Computer-Aided Rational Solvent Selection for Pharmaceutical Crystallization

by

Jie Chen

B.S. Chemical Engineering, Tsinghua University, 2004
M.S.CEP, Massachusetts Institute of Technology, 2008

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2010
[June 2010]

Author .....
Department of Chemical Engineering
May 3, 2010

Certified by .....
Bernhardt L. Trout
Professor of Chemical Engineering
Thesis Supervisor

Accepted by .....
William M. Deen
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Computer-Aided Rational Solvent Selection for Pharmaceutical Crystallization

by

Jie Chen

Submitted to the Department of Chemical Engineering
on May 3, 2010, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

Abstract

Solvents play an important role in crystallization, a commonly used separation and purification technique in the pharmaceutical, chemical and food industries. They affect crystal properties such as particle size distribution, morphology, polymorphism etc. and therefore have consequences for the downstream processing of the solid material. Current solvent selection techniques for solution crystallization remain *ad hoc* and typically do not have a theoretical underpinning. Elucidation of the interactions between solvent and solute molecules and the mechanism underlying the solvent effects on each aspect of the crystal properties would be a major aid for the rational selection of solvents and also the development of crystallization processes. In this work we studied the effect of solvent on the polymorphism and morphology of organic crystals using molecular modeling techniques.

The two most important contributions of this thesis are listed below.

1. We studied the self-assembly of solute molecules in solutions prior to nucleation for two polymorphic systems, tetrolic acid and glycine, using molecular dynamics simulations. We tested the existence of a link between the structure of the clusters formed in solution and the polymorphism of the crystals. Our results show that the link hypothesis succeeds in explaining the polymorph selection of tetrolic acid from different solvents. However it fails for glycine and thus should be used with caution.

2. We designed a computer-aided rational solvent selection procedure for improving the morphology of needle-like 2,6-dihydroxybenzoic acid form 2 crystal. We also experimentally grew 2,6-dihydroxybenzoic acid form 2 crystals using the solvent mixture suggested by computer simulations, which do exhibit reduced aspect ratios. This computer-aided selection procedure can not only quickly identify an effective solvent or solvent mixture, but also provide mechanistic understandings of the underlying chemistry. It can also be extended to improve the morphology of other needle-like organic crystals easily.

Thesis Supervisor: Bernhardt L. Trout
Title: Professor of Chemical Engineering

Dedication

This thesis is dedicated to my parents Huiyu Qiu and Jianlin Chen.

Acknowledgements

I would first of all like to thank Professor Bernhardt L. Trout for his constant support and help during the course of this thesis. He has always been very accessible and has given me immense freedom to pick and choose my projects. His insights into molecular simulations and pharmaceutical processing always provided me an overall guidance when I was lost. Our brain storming session gave me many new ideas when I was stuck. I truly appreciate having him as my advisor.

I would also like to thank my committee members, Professor Arup Chakraborty, Professor Michael J. J. Cima, Professor Collin M. M. Stultz. Their enthusiastic involvement and helpful discussions through out the course of the project helped me immensely. I am grateful to Professor Allan S. Myerson from Illinois Institute Technology for many helpful discussions about the glycine project. His experiences and insights into nucleation and crystallization problems inspired me with many interesting ideas.

There have been many present and former group members who have helped me get started with the project and then helped me at various times along the way. I would like to thank the former Trout group members Dr. Ning Shan and Dr. Gregg T. Beckham for helping me start with the molecular simulations and helping me define my project. I would like to thank the current Trout group members Dr. Keith Chadwick and Ying Diao for helping with the crystallization experiments. I would also like to thank the rest of Trout group members for their helpful inputs at the group meetings and their company both at and outside work. Last but not the least I would like to thank Gwen Wilcox, Angelique Scarpa, and Stephanie Bright for making life easier at MIT.

Finally I would like to thank all my friends who have made my time at MIT fun and memorable. I want to especially thank Sandeep for his constant support during my study at MIT and generally in life.

This work was funded by the Singapore-MIT Alliance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Crystallization from solution is a commonly used separation and purification process in a wide variety of industries, e.g. pharmaceutical, chemical, and food. Crystallization may be defined as a phase change in which a crystalline product is obtained from a solution. It is induced by changing the state of the system in some way that reduces the solubility of the crystallizing species. A crystallization event involves four steps: creation of supersaturation, induction, nucleation, and crystal growth. Supersaturation refers to a solution containing more dissolved material than could be dissolved by the solvent under normal/equilibrium conditions. It can be created by reducing the temperature, increasing the pressure, adding anti-solvents, etc. Crystals do not appear right way after the creation of supersaturation. Supersaturated solutions can be stable for minutes, hours and sometimes even days depending on the system. This period of time is called the induction time. Nucleation is the commencement of a new phase. Classical nucleation theory says that thermal fluctuations give rise to the appearance of small nuclei of a second phase and occasionally produce a long chain of favorable energetic fluctuations, thereby creating a nucleus exceeding the critical size.[1] Although this second phase has favorable lower free energy than the initial phase, there is a free energy penalty associated with the creation of an interface. The free energy, ΔG,

of the second phase is therefore the sum of a negative volume term and a positive surface term. For a spherical nucleus, ΔG is defined by equation 1.1:

$$\Delta G = -\frac{4}{3}\pi R^3 \Delta G_V + 4\pi R^2 \gamma \tag{1.1}$$

where R is the radius of the nucleus, ΔG_V is the bulk free energy difference per unit volume between the first and second phases, and γ is the surface free energy of the second phase per unit area. A critical nucleus usually consists of hundreds of atoms and a few nano meters in size, which cannot be detected by either human eyes or any advanced spectroscopic technology currently available. After nucleation, the nuclei continue to grow until the solution reaches its solubility equilibrium again. This is known as the crystal growth step.

In addition to product purity, it is often necessary to control the external shape and size of the crystals and to produce the desired polymorph and/or optical isomer. Solvent is an important process parameter in solution crystallization.[2–5] It affects the solubility, yield, particle size distribution, morphology and polymorphism of the crystal product, all of which have consequences for the down-stream processing of the solid material.[6] Solvent screening is a standard procedure in the pharmaceutical industry in searching for all the potential polymorphic forms in the early stage of drug development. Solvent/additive effects on crystal morphology have also been extensively studied and reported in literature.[7,8] However, current solvent selection for solution crystallization still adopts a trial and error approach. A typical selection process may involve starting with a list of solvent candidates, performing crystallization experiments, studying the crystal properties and then revising the solvent list based on the experimental results. This cycle will be repeated until the right solvent giving the desired crystal properties is found. This process is expensive and time consuming, especially when solvent mixtures are introduced. Elucidating of the interactions between solvent and solute molecules, understanding the mechanism underlying the solvent effects on each aspect of the crystal properties and moreover developing a model to predict them before carrying out crystallization experiments would be a major aid for the rational selection of solvents and also in the development of crystallization processes.

Molecular modeling is the term used to describe the techniques that employ quantum mechanics and statistical mechanics in conjunction with computer simulations to study the chemical and physical properties of materials. In recent years, molecular modeling has emerged as a useful tool in the solution of a number of crystallization problems.[9,10] In particular, modeling allows more focused experimentation based on structural and energetic calculations rather than intuition and trial and error.

1.2 Motivation and Outline of Thesis

The overall goal of this thesis is to apply molecular modeling techniques to gain mechanistic understandings of how solvents affect the properties of the crystals obtained from solution crystallization and to further aid in the rational design of crystallization processes.

We focused on solvent effects on the polymorphism and morphology of organic crystals. In Chapter 2, we shall overview some of the basics in molecular modeling. In Chapter 3, we discuss the self-assembly of solute molecules in solution prior to nucleation for two polymorphic systems, tetrolic acid and glycine. Using molecular simulation techniques, we tested the existence of a link between the structure of the solute clusters formed in solution and that of the polymorphic form of the crystals. Our results will show that the link hypothesis succeeded in explaining the polymorph selection of tetrolic acid from different solvents, but failed for glycine. Thus the link hypothesis should be used with caution in explaining the polymorph selection by solvents. In Chapter 4, we shall describe a computer-aided rational solvent selection procedure designed for improving the morphology of needle-like 2,6-dihydroxybenzoic acid form 2 crystals and also show that crystals grown from the solvent mixture suggested from the computer simulations do exhibit reduced aspect ratios. This computer-aided selection procedure can not only quickly identify an effective solvent or solvent mixture but can, as will be demonstrated, provide mechanistic understandings of the underlying chemistry. A solvent selection framework that can be applied to improve the morphology of other needle-like organic crystals will be outlined. In Chapter 5, we will present some of the salient conclusions of the thesis and possible future directions that the

present work can be extended in.

Chapter 2

Basics of Molecular Modeling

2.1 Introduction

In recent years molecular modeling methods based on computer simulations have become a useful tool in solving various scientific and engineering problems. Moreover, with the development of powerful supercomputers and advanced algorithms, the impact of molecular modeling has increased tremendously in the past few years and is expected to increase even more so in the future. Molecular modeling provides a bridge between the microscopic details of a system (the motion of atoms, the interactions between molecules, the coordination numbers, etc.) and the macroscopic properties of experimental interest (the equation of state, transport coefficients, structural order parameters and so on). Information under extremes of temperature and pressure, which may be difficult or impossible to gather experimentally, can be studied using computer modeling. Quite subtle details of molecular motion and structure in heterogeneous catalysis, fast ion conduction or enzyme action are difficult to probe experimentally but can be extracted readily from computer modeling. To some extent, computer-based modeling methods have filled the long existing gap between experiment and theory.

2.2 Two Types of Molecular Modeling

The forces of interaction between particles can be classified into four major categories: gravitational, strong nuclear, weak nuclear and electromagnetic.[11] The gravitational force is extremely long range and usually does not contribute to the intermolecular interaction. The strong nuclear forces and the weak nuclear forces are significant over the range 10^{-4} nm. Since molecular dimensions are typically of the order of 5×10^{-1} nm, these nuclear forces cannot contribute to the intermolecular forces. Consequently, intermolecular forces must be of electromagnetic origin.

Before performing molecular modeling we need a mathematical model to describe the intra and intermolecular interactions. There are mainly two types of models reported in the literature, the quantum mechanic and force field approaches.

2.2.1 Quantum Mechanics Approach

Quantum mechanics (QM) is a set of scientific principles describing the wave-particle duality of energy and matter at the atomic scale. In quantum mechanics, the properties of the particle can be described as the properties of a wave. Its quantum state can be represented as a wave of arbitrary shape and extending over space as a wave function. Therefore there is no clear definition of atoms and bonds. QM is widely used to study chemical reactions and to calculate interaction energies between molecules with high accuracy. The Schrödinger equation is the fundamental equation of quantum mechanics. Its solutions are called wave functions, which give a complete description of any system. In principle it is possible to describe all chemical systems using quantum theory. In practice, only the simplest chemical systems may realistically be investigated in purely quantum mechanical terms and approximations must be made for most practical purposes (e.g., Hartree-Fock, MP2, density functional theory etc.). Quantum calculations are usually computationally expensive and thus are mostly applied to small systems. In crystallization research QM has been mainly used in three areas:

1. Calculating the solvation free energy of a molecule which is further utilized to estimate its solubility.[12]

2. Predicting/optimizing the structure of a molecule in the gas phase which serves as the input for crystal structure prediction.[13]

3. Developing a force field for a novel molecule. The concept of a force field will be discussed next.

2.2.2 Force Field Approach

A force field refers to the functional form and parameter sets used to describe the potential energy of a system of particles. Force field functions and parameter sets are derived from both experimental work and high-level quantum mechanical calculations. The basic functional form of a force field consists of both bonded terms relating to atoms linked by covalent bonds and nonbonded terms describing the long range electrostatic and van der Waals forces. The specific decomposition of the terms depends on the force field but a general form for the total energy in an additive force field can be written as

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{vanderWaals} \tag{2.1}$$

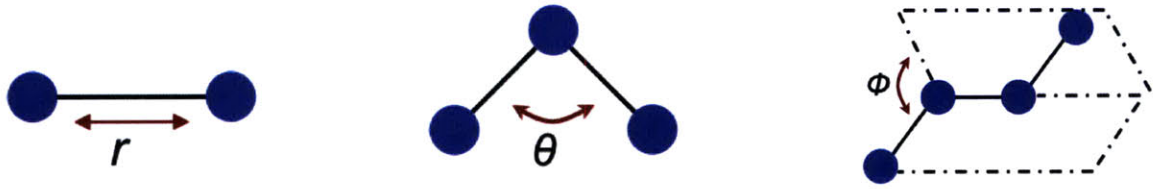
The first three terms belong to the bonded contributions which correspond to the bond stretching, angle bending and the torsion of a dihedral angle (shown in Figure 2-1). The other two terms belong to the nonbonded contributions, the electrostatic and van der Waals forces.

The bond and angle terms are usually modeled as harmonic oscillators in force fields that do not allow bond breaking. The functional form for the dihedral term is highly variable. As an example, the forms of all these three terms can be described by

$$E_{bond} = \sum_{bonds} K_r(r - r_{eq})^2 \tag{2.2}$$

$$E_{angle} = \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \tag{2.3}$$

$$E_{dihedral} = \sum_{dihedrals} K_\phi(1 + \cos n\phi) \tag{2.4}$$



(a) bond streching (b) angle bending (c) dihedral torsion

Figure 2-1: Bonded contributions in the force field approach.

The nonbonded terms are the most computationally intensive because they include many more interactions per atom. A popular choice is to limit interactions to pairwise energies. The van der Waals term is usually computed with a Lennard-Jones or Buckingham potential. The electrostatic term is usually computed with Coulombs law. Both can be scaled by a constant factor to account for electronic polarizability and produce better agreement with experimental observations. For example, the functional form for the nonbonded energies in a commonly used force field, OPLS force field,[14] is

$$E_{nonbonded} = \sum_j \sum_{i<j} \left[\frac{q_i q_j}{4\pi r_{ij}} + \varepsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] \tag{2.5}$$

Where the first and the second terms correspond to the electrostatic and van der Waals contributions respectively.

