: Alchemical Simulations .....	11
2.3: Small Organic Crystals .....	12
2.3.1: Computational Study of Active Pharmaceutical Ingredients.....	12
2.3.2: Crystal Structure Prediction.....	15
2.3.3: Crystal Polymorphism .....	16
CHAPTER 3: ABSOLUTE ORGANIC CRYSTAL THERMODYNAMICS .....	19
3.1: Methods.....	20
3.1.1: Model System Selection and AMOEBA Parameterization.....	20
3.1.2: Crystal Lattice Potential Energy.....	22
3.1.3: End-State Approximation to Thermodynamic Crystal Stability .....	24
3.1.4: Growth of the Asymmetric Unit into a Crystal via Alchemy (GAUCHE) .....	25
3.1.5: Multiple Molecules in the Asymmetric Unit.....	29
3.2: Results .....	29
3.2.1: Lattice Energies for AMOEBA and Dispersion Corrected DFT.....	29
3.2.2: Alchemical Stochastic Dynamics.....	31
3.2.3: AMOEBA Structures and Hydrogen-Bonding Compared to Experiment.....	36
3.2.4: A Comparison of Lattice Energy, End-State, and Alchemical Approaches .....	37
CHAPTER 4: RELATIVE ORGANIC CRYSTAL THERMODYNAMICS.....	39
4.1: Methods.....	39
4.1.1: Lattice Energies.....	39

4.1.2: Relative Free Energy Difference between Two Paracetamol Polymorphs .....	40
4.1.3: Entropic and Enthalpic Contributions to Free Energy .....	43
4.2: Results .....	44
4.2.1: Lattice Energies.....	44
4.2.2: Relative Free Energy Difference between Two Polymorphs.....	46
4.2.3: Enthalpic and Entropic Contribution to Free Energy .....	50
CHAPTER 5: CONCLUSION .....	52
5.1: Absolute Organic Crystal Thermodynamics.....	52
5.2: Relative Organic Crystal Thermodynamics.....	53
REFERENCES .....	56

## LIST OF TABLES

Table 1. Compounds studied and their associated CSD reference codes, space groups and unit cell parameters. Roman numerals following paracetamol and methyl paraben correspond to polymorph. ....	22
Table 2. Molecular weight, number of molecules per asymmetric unit, unit cell volume, number of unit cell molecules, volume per molecule and experimental temperature for each crystal studied. ....	22
Table 3. Comparison of the lattice energy between AMOEBA and dispersion corrected DFT (D-DFT) evaluated using the B3LYP functional and 6-31G* basis set (kcal/mol). Energies were computed after minimization of the experimental coordinates in the respective potential. ....	31
Table 4. Free energy values $\Delta G_{AU}$ (kcal mol <sup>-1</sup> ) computed from five independent 200 ns simulations of the asymmetric unit along an alchemical path between vapor and crystalline states. The mean and standard deviation are given for each compound. ....	32
Table 5. Shown are the free energy values (kcal/mol) for each simulation step in the thermodynamic cycle ( <b>Figure 4</b> ) used to compute the free energy change $\Delta G_{AU \rightarrow UC}$ of moving from a simulation of the asymmetric unit to a simulation of the unit cell. ....	34
Table 6. Calculated and experimental absolute sublimation free energies for each compound (kcal/mol). ....	35
Table 7. Lattice energies of the AMOEBA potential from 1) the minimized experimental coordinates and 2) the lowest potential energy found via minimization of simulation snapshots (kcal/mol). ....	36
Table 8. Hydrogen-bond distances (Å) for the experimental crystal structure, for AMOEBA minimization of the experimental structure, and for AMOEBA minimization of the most favorable snapshot from the alchemical simulations. ....	37
Table 9. Negative of the AMOEBA lattice energy, the negative of the end-state approximation $\Delta G_{ES}^o$ for deposition free energy, the GAUCHE prediction for sublimation free energy $\Delta G_{GAUCHE}^o$ and the experimental sublimation free energy (kcal/mol). ....	38
Table 10. Crystallographic information for the two stable paracetamol polymorphs including CSD reference code, space group, unit cell parameters, number of molecules in a unit cell, volume per molecule (Å <sup>3</sup> ) and experimental temperature (Kelvin). The $\alpha$ and $\gamma$ lattice angles are 90 degrees for each crystal. ....	40
Table 11. Crystallographic information determined experimentally at extreme temperatures. The $\alpha$ and $\gamma$ lattice angles are 90 degrees for each crystal. These parameters were fixed during lattice energy determination. Listed are CSD code, space group, a-, b-, c-axis (Å), $\beta$ (degrees), experimental temperature and density. ....	40

Table 12. Comparison of the lattice potential energies between AMOEBA, <sup>1</sup> Dreidig (Beyer et al., 2001) and <sup>2</sup> dispersion corrected DFT with the B3LYP functional and 6-31G** basis set (Li & Feng, 2006) (kcal/mol).....	45
Table 13. Lattice energies of crystal structures, optimized using the AMOEBA force field. The initial structures were experimentally determined at high and low extreme temperatures as shown in the table. The vapor energy in each case was -13.35. All energies are in kcal/mol.....	46
Table 14. Five simulations are reported for each step of the relative polymorph free energy thermodynamic cycle that changes paracetamol from polymorph I to polymorph II for three temperatures (100, 200, 288.15, 298.15 and 308.15 Kelvin). .....	47
Table 15. Two experimental decompositions (Perlovich et al., 2007; Sacchetti, 2000) of the relative free energy difference to change polymorph I into polymorph II at room temperature are compared to finite difference (FD) and enthalpy calculated simulation approaches, all reported in kcal/mol. ....	51

## LIST OF FIGURES

Figure 1. Thermodynamic Cycle which shows two different paths to calculating the free energy of solution: the black arrow shows the direct phase change from crystal to solvated phase and the green goes through the vapor phase. For different polymorphs, solubility changes due to change in sublimation; thus, only sublimation value needs to be recalculated (Schnieders et al., 2012) .....	14
Figure 2. Shown are the chemical structures of the five molecules studied: acetanilide, paracetamol, methyl paraben, ethyl paraben and phenacetin .....	21
Figure 3. Shown is a potential of mean force for the alchemical path between vapor and crystalline states for paracetamol based on the OSRW method. Enhanced sampling with OSRW speeds convergence of numerical thermodynamic integration that is used to compute $\Delta G_{AU}$ . .....	26
Figure 4. Shown is a thermodynamic cycle for computing the free energy change $\Delta G_{AU \rightarrow UC}$ between simulation systems composed of 1) an asymmetric unit and 2) a unit cell. In this case, the P2 <sub>1</sub> /c methyl paraben asymmetric unit is expanded into a P1 unit cell. ....	28
Figure 5. Shown is the convergence of five independent simulations of the phenacetin asymmetric unit deposition free energy ( $\Delta G_{AU}$ ). .....	32
Figure 6. Shown is the convergence of five independent simulations of the ethyl paraben asymmetric unit deposition free energy ( $\Delta G_{AU}$ ). .....	33
Figure 7. Shown is a diagram for an efficient five-step thermodynamic cycle for computing the relative free energy change between two polymorphs ( $\Delta \Delta G_{I-II}$ ). The steps include adding harmonic restraints to polymorph I ( $\Delta G_1$ ), turning off polymorph I intermolecular interactions to reach a plane-restrained vapor I ( $\Delta G_2$ ), transitioning to a plane-restrained vapor II ( $\Delta G_3$ ), turning on polymorph II intermolecular interactions ( $\Delta G_4$ ) and turning off harmonic restraints ( $\Delta G_5$ ). .....	42
Figure 8. The relative free energy difference between polymorph I and polymorph II as a function a temperature is plotted for two experimental approaches and from molecular dynamics simulation using the AMOEBA force field. Experimental data from Sacchetti includes 95% confidence interval (Sacchetti, 2000), and Perlovich data is an extrapolation based on Perlovich and Sacchetti data (Perlovich et al., 2007). .....	49
Figure 9. Five components of relative free energy difference between polymorph I and polymorph II described in Figure 7 as a function of temperature. All values are reported in kcal/mol. ....	50

## CHAPTER 1: INTRODUCTION

The thermodynamic stability of organic crystals provides important physical insights into drug formulation and crystal engineering. Crystal structure affects the intermolecular interactions between molecules in the solid phase, which in turn affects the solubility of the active pharmaceutical ingredient(s). Solubility, along with permeability, affects the bioavailability of the molecule. Each of the multiple solid phases of a compound, also known as polymorphs, can have different desirable physical properties, such as superior compaction properties or higher solubility. The solubility of organic molecules is of critical importance to the pharmaceutical industry, however, robust computational methods to predict this quantity from first principles are lacking. Solubility can be computed from a thermodynamic cycle that decomposes standard state solubility into the sum of solid-vapor sublimation and vapor-liquid solvation free energies  $\Delta G_{\text{solubility}}^{\circ} = \Delta G_{\text{sub}}^{\circ} + \Delta G_{\text{solv}}^{\circ}$ .

This thesis approaches the crystal thermodynamics problem computationally. Motivations for a simulation approach include the goals of reducing the cost and time of drug development by providing accurate thermodynamic insights in the early phases of the formulation process. Using an advanced polarizable molecular mechanics force field called AMOEBA (Atomic Molecular Optimized Energetics for Biomolecular Applications) and fast converging variant of metadynamics called OSRW (Orthogonal Space Random Walk), we defined thermodynamic paths to accurately calculate both the absolute and relative thermodynamics of organic crystals. These methods are implemented in the Force Field X program (<http://ffx.biochem.uiowa.edu>).



The GAUCHE (Growth of the Asymmetric Unit into a Crystal via Alchemy) approach used for the absolute crystal thermodynamics takes advantage of simulating the asymmetric unit followed by expansion of the system to a unit cell to reduce the restraints on the system and capture favorable entropy. Our approach does not require *a priori* knowledge of the atomic coordinates of the crystal, which had been a limiting factor for analogous sublimation/deposition simulation methods. Upon testing our method on five drug-like molecules, GAUCHE shows a significant improvement in accuracy compared to other commonly used approximations of free energy. Non-sampling, single snapshot mechanisms such as the end-state approximation or the use of lattice potential energy predict absolute sublimation free energy to root-mean-squared-error (RMSE) of 11.34 and 8.19 kcal/mol, respectively, compared to the experiment. Our GAUCHE method, in contrast, is accurate to 1.70 kcal/mol RMSE against experiment.

The relative thermodynamics approach utilizes similar thermodynamic steps as the GAUCHE expansion mechanism. By using harmonic restraints on the experimental coordinates, the sampling space is reduced compared to absolute sublimation/deposition free energy simulations. The reduction in phase space dramatically reduces computational cost and statistical uncertainty. While the standard deviation of the sublimation/deposition free energy for paracetamol form I is 0.86 kcal/mol, the standard deviation for the relative thermodynamic stability differences between polymorphs I and II is only 0.40 kcal/mol. The exploration of the free energy difference between paracetamol polymorphs as a function of temperature generates a similar parabolic trend *in silico* as that of experiment. Under ambient conditions, our free energy results show the same thermodynamic stability ranking as the experimental results, which is the

opposite ordering compared to both AMOEBA and density functional theory with dispersion corrections (DFT-D) potential energy methods (Li & Feng, 2006). Disagreement between potential and free energy based ordering of relative polymorph stability demonstrates the importance of calculating free energy despite its higher computational cost. These results highlight the need for more rigorous free energy based approaches to crystal structure prediction, instead of widely adopted potential energy strategies.

This thesis outlines the development of accurate free energy simulation procedures that can be used by solid-state organic chemists, crystal engineers and pharmaceutical formulation scientists to gain insight into the thermodynamics of their systems. Validation on sample cases showed the proposed methods approach chemical accuracy when compared to published experimental values. The next chapter is dedicated to providing details about the underlying polarizable AMOEBA potential energy function and alchemical sampling procedures that form the basis for the organic crystal simulations. The third chapter outlines the methods and results for our absolute GAUCHE approach while the fourth chapter describes our approach to computing relative polymorph stability. The final chapter contains the conclusions and implications of our findings.

## CHAPTER 2: BACKGROUND

### 2.1: Force Fields

Computational simulations of organic compounds require a physical description of molecular energetics. This is commonly known as a force field, which encompasses both the functional form used to describe bonded and non-bonded interactions and corresponding parameters for each chemical functional group. Molecular mechanics (MM) force fields are derived in part from quantum mechanical (QM) calculations. QM is currently too computationally expensive to calculate thermodynamics, which requires millions or more of energy evaluations. On the other hand, MM force fields based on fixed atomic partial charges are computationally inexpensive relative to QM, but are often too simple to calculate thermodynamic values to experimental accuracy. Classical force fields can be more transferable between environments by explicitly including polarization and more accurate via the use of higher order permanent multipoles. A drawback to polarizable atomic multipole MM force fields is their computational expense relative to established fixed partial charge models, however, their accuracy is essential for chemically accurate thermodynamics (i.e. mean errors of ~1 kcal/mol). Consequently, this work focused on establishing approaches based on a polarizable force field called AMOEBA for thermodynamic calculations and a QM method called DFT-D for comparison purposes.

#### 2.1.1: Atomic Multipole Optimized Energetics for Biomolecular Applications

The Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) polarizable force field was first developed for a water model by the Ponder lab at

Washington University in St. Louis (Pengyu Ren, Wu, & Ponder, 2011). The goal of modeling biomolecules at a lower computational cost than QM calculations, but with similar energetic accuracy, motivated development of the polarizable AMOEBA force field (Ponder et al., 2010). The features of AMOEBA that improve accuracy compared to fixed atomic charge force fields are explicit inclusion of polarization via induced dipoles and fixed atomic multipoles truncated at quadrupole. Polarization allows transferability of the model between phases. For the purposes of modeling the thermodynamic stability of organic crystals, transferability between vapor and crystalline phases is critical during the sublimation phase transition. The Poltype tool for automatic parameterization of organic molecules via the polarizable AMOEBA force field was developed at The University of Texas at Austin (J. Wu, Chatterjee, & Ren, 2012). The AMOEBA functional form is given by

$$U=U_{\text{bond}}+U_{\text{angle}}+U_{\text{b}\theta}+U_{\text{oop}}+U_{\text{torsion}}+U_{\text{vdW}}+U_{\text{ele}}^{\text{perm}}+U_{\text{ele}}^{\text{ind}}$$

**Eq. 1**

where the first five terms are covalent bonding terms and the last three terms represent non-bonded interactions. The five bonded terms include bond-stretching, angle-bending, bond-angle cross term, out-of-plane bending and torsional rotation. The three non-bonded terms are van der Waals, permanent electrostatics and polarization energy.

AMOEBA bonded interactions have anharmonic functional forms and are represented as shown below

$$U_{\text{bond}}=K_b(b-b_0)^2[1-2.55(b-b_0)+3.793125(b-b_0)^2]$$

**Eq. 2**

$$U_{\text{angle}} = K_{\theta}(\theta - \theta_0)^2[1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5}(\theta - \theta_0)^2 - 7.0 \times 10^{-7}(\theta - \theta_0)^3 + 2.2 \times 10^{-8}(\theta - \theta_0)^4]$$

**Eq. 3**

$$U_{b\theta} = K_{b\theta}[(b - b_0) + (b' + b'_{\theta})](\theta - \theta_0)$$

**Eq. 4**

$$U_{\text{torsion}} = \sum_n K_{n\phi}[1 + \cos(n\phi \pm \delta)]$$

**Eq. 5**

$$U_{\text{oop}} = K_{\chi}\chi^2$$

**Eq. 6**

where the bond lengths, angles and energies are in units of Å, degrees and kcal/mol, respectively. Deviations from the ideal bond lengths and angles give rise to the first three terms shown above.