All force fields are based on numerous approximations and derived from different types of experimental data. Therefore they are called empirical. A flowchart of developing the force field parameters for a novel molecule is shown in Figure 2-2. The advantage of using

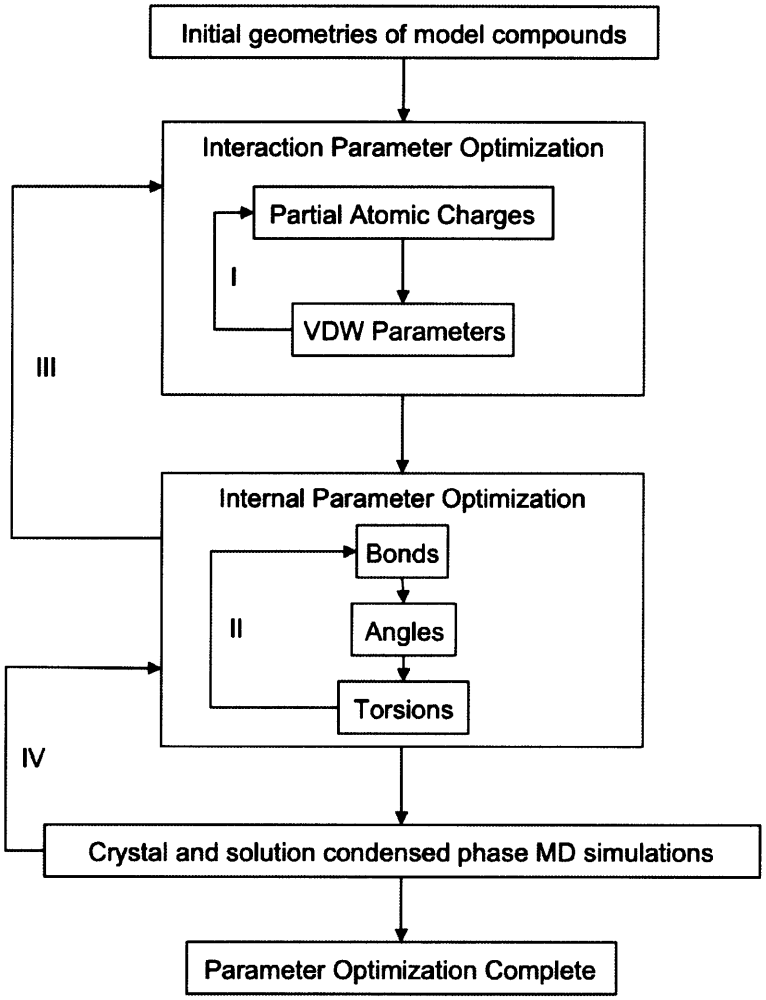


Figure 2-2: A flowchart of developing the force field parameters for a novel molecule.

a force field is that the energy evaluation becomes significantly cheaper computationally than the first principal quantum calculation. Therefore the simulation of bigger systems, e.g. proteins, polymers and crystals, becomes feasible. Moreover, these force fields are developed based on functional groups/residues and can be assembled together easily when studying a new molecule. For example, we do not need to develop the force field parameters from scratch for any new protein we would want to study because most force fields used for protein simulations are developed based on the amino acids. Once we know the sequence of the protein of interest, we can easily assemble the required parameters together.

There are a wide variety of force fields available in the literature. They are developed and

optimized for specific systems. The OPLS/OPLS-AA force field,[14] developed by Jorgensen *et al.* at Yale University, has a very good library of common organic solvents and can also be used for protein simulations. CHARMM,[15] AMBER,[16] GROMOS,[17] etc. are the force fields commonly used for macromolecule simulations: proteins, sugars, DNAs and so on. COMPASS[18,19] is a powerful force field that supports atomistic simulations of condensed phase materials. Considering the number of force fields, we will not list all available ones here. Force field development is a very active research area. Obtaining an accurate force field is the prerequisite of any meaningful calculation. Current effort has been placed in developing force fields that can account for the polarizability,[20–22] in which a particle's charge is influenced by the electrostatic interactions with its neighbors.

2.3 Techniques in Molecular Modeling

2.3.1 Energy Evaluation

Once we have the force field necessary to describe the interactions within and between molecules, we can evaluate the energy of the system. It is just a one-step calculation, which can be expressed as the pairwise summation of the interaction energies between the atoms of interest. This type of calculation can be used to evaluate the lattice energy of a crystal once its structure is solved, which can be further used either to rank the thermodynamic stability of newly discovered polymorphs or to check the validity of the force field when the stability ranking is experimentally determined.

2.3.2 Energy Minimization

Energy minimization is usually performed by changing the positions of atoms in the system to find a local/global minimum. It can be mathematically expressed as

$$E_{min} = \min_{\vec{q}_1,\ldots,\vec{q}_n} E\left(\vec{q}_1,\ldots,\vec{q}_i,\ldots,\vec{q}_n\right) \tag{2.6}$$

where $\vec{q}_i$ is the position of the i^{th} atom in the system. This type of calculation can be useful in searching for new polymorphic forms, helping to solve crystal structures and removing bad initial contacts for molecular dynamics simulations.

2.3.3 Molecular Dynamics Simulations

Molecular Dynamics simulation is a technique for computing the equilibrium and transport properties of a classical many-body system, where the motion of the constituent particles obeys the laws of classical mechanics. If we consider a system of atoms with Cartesian coordinates, the equation governing the motion becomes

$$m_i \frac{d^2}{dt^2} \vec{q}_i = \vec{F}_i \tag{2.7}$$

where m_i and $\vec{q}_i$ are the mass and position of the i^{th} atom and $\vec{F}_i$ is the force on that atom.

Molecular dynamics simulations are in many respects very similar to real experiments. Firstly, we prepare a sample: we select a model system consisting of N particles and we solve Newton's equations of motion for this system until the properties of the system no longer change with time. After equilibration we perform the actual measurement. In a real experiment, we connect our sample to an instrument and measure the property of interest over a certain time interval. If the measurements are subject to statistical noise then the longer we average the more accurate our measurement becomes. It is exactly the same in molecular dynamics. To measure an observable quantity in a molecular dynamics simulation we must first of all be able to express this observable as a function of the positions and momenta of the particles in the system. For example, the instantaneous interaction energy between two proteins A and B is fully determined by the position of the atoms in these two molecules:

$$E_{AB}(t) = f\left(\vec{q}_{A1}(t), \ldots, \vec{q}_{Ai}(t), \ldots, \vec{q}_{Am}(t), \vec{q}_{B1}(t), \ldots, \vec{q}_{Bj}(t), \ldots, \vec{q}_{Bn}(t)\right) \tag{2.8}$$

where $\vec{q}_{Ai}$ and $\vec{q}_{Bj}$ are the positions of the i^{th} and j^{th} atoms in molecule A and B respectively

at time instance t. As the position of atom fluctuates due to thermal motion, the instantaneous interaction energy also fluctuates. To reduce the statistical error of the measurement just as we do in real experiments, we will take many samples and calculate the average:

$$\langle E_{AB} \rangle = \frac{\sum_{k=1}^{K} E_{AB}(t_k)}{K} \tag{2.9}$$

where K is the total number of samples extracted in a molecular dynamics simulation.

2.3.4 Monte Carlo Simulation

Monte Carlo simulation methods are especially useful in studying systems with a large number of correlated degrees of freedom such as fluids, disordered materials, strongly coupled solids and macromolecules. Different from the molecular dynamics method, the Monte Carlo approach relies on statistical mechanics rather than trying to reproduce the dynamics of a system. It uses a large amount random numbers and generates states according to appropriate Boltzmann probabilities. Consequently, it can be applied to study those events that are impossible/inefficient to simulate directly using MD simulations due to either the difficulty in setting up a MD system or the extremely long time scale of observing the event of interest. For example, finding a phase transition point, e.g. melting temperature, is much easier using the Monte Carlo approach than molecular dynamics. Details of the Monte Carlo method can be found in many molecular simulation books.[23]

Chapter 3

Computational Study of the Self-Assembly of Solute Molecules in Solution and Its Implications for the Polymorphic Outcome of Solution Crystallization

3.1 Introduction

Polymorphism is the ability of a solid material to exist in more than one form or crystal structure, whilst retaining the same chemical composition.[24] Polymorphism exists as a result of the difference in crystal packing or the existence of different conformers of the same molecule, called packing polymorphism and conformational polymorphism respectively. The existence of other crystal types as the result of hydration or solvation is called pseudo-polymorphism. Figure 3-1 shows the polymorphs of the molecule 5-Methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, commonly known as ROY. It has ten known polymorphs, seven of them with solved structures.[25] This phenomenon leads to significant

variability in the properties of products in the pharmaceutical, chemical and food industries and continues to pose a challenge to scientists and engineers in producing crystal products of consistent quality. An undesired polymorphic form of the drug product, for instance, can lead to different bioavailability in the target organism, which could render the drug useless, or increase its potency to a dangerous limit.[26] One well-known example is the case of antiviral drug Ritonavir.[27–29] A more stable polymorphic form discovered after the drug was commercialized led to the withdraw of the drug product from the market and resulted in tremendous economical loss. In the pharmaceutical industry, a very large number of pharmaceuticals exhibit the phenomenon of polymorphism. 70% of barbiturates, 60% of sulfonamides and 23% of steroids exist in different polymorphic forms. Hence, the accurate control of the polymorphism of drug products is one of the most important considerations in crystallization process design.

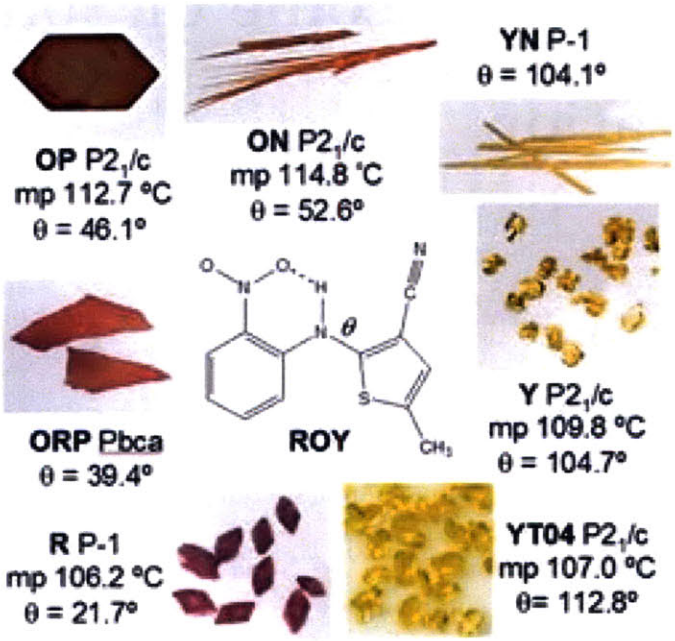


Figure 3-1: Polymorphs of ROY.

3.2 Solvent Effect and the Link Hypothesis

It is widely know that organic molecules can crystallize into different polymorphic forms from different solvents. Numerous examples have been reported in the literature, including oxalyl dihydrazide,[30] L-glutamic acid,[31] tetrolic acid,[32] 2,6- dihydroxybenzoic acid[33] etc. Solvent screening is a standard procedure in pharmaceutical industry to search for all the potential polymorphic forms in the early stage of drug development. In a recent review paper about solvent effects on polymorphism, Blagden and Davey pointed out three important areas in understanding polymorph selection from solution crystallization within a framework of structural and kinetic parameters:[34]

1. an understanding of the structural similarities and differences between polymorphs.
2. an appreciation of the interplay between thermodynamic and kinetic factors.
3. a knowledge of the fundamental crystallization growth unit and its relationship to the structural synthon.

The first point emphasizes the importance of understanding the structural characteristics of different polymorphs, including cell symmetry, subtle changes in the molecular conformations, hydrogen bonding schemes, etc. It is obvious that the first point serves as the basis for all the other studies. In the second point, the interplay between kinetics and thermodynamics is essentially summarized in Ostwald's rule of stages,[35] which indicates that in a polymorphic system the crystallization processes may begin with the nucleation of the least stable form and finish with the most stable form. Thus, the primary nucleation stages are interspersed with polymorphic transformations that often involve a growth and dissolution process from a metastable phase to a more stable phase. The situation for a dimorphic system in which the solution-mediated transformation process is possible is shown schematically in Figure 3-2. Initially, the nucleation and then growth of the metastable form occurs until its solubility is reached. The nucleation of the stable form occurs later and the subsequent growth is driven by the dissolution of the metastable form. The role of solvent in polymorphic transformation lies on its ability to modify the nucleation rate of both forms and the transformation

rate of the metastable form to the more stable form. Kinetic studies were conducted for the crystallization of 2,6-dihydroxybenzoic acid form chloroform and toluene[36] rescpectively. Results showed chloroform facilitates the transformation from the metastable to the stable form more effectively than toluene by accelerating the nucleation rate of the stable form. In the third point, emphasis has been placed on building a link between the growth synthons[3,37] formed through molecular self-assembly in the solution and the structural synthons packed in the crystal. There is a notion of a crystal as a supramolecular entity created from building blocks, also known as structural synthons.[38] Structural synthons can be a single molecule or an intermolecularly bonded group of molecules. The link hypothesis says that the most stable growth synthon formed by self-assembly of solute molecules in solutions has the highest probability to crystallize out as the structural synthon in the solid form, with the mediation of solvents. Obviously, this hypothesis is trying to explain polymorph formation from a thermodynamic point of view, whilst kinetic factors like supersaturation, the rate of creating supersaturation, nucleation, etc. are put aside.

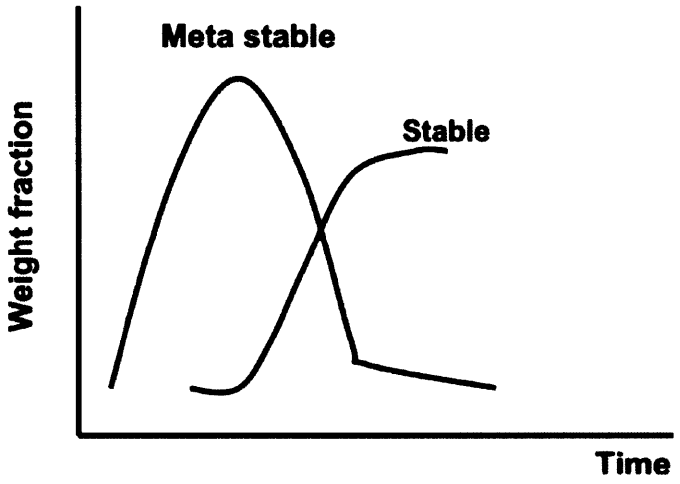


Figure 3-2: The kinetics of a dimorphic phase transformation.

Although intuitively the most direct way of truly understanding polymorph selection by solvents is to study the nucleation, it is a very challenging problem both experimentally and computationally. The induction time to nucleation is usually on the order of hours,

sometimes even days, which makes it impossible to study using classic molecular dynamics, which can only simulate nano-second time scale events of a modest sized system of ~10000 atoms. Recently, with the development of advanced simulation techniques for more efficient sampling of rare events (e.g. transition path sampling[39]), a few computational studies of the nucleation of Lennard-Jones fluids[40] and the nucleation of the stable form of a molecular crystal, terephthalic acid, from its metastable form[41] were reported. This progress in computational methods, although promising, is still not sophisticated enough for studying the nucleation of a real molecular crystal from solution, which requires a comprehensive consideration of the large number of degrees of freedom introduced by solvents, local orientation of the molecule, multiple molecules in the asymmetric unit cell, etc. Experimentally detecting the occurrence of nucleation is extremely difficult due to the limitation in the resolution of currently available instruments. Critical nuclei usually consist of hundreds molecules and are typically of the size of a few nano meters. One can only infer the nucleation mechanism from the post-nucleation observation (crystals harvested) or from the pre-nucleation solution chemistry. The latter approach falls in line with the link hypothesis mentioned earlier and will be discussed more next.