The van der Waals (vdW) interaction in AMOEBA has a buffered 14-7 functional form that is softer than the standard Lennard-Jones 6-12 potential and is given by

$$U_{\text{vdW}}(ij) = \epsilon_{ij} \left( \frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left( \frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right)$$

**Eq. 7**

where the potential is a function of separation distance,  $R_{ij}$ , between two atoms  $i$  and  $j$  and  $\rho_{ij}$  is defined as  $R_{ij}/R_{ij}^0$  where  $R_{ij}^0$  is the minimum energy distance. Combining rules for

heterogeneous atom pairs are given by  $R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2}$  for the minimum energy distance

and by  $\epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2}$  for the interaction strength.

The permanent electrostatic potential energy between atoms  $i$  and  $j$  separated by a distance  $r_{ij}$  is represented as  $U_{\text{ele}}^{\text{perm}}(r_{ij}) = M_i^T T_{ij} M_j$  where

$$M_i = [q_i, d_{ix}, d_{iy}, d_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz}, Q_{iyx}, Q_{iyy}, Q_{iyz}, Q_{izx}, Q_{izy}, Q_{izz}]^T$$

**Eq. 8**

$$T_{ij} = \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & L \\ \frac{\partial}{\partial x_j} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & L \\ \frac{\partial}{\partial y_j} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & L \\ \frac{\partial}{\partial z_j} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & L \\ M & M & M & M & 0 \end{bmatrix} \left( \frac{1}{r_{ij}} \right)$$

**Eq. 9**

where  $M_i$  is a permanent atomic multipole composed of charge,  $q_i$ , dipoles,  $d_{i\alpha}$ , and quadrupoles,  $Q_{i\beta\gamma}$ . Polarization in the AMOEBA force field is handled by smearing each induced dipole as described by Thole (Thole, 1981)

$$\rho = \frac{3a}{4\pi} e^{-au^3}$$

**Eq. 10**

where  $u = r_{ij} / (\alpha_i \alpha_j)^{1/6}$  is the effective distance as a function of linear separation  $r_{ij}$  and dimensionless atomic polarizabilities ( $\alpha$ ) at sites  $i$  and  $j$ . The polarization term is

responsible for induced polarization energy in **Eq. 1**. The mutual induction scheme in Thole's model leads to each induced dipole further polarizing all other sites, which continue to mutually polarize each other until every site has converged.

The AMOEBA polarizable force field has been implemented in many software platforms. The work featured here is an implementation of the AMOEBA force field in the Force Field X (FFX) platform (<http://ffx.biochem.uiowa.edu>). AMOEBA's polarizability and higher-order multipoles add transferability and accuracy to the computational modeling of chemicals.

### 2.1.2: Dispersion Corrected Density Functional Theory

Density functional theory with dispersion correction (DFT-D) is an electronic structure method that is currently a gold standard for condensed phase calculations. However, parameterization of the dispersion correction and imperfections in available functionals introduce uncertainty. Studies of organic solids often rely on DFT-D to capture the many body effects (Reilly & Tkatchenko, 2014). There are alternative quantum mechanical approaches with more reliable electron correlation, such as second-order Møller-Plesset perturbation theory (MP2) against which is the level of theory AMOEBA electrostatics are parameterized. Due to the complexity of the condensed phase MP2 calculations, DFT-D remains the most commonly used method for calculating the potential energy of organic crystal systems. As mentioned previously, the high computational costs associated with the use of quantum mechanics results in conformational sampling being impractical. Therefore, our molecular mechanics approach uses quantum theory during parameterization and validation, followed by classical simulation to sample phase space.

## 2.2: Statistical Mechanics

Force fields allow scientists to computationally simulate physically accurate models of chemicals. However, chemical compounds are not static systems, such that an additional layer of sophistication is needed to sample from a thermodynamic distribution defined by the force field and state variables such as volume, temperature, pressure, and particle number. Algorithms capable of sampling from a defined distribution include molecular dynamics and Metropolis Monte Carlo.

### 2.2.1: Molecular Dynamics

Molecular dynamics (MD) simulates the motions of molecules over a period of time. The force field, unit cell parameters and space group define the potential energy surface of the system. The potential energy and forces on constituent atoms can be determined for the current atomic coordinates, which are used to integrate Newton's equation of motion. Three state variables specify the thermodynamic ensemble of the system, including the NVT ensemble used in this work via constant volume ( $V$ ), constant number of particles ( $N$ ) and constant temperature ( $T$ ). Under the NVT ensemble, pressure and total potential energy fluctuate. In addition to simulating crystals, MD is commonly used to simulate and observe the folding of proteins, protein-ligand interactions and the dynamics of liquids. Ensemble averages of any system property can be calculated from an MD trajectory so long as sampling of phase space converges.

### 2.2.2: Metadynamics & Orthogonal Space Random Walk

To accelerate MD sampling convergence, the metadynamics family of algorithms adds a time dependent bias to the potential energy surface to overcome energy barriers. The



biasing term, which is composed of a sum of Gaussian hills, approaches the negative of the free energy as a function  $\lambda$  as the algorithm converges

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

**Eq. 11**

where  $f_m(\lambda)$  is the biasing term. The Orthogonal Space Random Walk (Zheng, Chen, & Yang, 2009) approach is a variant of metadynamics that employs a two-dimensional bias. The algorithm is named due to the biasing potential flattening the potential energy surface along both the thermodynamic path variable  $\lambda$  and the derivative of the potential energy with respect to  $\lambda$  ( $F_\lambda$ )

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

**Eq. 12**

The biasing term is a Gaussian-shaped repulsive potential centered at  $\lambda$  and  $F_\lambda$  for a given time step (Schnieders et al., 2012)

$$g_m(\lambda, F_\lambda) = \sum_{t_i} h \exp\left(-\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2} - \frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right)$$

**Eq. 13**

Using the thermodynamic integration formula **Eq. 15**, the free energy difference between the initial state and the final state, specified by two  $\lambda$  values, can be calculated as

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

**Eq. 14**

$$G_{AMOEB A}(\lambda) = \int_{\lambda_i}^{\lambda} \left. \frac{dG}{d\lambda} \right|_{\lambda'} d\lambda'$$

**Eq. 15**

### 2.2.3: Alchemical Simulations

Transmutation of an atom or a functional group into another during computational simulation is generally referred to as an alchemical transformation (Straatsma & McCammon, 1992). Such alchemical transitions between molecules are facilitated by soft-coring the atoms of interest. The atomic soft-core approach allows the potential to turn on and off non-bonded interactions smoothly, which dramatically improves convergence of free energy differences via thermodynamic integration. An alchemical transition between two phases is possible through an unphysical path that smoothly decouples intermolecular interactions. The calculation of free energy using alchemical simulations is often achieved by expressing non-bonded interactions as a function of a state variable such as  $\lambda$  in Sect. 2.2.2. For AMOEBA, the non-bonded vdW, multipolar real space Ewald summation, permanent reciprocal space Ewald summation and induced energy are expressed

$$U_{vdw}(\lambda, r) = \lambda^{\beta} \varepsilon_{ij} t_1 t_2$$

**Eq. 16**

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

**Eq. 17**

$$\phi_{recip}(\lambda) = \phi_{recip}^{total} + (\lambda - 1)\phi_{recip}^{solute}$$

**Eq. 18**

$$U_{\text{pol}}(\lambda) = \begin{cases} U_{\text{pol(v)}}^{\text{solute}} + U_{\text{pol}}^{\text{condensed}} & , 0 \leq \lambda < \lambda_{\text{pol}}^{\text{start}} \\ \lambda_{\text{pol}}^3 U_{\text{pol}}^{\text{total}} + (1 - \lambda_{\text{pol}}^3)(U_{\text{pol(v)}}^{\text{solute}} + U_{\text{pol}}^{\text{condensed}}) & , \lambda_{\text{pol}}^{\text{start}} \leq \lambda \leq 1 \end{cases}$$

**Eq. 19**

respectively. Detailed explanation of these equations can be found in (Schnieders et al., 2012), but for the purposes of this thesis, the need to express non-bonded interactions as a function of  $\lambda$  for alchemical simulations is all that is necessary.

### 2.3: Small Organic Crystals

Small organic crystals are the systems of interest for this thesis. One application of computational methods for this type of system is drug development and formulation. Classical drug development relies on the discovery and formulation of small biologically active organic compounds. Many organic compounds have complex structures, which lead to difficulties in both experimental characterization and simulation of these compounds. However, with sophisticated force fields, accurate modeling of organic solids has become an obtainable goal.

#### 2.3.1: Computational Study of Active Pharmaceutical

##### Ingredients

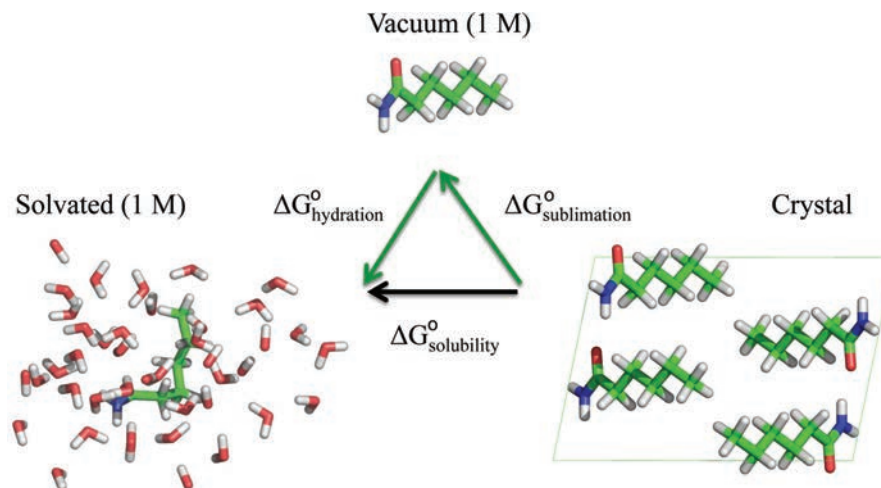
Active pharmaceutical ingredient (API) refers to the biologically active molecule that causes medicinal effects in the body. Drug development involves thorough experimental characterization of an API's physical properties, including considerations of bioavailability, drug delivery mechanism, industrially viable large-scale production of API, tabletability and desired biological effects. Two properties that are relevant to organic crystals featured in this thesis are thermodynamic stability and solubility.

The bioavailability and pharmacokinetics of a drug depend in part on its formulation, which often must overcome the low solubility of pharmaceutical compounds. Therefore, accurate computational methods to predict solubility thermodynamics could aid drug development and promote optimal formulations. To this end, the solubility equilibrium can be decomposed using a thermodynamic cycle that defines standard state solubility via the sum of solid-vapor sublimation  $\Delta G_{\text{sub}}^{\circ}$  and vapor-liquid solvation free energy  $\Delta G_{\text{solv}}^{\circ}$  terms as shown in **Figure 1** (David S. Palmer et al., 2008). The decomposition of the free energy of solubility not only makes the calculation computationally affordable, but it also provides an insight into the cause of insolubility: high affinity of the intermolecular interactions in crystalline form or hydrophobicity. Also, only the additional sublimation/deposition calculations are needed for polymorphs since  $\Delta G_{\text{solv}}^{\circ}$  stays the same for a given compound.

$$\Delta G_{\text{solubility}}^{\circ} = \Delta G_{\text{sub}}^{\circ} + \Delta G_{\text{solv}}^{\circ}$$

**Eq. 20**

Alchemical simulation methods to predict  $\Delta G_{\text{solv}}^{\circ}$  described over the last few decades (W. Jorgensen, 1985; Lybrand, Ghosh, & McCammon, 1985) can now be performed efficiently (Paluch, Shah, & Maginn, 2011) and with high statistical precision (Shirts, Pitera, Swope, & Pande, 2003), although work to overcome force field limitations is ongoing (Mobley, Liu, Cerutti, Swope, & Rice, 2012; Pengyu Ren et al., 2011).



**Figure 1.** This thermodynamic cycle shows two different paths for calculating the free energy of the solubility process: the black arrow shows the direct phase change from crystal to solvated phase and the green arrows goes through the vapor phase. For different polymorphs, solubility changes due to change in sublimation, but solvation is constant; thus, only sublimation values need to be recalculated (Schnieders et al., 2012)

On the other hand, current methods to predict  $\Delta G_{\text{sub}}^{\circ}$  rely on *a priori* knowledge of the crystalline atomic coordinates (D. S. Palmer, McDonagh, Mitchell, van Mourik, & Fedorov, 2012; Salahinejad, Le, & Winkler, 2013), on direct experimental measurements of crystalline properties such as fugacity (Paluch & Maginn, 2013), or on various heuristic approaches (W. L. Jorgensen & Duffy, 2002). Although crystal structure prediction (CSP) has made significant progress over the last decade, as indicated by the successes and increasing difficulty of the first blind CSP test (Lommerse et al., 2000) to the fifth (Bardwell et al., 2011), a unifying limitation of prediction methods has been their reliance on ranking structural candidates by potential energy, rather than on free energy. This striking approximation, the neglect of entropic differences between putative polymorphs, may limit CSP accuracy to the same degree as limitations in potential energy functions. For example, there is experimental evidence that the polymorphs of paracetamol differ in their entropic contribution to sublimation free energy at 298 K by

0.5 kcal/mole (Perlovich, Volkova, & Bauer-Brandl, 2007). As the size and complexity of organic molecules increase towards large pharmaceutical compounds, peptides (Colletier et al., 2011; Nelson et al., 2005) or even proteins, neglect of entropy becomes unreasonable.

### 2.3.2: Crystal Structure Prediction

Cambridge Crystallography Data Center (CCDC) Cambridge Structural Database (CSD) was established in 1965. As the name indicates, CSD holds the experimental results of the solved crystal structures, and all of the experimental structures used in this thesis are obtainable from this database. In many cases, there are multiple crystal structures for a given compound, and there are also variations in lattice parameters for a given crystalline form depending on the temperature at which structure is determined. The experimental determination of crystal structure requires significant experimental effort; therefore, the development of computational prediction of crystal structure has been advancing significantly in the last few decades.

Most widely used mechanisms for crystal structure prediction (CSP) and polymorph prediction rely on searching a defined potential energy landscape (Price, 2008). Potential energy approaches based on molecular mechanics (MM), quantum mechanics (QM) and hybrid QM/MM techniques have made a tremendous amount of progress and shown some success (Bardwell et al., 2011; Graeme M. Day et al., 2009; G. M. Day et al., 2005). However, the use of potential energy instead of free energy suffers from a critical limitation, the neglect of entropic differences between polymorphs (Price, 2008). With respect to absolute deposition phase transition free energy, the potential energy approach over-stabilizes the crystal compared to its true thermodynamic stability (Park et al.,

2014). Although a variety of end-state techniques, such as a normal modes approximation (Price, 2008), attempt to account for entropic effects, the estimated entropic terms are insufficient relative to sampling. Estimating thermodynamic stability with potential energy assumes that a single crystal structure is the most stable forms across all temperatures and pressures. It has been shown that some predicted structures have lower lattice energies than the experimental structures, which suggests either the energy models need improvement or the minimum lattice energy structures are not minimum free energy structures (Beyer & Price, 2000; Park et al., 2014). Beyond potential-energy-guided polymorph predictions, advances have been made using technique called metadynamics to explore the free energy surface, which can account for temperature and pressure effects (Raiteri, Martoňák, & Parrinello, 2005). However, the exact rank order of thermodynamic stability of the structures still relies on potential energies rather than free energies (Raiteri et al., 2005). For quantum-mechanics-based calculations, additional force field terms, such as many-body dispersion, are utilized to estimate some of the entropic effects without molecular dynamics sampling (Reilly & Tkatchenko, 2014). Correction terms are added because sampling with QM force field is impractical due to the expensive computational cost of energy evaluations.

### 2.3.3: Crystal Polymorphism

Polymorphism in crystals refers to the existence of multiple stable crystalline packing motifs. One of the best-known examples of chemicals with different solid phases is carbon: graphite and diamond. Since graphite and diamond are composed of only carbon, they belong to a subset of polymorphism called allotropism. These two phases have very different mechanical properties. Similar to carbon, small organic molecules also have

polymorphs, which have different properties. Such differences can affect solubility and ability to be compacted into a tablet. However, experimental determination of polymorph thermodynamic properties and other characterizations are challenging tasks, due to difficulties in producing and maintaining polymorph purity during experimental procedures (Perlovich et al., 2007)

Paracetamol (acetaminophen) is a widely used analgesic and one of the rare systems for which stability has been measured experimentally for the most stable two of its three known polymorphs (Perlovich et al., 2007). Form II, while being less abundant than form I, has a superior compaction property than form I, making it desirable in the pharmaceutical industry (Joiris, Martino, Berneron, Guyot-Hermann, & Guyot, 1998). Consequently, significant experimental and computational efforts have been dedicated to study this system. Paracetamol polymorph I is the most thermodynamically stable at ambient temperature and pressure out of the three known polymorphs. Although polymorph I is more stable than polymorph II under ambient conditions, experimental estimation of lattice energy derived from the fusion process at low temperature indicates that polymorph II is more favorable than form I (Espeau et al., 2005; Perlovich et al., 2007). Sublimation thermodynamics for polymorph II was experimentally determined at 391 K and extrapolated to 298 K using fusion thermodynamics. The parabolic data trend of switching thermodynamic stability ranking as a function of temperature is similar to the trend found based on difference in heat capacity  $2.39 \text{ cal mol}^{-1} \text{ K}^{-1}$  (Sacchetti, 2000). There is a decrease in entropy from polymorph I to polymorph II, which is an unfavorable change in crystal stability that cannot be calculated by lattice energy. Most stability determination is limited to estimating rank order of stability by correlating



polymorph's relative abundance to thermodynamic stability. However, accurate calculations of the relative stability of polymorphs are critical for determining the most stable crystal structure and material properties of each polymorph, as well as for validation of computational methods.

## CHAPTER 3: ABSOLUTE ORGANIC CRYSTAL THERMODYNAMICS

We propose a simulation procedure to divide the absolute organic crystal thermodynamics calculation (i.e. sublimation/deposition free energy) into accurate but computationally affordable steps by expanding the asymmetric unit to a unit cell. Beginning from the asymmetric unit (AU) in vapor offers the appealing advantages of 1) inexpensive force evaluations on  $N_{AU}$  molecules relative to a simulation system composed of one or more copies of the  $N_{UC}$  unit cell (UC) molecules, 2) reduced degrees of freedom during the alchemical phase transition to accelerate sampling convergence, and 3) avoidance of the need for *a priori* knowledge of crystalline atomic coordinates. GAUCHE decomposes standard state deposition free energy  $\Delta G_{dep}^{\circ}$  into a sum of two simulation terms and an entropic correction  $\Delta G_{Vol}$  to move from a 1 M standard state vapor to the molar volume of the asymmetric unit ( $V_{AU}$ )

$$\Delta G_{dep}^{\circ} = -\Delta G_{sub}^{\circ} = \Delta G_{Vol} + \Delta G_{AU} + \Delta G_{AU \rightarrow UC}$$

**Eq. 21**

where

$$\Delta G_{Vol} = k_B T \ln(V_{AU}/V_{1M})$$

**Eq. 22**

and  $k_B T$  is Boltzmann's constant multiplied by temperature in degrees Kelvin. First, the deposition free energy for a system composed of only  $N_{AU}$  asymmetric unit molecule(s)  $\Delta G_{AU}$  is computed, beginning from an arbitrary conformation in vacuum (Schnieders et al., 2012). Although the current work is based on NVT simulations and requires knowledge of the crystal space group and unit cell parameters, no *a priori* knowledge of

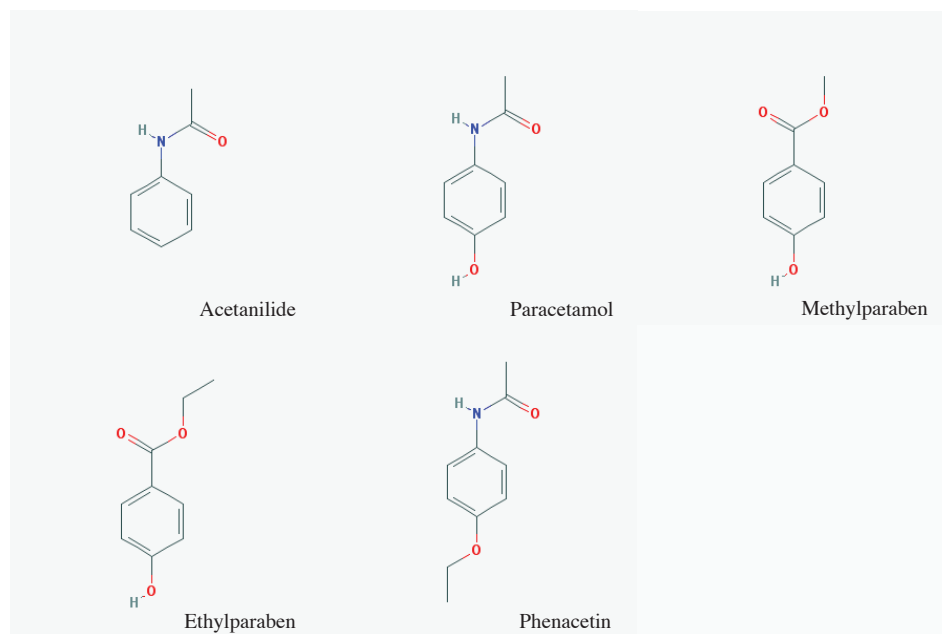
crystalline atomic coordinates is needed. Second, GAUCHE computes the change in free energy  $\Delta G_{AU \rightarrow UC}$  to expand the asymmetric unit into a larger system composed of all  $N_{UC}$  molecules in the unit cell. This latter step accounts for the favorable free energy of removing the constraint that at every dynamics step, each symmetry mate of the asymmetric unit adopts an identical conformation and has identical intermolecular interactions. For example, a unit cell end state permits favorable coupled motions between asymmetric units.

### 3.1: Methods

#### 3.1.1: Model System Selection and AMOEBA

##### Parameterization

Five molecules shown in **Figure 2** were selected from approximately 30 compounds with experimentally measured sublimation free energy based on the following criteria: 1) molecular weight larger than 130 g/mol, 2) all atoms must be from the set [H, C, N, O], and 3) the absence of bicyclic or polycyclic ring systems.



**Figure 2.** Shown are the chemical structures of the five molecules studied: acetanilide, paracetamol, methyl paraben, ethyl paraben and phenacetin

The compounds include acetanilide (Brown, 1966), paracetamol (polymorph I) (Haisa, Kashino, Kawai, & Maeda, 1976), methyl paraben (polymorph II) (Nath, Aggarwal, & Nangia, 2011), ethyl paraben (Lin, 1986) and phenacetin (Patel, Patel, & Singh, 1983). Some functional groups, such as halogens (Mu et al., 2014) and polycyclic rings, are a current focus of AMOEBA development efforts and will be examined in future work on organic crystal thermodynamics. Crystallographic space group and unit cell information was obtained from the Cambridge Structural Database (CSD) and is provided in **Table 1** and **Table 2**. AMOEBA force field parameters were prepared for the five molecules using the PolType program (J. C. Wu, Chattré, & Ren, 2012), which in turn relies on Gaussian (M. J. Frisch et al., 2009), GDMA (Stone, 2005; Stone & Alderton, 1985) and TINKER (Ponder, 2009).

**Table 1.** Compounds studied and their associated CSD reference codes, space groups and unit cell parameters. Roman numerals following paracetamol and methyl paraben correspond to polymorph.

Compound	CSD Code	Space	<i>a</i>	<i>b</i>	<i>c</i>	$\alpha$	$\beta$	$\gamma$
		Group						
Acetanilide	ACANIL	Pbca	19.64	9.48	7.98	90	90.0	90
Paracetamol I	HXACAN01	P21/c	11.72	9.40	12.93	90	147.0	90
Methyl Paraben II	CEBGOF03	P21/c	4.82	14.63	10.24	90	99.8	90
Ethyl Paraben	FEGLEI	P21/c	13.76	13.18	11.58	90	125.5	90
Phenacetin	PYRAZB10	P21/c	13.25	9.65	7.81	90	104.9	90

**Table 2.** Molecular weight, number of molecules per asymmetric unit, unit cell volume, number of unit cell molecules, volume per molecule and experimental temperature for each crystal studied.

Compound	Mol. Weight (g/mol)	AU Molecules	UC Vol. ( $\text{\AA}^3$ )	Z	Vol./Z ( $\text{\AA}^3$ )	Temp. (K)
Acetanilide	135.16	1	1486.1	8	185.8	297
Paracetamol I	151.16	1	776.3	4	194.1	297
Methyl Paraben II	152.15	1	711.3	4	177.8	100
Ethyl Paraben	166.17	2	1710.2	8	213.8	297
Phenacetin	179.22	1	965.0	4	241.3	297

### 3.1.2: Crystal Lattice Potential Energy

Beginning from the experimental coordinates, each molecule was optimized in FFX (Fenn & Schnieders, 2011; Schnieders et al., 2012; Schnieders, Fenn, & Pande, 2011) to compare with dispersion corrected density functional theory (DFT-D) lattice energies. During minimization, space group symmetry was enforced. The van der Waals cut-off was set to 12.0  $\text{\AA}$  and smoothly tapered to zero interaction energy using a multiplicative switch beginning at 10.8  $\text{\AA}$ . Polarizable AMOEBA electrostatics were evaluated using a version of the smooth (Essmann et al., 1995) particle-mesh Ewald (Darden, York, & Pedersen, 1993) (PME) algorithm for multipoles (Sagui, Pedersen, & Darden, 2004) that explicitly supports space group symmetry (Schnieders et al., 2011) and the self-consistent field calculation was converged to  $10^{-5}$  RMS Debye. High precision PME parameters

included a real-space cutoff of 9.0 Å, a mesh density of 2.0 grid points per Å, 8<sup>th</sup> order B-splines and an Ewald parameter of 0.45. This allowed minimization to reach an RMS gradient convergence criterion of 10<sup>-4</sup> (kcal/mole/Å). A second minimization was performed in vacuum, beginning from the crystal minimum, to provide a reference energy  $E_{vac}$ , which is subtracted from the potential energy per molecule at the crystal minimum  $E_{cryst}$  to give the lattice potential energy

$$E_{lattice} = E_{cryst} - E_{vac}$$

**Eq. 23**

Periodic solid-state *ab initio* calculations were performed in the program suite CRYSTAL'09 (Roberto Dovesi et al., 2005; R. Dovesi et al., 2009), 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, the hybrid B3LYP (Becke, 1993; Lee, Yang, & Parr, 1988) Hamiltonian and a dispersion correction were used (Grimme, 2011). 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 (R. Dovesi et al., 2009) and values of 7, 7, 7, 7, and 16 were used. The reciprocal space integration was performed by sampling the Brillouin zone with the 6x6x6 Pack–Monkhorst net (Monkhorst & Pack, 1976). Structure optimizations were performed using analytical energy gradients with respect to atomic coordinates with cell parameters fixed (Civalleri, D'Arco, Orlando, Saunders, & Dovesi, 2001; Doll, 2001; Doll, Saunders, & Harrison, 2001), within a quasi-Newton scheme

based on the Broyden–Fletcher–Goldfarb–Shanno scheme for Hessian updating (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). 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\* (Ditchfield, Hehre, & Pople, 1971) basis set was used, and condensed phase energies were corrected for BSSE via the counterpoise method (van Duijneveldt, van Duijneveldt-van de Rijdt, & van Lenthe, 1994).

### 3.1.3: End-State Approximation to Thermodynamic Crystal

#### Stability

Using the optimized vapor phase structure, normal modes were computed in TINKER (Ponder, 2009). The normal modes are converted to a vibrational free energy using the Einstein model of independent quantum harmonic oscillators

$$G_{\text{vib}} = \sum_{i=1}^{3N-6} \left( \frac{hv_i}{2} \right) \coth \left( \frac{hv_i}{2k_B T} \right) - \sum_{i=1}^{3N-6} \left[ \left( \frac{hv_i}{2} \right) \coth \left( \frac{hv_i}{2k_B T} \right) - 2k_B T \ln \left( \coth \left( \frac{hv_i}{2k_B T} \right) \right) \right]$$

**Eq. 24**

where  $v_i$  is the  $i^{\text{th}}$  vibrational frequency,  $h$  is Planck's constant,  $N$  is the number of atoms in the molecule and  $3N-6$  is the total number of vibrational modes. For the crystalline phase of the molecule, the unit cell parameters as well as the conformation of the molecules were optimized. During crystal minimization in TINKER, symmetry constraints were applied to the unit cell parameters but not the molecules. A convergence criterion of 0.5 kcal/mol/Å was used. The same Einstein model was used for deriving the crystal vibrational free energy using normal modes calculated from the entire unit cell instead of a single molecule. Therefore, the number of molecules in the unit cell was used

to normalize the crystalline vibrational contribution. Lattice energy and the change in vibrational free energy were combined with translational and rotational terms (Tidor & Karplus, 1994) to yield an end-state estimate of deposition free energy  $\Delta G_{ES}^0$  as

$$\Delta G_{ES}^0 = \Delta G_{\text{trans}} + \Delta G_{\text{rot}} + \Delta G_{\text{vib}} + E_{\text{lattice}}$$