The self-assembly of solute molecules in solution happens much faster than the nucleation. It usually takes place in the time scale of nano seconds. Both experimental and computational approaches have been applied to study this process, to investigate the existence of the link, and to gain insights into the solvent mediated polymorph selection. Davey *et al.* studied the self-assembly of tetrolic acid in various solvents using FTIR Spectroscopy.[32] Carboxylic acid dimers (growth synthon) formed by a pair of tetrolic acid molecules were detected in a tetrolic acid chloroform solution, from which a dimer based (structural synthon) crystal structure was obtained upon crystallization. The dimer motifs were absent in a tetrolic acid ethanol solution, from which a catemer based (structural synthon) crystal structure was obtained. They firstly proposed the existence of a link between the solution chemistry and the solid-state polymorphic outcome, which is schematically shown in Figure 3-3. This hypothesis was also successful in explaining the crystal structures obtained for various other systems, e.g.

2,6-dihydroxylbenzoic acid[33] and 5-fluorouracil.[42] Exceptions have been reported as well, i.e. mandelic acid[43] and anthranilic acid,[44] where the link between solution chemistry and crystallization output is absent. There is sometimes a debate as to whether such a link exists when the structures of the clusters in solution are inferred using different measurements, e.g. glycine.[45–48] It is clear that the understanding of the clusters formed in solution using a solely experimental approach is still limited. Molecular dynamics (MD) is a powerful tool that can track the motion of each atom in the system. It has been used previously to study the molecular self-assembly and precursor formation in a variety of solutions.[49,50] This approach can provide understandings of the solute-solvent interactions at atomic level. Moreover, MD can be implemented to study those processes where direct access using experimental approaches is still very difficult. Hence, it is a good complementary tool for experimental techniques.

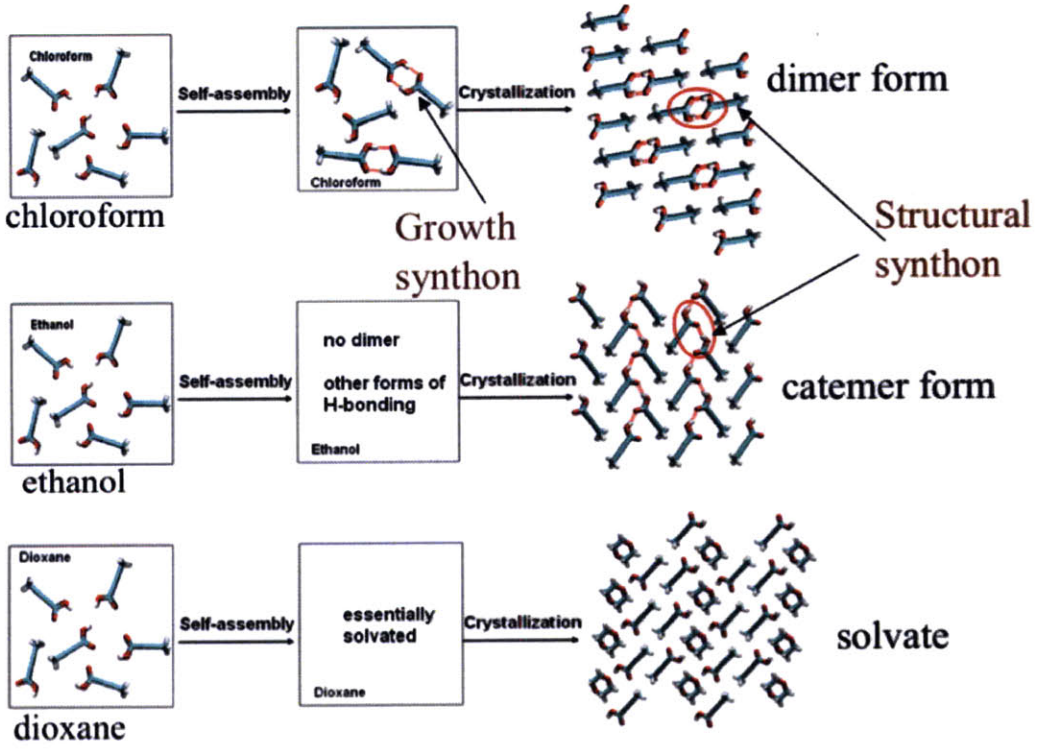


Figure 3-3: The link hypothesis for tetrolic acid. Red circles highlight the structural synthons in forms I and II. TTA molecules are in Licorice representation: O in red, H in white, C in cyan and hydrogen bonds in red dashed lines.

3.3 Objectives

In this work, we have studied the self-assembly processes of two polymorphic organic crystals, tetrolic acid and glycine, in solution. Tetrolic acid (TTA) was selected as the model compound because it crystallizes into different polymorphic forms from different solvents and is a good supporting example for the link hypothesis.[32] Glycine was selected as the other model compound because of the controversy surrounding the existence of the link,[45–48] which will be described in detail in the glycine section. The objective of this study was to gain fundamental understandings of

1. the interplay between solute and solvent molecules,

2. the effects of solvent on the formation of various growth synthons and their relative stabilities,

3. the relationship between the structure of the solute clusters in solution and that of the polymorph formed during crystallization.

Additionally, we wanted to provide some guidance for the rational selection of solvents for solution crystallization to obtain the desired polymorph.

3.4 Tetrolic Acid

3.4.1 Polymorphism of Tetrolic Acid

Tetrolic acid (TTA) crystallizes in both a dimer-based, centric $P\bar{1}$ structure (form I) and a catemeric $P2_1$ chain structure (form II).[51] Form II is the stable form at room temperature, whilst form I is more stable at higher temperatures, above 354 K.[52] So the polymorphs are enantiotropically related. The conformational difference between TTA molecules in forms I and II is very small, thus the polymorphism studied in this work is packing polymorphism. TTA is also reported to form a solvate with dioxane.[32] Detailed crystallographic information is listed in Table 3.1. Previous studies on TTA precursors in solution by Gavezzotti et al.[49] using molecular simulation tools showed that carboxylic acid dimers were the most persistent configuration in carbon tetrachloride, whilst the dimers broke immediately when the solvent was switched from carbon tetrachloride to water. This preliminary computational study was carried in a very dilute system without accounting for the multi-body interactions between solute molecules. Moreover, the relative stability of various synthons in solution and the mechanism through which solvent molecules mediate the self-assembly process are still not clear.

3.4.2 Computational Details

3.4.2.1 Force Field Validation

TTA was modeled using the CHARMM[15] force field. Bond and angle equilibrium parameters were extracted by averaging the experimental X-ray crystal structures of forms I and II. The four solvents studied were carbon tetrachloride, chloroform, ethanol and dioxane. TTA is known to crystallize into form I, form II and a solvate from chloroform, ethanol and dioxane respectively. Carbon tetrachloride was chosen due to its strong apolarity, although no TTA crystals were experimentally obtained from this solvent currently. The dielectric constants of these solvents are listed in Table 3.2 as indictors of their polarity.[53] Hydrogen bond parameters[54] for these four solvents are listed in Table 3.2 as well. The α parameter is a

	crystal type	lattice parameters		reported solvents	refcode
form I[52]	triclinic	a = 7.320 Å b = 5.099 Å c = 7.226 Å	$\alpha = 83.91$ $\beta = 117.46$ $\gamma = 112$ [a]	chloroform	TETROL
form II[52]	monoclinic	a = 7.887 Å b = 7.121 Å c = 3.937 Å	$\alpha = 90$ $\beta = 100.18$ $\gamma = 90$	ethanol	TETROL01
solvate[32]	triclinic	a = 4.1673 Å b = 6.5063 Å c = 12.4424 Å	$\alpha = 97.557$ $\beta = 92.948$ $\gamma = 91.670$	dioxane	XAKGEW

Table 3.1: Polymorph information of tetrolic acid.
[a] Crystal data are directly obtained from the Cambridge Crystallographic Data Center (CCDC). Thus, the number of significant digits is not consistent.

measure of the molecule's hydrogen bond donor ability, which is related to the free energy change of the hydrogen bond formed between the molecule of interest and some reference base; the β parameter is a measure of the hydrogen accepter ability and is related to the free energy change of the hydrogen bond formed between the molecule of interest and some reference acid. For example, ethanol ($\alpha = 2.7, \beta = 5.8$) is a strong hydrogen bond donor as well as a strong hydrogen bond acceptor, while dioxane ($\alpha = 0.9, \beta = 5.3$) is only a strong hydrogen bond acceptor. Since chloroform and carbon tetrachloride are not well parameterized in the CHARMM force field, all solvents were modeled using the OPLS-AA force field[14] with one modification: the arithmetic average instead of the geometric average was used for combining the Lennard-Jones radii, thus the same potential energy functional forms could be applied for both TTA and solvents.

This modified CHARMM-OPLS forced field was carefully tested. The density and the heat of vaporization of these four solvents at 298 K and 1 atm were calculated using MD and compared to experimental values to validate the modification in the combination rule for LJ radius. A 2 ns MD trajectory of a cubic solvent box was obtained with time step of 1 fs in the NPT ensemble at 298 K and 1 atm. Periodic boundary conditions were applied. A

14 $\AA$ cutoff was used for the van der Waals interactions and the smooth particle mesh Ewald method[55] was applied to the electrostatic interactions. (This setting was used in all MD simulations in this study if periodic boundary conditions were present.) All atoms were fully flexible during the simulations, and all calculations were performed using the CHARMM program package. The last 1.5 ns run was used to calculate the density with a sampling interval of 1 ps and the statistical error was estimated using the method described by Allen.[56] The heat of vaporization was calculated using the formula:

$$\Delta E_{vap} = E(g) - E(l) \tag{3.1}$$

$$\Delta H_{vap} = \Delta E_{vap} + RT \tag{3.2}$$

where $E(g)$ and $E(l)$ are the potential energies of the gas and liquid. The gas phase was treated as an ideal gas. $E(g)$ represents the average potential energy of a single molecule in a vacuum (essentially the intramolecular energy of the model compound), computed from a 1 ns MD simulation in vacuum. Langevin dynamics was applied, with a friction coefficient of 500.0. The calculated and experimental values[57,58] of the density and heat of vaporization of these solvents are listed in Table 3.2. There are only minor differences between the calculated and experimental values, which confirmed that the modified OPLS-AA force field (with the arithmetic average combination rule) was adequate in modeling these four organic solvents.

The suitability of using the CHARMM force field to model TTA was verified by the percentage changes of lattice parameters (PCLPs) and the root mean squared difference (RMSD) of a crystal supercell in a 2 ns MD simulation at 20 K and 1 atm. Low temperature MD simulation in the NPT ensemble is similar to a potential energy optimization in allowing the system to relax to a local minimum. The experimental X-ray crystal structure was used as the reference state. This simulation was performed for both TTA polymorphic forms I and II, and the results are plotted in Figure 3-4a and c. The validation of mixing the CHARMM and the OPLS-AA force fields was tested by performing the same MD simulation for the solvate form of TTA (shown in Figure 3-4e). The interactions between TTA and the other

three solvents were not tested due to the lack of comparable experimental data. The PCLPs for all these three forms of TTA crystal are within 6.5%, and the RMSDs over the whole MD simulation are less than 1.4 $\mathring{A}$, which suggests that the modified CHARMM-OPLS force field is adequate in modeling TTA crystals. We also performed the same simulations using purely the OPLS-AA force field (geometric average combination rule used for LJ radii) for TTA and dioxane, and results are plotted in Figure 3-4b, d and f. Compared to the modified CHARMM-OPLS force field, the pure OPLS-AA force field gives similar RMSD values, decreased PCLP values for the lattice parameter a of the form II and the lattice parameter b of the solvate, but increased PCLP values for the others cell parameters. Thus, we concluded that it was better to use the modified CHARMM-OPLS force field to model TTA and solvents in the following studies.

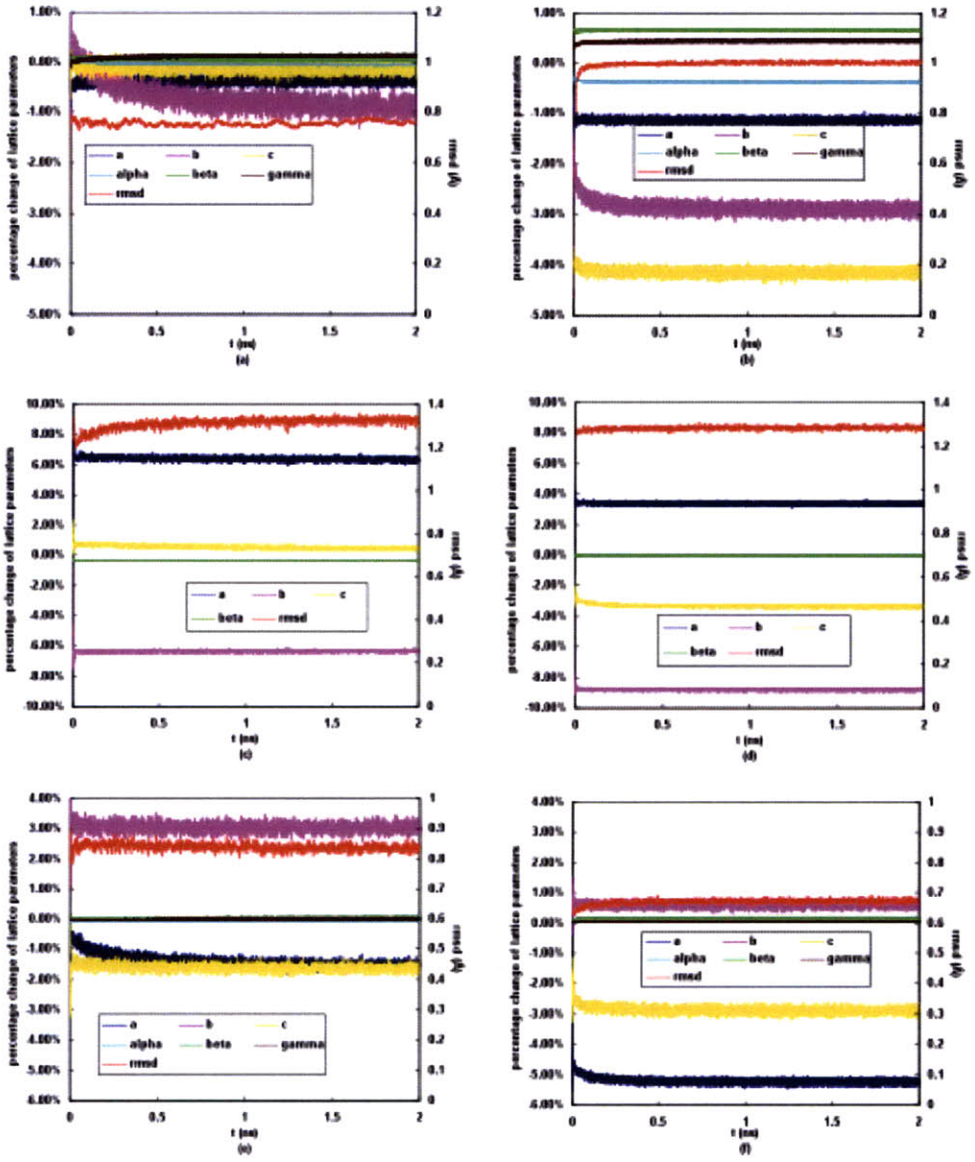


Figure 3-4: Percentage changes of the lattice parameters and RMSDs of TTA crystal supercells in 2 ns MD simulations at 20 K and 1 atm. a) a box of form I crystal with $6a \times 6b \times 6c$ cells, the modified CHARMM-OPLS force field applied; b) a box of form I crystal with $6a \times 6b \times 6c$ cells, the OPLS-AA force field applied; c) a box of form II crystal with $6a \times 6b \times 12c$ cells, the modified CHARMM-OPLS force field applied, alpha and gamma parameters do not change over the whole simulation run, thus, are not shown here; d) a box of form II crystal with $6a \times 6b \times 12c$ cells, the OPLS-AA force field applied; (e) a box of solvate with $8a \times 6b \times 3c$ cells, the modified CHARMM-OPLS force field applied; (f) a box of solvate with $8a \times 6b \times 3c$ cells, the OPLS-AA force field applied.