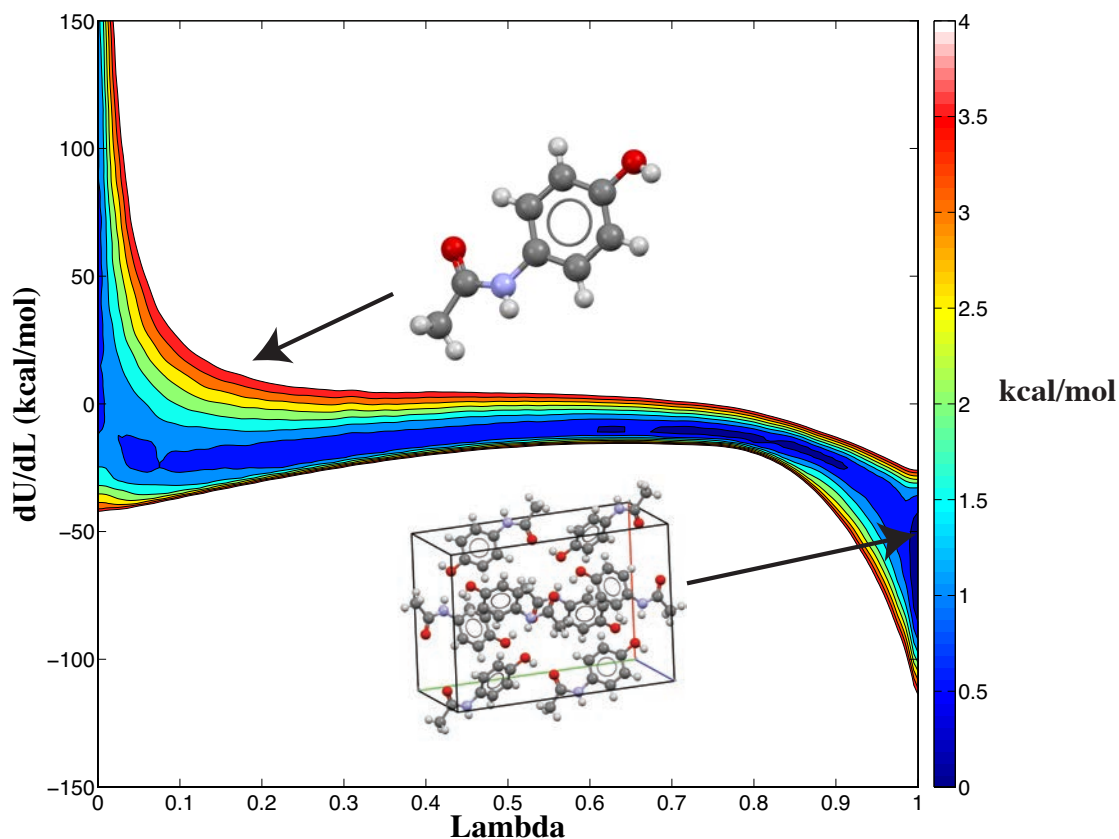
**Eq. 25**

### 3.1.4: Growth of the Asymmetric Unit into a Crystal via

#### Alchemy (GAUCHE)

The first step of GAUCHE is simulation of the sublimation-deposition phase transition based on a system consisting of only the molecule(s) in the asymmetric unit, which smoothly samples an alchemical path between vapor and crystalline phases to yield  $\Delta G_{AU}$ . This alchemical thermodynamic path provides a smooth transition between vacuum and the crystalline state by slowly turning on intermolecular interactions with symmetry mates. The overall behavior of the alchemical transformation for paracetamol under the OSRW sampling methodology can be observed in Figure 2, which shows a two dimensional potential of mean force that is a function of  $\lambda$  and  $\partial U/\partial\lambda$ . OSRW applies a time-dependent bias that flattens the surface along  $\lambda$  and  $\partial U/\partial\lambda$  until a random walk is achieved, thereby accelerating convergence of thermodynamic integration. In vapor ( $\lambda=0$ ), many iso-energetic rigid body coordinates would result in high-energy steric clashes in the crystalline phase ( $\lambda=1$ ), which explains the unfavorable (large positive) values of  $\partial U/\partial\lambda$  sampled in the vapor region. Once the alchemical simulation reached values of  $\lambda$  near 1, only conformations consistent with reasonable crystal packing are sampled, which have favorable (negative) values of  $\partial U/\partial\lambda$ . We note that each simulation sampled the AMOEBA crystalline free energy minimum multiple times.





**Figure 3.** Shown is a potential of mean force for the alchemical path between vapor and crystalline states for paracetamol based on the OSRW method. Enhanced sampling with OSRW speeds convergence of numerical thermodynamic integration that is used to compute  $\Delta G_{AU}$ .

Constraining a crystal to have perfect symmetry within the unit cell eliminates the possibility of coupled motions between symmetry mates, which is a favorable free energy contribution. The 2<sup>nd</sup> step of GAUCHE uses a thermodynamic cycle (**Figure 4**) to compute the change in free energy  $\Delta G_{AU \rightarrow UC}$  along a path between the asymmetric unit simulation system and one that includes the entire unit cell. The first step of this thermodynamic cycle computes the free energy  $\Delta G_{AU \rightarrow RAU}$  to add a harmonic restraint to each atom  $i$  in the asymmetric unit between its current simulation coordinates  $\mathbf{r}_i$  and those of the energy minimized crystal  $\mathbf{r}_{min,i}$

$$E_{\text{restraint}} = \sum_{i=1}^{N_{\text{atoms}}} \frac{k}{2} |\mathbf{r}_i - \mathbf{r}_{\text{min},i}|^2$$

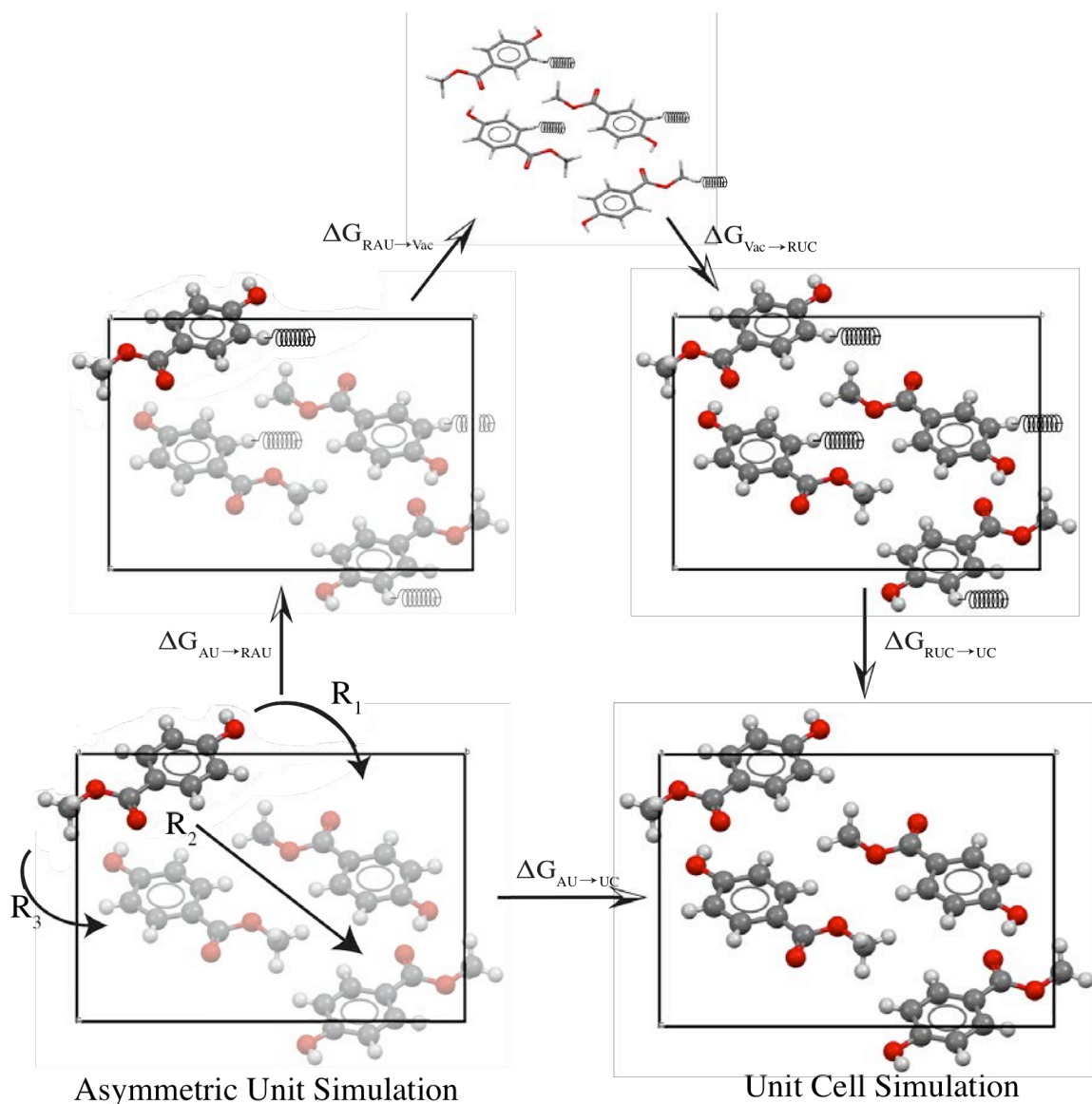
**Eq. 26**

with  $k$  chosen as 20 kcal/mol/Å. The 2<sup>nd</sup> step of the thermodynamic cycle computes the free energy change  $\Delta G_{\text{RAU} \rightarrow \text{vac}}$  to transfer the restrained asymmetric unit (RAU) molecules into vapor by alchemical annihilation method of turning off both intermolecular and intramolecular non-bonded interactions. The 3<sup>rd</sup> step computes the free energy change  $\Delta G_{\text{vac} \rightarrow \text{RUC}}$  to move from a restrained unit cell in vapor into the crystalline phase by alchemically restoring all non-bonded interactions. The final step of the thermodynamic cycle computes the free energy to remove the harmonic restraints from all atoms in the unit cell  $\Delta G_{\text{RUC} \rightarrow \text{UC}}$ , where the UC restraints are analogous to the AU restraints defined in **Eq. 26**. For the final two steps, the simulation system is composed of  $N_{\text{symm}}$  copies of the asymmetric unit, where  $N_{\text{symm}}$  is the number of symmetry operators of the space group. Therefore, the calculated free energy of these steps is normalized by  $N_{\text{symm}}$  to achieve the energy per asymmetric unit. Overall, the free energy change in moving from an asymmetric unit simulation to a unit cell simulation is given by

$$\Delta G_{\text{AU} \rightarrow \text{UC}} = \Delta G_{\text{AU} \rightarrow \text{RAU}} + \Delta G_{\text{RAU} \rightarrow \text{vac}} + \Delta G_{\text{vac} \rightarrow \text{RUC}} + \Delta G_{\text{RUC} \rightarrow \text{UC}}$$

**Eq. 27**

Instead of choosing the simulation system to be a single unit cell, a system with arbitrarily many copies of the unit cell would allow for coupled motions between unit cells. However, this is not explored in the current work.



**Figure 4.** Shown is a thermodynamic cycle for computing the free energy change  $\Delta G_{AU \rightarrow UC}$  between simulation systems composed of 1) an asymmetric unit and 2) a unit cell. In this case, the  $P2_1/c$  methyl paraben asymmetric unit is expanded into a  $P1$  unit cell.

For all steps, the free energy change was determined as the mean of five independent canonical ensemble (NVT) simulations per compound. (Schnieders et al., 2012) For each individual trajectory, stochastic dynamics (SD) was performed using a 1.0 femtosecond time step. Asymmetric unit deposition simulations  $\Delta G_{AU}$  were 200 nanoseconds in length, yielding a  $\mu\text{sec}$  of sampling per compound in aggregate. The components of the

$\Delta G_{AU \rightarrow UC}$  thermodynamic cycle converged after 10-50 nanoseconds. For all simulations, the volume was held constant during the simulations based on fixing the unit cell parameters at their experimental values. SD provided temperature fluctuations around a target of 298.15 K. Each  $\Delta G_{AU}$  alchemical simulation began at the experimental coordinates of the molecule in the crystalline state ( $\lambda=1$ ). We note that equivalent results can be obtained by starting each simulation from an arbitrary conformation in vacuum ( $\lambda=0$ ), although sampling convergence is superior in the former case.

### 3.1.5: Multiple Molecules in the Asymmetric Unit

Our approach also handles the case of multiple molecules within an asymmetric unit, which is necessary for salts, co-crystals and cases of non-crystallographic symmetry (NCS). This is accomplished by smoothly turning off all intermolecular interactions between molecules in the asymmetric unit as they transition from the crystalline state ( $\lambda=1$ ) into the vapor state ( $\lambda=0$ ) during alchemical simulations of  $\Delta G_{AU}$ . The effect in vapor is that the asymmetric unit molecules can pass through each other, while softcore repulsion gently removes molecular overlaps for  $\lambda > 0$ . For example, because there are two copies of ethyl paraben in the asymmetric unit due to NCS, the computed values of  $\Delta G_{AU}$  and  $\Delta G_{AU \rightarrow UC}$  must be normalized by a factor of two.

## 3.2: Results

### 3.2.1: Lattice Energies for AMOEBA and Dispersion

#### Corrected DFT

Currently, *ab initio* dynamics are much too expensive to reach the microsecond time scale required for convergence of GAUCHE as described here. On the other hand, our implementation of the polarizable AMOEBA force field, which models permanent

electrostatics on each atom using an ideal point multipole through quadrupole order and an ideal point induced dipole (P. Ren & Ponder, 2002), achieves 200 nanoseconds of sampling on a single compute node in a few days. However, it is important to quantify how reliable AMOEBA is for computing lattice potential energies relative to electronic structure methods such as dispersion corrected DFT. Shown in **Table 3** are lattice energies for the five compounds studied here based on AMOEBA and dispersion corrected B3LYP/6-31G\* with and without zero point energy (ZPE). AMOEBA lattice energies were systematically less favorable by 1.3 kcal/mol than dispersion corrected B3LYP/6-31G\* when the ZPE is not considered. After inclusion of ZPE changes between vacuum and condensed phase, AMOEBA lattice energies become systematically too favorable by 2.2 kcal/mol. This behavior is reasonable for a polarizable force field such as AMOEBA, whose empirical repulsion-dispersion energy term was fit to reproduce experimental liquid densities and heats of vaporization and thereby implicitly includes ZPE changes between vapor and condensed phase to some degree (P. Ren & Ponder, 2003, 2004; Pengyu Ren et al., 2011). On the other hand, substantial relative differences in ZPE energy between polymorphs (Rivera, Allis, & Hudson, 2008) could become a limiting factor to the accuracy of classical multipolar force fields (Gresh, Cisneros, Darden, & Piquemal, 2007; Piquemal, Cisneros, Reinhardt, Gresh, & Darden, 2006) for some crystal structure prediction applications.

**Table 3.** Comparison of the lattice energy between AMOEBA and dispersion corrected DFT (D-DFT) evaluated using the B3LYP functional and 6-31G\* basis set (kcal/mol). Energies were computed after minimization of the experimental coordinates in the respective potential.

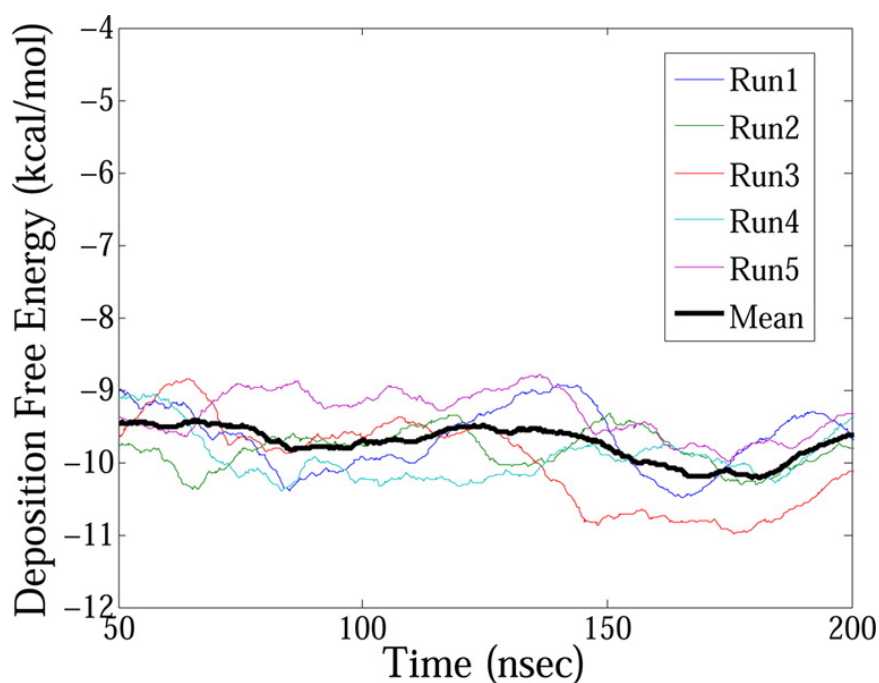
<b>Compound</b>	<b>AMOEBA</b>	<b>DFT-D</b>	<b>Abs. <math>\Delta E</math></b>	<b>DFT-D + ZPE</b>	<b>Abs. <math>\Delta E</math></b>
Acetanilide	-22.56	-24.20	1.64	-20.79	1.77
Paracetamol I	-27.69	-31.34	3.65	-27.19	0.50
Methylparaben II	-22.05	-23.10	1.05	-19.56	2.49
Ethylparaben	-24.58	-21.44	3.14	-18.64	5.94
Phenacetin	-26.12	-28.34	1.22	-24.92	1.20
<b>Mean</b>	-24.40	-25.68	2.34	-22.22	2.38

### 3.2.2: Alchemical Stochastic Dynamics

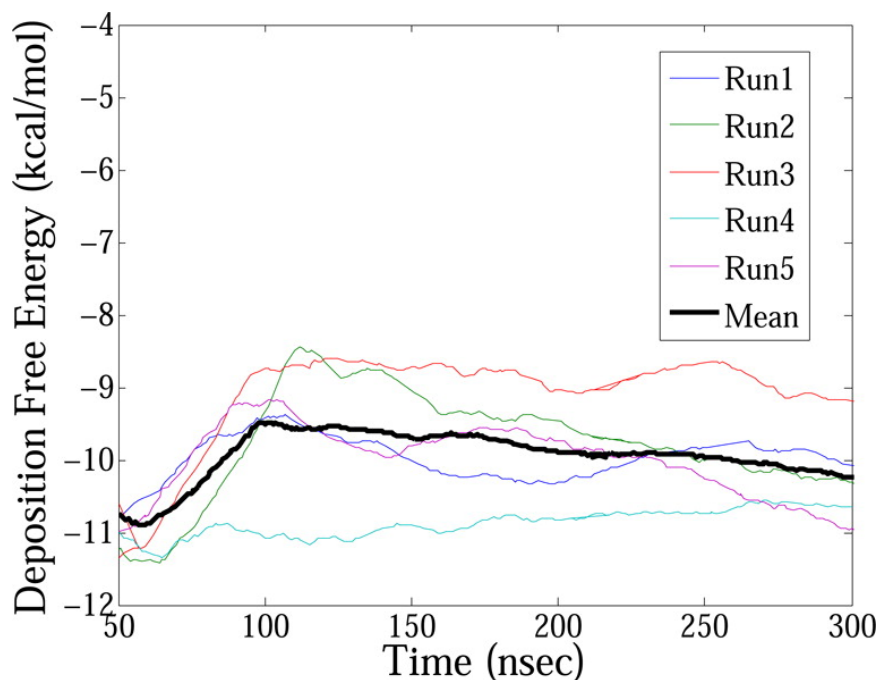
The deposition free energy was computed from two sets of simulations. First, the free energy difference  $\Delta G_{AU}$  between the vapor and crystalline states was computed from a simulation composed of only asymmetric unit molecule(s). The results for five independent simulations per crystal are given in **Table 5**. The largest statistical uncertainty is for ethyl paraben, which has the largest asymmetric unit as it contains two NCS copies of ethyl paraben. Overall, the average of the statistical uncertainty is only 0.55 kcal/mol, which is less than 5% of the average deposition free energy. Convergence of the five independent simulations for phenacetin and ethyl paraben can be observed in **Figure 5** and **Figure 6**, respectively.

**Table 4.** Free energy values  $\Delta G_{AU}$  (kcal mol<sup>-1</sup>) computed from five independent 200 ns simulations of the asymmetric unit along an alchemical path between vapor and crystalline states. The mean and standard deviation are given for each compound

Compound	Simulation					Mean	$\sigma$
	1	2	3	4	5		
Acetanilide	-12.32	-12.64	-12.25	-12.78	-12.09	-12.42	0.29
Paracetamol I	-14.57	-14.64	-13.52	-13.53	-13.66	-13.98	0.57
Methyl paraben II	-9.68	-9.83	-9.39	-10.12	-10.05	-9.81	0.30
Ethyl paraben	-13.47	-13.19	-10.51	-10.86	-11.40	-11.89	1.36
Phenacetin	-14.68	-14.14	-14.60	-14.46	-14.45	-14.47	0.21
<b>Mean</b>							0.55



**Figure 5.** Shown is the convergence of five independent simulations of the phenacetin asymmetric unit deposition free energy ( $\Delta G_{AU}$ ).



**Figure 6.** Shown is the convergence of five independent simulations of the ethyl paraben asymmetric unit deposition free energy ( $\Delta G_{AU}$ ).

The second set of simulations are combined to compute the favorable free energy  $\Delta G_{AU \rightarrow UC}$  to expand the crystalline state from a simulation system composed of only the  $N_{AU}$  asymmetric unit molecule(s) into a simulation system composed of all  $N_{UC}$  molecules in the unit cell. The latter unit cell simulation system permits entropic degrees of freedom, such as coupled motions between symmetry mates, which are impossible in the former. The results for each simulation needed to complete the thermodynamic cycle given in Eq. 27 and diagrammed in Figure 4 are given in Table 5. The first three contributions, including  $\Delta G_{AU \rightarrow RAU}$ ,  $\Delta G_{RAU \rightarrow Vac}$  and  $\Delta G_{Vac \rightarrow RUC}$ , converge in approximately 20 nsec, which is an order of magnitude faster than the alchemical vapor-solid phase transition. Their statistical uncertainty is generally less than 0.25 kcal/mol. However, the free energy to remove symmetry restraints from the unit cell molecules  $\Delta G_{RUC \rightarrow UC}$  proved to require sampling of at least 50 nsec, although this is still considerably less than  $\Delta G_{AU}$ .



**Table 5.** Shown are the free energy values (kcal/mol) for each simulation step in the thermodynamic cycle (**Figure 4**) used to compute the free energy change  $\Delta G_{AU \rightarrow UC}$  of moving from a simulation of the asymmetric unit to a simulation of the unit cell.

Compound	Component	Simulation					Mean	$\sigma$
		1	2	3	4	5		
Acetanilide	$\Delta G_{AU \rightarrow RAU}$	9.44	10.05	10.66	9.57	10.32	10.01	0.51
	$\Delta G_{RAU \rightarrow Vac}$	29.48	29.46	29.47	29.47	29.47	29.47	0.01
	$\Delta G_{Vac \rightarrow RUC}$	-29.48	-29.48	-29.48	-29.48	-29.47	-29.48	0.00
	$\Delta G_{RUC \rightarrow UC}$	-9.88	-10.14	-10.88	-9.95	-10.43	-10.26	0.41
	$\Delta G_{AU \rightarrow UC}$						-0.25	0.65
Paracetamol I	$\Delta G_{AU \rightarrow RAU}$	8.78	8.71	8.69	8.67	8.73	8.72	0.04
	$\Delta G_{RAU \rightarrow Vac}$	32.90	32.92	32.91	32.92	32.91	32.91	0.01
	$\Delta G_{Vac \rightarrow RUC}$	-33.01	-33.01	-33.01	-33.02	-33.00	-33.01	0.01
	$\Delta G_{RUC \rightarrow UC}$	-9.89	-10.39	-11.38	-9.95	-9.85	-10.29	0.64
	$\Delta G_{AU \rightarrow UC}$						-1.67	0.64
Methyl paraben II	$\Delta G_{AU \rightarrow RAU}$	8.28	8.44	8.40	8.33	8.35	8.36	0.06
	$\Delta G_{RAU \rightarrow Vac}$	26.14	26.16	26.14	26.14	26.15	26.14	0.01
	$\Delta G_{Vac \rightarrow RUC}$	-26.22	-26.22	-26.22	-26.22	-26.22	-26.22	0.00
	$\Delta G_{RUC \rightarrow UC}$	-9.76	-9.47	-9.92	-11.58	-9.41	-10.03	0.89
	$\Delta G_{AU \rightarrow UC}$						-1.74	0.89
Ethyl paraben	$\Delta G_{AU \rightarrow RAU}$	9.73	9.99	10.18	10.06	10.06	10.00	0.17
	$\Delta G_{RAU \rightarrow Vac}$	24.14	24.13	24.13	24.13	24.14	24.13	0.01
	$\Delta G_{Vac \rightarrow RUC}$	-24.14	-24.15	-24.14	-24.15	-24.14	-24.14	0.00
	$\Delta G_{RUC \rightarrow UC}$	-11.59	-10.53	-11.73	-11.27	-10.82	-11.19	0.51
	$\Delta G_{AU \rightarrow UC}$						-1.19	0.54
Phenacetin	$\Delta G_{AU \rightarrow RAU}$	12.51	12.36	12.83	12.24	12.67	12.52	0.23
	$\Delta G_{RAU \rightarrow Vac}$	30.20	30.21	30.21	30.18	30.20	30.20	0.01
	$\Delta G_{Vac \rightarrow RUC}$	-30.20	-30.20	-30.20	-30.20	-30.20	-30.20	0.00
	$\Delta G_{RUC \rightarrow UC}$	-15.93	-16.27	-13.56	-15.72	-16.51	-15.60	1.18
	$\Delta G_{AU \rightarrow UC}$						-3.08	1.20

Summing volume entropy  $\Delta G_{Vol}$ , asymmetric unit deposition  $\Delta G_{AU}$  and asymmetric unit expansion into the unit cell  $\Delta G_{AU \rightarrow UC}$  yields the standard state deposition free energy  $\Delta G_{dep}^0$  (**Eq. 21**) based on GAUCHE. Results for each compound are given in **Table 6**, along with a comparison to experimental sublimation free energy ( $\Delta G_{dep}^0 = -\Delta G_{sub}^0$ ).

Overall, the sublimation free energy values from GAUCHE exhibit a small mean signed error of -0.44 kcal/mol relative to experiment, which indicates the simulations slightly under-estimate crystal stability. Although the favorable free energy to further expand the P1 unit cell end state to a replicated super cell end state (i.e. a system built from 2 x 2 x 2 = 8 unit cells) may correct this, such a minor gain in accuracy did not justify the added computational expense in this case. The combined statistical uncertainty of the deposition free energy calculation is on average 1 kcal/mol, and for two of the five compounds, their uncertainties are larger than the magnitude of the error. The mean unsigned error of 1.24 kcal/mol and the root-mean-square error (RMSE) of 1.47 kcal/mol are comparable to that described for AMOEBA solvation free energy calculations,(Pengyu Ren et al., 2011; Y. Shi, Wu, Ponder, & Ren, 2010) which are promising results in the context of solubility predictions. The magnitude of the errors and the statistical uncertainty implies that the reduction in the statistical uncertainty may increase the accuracy of GAUCHE to be less than 1.0 kcal/mol. For the first time, GAUCHE offers an alchemical simulation protocol to complement existing solvation free energy methods and complete the solubility thermodynamic cycle.

**Table 6.** Calculated and experimental absolute sublimation free energies for each compound (kcal/mol).

<b>Compound</b>	$\Delta G_{Vol}$	$\Delta G_{AU}$	$\Delta G_{AU \rightarrow UC}$	<b>Calc.</b> $\Delta G_{sub}^o$	<b>Expt</b> $\Delta G_{sub}^o$	<b>Abs.</b> $\Delta E$	$\sigma$
Acetanilide	1.30	-12.42	-0.25	11.37	11.57	0.20	0.72
Paracetamol I	1.27	-13.98	-1.67	14.38	16.23	1.85	0.86
Methyl paraben II	1.32	-9.81	-1.74	10.23	11.98	1.75	0.94
Ethyl paraben	1.21	-11.89	-1.19	11.87	12.27	0.40	1.46
Phenacetin	1.14	-14.47	-3.08	16.41	14.39	2.02	1.22
<b>Mean</b>						1.24	1.04

### 3.2.3: AMOEBA Structures and Hydrogen-Bonding

#### Compared to Experiment

Optimization of stochastic dynamics snapshots from the OSRW/AMOEBA simulations produced conformations that had slightly more favorable potential energies relative to local optimization of experimental crystal structures. As shown in **Table 7**, the potential energy difference between the optimized experimental structures and the lowest potential energy found from optimizing stochastic dynamics snapshots was quite small (0.17 kcal/mol or less), except for phenacetin. In this case, the AMOEBA crystal minimum exhibits a 180° torsional flip relative to the experimental crystal structure for an aromatic ether functional group. Although ether groups have been studied during development of the AMOEBA force field (Pengyu Ren et al., 2011), further systematic work on aromatic ether may be needed.

**Table 7.** Lattice energies of the AMOEBA potential from 1) the minimized experimental coordinates and 2) the lowest potential energy found via minimization of simulation snapshots (kcal/mol).

Compound	Expt. Minimized	Snapshot Minimized	$\Delta E$
Acetanilide	-22.556	-22.725	-0.169
Paracetamol I	-27.694	-27.696	-0.002
Methyl paraben II	-22.063	-22.096	-0.033
Ethyl paraben	-24.580	-24.598	-0.018
Phenacetin	-26.124	-27.247	-1.123

Overall, hydrogen bond distances for the AMOEBA minimum energy structures are nearly identical to those of the experimental crystal structures, as shown in **Table 8**. The mean N-H...O distance for the experimental crystal structures and AMOEBA minima were 2.943 Å and 2.946 Å, respectively. The mean O-H...O distance for the experimental crystal structures and AMOEBA minima were 2.738 Å and 2.754 Å, respectively. The

discrepancies for N-H..O and O-H..O of 0.003 Å and 0.012 Å, respectively, are of similar magnitude to experimental uncertainty and refinement artifacts.

**Table 8.** Hydrogen-bond distances (Å) for the experimental crystal structure, for AMOEBA minimization of the experimental structure, and for AMOEBA minimization of the most favorable snapshot from the alchemical simulations.