	dielectric constant [a]	hydrogen bond parameter α [b]	hydrogen bond parameter β [b]	density calc (g/ml)	density exp [c] (g/ml)	ΔH_{vap} calc (kJ/mol)	ΔH_{vap} exp [d] (kJ/mol)
ethanol	24.3 (298K)	2.7	5.8	0.767 ± 0.001	0.789	42.01 ± 0.17	42.30
chloroform	5.0 (273 K)	2.2	0.8	1.462 ± 0.001	1.498	31.38 ± 0.08	31.25
carbon tetrachloride	2.24 (293 K)	1.4	0.6	1.618 ± 0.002	1.594	32.17 ± 0.13	32.43
dioxane	2.2 (298 K)	0.9	5.3	1.002 ± 0.001	1.033	42.09 ± 0.08	40.42

Table 3.2: Physical and chemical properties of carbon tetrachloride, chloroform, ethanol, and dioxane.
[a] Experimental values of dielectric constant from ref.[53]
[b] Obtained from ref.[54] α parameter is a measure of the hydrogen bond donor ability, β parameter is a measure of the hydrogen bond acceptor ability.
[c] Experimental values of density at 298 K and 1 atm from ref.[57]
[d] Experimental values of heat of vaporization at 298 K and 1 atm from ref.[58]

3.4.2.2 Solvation Free Energy Calculation

The solvation free energies of TTA in different solvents from its gas phase at 298 K and 1 atm were used to characterize how strongly the solute molecules interact with the solvent molecules. The thermodynamic integration (TI) method was applied to calculate the solvation free energies. The initial state ($\lambda = 0$) is defined as one TTA molecule fully solvated in a $\sim 30\mathring{A} \times \sim 30\mathring{A} \times \sim 30\mathring{A}$ solvent box. The system was equilibrated at the target temperature and pressure. System details are listed in Table 3.3. The final state ($\lambda = 1$) is defined as turning off the electrostatic and van der Waals interactions between the solute and solvent molecules. Ten windows were used from the initial state to the final state with $\delta\lambda = 0.1$. For each window, the system was equilibrated for 100 ps and sampled for 500 ps.

	carbon tetrachloride solution	chloroform solution	ethanol solution	dioxane solution
no. of solvent molecules	163	190	243	154
no. of TTA molecules	1	1	1	1
equilibrated cell size	29.5 $\mathring{A}$	30.2 $\mathring{A}$	31.8 $\mathring{A}$	29.1 $\mathring{A}$

Table 3.3: System details of the MD simulations for the solvation free energy calculations of tetrolic acid.

3.4.2.3 Free Energy Change of the Dimerization Reaction

The formation of a carboxylic acid dimer (product) from two fully solvated TTA molecules (reactant) was studied in these four solvents. The free energy change of the dimerization reaction (Figure 3-5) was used to characterize the stability of the carboxylic acid dimer, which was calculated using the MD umbrella sampling method.[59–61] The order parameter (OP) used to study this reaction was defined by averaging two distances between the hydroxyl hydrogen and carbonyl oxygen of a pair of TTA molecules, which is schematically shown in Figure 3-6. Two TTA molecules were inserted into a $\sim 30\mathring{A} \times \sim 30\mathring{A} \times \sim 30\mathring{A}$ pre-equilibrated solvent

box, same as the solvent box used in the solvation free energy calculations (Table 3.3). To overcome diffusion, one TTA molecule was confined in the center of the simulation box using a harmonic restraint with a force constant of 836.8 kJ/(mol·$\AA^2$). A harmonic functional form of an umbrella potential was used, as shown below

$$U = k_u(\delta - \delta_0)^2 \tag{3.3}$$

Where, k_u is the harmonic force constant with unit kJ/(mol·$\AA^2$), δ_0 is the equilibrium point of the sampling window.

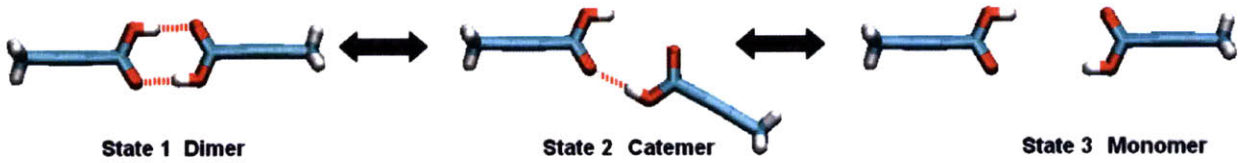
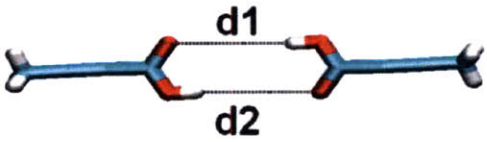


Figure 3-5: Dimerization reaction of tetrolic acid.



$$OP = \frac{d1 + d2}{2}$$

Figure 3-6: Definition of the order parameter (OP) used in the umbrella sampling of the dimerization reaction of tetrolic acid.

3.4.2.4 Radial Distribution Function Calculations

Radial distribution functions (RDFs)[42] were used to provide more detailed structural information about the interactions between TTA and solvent molecules. It describes how the

atomic density varies as a function of the distance from one particular atom (Figure 3-7), and can be expressed as

$$g_{12}(r) = \frac{N_2(r)}{\rho^* 4\pi r^2 \delta r} \tag{3.4}$$

where $N_2(r)$ is the number of particle 2 in the shell of thickness δr and distance r away from particle 1, and ρ^* is the density of particle 2 in the bulk. To calculate the RDFs, a tetrolic acid molecule is inserted into a pre-equilibrated solvent box. More system information is shown in Table 3.4. Again to overcome diffusion, the TTA molecule was confined in the center of the simulation box using a harmonic restraint with a force constant of 836.8 kJ/(mol·$\AA^2$). A 2 ns MD trajectory with time step 1 fs was harvested. Various RDFs were calculated throughout the duration of the production run (the last 1.5 ns) by evaluating a histogram of the distances between the specific atom in solvent molecule and the interested atom in TTA every 0.1 ps. All atoms used in RDF calculations are labeled in Figure 3-8. RDFs reported are those obtained by averaging all of the histograms and normalizing the RDF value in the bulk to 1.

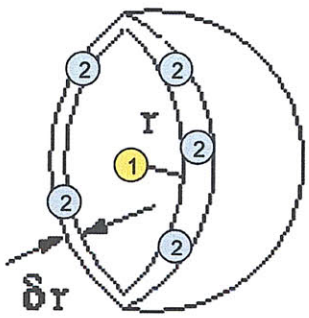


Figure 3-7: Definition of the radial distribution function between atom 1 and atom 2.

3.4.2.5 Dimer Fraction of the Hydrogen Bond Network

MD simulations for tetrolic acid solutions of various concentrations (Table 3.5) were conducted to study the structure of the clusters formed in solution. Previous work from Davey *et al.*[43] showed that TTA is highly soluble in chloroform, ethanol and dioxane. The highest TTA concentrations used in their experiments were 3.0 M in chloroform, 7.9 M in ethanol,

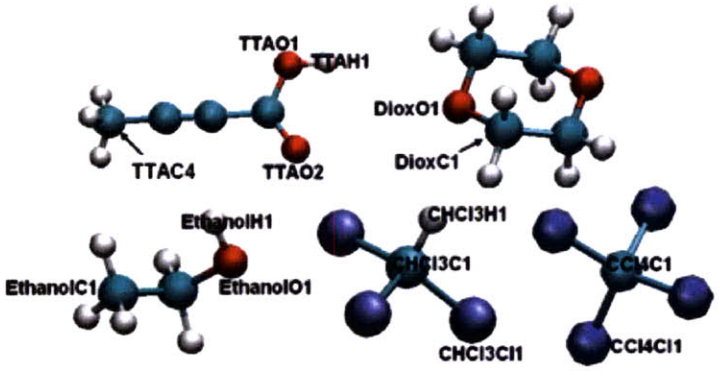


Figure 3-8: Atom labels in the RDF calculations of tetrolic acid.

	carbon tetrachloride solution	chloroform solution	ethanol solution	dioxane solution
no. of solvent molecules	416	390	464	247
no. of TTA molecules	1	1	1	1
equilibrated cell size	40.8 $\AA$	37.4 $\AA$	35.9 $\AA$	32.4 $\AA$

Table 3.4: System details of the MD simulations for the RDF calculations.

and 4.8 M in dioxane, which were all reported to be undersaturated. The TTA concentrations simulated in this work are in line with the concentrations used in Davey's experiments. We still call these TTA-solvent mixtures solutions as they do, although some highly concentrated solutions might be also called binary mixtures. Note that no accurate solubility data of TTA has been reported so far. Thus, it is unknown whether the highest concentrations used in the MD simulations were above or below the solubility line. The distance cutoff for the hydrogen bond formed between the hydroxyl hydrogen of a TTA molecule and the oxygen (either carbonyl oxygen or hydroxyl oxygen) of a neighboring TTA molecule was 2.35 $\AA$, independent of orientation, which was obtained by adding the average length of such hydrogen bond formed in carbon tetrachloride (1.94 $\AA$) and three times its standard deviation. The dimer fractions were used to characterize the overall hydrogen bond networks in various solutions. It was calculated using the formula:

$$\text{dimer fraction} = \frac{\text{total no. of hbonds in carboxylic dimers}}{\text{total no. of hbonds in all tetrolic acid molecules}} \tag{3.5}$$

For example, if there are only three TTA molecules in the system, and they exist in a conformation shown in Figure 3-9, there are two hydrogen bonds in the carboxylic dimer and three hydrogen bonds between the tetrolic acid molecules. Thus, the dimer fraction is 0.667. The dimer fractions in the crystal structures of both forms I and II were calculated as well, using both the experimental X-ray structures directly obtained from the literature and the optimized structures obtained from the low temperature MD simulations. In both cases, the dimer fraction is 1 for form I (the dimer form) and 0 for form II (the catemer form).

3.4.3 Results and Discussions

3.4.3.1 The Interplay between Tetrolic Acid and Solvent Molecules

The solvation free energies (ΔG_{solv}) of TTA in carbon tetrachloride, chloroform and ethanol are 37.78, 41.42 and 45.94 kJ/mol respectively, as listed in Table 3.6. It is clear that ΔG_{solv} increases as the solvent polarity increases, as measured by the increase in the dielectric

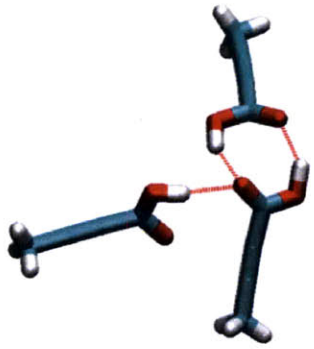


Figure 3-9: Illustration for dimer fraction calculation.

	solvent	no. of TTA molecules	no. of solvent molecules	concentration (mol/L)
1	carbon tetrachloride	339	257	6.17
2	carbon tetrachloride	225	436	3.58
3	carbon tetrachloride	113	632	1.56
4	carbon tetrachloride	75	708	0.99
5	chloroform	339	339	5.87
6	chloroform	225	473	3.84
7	chloroform	113	612	1.88
8	chloroform	75	677	1.21
9	ethanol	339	369	6.47
10	ethanol	225	508	4.44
11	ethanol	113	718	2.12
12	ethanol	75	810	1.35
13	dioxane	339	203	7.07
14	dioxane	225	290	4.89
15	dioxane	113	386	2.51

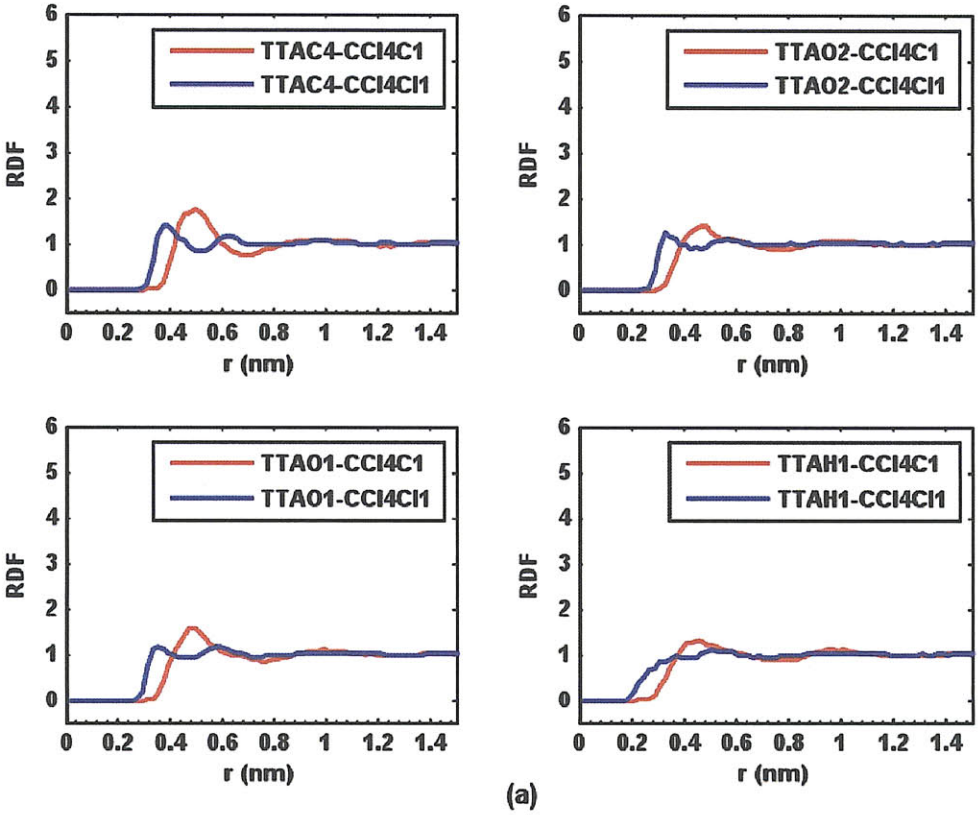
Table 3.5: System details of the MD simulations in the study of dimer composition of the hydrogen bond network of tetrolic acid.