Compound	H-Bond	Expt.	Expt. Minimized	Snapshot Minimized
Acetanilide	N-H..O	2.943	3.030	2.901
Paracetamol I	N-H..O	2.934	2.971	2.972
	O-H..O	2.663	2.728	2.728
Methyl paraben II	O-H..O	2.687	2.727	2.727
	O-H..O	2.864	2.799	2.808
Phenacetin	N-H..O	2.953	3.086	2.964
Mean	N-H..O	2.943	3.029	2.946
Mean	O-H..O	2.738	2.751	2.754

### 3.2.4: A Comparison of Lattice Energy, End-State, and

#### Alchemical Approaches

We conclude the results section with a brief comparison of lattice energy, end-state and alchemical approaches for predicting the stability of organic crystals as shown in **Table 9**. Lattice potential energy is the simplest metric and is often used for polymorph prediction; however, it completely neglects entropic effects and the temperature dependence of stability. On average, this approach overestimates stability by 11.3 kcal/mol at room temperature. An end-state approach based on single structure estimates of vibrational entropies offers an inexpensive, but approximate framework to begin accounting for entropic changes between vapor and crystalline end states. In this case, the stability is underestimated on average by 8.2 kcal/mole. Finally, GAUCHE offers a rigorous way to account for entropic contributions to the stability of organic crystals as a

function of temperature. Although more expensive than lattice potential energy and end-state approaches by a factor of  $\sim 10^9$ , the mean signed error is below 0.5 kcal/mol.

**Table 9.** Negative of the AMOEBA lattice energy, the negative of the end-state approximation  $\Delta G_{ES}^0$  for deposition free energy, the GAUCHE prediction for sublimation free energy  $\Delta G_{GAUCHE}^0$  and the experimental sublimation free energy (kcal/mol).

<b>Compound</b>	<b><math>-E_{Lattice}</math></b>	<b><math>-\Delta G_{ES}^0</math></b>	<b><math>\Delta G_{GAUCHE}^0</math></b>	<b>Expt.</b>
Acetanilide	22.56	4.17	11.37	11.57
Paracetamol I	27.69	7.58	14.38	16.23
Methylparaben II	22.06	4.58	10.23	11.98
Ethylparaben	24.58	3.20	11.87	12.27
Phenacetin	26.12	6.11	16.41	14.39
<b>RMSE</b>	11.34	8.19	1.70	

## CHAPTER 4: RELATIVE ORGANIC CRYSTAL THERMODYNAMICS

For applications such as protein-ligand binding affinity and protein folding stability, it has been established that relative thermodynamic calculations are more computationally affordable and have lower statistical uncertainty than absolute thermodynamic calculations. Thus, there should exist a relative thermodynamic path between polymorphs that is substantially more efficient to compute compared to simply taking the difference between two absolute sublimation free energy calculations. To calculate the difference in sublimation free energy for two paracetamol polymorphs, a relative thermodynamic calculation method is a logical choice over an absolute method. From the absolute thermodynamics study in Chapter 3, it became clear that using two absolute calculations to compare the difference in thermodynamic stability of two polymorphs of paracetamol would be quite expensive to achieve low statistical uncertainty. For example, the absolute error for paracetamol form I was  $1.85 \text{ kcal mol}^{-1}$  with standard deviation of  $0.86 \text{ kcal mol}^{-1}$ , but the experimental free energy difference between form I and II of paracetamol is only  $0.93 \text{ kcal mol}^{-1}$  (Perlovich et al., 2007). Consequently, there is a practical need to find a relative thermodynamic path that efficiently calculates the change in free energy between two crystal polymorphs.

### 4.1: Methods

#### 4.1.1: Lattice Energies

Optimized lattice potential energies using AMOEBA were determined for two paracetamol polymorphs. To determine the sensitivity of crystal stability rank order as a

function of temperature (20, 100, 298, 330 and 360K), we calculated the lattice energies for experimentally determined paracetamol polymorphs with unit cell parameters fixed.

$$U_{\text{lattice}} = U_{\text{cryst}} - U_{\text{vac}}$$

**Eq. 28**

**Table 10.** Crystallographic information for the two stable paracetamol polymorphs including CSD reference code, space group, unit cell parameters, number of molecules in a unit cell, volume per molecule ( $\text{\AA}^3$ ) and experimental temperature (Kelvin). The  $\alpha$  and  $\gamma$  lattice angles are 90 degrees for each crystal.

Compound	CSD Code	Space	<i>a</i>	<i>b</i>	<i>c</i>	$\beta$	<i>Z</i>	Vol/ <i>Z</i>	Temp
		Group							
Paracetamol I	HXACAN01	P21/c	11.72	9.40	12.93	147.0	4	194.07	298
Paracetamol II	HXACAN23	Pbca	11.84	7.41	17.16	90	8	188.06	298

**Table 11.** Crystallographic information determined experimentally at extreme temperatures. The  $\alpha$  and  $\gamma$  lattice angles are 90 degrees for each crystal. These parameters were fixed during lattice energy determination. Listed are CSD code, space group, a-, b-, c-axis ( $\text{\AA}$ ),  $\beta$  (degrees), experimental temperature and density.

	CSD Code	Space	<i>a</i>	<i>b</i>	<i>c</i>	$\beta$	Temp (K)	$\rho$ ( $\text{g/cm}^3$ )
		Group						
Form I <sub>low</sub>	HXACAN13	P21/c	11.55	9.17	12.67	146.4	20	1.35
Form I <sub>high</sub>	HXACAN19	P21/c	11.71	9.37	12.87	146.9	330	1.30
Form II <sub>low</sub>	HXACAN21	Pbca	11.78	7.20	17.18	90	100	1.38
Form II <sub>high</sub>	HXACAN24	Pbca	11.85	7.48	17.16	90	360	1.32

#### 4.1.2: Relative Free Energy Difference between Two

##### Paracetamol Polymorphs

The five-step method to calculate the relative free energy difference between two polymorphs of paracetamol was inspired by the GAUCHE method previously developed by our group (Park et al., 2014). For example, the technique of adding harmonic restraints to promote rapid convergence of each step in the thermodynamic cycle was leveraged.

The equation for our harmonic restraints is given by

$$E_{\text{restraint}} = \sum_{i=1}^{N_{\text{atoms}}} \frac{k}{2} |\mathbf{r}_i - \mathbf{r}_{\text{min},i}|^2$$

**Eq. 29**

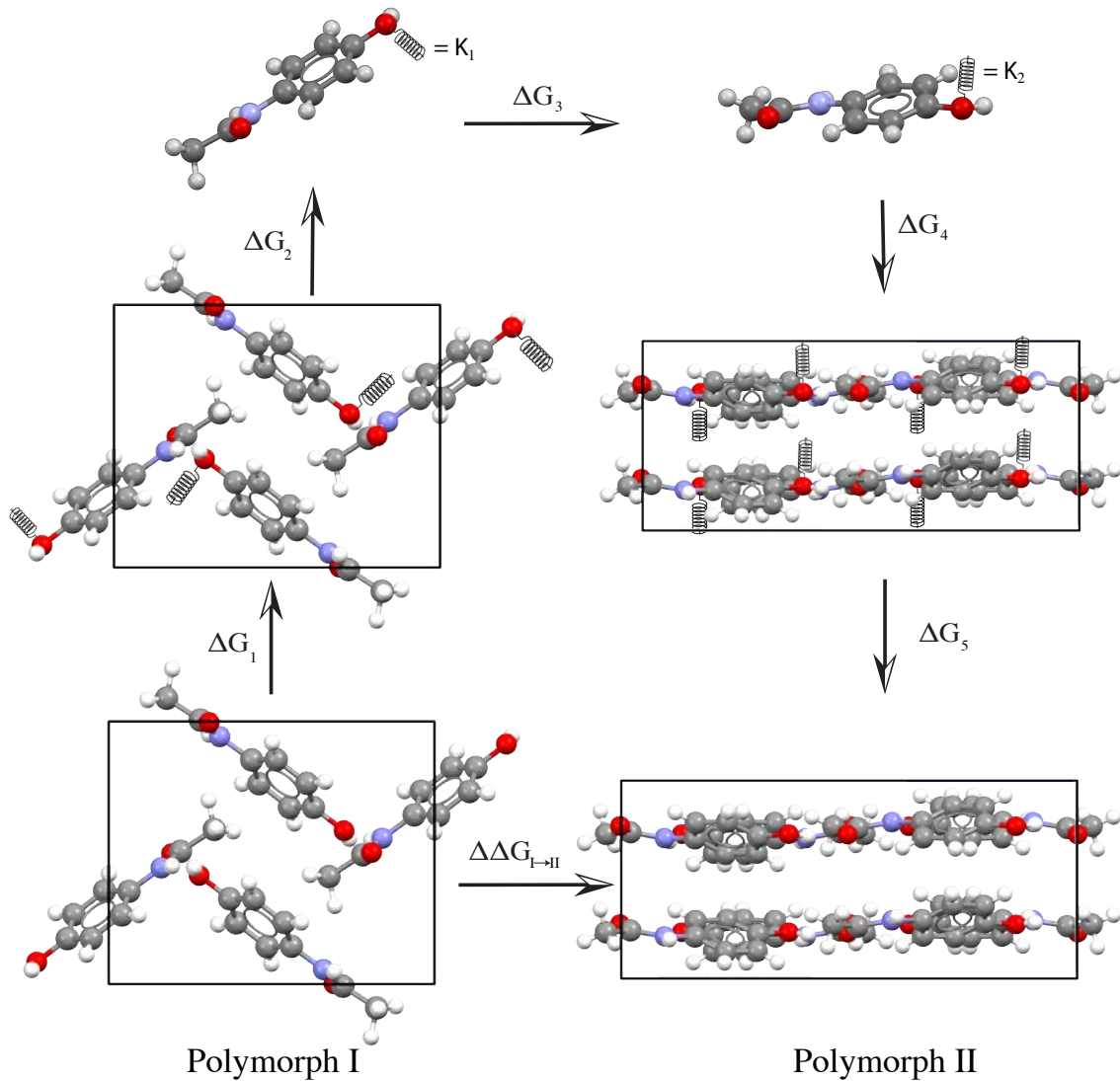
where  $k$  is a spring constant with magnitude  $20 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ ,  $\mathbf{r}_i$  is a vector of the current atomic coordinates for atom  $i$  and  $\mathbf{r}_{\text{min},i}$  is a vector of the initial atomic coordinates.

The crystalline conformations of the paracetamol polymorphs are not identical. However, to complete the thermodynamic cycle between the polymorphs, a common reference state is needed. Vapor energetics, which rely solely on the intramolecular forces of the molecule, cannot include heavy atom restraints based on each crystal polymorph conformation since the two conformations are not identical. Thus, instead of adding restraints to every heavy atom, we add restraints to the planar portion of the molecule. In the three-dimensional space, three points that are not collinear define a plane. Using this concept, we devised a three-atom restraint for the vapor reference state. By selecting three atoms of a given molecule, we could converge the simulation efficiently while sampling restrained ensemble of vapor conformations for each polymorph. In defining of a plane, three atoms in the aromatic ring of paracetamol were chosen. Although other atoms can be used to restrain the molecule, selecting the aromatic carbons was a natural choice given the planarity of this functional group. Several other thermodynamic paths using alternative restraints were explored and were discovered to be less efficient. In particular, with a restraint on only the molecule's center of mass, its rotational degrees of freedom remain in vapor and lead to a similar sampling challenge to that overcome by the GAUCHE method during transition into the crystalline environment.



$$\Delta G_{P1 \rightarrow P2} = \Delta G_{P1 \rightarrow RP1} + \Delta G_{RP1 \rightarrow RV1} + \Delta G_{RV1 \rightarrow RV2} + \Delta G_{RV2 \rightarrow RP2} + \Delta G_{RP2 \rightarrow P2}$$

Eq. 30



**Figure 7.** Shown is a diagram for an efficient five-step thermodynamic cycle for computing the relative free energy change between two polymorphs ( $\Delta\Delta G_{I-II}$ ). The steps include adding harmonic restraints to polymorph I ( $\Delta G_1$ ), turning off polymorph I intermolecular interactions to reach a plane-restrained vapor I ( $\Delta G_2$ ), transitioning to a plane-restrained vapor II ( $\Delta G_3$ ), turning on polymorph II intermolecular interactions ( $\Delta G_4$ ) and turning off harmonic restraints ( $\Delta G_5$ ).

### 4.1.3: Entropic and Enthalpic Contributions to Free Energy

In this thesis, two complementary mechanisms are used to calculate the entropic and enthalpic contributions to the free energy change. One method exclusively solves for enthalpy by calculating the ensemble average of the potential energy over an MD trajectory, while the second solves for entropy using the finite difference method.

The enthalpy of sublimation can be defined as the difference in the ensemble average of the crystal and vapor potential energies

$$\Delta H_{sub} = \langle U_{cryst} \rangle - \langle U_{vap} \rangle$$

**Eq. 31**

Potential energies for crystal and vapor phases were collected during 5 nanoseconds MD trajectories. The last 4 nanoseconds were divided into five windows (0.8 nanoseconds each), and the ensemble average was calculated for each window. The statistical uncertainties were determined by calculating the standard deviations of ensemble averages of five windows.

The relative enthalpy difference was determined by calculating the difference between the sublimation enthalpy of two crystalline forms. The entropic contribution to the relative free energy is calculated by subtracting the relative free energy from the relative enthalpy change.

$$\Delta\Delta H_{P1 \rightarrow P2} = \Delta H_{sub,P2} - \Delta H_{sub,P1}$$

**Eq. 32**

$$T\Delta S = \Delta H - \Delta G$$

**Eq. 33**

The change in entropy as a function of change in free energy (Kubo, Gallicchio, & Levy, 1997) is

$$\Delta S = - \left( \frac{\partial \Delta A}{\partial T} \right)_{N,V}$$

**Eq. 34**

The Helmholtz free energy calculated via NVT molecular dynamics simulation is equal to the Gibbs free energy in NPT simulation under the assumption that the finite size effects are relatively insignificant (Yue Shi, Zhu, Martin, & Ren, 2012). Centered finite-difference calculation of  $\Delta S$  (Smith & Haymet, 1993) is given by

$$\Delta S \approx - \left( \frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{2\Delta T} \right)$$

**Eq. 35**

where  $\Delta T$  was chosen to be 10 K. Once entropy is solved, the enthalpic contributions can be calculated as the sum of free energy and entropy.

$$\Delta H = \Delta G + T\Delta S$$

**Eq. 36**

## 4.2: Results

### 4.2.1: Lattice Energies

Quantum calculations from the literature have shown that DFT-D with the B3LYP functional and 6-31G\*\* basis set ranks the stability of the two paracetamol polymorphs in the wrong order by 1.06 kcal/mol (Li & Feng, 2006) compared to experimental free energy differences that show form I is more stable by 0.93 kcal/mol than form II (Perlovich et al., 2007). Local optimization of the experimental crystal structure using the polarizable AMOEBA force field ranks the polymorphs in the same order as that of DFT-D, although in this case the lattice energy difference is approximately half the magnitude

(0.52 kcal/mol). AMOEBA potential energies calculated from the structures found at extreme temperatures show only a few hundredths of a kcal/mol difference, but the stability ranking based on lattice energies consistently favor form II over form I. Alternatively, a CSP procedure that uses the Dreidig force field that includes higher-order fixed multipoles shows the same stability ranking as that of the experiment with just the potential energy (Beyer, Day, & Price, 2001). Thus, inclusion of entropic effects might over-stabilize Form I relative to Form II. Although both DFT-D and AMOEBA have some limitations, their lattice-energy-based ranking raises an interesting question of whether their incorrect stability ranking is an inherent flaw of the force field or due to the approximation of calculating thermodynamic stability using lattice energy rather than free energy. Thus, for potential functions that reverse the stability without consideration of entropy, can the correct order under ambient conditions be recovered by calculating free energy? For many condensed phase systems, it is widely appreciated that relative thermodynamic stability at 298 K differs from that at 0 K, although a change in stability ranking as a function of temperature has not been previously demonstrated computationally for polymorphs.