	ΔG_{solv} (kJ/mol)
ethanol	45.95 ± 1.26
chloroform	41.42 ± 0.92
carbon tetrachloride	37.78 ± 0.96
dioxane	47.15 ± 1.13

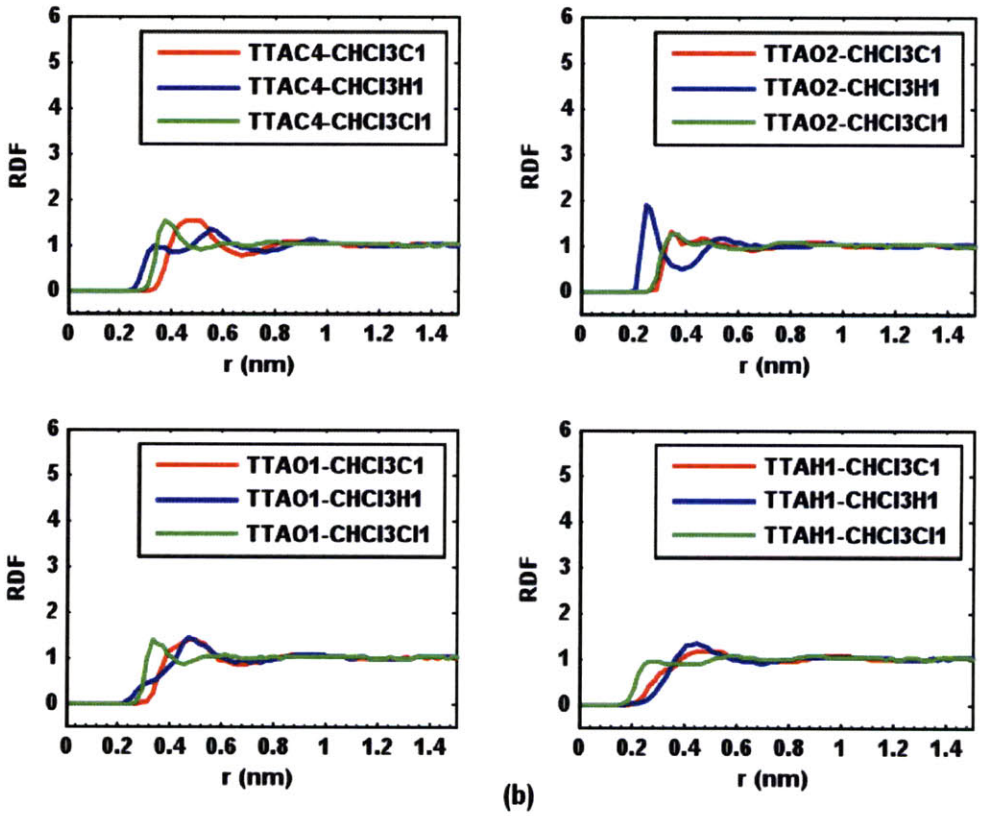
Table 3.6: Calculated solvation free energies of TTA in ethanol, chloroform, carbon tetrachloride, and dioxane at 298 K and 1 atm.

constant. This is mainly because the carboxylic acid group in tetrolic acid will be better solvated by polar solvents rather than non-polar solvents. However, ΔG_{solv} in dioxane is an exception, being comparable to or even higher than that in ethanol, although the polarity of dioxane is close to that of carbon tetrachloride. It is clear that solvent polarity can not fully explain the solvation free energies of TTA. More detailed structural characteristics of the solvent must be considered as well, which is reflected in the RDF plots in Figure 3-10. The peaks near $r = 0.17$ nm in TTAH1-EthanolO1 curve and TTAH1-DioxO1 curve indicate that strong interactions are formed between the hydroxyl hydrogen in TTA and the oxygen atom in ethanol and dioxane through hydrogen bonding. Ethanol and dioxane are both strong hydrogen bond acceptors as indicated by their β parameters. These interactions contribute greatly to the large solvation free energy values of tetrolic acid in ethanol and dioxane. It should also be noted that there is a peak near $r = 0.18$ nm in the TTAO2-EthanolH1 curve, whilst no peak exists near this position in the RDFs between TTAO2 and dioxane. This peak corresponds to the hydrogen bond formed between carbonyl oxygen of TTA and hydroxyl hydrogen of ethanol. This is consistent with the fact that ethanol is a strong hydrogen bond donor ($\alpha = 2.7$), while dioxane is not ($\alpha = 0.9$). The RDF curves obtained from carbon tetrachloride are significantly broader than those obtained from dioxane and ethanol. There are no significant preferential interactions between TTA and carbon tetrachloride. The carboxylic acid group is poorly solvated in carbon tetrachloride, which is a poor hydrogen bond donor/acceptor as indicated by its small α and β parameter values (Table 3.2). In chloroform, there is a clear peak near $r = 0.25$ nm on the TTAO2-CHCl3H1 curve, which

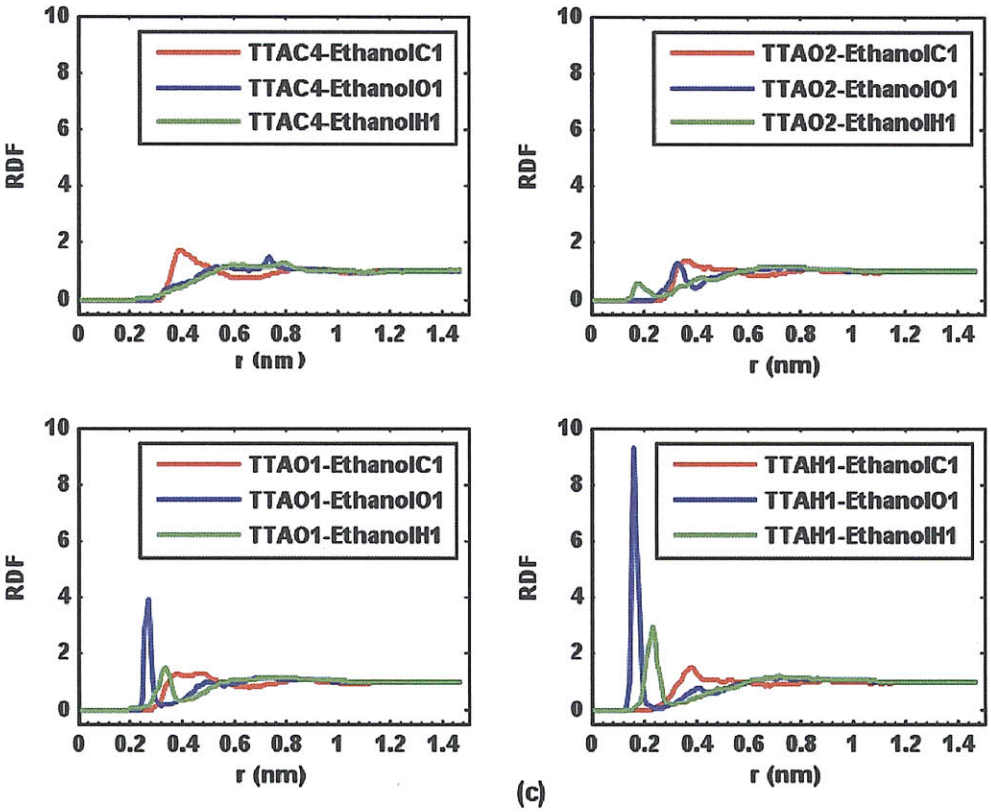
is shifted to the right compared to the first peak on the TTAO2-EthanolH1 curve. This peak corresponds to a weak hydrogen bonding interaction between the carbonyl oxygen of TTA and the hydrogen of chloroform, as a result of the weak hydrogen bond donor ability of chloroform. Besides that, there are no strong interactions between TTAO1 and chloroform, as indicated by the absence of clear peaks on all the TTAO1-CHCl3 curves. All these suggest that the solute-solvent interactions in chloroform are stronger than those in carbon tetrachloride, but weaker than those in ethanol and dioxane, which is consistent with their solvation free energy values.


TTAC4-CCl4C1
TTAC4-CCl4Cl1
TTAO2-CCl4C1
TTAO2-CCl4Cl1
TTAO1-CCl4C1
TTAO1-CCl4Cl1
TTAH1-CCl4C1
TTAH1-CCl4Cl1
RDF
r (nm)
0
1
2
3
4
5
6
0.2
0.4
0.6
0.8
1.2
1.4

(a)


TTAC4-CHCl3C1
TTAC4-CHCl3H1
TTAC4-CHCl3Cl1
TTAO2-CHCl3C1
TTAO2-CHCl3H1
TTAO2-CHCl3Cl1
TTAO1-CHCl3C1
TTAO1-CHCl3H1
TTAO1-CHCl3Cl1
TTAH1-CHCl3C1
TTAH1-CHCl3H1
TTAH1-CHCl3Cl1
RDF
r (nm)
0 0.2 0.4 0.6 0.8 1 1.2 1.4
0 1 2 3 4 5 6

(b)


TTAC4-EthanolC1
TTAC4-EthanolO1
TTAC4-EthanolH1
TTAO2-EthanolC1
TTAO2-EthanolO1
TTAO2-EthanolH1
TTAO1-EthanolC1
TTAO1-EthanolO1
TTAO1-EthanolH1
TTAH1-EthanolC1
TTAH1-EthanolO1
TTAH1-EthanolH1
RDF
r (nm)
0
2
4
6
8
10
0.2
0.4
0.6
0.8
1
1.2
1.4

(c)

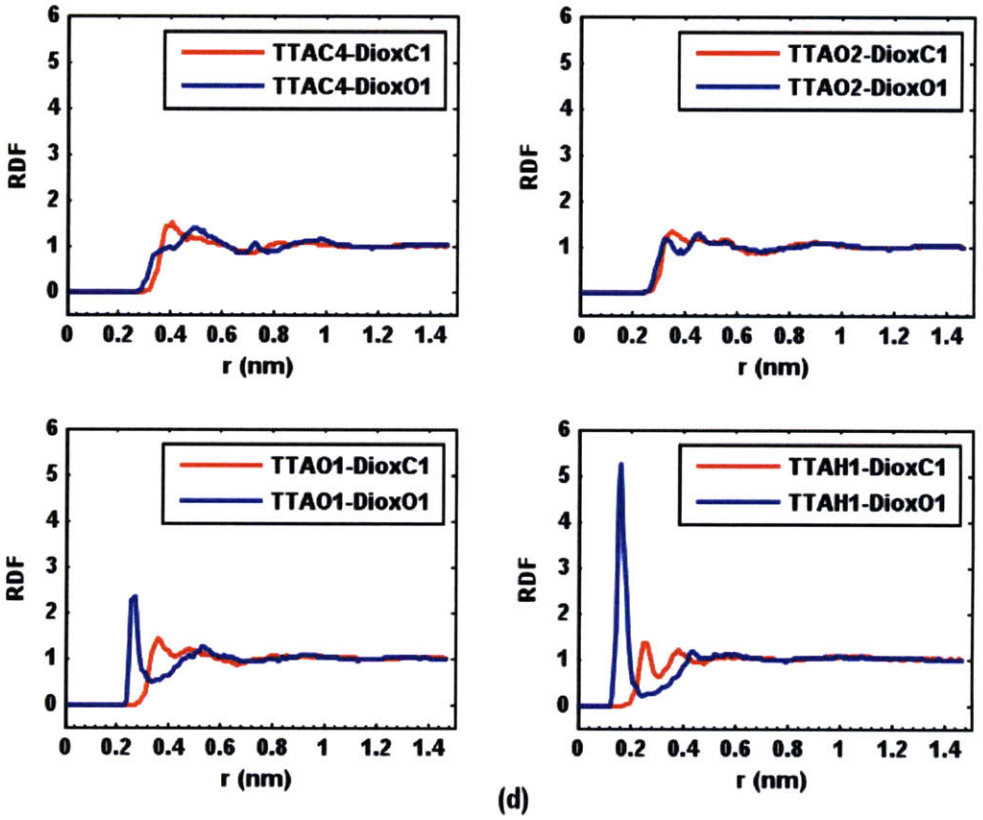


Figure 3-10: RDFs of the distances between the atoms in TTA and the atoms in various solvents at 298 K and 1 atm. (a) carbon tetrachloride, (b) chloroform, (c) ethanol and (d) dioxane.

3.4.3.2 Dimerization Reaction

As discussed in the computational details section, MD umbrella sampling was used to obtain free energy profile of the dimerization reaction. Figure 3-11 shows the free energy profiles of the dimerization reaction obtained from four different solvents. In the free energy curve obtained from the carbon tetrachloride solution (Figure 3-11a), there is a dip near OP = 1.94 $\mathring{A}$, which corresponds to the carboxylic dimer basin. For convenience, it is labeled state 1. When the OP is greater than 5.0 $\mathring{A}$, the free energy curve is almost flat. This corresponds to two completely separated (fully solvated) TTA molecules, labeled state 3. The dip near OP = 3.6 $\mathring{A}$ corresponds to the intermediate catemer motif, where only one hydrogen bond is formed between two tetrolic acid molecules, labeled state 2. The basin positions do not shift significantly when switching solvents. In carbon tetrachloride, state 1 is the most thermodynamically stable state among the three states. Noting that the OPs are not necessarily the true reaction coordinates, the barriers reported for the dimerization reactions may be underestimated. The true reaction coordinate could be obtained using likelihood maximization,[62,63] an approach that uses information theory to determine the best reaction coordinate from a list of trial coordinates. However, this would be another study in itself and we think that important qualitative insight can be gleaned from the OP results. There is almost no free energy barrier from state 3 to state 1.

In chloroform (Figure 3-11b), states 1 and 2 are nearly thermodynamically equally stable with a free energy difference of only 1.3 kJ/mol. There is no energy barrier for the formation of catemers from fully solvated molecules and there is an energy barrier of around 2.9 kJ/mol for the formation of carboxylic dimers from catemers, which suggests that the dimerization reaction for TTA in chloroform is less favorable than that in carbon tetrachloride. This may be due to the weak hydrogen bond between the hydrogen atom in chloroform and the oxygen atom in tetrolic acid. However, the 2.9 kJ/mol energy barrier is not very high and leaves states 1 and 2 both kinetically accessible. The polymorph resulting from the crystallization of TTA in chloroform might be a kinetically controlled process. At 298 K the dimer based structure (form I), the metastable form, nucleates first and does not transform into the more

stable catemer based structure (form II) in the time scale of the crystallization experiment.

In ethanol (Figure 3-11c), there are no clear dips to define the dimer and catemer basins. If we use the same order parameter value to define the carboxylic acid dimer, we can see that the free energy barrier to go from state 3 to state 1 is approximately 23.4 kJ/mol. State 1 becomes thermodynamically unfavorable, as does state 2. As discussed in the solvation free energy calculation section, the tetrolic acid molecule is well solvated in ethanol. The strong interactions between ethanol and tetrolic acid make the solute-solute interactions unfavorable. Thus forming the catemer motif is unfavorable and a dimer motif even more so.

In dioxane (Figure 3-11d), there are dips near 1.94 $\AA$ and 3.60 $\AA$, which correspond to the dimer and catemer motifs respectively. The dimer motif is less favorable than the catemer motif and the catemer motif is less favorable than two fully solvated molecules. Unlike those for the ethanol solution, the free energy differences between the labeled three states are much smaller, although ΔG_{solv} of TTA in dioxane is comparable to that in ethanol. The reason for this is that both oxygen atoms of TTA form strong interactions with ethanol due to ethanols strong hydrogen bond donor and acceptor abilities, whilst only the hydroxyl oxygen of TTA forms strong interactions with dioxane, due to dioxanes strong hydrogen bond acceptor ability but poor donor ability.

To summerize, the interactions between solute and solvent molecules determine the relative thermodynamic stability of different growth synthons as well as the barriers on the pathways connecting these synthons. This analysis used certain OPs in the dimerization reaction, but does not assume that the OPs are the true reaction coordinates. The free energy differences between those three states on the dimerization pathway estimated using the OP approach and their relative populations are listed in Table 3.7.

	ΔG_{32}	$K_{32}\left(\frac{[D]}{[C]}\right)$	$\Delta G_{32}^{\neq}$	ΔG_{21}	$K_{21}\left(\frac{[C]}{[M]^2}\right)$	$\Delta G_{21}^{\neq}$	ΔG_{31}	$K_{31}\left(\frac{[D]}{[M]^2}\right)$
carbon tetrachloride	-3.8	4.6	0	-6.7	14.9	~0	-10.5	68.0
Chloroform	-3.3	3.9	0	-1.3	1.7	2.9	-4.6	6.4
Ethanol	7.1	6E-02	7.1	16.3	1E-03	16.3	23.4	8E-05
Dioxane	2.9	0.3	2.9	5.4	0.1	~5.4	8.4	0.03

Table 3.7: Free energy differences between the three states on the dimerization pathway and their relative population. State 1, 2 and 3 are carboxylic acid dimer (2 hydrogen bonds formed), catemer (1 hydrogen bond formed) and two fully solvated TTA monomers (0 hydrogen bond formed) respectively. [D], [C] and [M] represent the concentrations of dimer, catemer and monomer respectively. ΔG_{ij} is the free energy difference between state i and state j, and $\Delta G_{ij}^{\neq}$ is the free energy barrier from state i to state j. They are all with the unit kJ/mol.

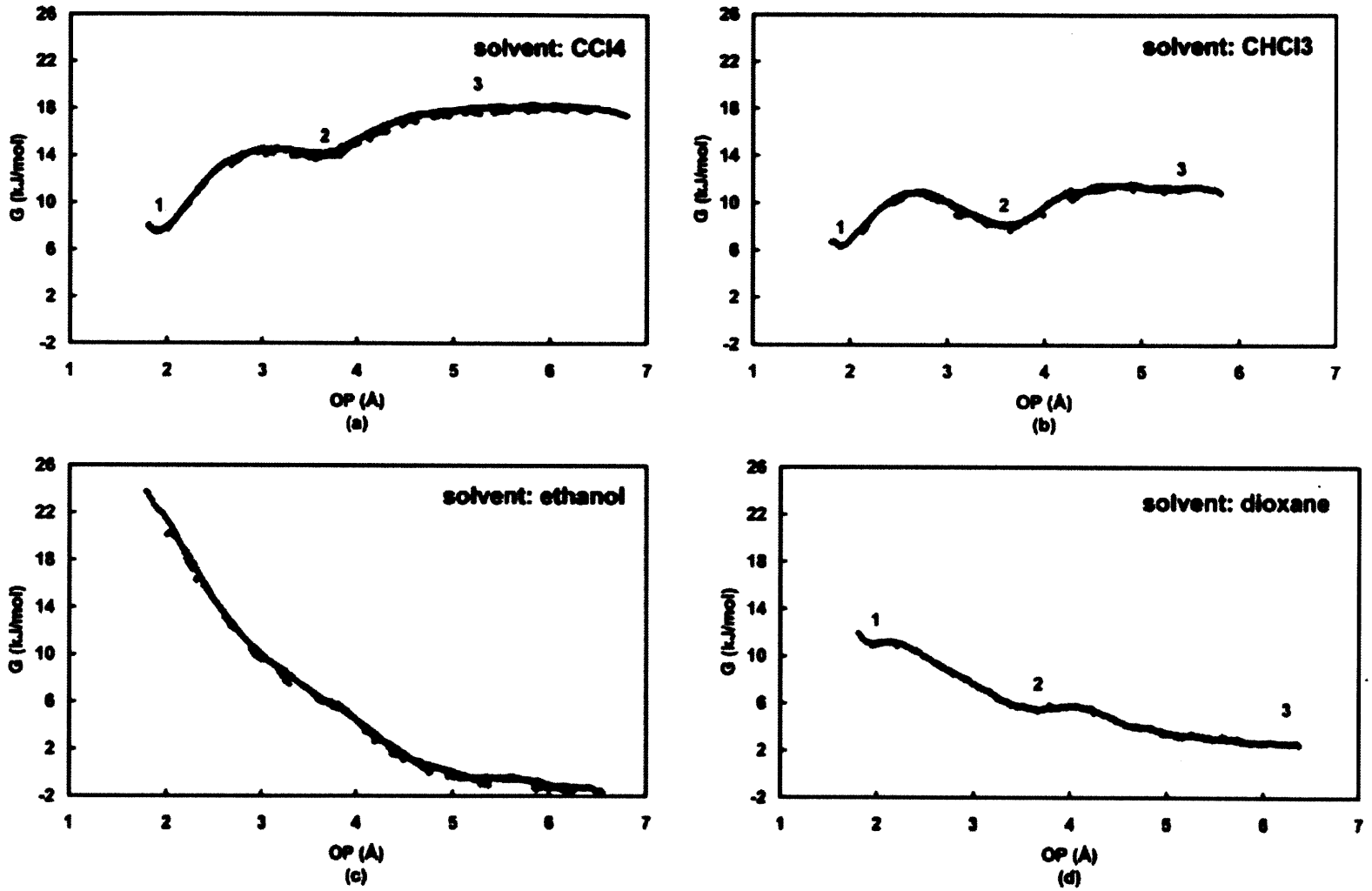
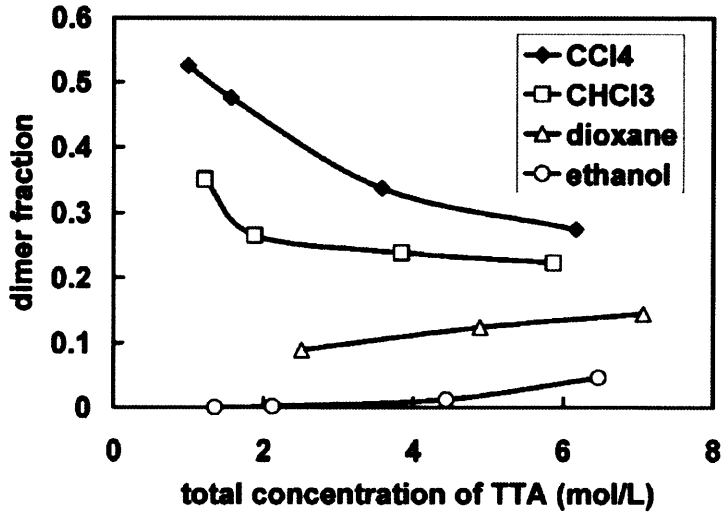


Figure 3-11: Free energy profiles for the TTA dimerization reation in various solvents at 298 K and 1 atm obtained using MD umbrella sampling. (a) carbon tetrachloride (b) chloroform (c) ethanol (d) dioxane.

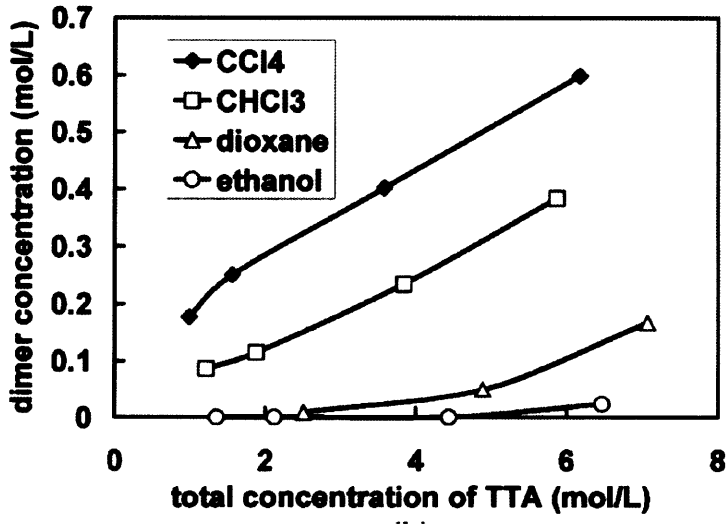
3.4.3.3 Dimer Fraction of Hydrogen Bond Network

As shown in Figure 3-12a, dimer growth synthons exist in all of the tetrolic acid solutions, no matter what polymorph is obtained from crystallization. However, the choice of solvent clearly affects the dimer composition of the hydrogen bond network of tetrolic acid. The dimer fraction is the highest in carbon tetrachloride solution and the lowest in ethanol solution, with the values for chloroform and dioxane in between. This is consistent with the order of ease for carboxylic acid dimer formation in these four solvents (as shown in Figure 3-11). Moreover, the dimer fraction curves decrease to the right in chloroform and carbon tetrachloride, from which the dimer based structure is obtained, but increase to the right in ethanol and dioxane, where give the catemer based structure and the solvate respectively. This fact can be used to help select the correct solvent to get the desired polymorph. One possible reason for this phenomenon is that stronger interactions between TTA molecules in chloroform and carbon tetrachloride cause the formation of larger clusters (trimer, tetramer, etc.), which reduces the opportunity to form a carboxylic acid dimer and thus decreases its composition in the entire hydrogen bond network. In another words, the hydrogen bond network expands faster than the dimer motif. Some snapshots of the conformations that form instead of the dimers are shown in Figure 3-13. Similar conformations were also reported previously by Gavezzotti[50] in his work on acetic acid in carbon tetrachloride. In ethanol and dioxane, the interactions between TTA molecules and the solvent molecules are much stronger. The chance for two TTA molecules to meet and form a dimer is very low when the solution concentration is low, but it increases as the total concentration increases. In conclusion, the decreasing dimer fraction curves relate to the dimer based polymorph. This is due to solute-solute interactions overcoming solute-solvent interactions. The opposite is the case for the increasing dimer fraction curves, which relate to the catemer based polymorph or solvate. This is the result of solute-solvent interactions overcoming solute-solute interactions. Whilst the dimer fraction curve of TTA from chloroform and carbon tetrachloride show a downward trend, the absolute concentration of dimer motifs still increases with the total TTA concentration in all the four solvents, as shown in Figure 3-12b. Moreover, the dimer motifs

are more abundant in carbon tetrachloride and chloroform than in ethanol and dioxane at any TTA concentration level, an observation which supports the link hypothesis.



(a)



(b)

Figure 3-12: TTA clusters in various solutions, (a) dimer fraction of the hydrogen bond network and (b) dimer concentration.

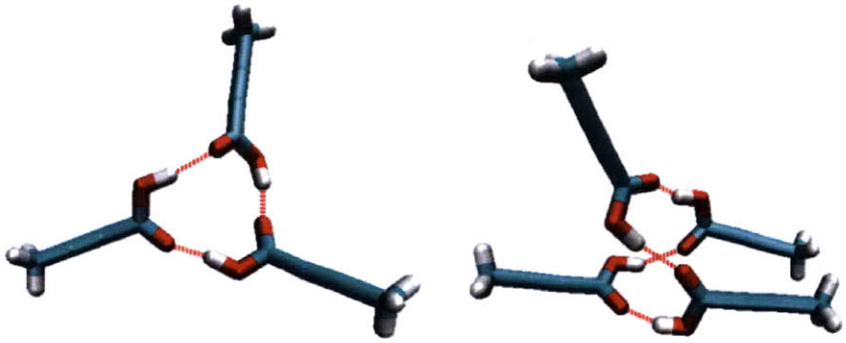


Figure 3-13: The clusters reducing TTA's potential to form carboxylic acid dimer (obtained from the MD simulations of TTA in carbon tetrachloride at 6.17 mol/L). Hydrogen bonds are shown in red dashed lines.

3.4.4 Conclusions of the Self-Assembly Study of Tetrolic Acid

There is indeed a link between the self-assembly of solute molecules in solution and the polymorphic outcome of crystallization for tetrolic acid. The interactions between solvent and solute molecules play an important role in the self-assembly process. The solvent can not only change the relative thermodynamic stability of the different growth synthons of tetrolic acid but also modify the energy barriers along the pathway to create these growth synthons. Weak interactions (low solvation free energy value) between the solvent and TTA molecules prompt two solute molecules to assemble into a carboxylic dimer, and also increase the dimer fraction of the overall hydrogen bond network formed among all TTA molecules. This explains why TTA crystals obtained from chloroform pack into a dimer based structure. We also expect that the TTA crystal obtained from carbon tetrachloride solution will be in a dimer based form. As the solute-solvent interactions become stronger (high solvation free energy value) the formation of a carboxylic dimer from two TTA molecules becomes both thermodynamically and kinetically unfavorable. In addition, the propensity of TTA molecules to form an extensive hydrogen bond network is reduced and the dimer fraction of the overall hydrogen bond network decreases significantly. Strong solute-solvent interactions cause TTA molecules to crystallize into either catemer based structure or a solvate. However,

we can not determine which one will be formed based solely on interactions in solution. To explain fully why TTA forms a solvate with dioxane further investigation into the interactions of the TTA and solvent molecules in the solid phase is required, a task beyond the scope of this study.

3.5 Glycine

3.5.1 Polymorphism of Glycine

Glycine, a simple organic molecule with only ten atoms, is of great interest to researchers for the study of polymorphism both experimentally and computationally. It has three forms, α, β, and γ, with thermodynamic stability in the order $\gamma > \alpha > \beta$.[64] α glycine is packed in centrosymmetirc dimers and crystallizes from aqueous solutions,[65] β gycine is packed through a twofold screw-symmetry axis and crystallizes from water-alcohol solutions[66] and γ glycine is packed in helical chains around a threefold screw axes parallel to the c-axis and crystallizes from acidic aqueous solutions and sometimes from aqueous solutions as well with a rate about 500 times slower than α glycine.[67] CCDC refcodes and some unit cell parameters for all three forms of glycine crystals are listed in Table 3.8 and their structures are plotted in Figure 3-14.