**Table 12.** Comparison of the lattice potential energies between AMOEBA, <sup>1</sup>Dreidig (Beyer et al., 2001) and <sup>2</sup>dispersion corrected DFT with the B3LYP functional and 6-31G\*\* basis set (Li & Feng, 2006) (kcal/mol).

Compound	AMOEBA			$U_{\text{lattice, Dreidig}}^1$	$U_{\text{lattice, DFT-D}}^2$
	$U_{\text{crystal}}$	$U_{\text{vapor}}$	$U_{\text{lattice}}$		
Paracetamol I	-41.04	-13.35	-27.69	-27.49	-31.94
Paracetamol II	-41.56	-13.35	-28.21	-26.59	-33.00
$\Delta U_{\text{lattice}}$			-0.52	0.90	-1.06

**Table 13.** Lattice energies of crystal structures, optimized using the AMOEBA force field. The initial structures were experimentally determined at high and low extreme temperatures as shown in **Table 11**. The vapor energy in each case was -13.35. All energies are in kcal/mol.

	CSD Code	Space Group	Temp (K)	$U_{\text{lattice}}$
Form I <sub>low</sub>	HXACAN13	P21/c	20	-27.83
Form I <sub>high</sub>	HXACAN19	P21/c	330	-27.76
Form II <sub>low</sub>	HXACAN21	Pbca	100	-28.19
Form II <sub>high</sub>	HXACAN24	Pbca	360	-28.14

#### 4.2.2: Relative Free Energy Difference between Two Polymorphs

By using a variant of metadynamics called Orthogonal Space Random Walk (OSRW), we defined a five-step procedure to calculate the relative free energy difference between two polymorphs. Starting from experimental crystal structures solved under ambient conditions, we construct and sample a thermodynamic path between the two polymorphs. Harmonic restraints are used on intermediate steps to promote efficient convergence of the simulations and to minimize statistical uncertainty. In the vapor phase, a planar restraint is applied using the coordinates of three atoms to avoid conformational sampling of translational and rotational degrees of freedom. Although the same force field is used for both the lattice energy calculations and free energy calculations, rank order of stability changes nonlinearly as a function of temperature. The relative free energy difference between the two polymorphs was determined at six different temperatures ranging from 0 to 308.15 K and compared with the experimental free energy differences. As shown in **Figure 8**, experimental thermodynamic differences show a parabolic trend for the relative free energy difference in the temperature range. The relative free energy difference between paracetamol polymorphs at 0 K is simply equated to the lattice

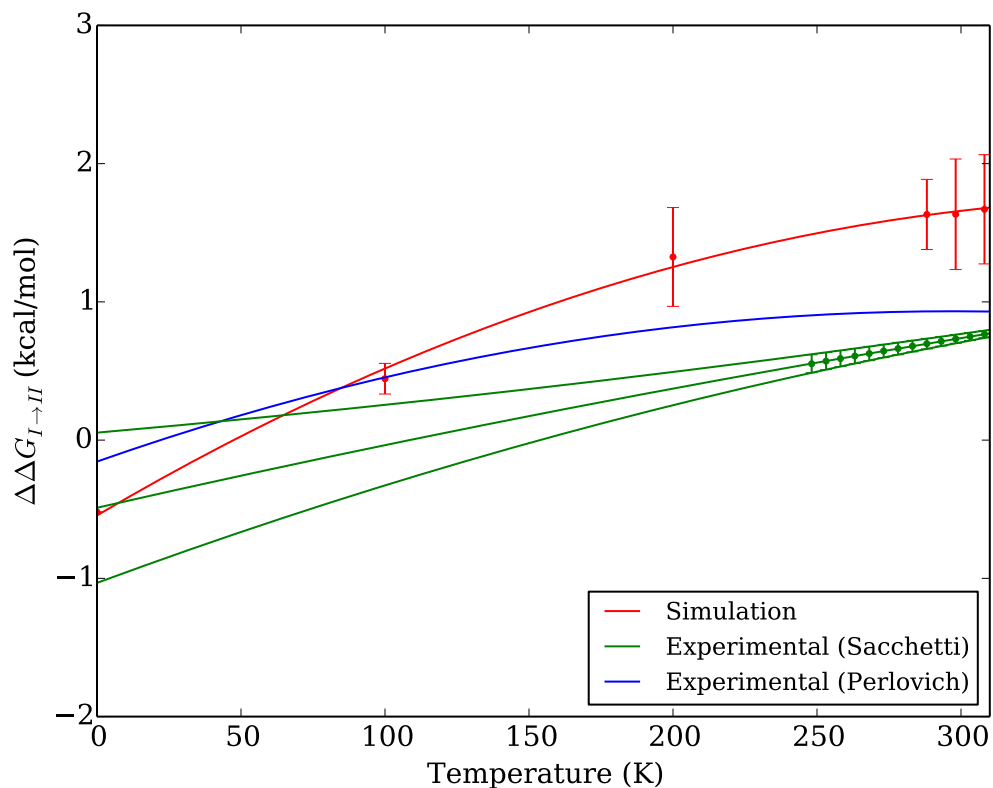
potential energy difference with all other physical contributions neglected. The relative thermodynamic difference computed using the path described above also shows a parabolic trend over the temperature range.

**Table 14.** Five simulations are reported for each step of the relative polymorph free energy thermodynamic cycle that changes paracetamol from polymorph I to polymorph II for three temperatures (100, 200, 288.15, 298.15 and 308.15 Kelvin).

Component	100	200	288.15	298.15	308.15
$\Delta G_{P1 \rightarrow RP1}$	1.97±0.10	4.13±0.34	5.55±0.23	5.74±0.39	5.80±0.39
$\Delta G_{RP1 \rightarrow RV1}$	33.35±0.00	32.27±0.02	31.24±0.02	31.12±0.02	31.00±0.05
$\Delta G_{RV1 \rightarrow RV2}$	-0.33±0.04	-0.36±0.03	-0.37±0.04	-0.39±0.05	-0.39±0.02
$\Delta G_{RV2 \rightarrow RP2}$	-1.39±0.03	-2.96±0.06	-4.31±0.07	-4.44±0.07	-4.54±0.05
$\Delta G_{RP2 \rightarrow P2}$	-33.16±0.01	-31.75±0.06	-30.48±0.06	-30.40±0.04	-30.21±0.05
$\Delta \Delta G_{P1 \rightarrow P2}$	0.44±0.11	1.33±0.36	1.65±0.25	1.63±0.40	1.66±0.40

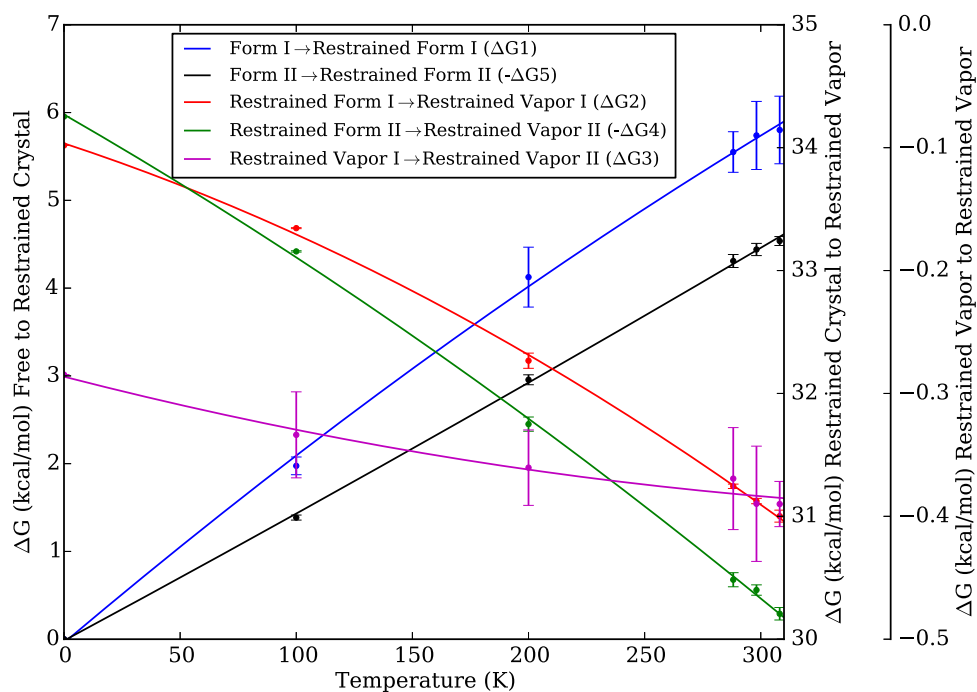
Relative free energy calculations are more efficient and exhibit lower statistical uncertainty than directly using the difference between absolute calculations for determination of the relative stability of polymorphs. The absolute error and the standard deviation for paracetamol form I were 1.85 and 0.86 kcal/mol, respectively. Although these uncertainty values are acceptable for absolute calculations, the error and the standard deviation need to be significantly smaller for relative free energy differences between polymorphs because the magnitude of the relative free energy between polymorphs is often on the order of ~0.5 kcal/mol. In the case of paracetamol, the experimental free energy difference between the two most stable polymorphs is only 0.93 kcal/mol. Consequently, the calculation of relative free energy by taking the difference between two absolute free energies becomes too expensive for statistically meaningful results.

The largest statistical uncertainty consistently came from the first step of the five-step procedure: transition from free to restrained form I. Since van der Waals interactions are turned off in the vapor phase, free crystal to restrained crystal has significantly more complexity in computational evaluation of energetics than restrained crystal to vapor transition. In addition to energetic complexity, there seems to be greater translational and vibrational degrees of freedom in form I than form II. Greater entropic freedom is expected based on calculating the volume occupied per molecule for each polymorph (**Table 10**). There is experimental and computational evidence that form I is more entropically favorable than form II. The greater the conformation degrees of freedom, the more difficult it is to converge a system to low statistical uncertainty.



**Figure 8.** The relative free energy difference between polymorph I and polymorph II as a function a temperature is plotted for two experimental approaches and from molecular dynamics simulation using the AMOEBA force field. Experimental data from Sacchetti includes 95% confidence interval (Sacchetti, 2000), and Perlovich data is an extrapolation based on Perlovich and Sacchetti data (Perlovich et al., 2007).





**Figure 9.** Five components of relative free energy difference between polymorph I and polymorph II described in **Figure 7** as a function of temperature. All values are reported in kcal/mol.

#### 4.2.3: Enthalpic and Entropic Contribution to Free Energy

Using the ensemble averages of potential energies and central finite difference methods, entropy and enthalpy were calculated (Yue Shi et al., 2012). Calculation of the change in enthalpy showed that almost all of the favorable free energy of form I was due to a favorable increase in entropy relative to form II. The relative free energy differences were calculated at six different temperatures, three of which were used to decompose free energy into enthalpic and entropic contributions at 298.15 K by the finite difference method. The temperatures were chosen to be 10 K above and below room temperature under the assumption that enthalpy and entropy are essentially constant over this temperature range.

Although the experimental literature and our simulation data do not agree on the exact magnitude of enthalpic and entropic contribution to the free energy difference at room temperature, the trends in the data are similar. The transition from paracetamol I to paracetamol II is thermodynamically unfavorable at room temperature, and a decrease in entropy is the most significant contribution to this unfavorable thermodynamics.

**Table 15.** Two experimental decompositions (Perlovich et al., 2007; Sacchetti, 2000) of the relative free energy difference to change polymorph I into polymorph II at room temperature are compared to finite difference (FD) and enthalpy calculated simulation approaches, all reported in kcal/mol.

Source	$\Delta\Delta G$	$\Delta\Delta H$	$-T\Delta\Delta S$
Perlovich	0.93±0.19	0.48±0.19	0.71±0.18
Sacchetti	0.73	-0.30	1.04
FD Simulation	1.63±0.40	1.08	0.55
Enthalpy Simulation	1.63±0.40	0.12±0.03	1.41

## CHAPTER 5: CONCLUSION

### 5.1: Absolute Organic Crystal Thermodynamics

The goal of the absolute organic crystal thermodynamics work was to describe an efficient alchemical method for computing organic crystal thermodynamics that is analogous to solvation free energy methods. Five organic molecules were chosen whose experimental sublimation free energy has been measured (**Table 1** & **Table 2**). Our method begins by computing the sublimation-deposition phase transition free energy  $\Delta G_{AU}$  based on a simulation system composed of only the asymmetric unit molecules (

**Table 4).** Next, we demonstrated a novel strategy to compute the free energy change between a crystalline simulation system with only  $N_{AU}$  asymmetric unit molecules and one with  $N_{UC}$  unit cell molecules. The magnitude of the correction  $\Delta G_{AU \rightarrow UC}$  (**Table 5, Eq. 27**) was the smallest for acetanilide, the smallest molecule studied, and the largest for phenacetin, the largest molecule studied. This is consistent with the premise that as the asymmetric unit of the crystal increases so does the importance of entropic contributions to stability. Overall, the GAUCHE approach achieved a RMSE of 1.47 kcal/mol and a mean signed and unsigned errors of -0.44 kcal/mol and 1.24 kcal/mol, respectively.

Compared to alchemical approaches that begin from the unit cell, GAUCHE has the appealing advantage of not requiring *a priori* knowledge of crystalline coordinates determined from X-ray crystallography experiments. However, this work focused on NVT simulations, which do require knowledge of the space group and unit cell parameters. A future direction is to extend GAUCHE to the NPT ensemble, and thereby eliminate reliance on all experimental information. For example, the most favorable deposition free energy from NPT simulations of all reasonable space groups (i.e. those with mirror planes eliminated for chiral molecules) would predict the space group of the most stable polymorph, its unit cell parameters and its structural ensemble.

## 5.2: Relative Organic Crystal Thermodynamics

The relative organic crystal thermodynamics method sought to rank the stability of organic polymorphs according to free energy rather than potential energy. Paracetamol was chosen due to the availability of experimental measurements of the relative sublimation thermodynamics for the its two most stable polymorphs as a function of temperature. Additionally, there are both experimental and computational disagreements

on the relative stability of the two polymorphs that need to be rationalized. Under ambient conditions, paracetamol form I is known to be more thermodynamically stable than form II by 0.93 kcal/mol. Both DFT-D and AMOEBA potential energies rank form II as more stable than form I, which is the opposite of room temperature experimental thermodynamics. However, since the potential energy ranking does not account for entropy, the magnitude of stability differences and even the rank order are expected to change as a function of temperature. We have demonstrated a novel approach to efficiently compute relative thermodynamics as a function of temperature for paracetamol forms I and II. Although the simulation magnitude of the free energy change modestly differed from the experimental magnitude, the parabolic trend over the temperature range from 0 to 308 K matched, as well as the rank order under ambient conditions. The exact temperature at which the stability ranks change is disagreed upon in the literature (Perlovich et al., 2007; Sacchetti, 2000). Some experimental results even indicate there is no change in stability ranking at any temperature under ambient pressure (Espeau et al., 2005).

Although experimental results are inconclusive, the overall free energy trend and favorable entropic change transitioning from form II to I are agreed upon in the experimental literature and our simulation data. Our data quantitatively demonstrate the contribution of entropy to calculating and ranking the stability of multiple solid phases. By defining a relative free energy method, we were able to reduce the statistical uncertainty by more than a factor of 2. From our absolute approach, the sublimation free energy of paracetamol form I had a standard deviation of 0.86 kcal/mol, which is 92% of the experimental free energy difference between the two polymorphs. Such a large

statistical uncertainty made the absolute approach impractical for calculating small stability differences. However, with the relative method, the standard deviation of the combined five steps was only 0.40 kcal/mol.

By demonstrating that lattice potential energy and lattice free energy rank polymorph stability differently, we can conclude that entropy is essential to both thermodynamic stability ranking and crystal structure prediction. Although CSP has traditionally explored a potential energy surface for its global minimum, this work makes clear that the most thermodynamically stable structure is not necessarily the lowest potential structure. For example, the AMOEBA potential energies for paracetamol polymorphs and the AMOEBA free energies show a reversal rank ordering that agrees with experiment.

## REFERENCES

- Bardwell, D. A., Adjiman, C. S., Arnautova, Y. A., Bartashevich, E., Boerrigter, S. X. M., Braun, D. E., & Zhitkov, I. K. (2011). Towards Crystal Structure Prediction of Complex Organic Compounds - a Report on the Fifth Blind test. *Acta Crystallographica. Section B: Structural Crystallography and Crystal Chemistry*, 67(6), 535-551.
- Becke, A. D. (1993). Density-functional thermochemistry. III. The Role of Exact Exchange. *The Journal of Chemical Physics*, 98(7), 5648-5652.
- Beyer, T., Day, G. M., & Price, S. L. (2001). The Prediction, Morphology, and Mechanical Properties of the Polymorphs of Paracetamol. *Journal of the American Chemical Society*, 123(21), 5086-5094.
- Beyer, T., & Price, S. L. (2000). Dimer or Catemer? Low-Energy Crystal Packings for Small Carboxylic Acids. *The Journal of Physical Chemistry B*, 104(12), 2647-2655.
- Brown, C. (1966). Further Refinement of the Crystal Structure of Acetanilide. *Acta Crystallographica*, 21(3), 442-445.
- Broyden, C. G. (1970). The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1), 76-90.
- Civalleri, B., D'Arco, P., Orlando, R., Saunders, V. R., & Dovesi, R. (2001). Hartree-Fock Geometry Optimisation of Periodic Systems with the Crystal Code. *Chemical Physics Letters*, 348(1-2), 131-138.
- Colletier, J.-P., Laganowsky, A., Landau, M., Zhao, M., Soriaga, A. B., Goldschmidt, L., & Eisenberg, D. (2011). Molecular Basis for Amyloid- $\beta$  Polymorphism. *Proceedings of the National Academy of Sciences*, 108(41), 16938-16943.
- Darden, T., York, D., & Pedersen, L. (1993). Particle-Mesh Ewald - An  $n \log(n)$  Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*, 98(12), 10089-10092.
- Day, G. M., Cooper, T. G., Cruz-Cabeza, A. J., Hejczyk, K. E., Ammon, H. L., Boerrigter, S. X. M., & Schweizer, B. (2009). Significant Progress in Predicting the Crystal Structures of Small Organic Molecules - a Report on the Fourth Blind Test. *Acta Crystallographica Section B*, 65(2), 107-125.
- Day, G. M., Motherwell, W. D. S., Ammon, H. L., Boerrigter, S. X. M., Della Valle, R. G., Venuti, E., & Verwer, P. (2005). A Third Blind Test of Crystal Structure Prediction. *Acta Crystallographica Section B*, 61(5), 511-527. doi:10.1107/S0108768105016563
- Ditchfield, R., Hehre, W. J., & Pople, J. A. (1971). Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics*, 54(2), 724-728.
- Doll, K. (2001). Implementation of Analytical Hartree-Fock Gradients for Periodic Systems. *Computer Physics Communications*, 137(1), 74-88.
- Doll, K., Saunders, V. R., & Harrison, N. M. (2001). Analytical Hartree-Fock Gradients for Periodic Systems. *International Journal of Quantum Chemistry*, 82(1), 1-13.
- Dovesi, R., Orlando, R., Civalleri, B., Roetti, C., Saunders Victor, R., & Zicovich-Wilson Claudio, M. (2005). CRYSTAL: a Computational Tool for the Ab Initio Study of the Electronic Properties of Crystals. *Zeitschrift für Kristallographie* (Vol. 220, pp. 571).
- Dovesi, R., Saunders, V. R., Roetti, R., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., & Llunell, M. (2009). CRYSTAL09 User's Manual. Torino: University of Torino.
- Espeau, P., Céolin, R., Tamarit, J.-L., Perrin, M.-A., Gauchi, J.-P., & Leveiller, F. (2005). Polymorphism of Paracetamol: Relative Stabilities of the Monoclinic and Orthorhombic Phases Inferred from Topological Pressure Temperature and

- Temperature-Volume Phase Diagrams. *Journal of Pharmaceutical Sciences*, 94(3), 524-539.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A Smooth Particle-Mesh Ewald Method. *Journal of Chemical Physics*, 103(19), 8577-8593.
- Fenn, T. D., & Schnieders, M. J. (2011). Polarizable Atomic Multipole X-ray Refinement: Weighting Schemes for Macromolecular Diffraction. *Acta Crystallographica Section D*, 67(11), 957-965.
- Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. *The Computer Journal*, 13(3), 317-322.
- Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109), 23-26.
- Gresh, N., Cisneros, G. A., Darden, T. A., & Piquemal, J. P. (2007). Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *Journal of Chemical Theory and Computation*, 3(6), 1960-1986.
- Grimme, S. (2011). Density Functional Theory with London Dispersion Corrections. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2), 211-228.
- Haisa, M., Kashino, S., Kawai, R., & Maeda, H. (1976). The Monoclinic Form of p-Hydroxyacetanilide. *Acta Crystallographica Section B*, 32(4), 1283-1285.
- Joiris, E., Martino, P., Berneron, C., Guyot-Hermann, A.-M., & Guyot, J.-C. (1998). Compression Behavior of Orthorhombic Paracetamol. *Pharmaceutical Research*, 15(7), 1122-1130.
- Jorgensen, W. (1985). Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.*, 83(6), 3050.
- Jorgensen, W. L., & Duffy, E. M. (2002). Prediction of Drug Solubility from Structure. *Advanced Drug Delivery Reviews*, 54(3), 355-366.
- Kubo, M. M., Gallicchio, E., & Levy, R. M. (1997). Thermodynamic Decomposition of Hydration Free Energies by Computer Simulation: Application to Amines, Oxides, and Sulfides. *The Journal of Physical Chemistry B*, 101(49), 10527-10534.
- Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B*, 37(2), 785.
- Li, T., & Feng, S. (2006). Empirically Augmented Density Functional Theory for Predicting Lattice Energies of Aspirin, Acetaminophen Polymorphs, and Ibuprofen Homochiral and Racemic Crystals. *Pharmaceutical Research*, 23(10), 2326-2332.
- Lin, X. (1986). The Structure of Ethyl Paraben. *Chinese Journal of Structural Chemistry*, 5, 281.
- Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., & Williams, D. E. (2000). A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Crystallographica Section B*, 56(4), 697-714.
- Lybrand, T. P., Ghosh, I., & McCammon, J. A. (1985). Hydration of Chloride and Bromide Anions: Determination of Relative Free Energy by Computer Simulation. *Journal of the American Chemical Society*, 107(25), 7793-7794.
- M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, & Fox, D. J. (2009). Gaussian 09 (Version A.02). Wallingford CT: Gaussian, Inc.
- Mobley, D., Liu, S., Cerutti, D., Swope, W., & Rice, J. (2012). Alchemical Prediction of Hydration Free Energies for SAMPL. *Journal of Computer-Aided Molecular Design*, 26(5), 551-562.



- Monkhorst, H. J., & Pack, J. D. (1976). Special Points for Brillouin-Zone Integrations. *Physical Review B*, 13(12), 5188-5192.
- Mu, X., Wang, Q., Wang, L.-P., Fried, S. D., Piquemal, J.-P., Dalby, K. N., & Ren, P. (2014). Modeling Organochlorine Compounds and the  $\sigma$ -Hole Effect Using a Polarizable Multipole Force Field. *The Journal of Physical Chemistry B*.
- Nath, N. K., Aggarwal, H., & Nangia, A. (2011). Crystal Structure of Methyl Paraben Polymorph II. *Crystal Growth & Design*, 11(4), 967-971.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riek, C., Grothe, R., & Eisenberg, D. (2005). Structure of the Cross- $\beta$  Spine of Amyloid-like Fibrils. *Nature*, 435(7043), 773-778.
- Palmer, D. S., Llinàs, A., Morao, I., Day, G. M., Goodman, J. M., Glen, R. C., & Mitchell, J. B. O. (2008). Predicting Intrinsic Aqueous Solubility by a Thermodynamic Cycle. *Molecular Pharmaceutics*, 5(2), 266-279.
- Palmer, D. S., McDonagh, J. L., Mitchell, J. B. O., van Mourik, T., & Fedorov, M. V. (2012). First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *Journal of Chemical Theory and Computation*, 8(9), 3322-3337.
- Paluch, A. S., & Maginn, E. J. (2013). Efficient Estimation of the Equilibrium Solution-Phase Fugacity of Soluble Nonelectrolyte Solids in Binary Solvents by Molecular Simulation. *Industrial & Engineering Chemistry Research*, 52(38), 13743-13760.
- Paluch, A. S., Shah, J. K., & Maginn, E. J. (2011). Efficient Solvation Free Energy Calculations of Amino Acid Analogs by Expanded Ensemble Molecular Simulation. *Journal of Chemical Theory and Computation*, 7(5), 1394-1403.
- Park, J., Nessler, I., McClain, B., Macikenas, D., Baltrusaitis, J., & Schnieders, M. J. (2014). Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. *Journal of Chemical Theory and Computation*, 10(7), 2781-2791.
- Patel, U., Patel, T. C., & Singh, T. P. (1983). Structure of Phenacetin, C<sub>10</sub>H<sub>13</sub>NO<sub>2</sub>. *Acta Crystallographica Section C*, 39(10), 1445-1447.
- Perlovich, G. L., Volkova, T., & Bauer-Brandl, A. (2007). Polymorphism of Paracetamol. *Journal of Thermal Analysis and Calorimetry*, 89(3), 767-774.
- Piquemal, J. P., Cisneros, G. A., Reinhardt, P., Gresh, N., & Darden, T. A. (2006). Towards a Force Field Based on Density Fitting. *Journal of Chemical Physics*, 124(10).
- Ponder, J. W. (2009). TINKER: Software Tools for Molecular Design (Version 5.0) [FORTRAN]. Saint Louis, MO. Retrieved from [dasher.wustl.edu](http://dasher.wustl.edu)
- Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., & Head-Gordon, T. (2010). Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B*, 114(8), 2549-2564.
- Price, S. L. (2008). Computed Crystal Energy Landscapes for Understanding and Predicting Organic Crystal Structures and Polymorphism. *Accounts of Chemical Research*, 42(1), 117-126.
- Raiteri, P., Martoňák, R., & Parrinello, M. (2005). Exploring Polymorphism: The Case of Benzene. *Angewandte Chemie International Edition*, 44(24), 3769-3773.
- Reilly, A. M., & Tkatchenko, A. (2014). Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Physical Review Letters*, 113(5), 055701.
- Ren, P., & Ponder, J. W. (2002). Consistent Treatment of Intramolecular Polarization in Molecular Mechanics Calculations. *Journal of Computational Chemistry*, 23(16), 1497-1506.
- Ren, P., & Ponder, J. W. (2003). Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *Journal of Physical Chemistry B*, 107(24), 5933-5947.

- Ren, P., & Ponder, J. W. (2004). Temperature and Pressure Dependence of the AMOEBA water model. *Journal of Physical Chemistry B*, 108(35), 13427-13437.
- Ren, P., Wu, C., & Ponder, J. W. (2011). Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *Journal of Chemical Theory and Computation*, 7(10), 3143-3161.
- Rivera, S. A., Allis, D. G., & Hudson, B. S. (2008). Importance of Vibrational Zero-Point Energy Contribution to the Relative Polymorph Energies of Hydrogen-Bonded Species. *Crystal Growth & Design*, 8(11), 3905-3907.
- Sacchetti, M. (2000). Thermodynamic Analysis of DSC Data for Acetaminophen Polymorphs. *Journal of Thermal Analysis and Calorimetry*, 63(2), 345-350.
- Sagui, C., Pedersen, L. G., & Darden, T. A. (2004). Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. *Journal of Chemical Physics*, 120(1), 73-87.
- Salahinejad, M., Le, T. C., & Winkler, D. A. (2013). Aqueous Solubility Prediction: Do Crystal Lattice Interactions Help? *Molecular Pharmaceutics*, 10(7), 2757-2766.
- Schnieders, M. J., Baltrusaitis, J., Shi, Y., Chatree, G., Zheng, L., Yang, W., & Ren, P. (2012). The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *Journal of Chemical Theory and Computation*, 8(5), 1721-1736.
- Schnieders, M. J., Fenn, T. D., & Pande, V. S. (2011). Polarizable Atomic Multipole X-Ray Refinement: Particle Mesh Ewald Electrostatics for Macromolecular Crystals. *Journal of Chemical Theory and Computation*, 7(4), 1141-1156.
- Shanno, D. F. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111), 647-656.
- Shi, Y., Wu, C., Ponder, J. W., & Ren, P. (2010). Multipole Electrostatics in Hydration Free Energy Calculations. *J. Chem. Comput.*, 32.
- Shi, Y., Zhu, C. Z., Martin, S. F., & Ren, P. (2012). Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. *The Journal of Physical Chemistry B*, 116(5), 1716-1727.
- Shirts, M. R., Pitera, J. W., Swope, W. C., & Pande, V. S. (2003). Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *Journal of Chemical Physics*, 119(11), 5740-5761.
- Smith, D. E., & Haymet, A. D. J. (1993). Free Energy, Entropy, and Internal Energy of Hydrophobic Interactions: Computer Simulations. *The Journal of Chemical Physics*, 98(8), 6445-6454.
- Stone, A. J. (2005). Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation*, 1(6), 1128-1132.
- Stone, A. J., & Alderton, M. (1985). Distributed Multipole Analysis - Methods and Applications. *Molecular Physics*, 56(5), 1047-1064.
- Straatsma, T. P., & McCammon, J. A. (1992). Computational Alchemy. *Annual Review of Physical Chemistry*, 43(1), 407-435.
- Thole, B. T. (1981). Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chemical Physics*, 59(3), 341-350.
- Tidor, B., & Karplus, M. (1994). The Contribution of Vibrational Entropy to Molecular Association: The Dimerization of Insulin. *Journal of Molecular Biology*, 238(3), 405-414.
- van Duijneveldt, F. B., van Duijneveldt-van de Rijdt, J. G. C. M., & van Lenthe, J. H. (1994). State of the Art in Counterpoise Theory. *Chemical Reviews*, 94(7), 1873-1885.

- Wu, J., Chattree, G., & Ren, P. (2012). Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theoretical Chemistry Accounts*, 131(3), 1-11.
- Zheng, L., Chen, M., & Yang, W. (2009). Simultaneous Escaping of Explicit and Hidden Free Energy Barriers: Application of the Orthogonal Space Random Walk Strategy in Generalized Ensemble Based Conformational Sampling. *The Journal of Chemical Physics*, 130(23), 234105.