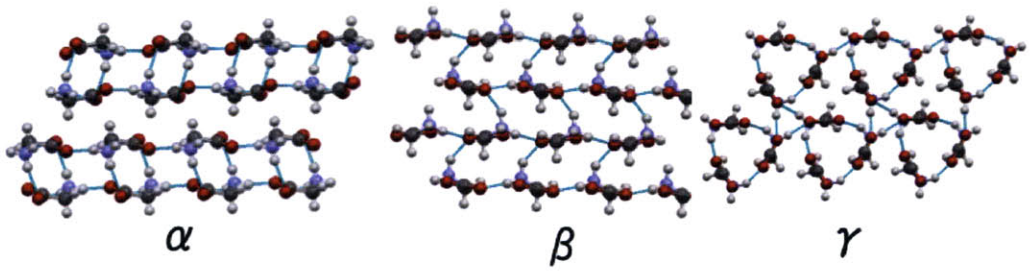


Figure 3-14: Crystal structures of α, β and γ glycine, viewed along the a-axis, a-axis and b-axis respectively.

3.5.2 The Link Hypothesis for Glycine and the Controversy

The link hypothesis suggests that molecules self-assemble into various structures in the solution phase and the most stable/populated one gets used as the structural synthon in the crystallization process.[32] Both experimental and molecular simulation approaches have

	α glycine	β glycine	γ glycine
refcode	GLYCIN02	GLYCIN	GLYCIN01
a	5.1020	5.077	7.037
b	11.9709	6.268	7.037
c	5.4575	5.380	5.483
alpha	90	90 90	
beta	111.70	113.20	90
gamma	90	90	120
no. of glycine per cell	4	2	3

Table 3.8: Refcodes and cell parameters of glycine crystals.

been applied to study the formation of glycine clusters in aqueous solutions and to explore the relationship between self-assembly in solution and the solid state outcome. However, this leads to a debate of whether the link hypothesis is valid to explain the polymorphism of glycine. The diffusion-coefficient measurements[68] and the small angle X-ray scattering (SAXS) experiments[45,46] of the supersaturated aqueous solutions of glycine performed by Myerson and coworkers suggested that the majority of glycine molecules exist as dimers in aqueous solutions and therefore it crystallizes as the cyclic-dimer based α form. Hamad and Catlow performed a molecular dynamics study of various aqueous glycine solutions (from undersaturated to supersaturated) using the AMBER force field and suggested that the glycine monomer is the dominating species in aqueous solutions.[47] The percentage of glycine molecules existing in monomers is almost twice as that in dimers in 3.6 mol/L glycine aqueous solutions at 20 ^oC. Huang and Yu performed freezing-point depression studies of glycine aqueous solutions and suggested that glycine exists mainly as monomers, not dimers, near 0 ^oC.[48] Their results showed that approximately 25% of glycine molecules exist as dimers in 2.92 mol/kg of H_2O glycine solution. Although the debate around what is the dominating species in aqueous glycine solution is still going on, there is a common shortfall in all these studies. They all focused only on the size of the clusters but not on the structural information. For example, they do not differentiate a cyclic dimer and an open chain dimer. From the link hypothesis, only the formation of more cyclic dimers in the solution leads to the

crystallization of α glycine, which utilizes the cyclic dimer as the packing unit, while the open chain dimer can be the building unit for all three polymorphic forms, as shown in Figure 3-15. Moreover, very little information about the glycine clusters in water-alcohol mixtures was gathered, which could, from another perspective, provide us additional knowledge of the applicability of the link hypothesis in understanding the polymorphism of glycine.

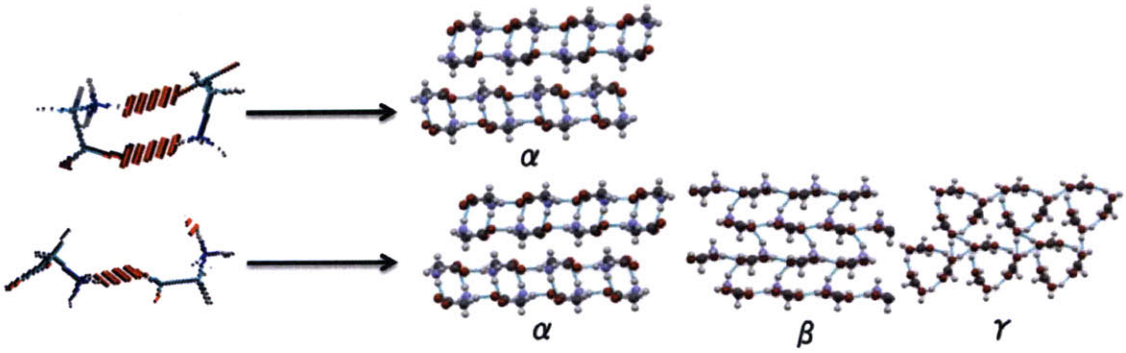


Figure 3-15: Link hypothesis for the polymorphic system of glycine. Cyclic dimer structure is only seen in α glycine,while open chain dimer structure can be found in all three forms. Therefore, only cyclic dimer is the corresponding growth synthon in solutions for α glycine by the argument of the link hypothesis.

3.5.3 Objectives

In this study, we revisited the self-assembly of glycine in solution using molecular dynamics. Our focus was placed on gathering some structural information on the glycine clusters formed in both pure and alcoholic aqueous solutions with the aim to clarify the controversy surrounding the existence of a link between the structure of the glycine clusters formed in solution and that of the polymorph obtained from solution crystallzation. Furthermore, we also studied the structure of those glycine clusters formed in the vicinity of the α crystal (010) face. This surface is composed of centrosymmetric double layers that are held firmly intact by hydrogen bonds between cyclic molecular pairs (as shown in Figure 3-14 α glycine structure) and is also of morphological significance. The objective was to investigate whether

the existence of an ordered crystal face would induce the solute molecules in the solution to order in a similar structure and thus facilitate the crystallization of that polymorphic form.

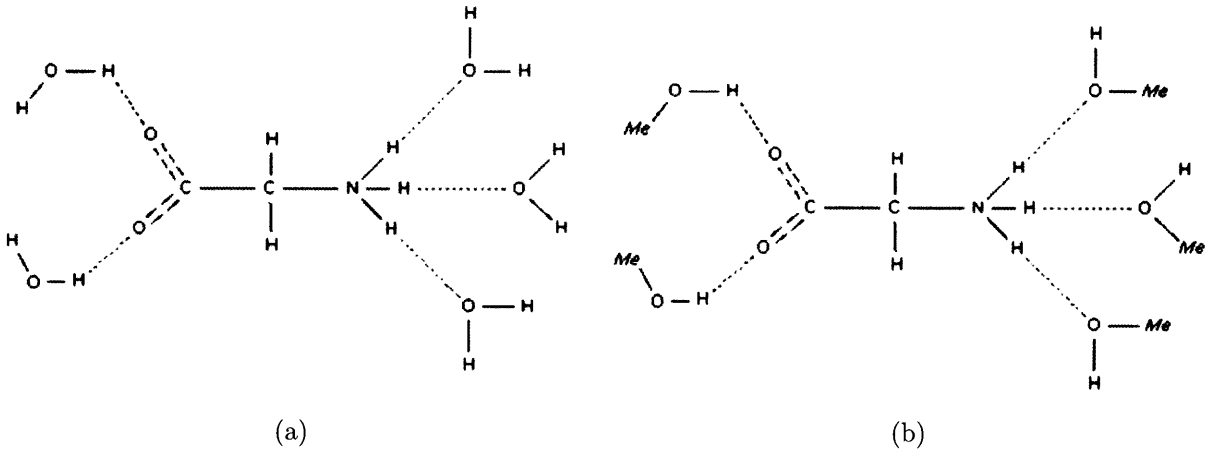
3.5.4 Force Field Development

In order to investigate the influence of solvent on the polymorphic outcome of glycine, a force field, which can not only reproduce the three glycine polymorphs at ambient temperature but also describe glycine-solvent interactions equally well, was required. Methanol molecules were modeled using the OPLS-AA potential,[14] which has a good library of organic solvents and is optimized for liquid simulation. Water molecules were modeled using the TIP3P potential[69] with the LJ parameters obtained from the OPLS-AA force field. To maintain the consistency among all the potentials, it would be ideal to model the glycine molecules using the OPLS-AA potential as well, if it satisfied the two criteria mentioned above. As noticed in many previous studies, the partial charges are the most influential parameters in modeling crystals. We carefully tested two charge sets for the glycine zwitterion. The first charge set uses the assembled OPLS-AA charges, with the charges for the NH3+ and COO- taken directly from the glycine terminal groups, the charges for the CH2 group hydrogen atoms taken from the glycine amino-acid residue, and the charge on C_α being adjusted to give a neutral ziwtterion. This approach was previously employed by Price *et al.* in their study of glycine crystals using the AMBER potential.[70] The second charge set uses the CHELPG charges of the glycine zwitterion obtained directly from *ab initio* calculations. Both of these two charge sets are directly available in the OPLS-AA force field. Here we will not look at the fundamental theories of how they were derived. Although the OPLS-AA potential was preferred for glycine to maintain consistency with those of methanol and water molecules, we also investigated another candidate potential for glycine, the AMBER potential, which was shown previously to be able to reproduce the crystal structures of the three polymorphs of glycine and to correctly predict their relative energy ranking.[70] We performed two tests to compare the OPLS-AA potential with assembled charges, the OPLS-AA with CHELPG charges and the AMBER potential.

In the first test, the ability of these three potentials to reproduce the crystal structure of all three polymorphic forms of glycine at ambient temperature was investigated. Molecular dynamics was applied to simulate crystal supercells in a NPT ensemble (298K and 1 atm), with periodic boundary conditions. All MD simulations in this work were performed using CHARMM. Simulation boxes of $6a \times 3b \times 6c$ unit cells for α (432 molecules), $7a \times 5b \times 7c$ unit cells for β (490 molecules), and $5a \times 5b \times 6c$ unit cells for γ (450 molecules) were built to ensure that all simulation box edges were at least 28 $\AA$, two times of the cutoff distance for nonbonded interactions. The particle mesh Ewald summation method was used to correct for the long range electrostatic interactions. When the AMBER potential was used, an arithmetic average combination rule was employed, the charge-charge interactions between atoms separated by 3 bonds (1-4 interactions) were scaled by 1.2 and the corresponding van der Waals interactions were scaled by 2. When the OPLS potential was used, a geometric average combination rule was employed and all the non-bonded 1-4 interactions were scaled by 2. The monoclinic crystal type was used for both α and β forms which each has 4 degrees of freedom (a, b, c, and beta) and the hexagonal crystal type was used for the γ form which has two degrees of freedom (a, c). All simulation boxes were sampled every 1 ps for 1 ns (with 1 fs time step), after an equilibration run for 1 ns. The percentage change of the lattice parameters (PCLP) and root mean squared difference (RMSD), compared to the experimental X-ray structures, were calculated using the average structure of the 1000 frames sampled every 1 ps in the production run. The relative energies of the three glycine forms were calculated by averaging the energies sampled every 1ps, normalizing using the number of molecules in the simulation box and resetting the lowest value to zero.

In the second test, the interaction energies of glycine-water and glycine-methanol were investigated. Hydrogen bond energies between glycine and water/methanol molecules were calculated following the approach developed by MacKerell.[71] In this approach, a water/methanol molecule was placed in the vicinity of a hydrogen bonding site of a glycine molecule with fixed orientation. The configurations of glycine and water molecules were obtained from the α glycine X-ray structure and MacKerell's work respectively. The configuration of methanol

was optimized using MP2/6-31G* in vacuum. The potential energy of the pair was optimized with regard to the distance between the water/methanol molecule and the glycine molecule with all the other internal coordinates fixed, as shown in Figure 3-16. The hydrogen bonding energy was calculated using the optimized configuration. This procedure was performed for all the hydrogen bonding sites of glycine including the two oxygen sites of the COO- group and the three hydrogen sites of the NH3+ group, using all the three empirical potentials. An *ab initio* method (MP2/6-31G*) was used as a reference.



(a) (b)

Figure 3-16: Schematic demonstration of the calculations of the interaction energies between the hydrogen bonding sites of glycine and (a) water/(b) methanol.

The MD calculated structures of the three glycine polymorphs clearly depend on the choice of charges and the choice of potential models, as suggested by the PCLP and RMSD values listed in Table 3.9. Although it succeeded in reproducing the stability order of the three polymorphs in Prices fixed cell angle simulations,[70] The AMBER potential failed in our simulations with the adjustable angle beta for α and β glycine. This additional degree of freedom, angle beta, which monoclinic unit cells have, gave lower energies values for α and β glycine than for γ glycine. This is inconsistent with the experimental fact that γ is the most stable form. A similar trend was also observed when the other two OPLS-AA potentials were used. Therefore, we could not draw any conclusion as to which potential is superior based on their ability to predict the relative stability of the three polymorphic forms. A close

		PCLP a	PCLP b	PCLP c	RMSD	beta (Å)	relative energies (kJ/mol)
OPLS with assembled charges	α form	-2.1%	2.6%	-3.7%	0.6%	0.5	0.0
	β form	-3.3%	4.8%	-4.8%	0.9%	0.8	4.2
	γ form	-0.3%	-0.3%	-1.6%		1.3	5.0
OPLS with CHELPG charges	α form	-1.2%	4.0%	-2.4%	0.3%	0.5	0.0
	β form	-2.1%	6.0%	-3.4%	0.6%	0.7	3.0
	γ form	0.5%	0.5%	-0.8%		0.5	3.3
AMBER	α form	-3.6%	7.8%	-2.9%	0.7%	1.1	0.0
	β form	0.5%	3.3%	-3.9%	0.4%	0.6	0.5
	γ form	-1.7%	-1.7%	-1.3%		0.3	0.9

Table 3.9: The percentage change of lattice parameters (PCLPs), the root mean squared difference (RMSD) and the relative energies of the three polymorphs of glycine modeled using three different sets of force field parameters.

examination of the PCLP values shows that all the three potentials gave similar deviations for most of the lattice parameters. The biggest deviation was the cell parameter b of α glycine, which increased by 7.8% when simulated using the AMBER potential. The RMSD is usually a good indicator of whether the ensemble average structure can preserve the local interactions of the experimental X-ray structure, such as hydrogen bonds. The OPLS-AA potential with CHELPG charges gave reasonable RMSD values for all three forms of glycine, while the OPLS potential with assembled charges and the AMBER potential gave significantly higher RMSD values for γ and α glycine respectively. With that, we conclude that the three potentials tested in this work perform similarly in simulating glycine crystals with the OPLS-AA potential with CHELPG charges being slightly better in preserving local structures.

Hydrogen bond energies between glycine and water/methanol molecules calculated using both the *ab initio* method and the empirical force field method are listed in Table 3.10. When glycine is modeled using the OPLS-AA potential with CHELPG charges, the differences between the hydrogen bond energies calculated using the empirial force field method and

the *ab initio* MP2 method are the smallest on average for both water and methanol.

Combining the results of the solid phase simulations of glycine crystals and the hydrogen bond energies between glycine and water/methanol, we decided to use the OPLS-AA potential with CHELPG charges for glycine.

	ab initio (MP2)	OPLS-AA with assembled charges		OPLS-AA with CHELPG charges		AMBER	
	energy (kJ/mol)	energy (kJ/mol)	deviation from MP2 (kJ/mol)	energy (kJ/mol)	deviation from MP2 (kJ/mol)	energy (kJ/mol)	deviation from MP2 (kJ/mol)
O1(gly)-H(wat)	-46.5	-59.4	-13.0	-49.8	-3.8	-56.5	-10.5
O2(gly)-H(wat)	-42.3	-56.1	-13.8	-46.9	-4.6	-53.6	-11.7
H1(N, gly)-O(wat)	-53.2	-62.0	-8.4	-52.7	0.4	-54.0	-0.4
H2(N, gly)-O(wat)	-49.4	-49.0	0.4	-41.9	7.5	-44.4	5.0
H3(N, gly)-O(wat)	-52.3	-52.3	0.0	-44.4	7.5	-47.3	5.0
O1(gly)-H(moh)	-48.1	-57.8	-10.0	-49.4	-1.7	-56.1	-8.0
O2(gly)-H(moh)	-44.8	-54.8	-10.5	-46.9	-2.1	-53.6	-8.8
H1(N, gly)-O(moh)	-53.6	-61.1	-7.5	-52.7	0.8	-54.4	-0.8
H2(N, gly)-O(moh)	-49.8	-51.1	-0.8	-44.0	5.9	-46.9	2.9
H3(N, gly)-O(moh)	-49.8	-52.7	-2.9	-46.0	4.2	-49.0	0.8

Table 3.10: Hydrogen bond energies between glycine and water/methanol, calculated using both the *ab initio* method and the empirical force field method.

3.5.5 Glycine Clusters in Solution

Molecular dynamics simulations for both aqueous glycine and 30% (mole fraction) methanol-water solutions at various concentrations (Table 3.11) were conducted at 298 K and 1 atm to study the structure of the glycine clusters formed in solutions. The reported experimental solubility of α glycine is 25g/100ml in water at 25 ^oC, and reduces to 2.9 g/100 ml in 50% (v/v) methanol-water mixtures. The concentrations we chose in this study range from undersaturation to supersaturation. The cyclic dimer fraction of the overall hydrogen bonds was used to characterize the hydrogen bond network among all the glycine molecules. It was defined as:

$$\text{cyclic dimer fraction} = \frac{\text{total no. of hbonds in cyclic dimers}}{\text{total no. of hbonds in all glycine molecules}} \tag{3.6}$$

The cyclic dimer fractions of α, β, and γ glycine crystals are 0.33, 0.0, and 0.0 respectively. The hydrogen bonds between glycine molecules are defined as the (C)O-H(N) distance being no larger than 2.2 $\AA$ and the N-H-O(C) angle being no less than 140 ^{o}, which is the same as the criteria used in the previous computational study performed by Hamad.[47] Each simulation box was equilibrated for 2 ns and sampled for 14 ns to calculate cyclic dimer fractions with periodic boundary conditions and the smooth particle mesh Ewald summation method to correct for the long rang electrostatic interactions. The time step used in the simulations was 1 fs and the samples were taken every 1000 steps (1 ps interval).

As shown in Figure 3-17, cyclic dimers exist in both pure and alcoholic solutions for the concentration range studied here. The choice of solvent clearly affects the dimer composition of the hydrogen bond network of glycine. The dimer fractions are higher in the methanol-water solutions than those in the aqueous solutions of the same concentration, which suggests that the glycine clusters in the methanol-water mixtures ressemble the structure of α glycine more so than those formed in pure water. This surprisingly contradicts the link hypothesis, which argues that the cyclic dimer should form more easily and be the more dominant structure in aqueous solutions rather than in alcoholic solutions, since a cyclic dimer based structure (α form) crystallizes from the former and an open chain based structure (β form)

glycine-water solutions			glycine-methanol-water solutions			
no. of glycine	no. of water	conc. (mol/L)	no. of glycine	no. of water	no. of methanol	conc. (mol/L)
81	4500	1.0	81	3150	1350	0.7
144	4000	1.9	144	2800	1200	1.4
216	4000	2.7	216	2800	1200	2.0
256	4000	3.1	256	2800	1200	2.4
324	4000	3.8	324	2800	1200	2.9

Table 3.11: The number of molecules in the simulation boxes for the study of glycine clusters in solutions.

crystallizes from the latter. It is completely opposite here. Moreover, the dimer fraction decreases as the concentration of glycine increases and the number of hydrogen bonds formed per glycine molecule increases with the concentration, as shown in Figure 3-18. This indicates that the hydrogen bond network among all glycine molecules tends to propagate faster in a form other than cyclic dimers and the cyclic dimer structure is less favored in supersaturated solutions. The fact that the number of hydrogen bonds formed per glycine molecue is larger in methanol-water solutions than in pure aqueous solutions again suggests that glycine molecules tend to aggregate more easily in the former. All these facts strongly point to that cyclic dimers are more favored in the methanol-water mixtures than in the pure water and the link hypothesis does not work here.

To further test the link hypothesis, we compared the thermodynamic stabilities of a cyclic dimer, an open chain dimer and two fully solvated glycine monomers in both aqueous and methanol-water solutions directly by performing the free energy calculations using an umbrella sampling technique which directs two fully apart glycine molecules to move closer along a particular axis (order parameter) by adding a harmonic biasing potential. The order parameter (OP) used to direct the sampling process is defined by the average of C1(O)-N2 and C2(O)-N1 distances. For each simulation, two glycine molecules were inserted into a pre-equilibrated solvent box. To adequately sample, one glycine molecule was confined in

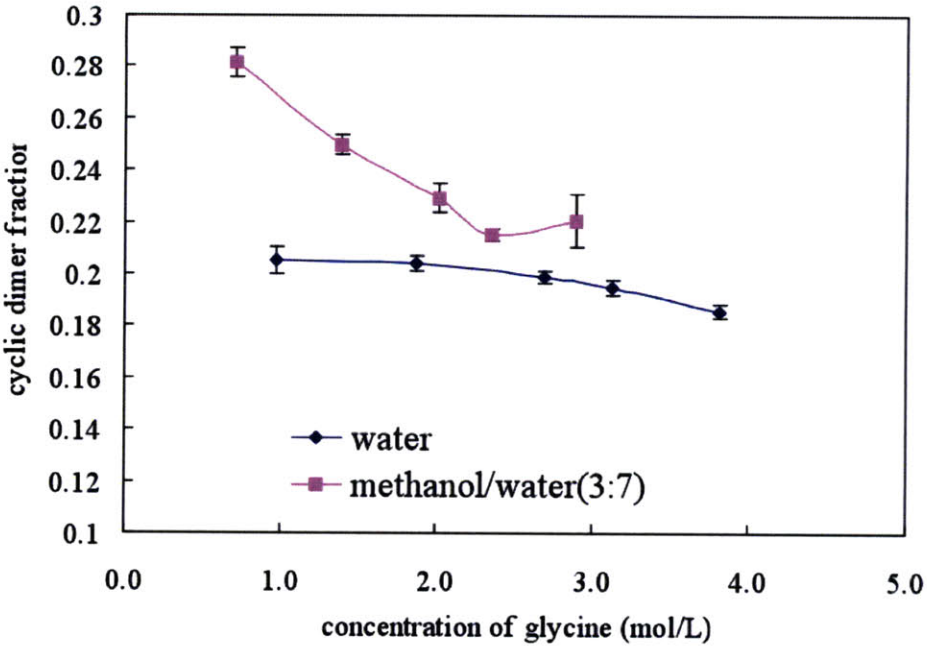


Figure 3-17: Cyclic dimer fraction of glycine hydrogen bond networks in solutions at 298 K and 1 atm. The distance cutoff for hydrogen bond is (N)H-O(C) < 2.2 $\AA$ and the angle cutoff for hydrogen bond is N-H-O(C) > 140 ^{o}.

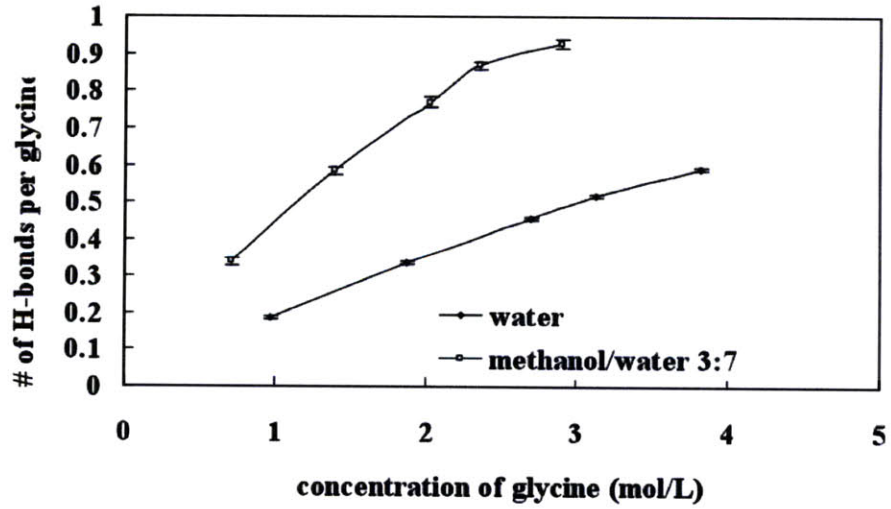


Figure 3-18: The number of hydrogen bonds per glycine molecule in various solutions at 298 K and 1 atm.

the center of the simulation box using a harmonic restraint with a force constant of 836.8 kJ/(mol·$\mathring{A}^2$). A harmonic functional form of an umbrella potential was used, which is the same as the one used in the tetrolic acid study.

Figure 3-19 shows the free energy profiles for the formation of a cyclic dimer from two fully solvated monomers in two different solvents obtained using MD umbrella sampling. On both curves, there is a dip near OP = 3.7 $\mathring{A}$, which corresponds to the cyclic dimer basin. When the OP is greater than 6.0 $\mathring{A}$, the free energy curve is almost flat. This corresponds to two completely separate (fully solvated) glycine molecules. The dip near OP = 5.2 $\mathring{A}$ corresponds to the intermediate open chain dimer, where only one hydrogen bond is formed between two glycine molecules. The basin positions do not shift significantly when switching solvents. In both water and methanol-water 3:7 mixture, the open chain dimer is more stable than both the cyclic dimer and the monomer and the free energy gain for the formation of an open chain dimer from two monomers is slightly larger in methanol-water mixtures than in pure water. This confirms our findings in the previous study of the cyclic dimer fractions that glycine molecules in concentrated solutions tend to aggregate more in methanol-water mixtures than in pure water and that glycine clusters adopt some structures other than the cyclic dimer form more easily. The free energy difference between the cyclic dimer and open chain dimer is about 0.9 and 2.5 kJ/mol in methanol-water mixtures and pure water respectively, which suggests that although cyclic dimer structure exist in both solvents, it is more readily formed in methanol-water mixtures than in pure water, not the other way around as suggested by the link hypothesis.

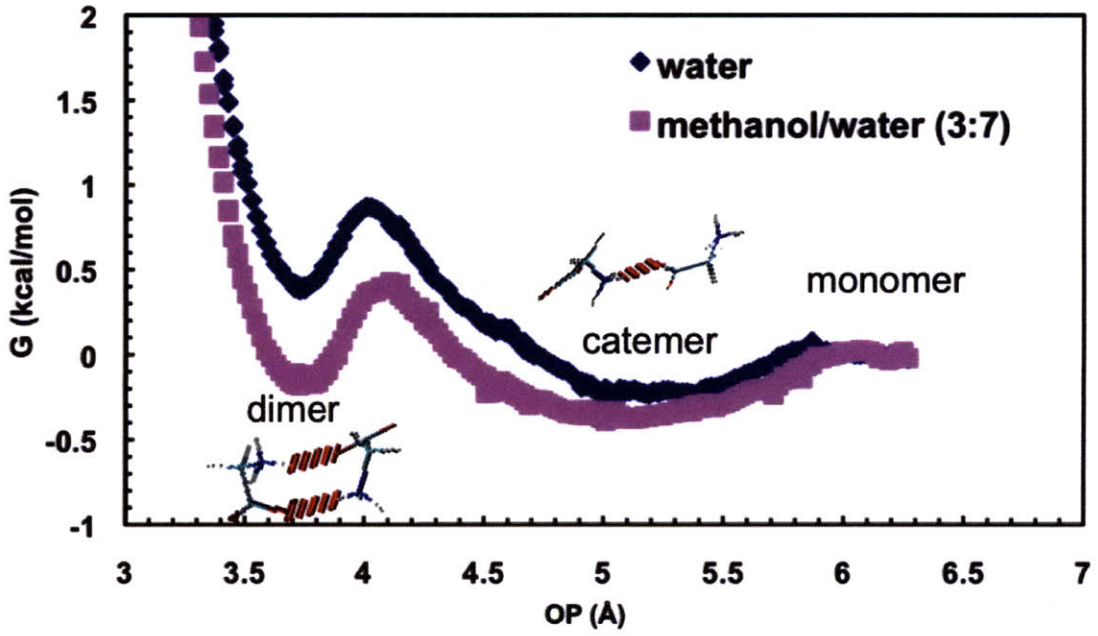


Figure 3-19: Free energy profiles for the glycine dimerization reaction in water and methanol-water 3:7 mixtures at 298 K and 1 atm obtained using the MD umbrella sampling method.

3.5.6 Glycine Clusters near the α Glycine Crystal Surface (010)

We also investigated the impact of the presence of the α glycine (010) crystal surface on the structure of glycine clusters formed in its vicinity. This crystal surface is composed of double layers of centrosymmetirc dimers, similar to the double layer structure of a membrane. The objective was to check if there is any increase in the cyclic dimer fraction of glycine clusters in solution when contracting with an ordered crystal surface. The study of the methanol-water solutions was not performed, since α glycine does not grow from methanol-water mixtures. A crystal-solution-crystal sandwich was set up, with the α glycine (010) crystal surface contacting with the glycine solution (as shown in Figure 3-20). Five glycine solutions were used here, with the number of glycine and water molecules exactly same as those used in the previous solution studies for direct comparison. Two $10a \times 2b \times 10c$ α glycine super cells were placed on each side of the solution box. To prevent the glycine molecules on crystal surface from diffusing into the solution phase, a harmonic restraint with a force constant 8.36 kJ/(mol·$\AA^2$) was applied to the root mean squared difference of the atoms other than hydrogens of the crystal to a reference structure. The reference structure was calculated by pre-equilibrating the crystal super cell at 298 K and 1 atm and extracting the average structure. A 16 ns MD simulation was performed for the simulation box in a NPT ensemble (298 K and 1 atm) with 3D periodic boundary conditions and the particle mesh Ewald summation method. After equilibration (the first 2 ns), the thickness of the solution phase in the y direction was around 56 $\AA$, which is larger than twice the cutoff distance for the none bonded interactions (14 $\AA$). The dimer fraction of the glycine molecules in the solution phase was calculated using the frames taken from the last 14 ns run with sampling interval 1ps.

The cyclic dimer factions of glycine molecules in aqueous solutions contacting with the α glycine (010) face are plotted in Figure 3-21, together with the dimer fractions without the crystal surface. It is clear that the presence of the crystal surface with the cyclic dimer double layer increases the cyclic dimer composition of glycine clusters formed in the solution phase. The increase is consistent for all the concentrations studied here. Although this

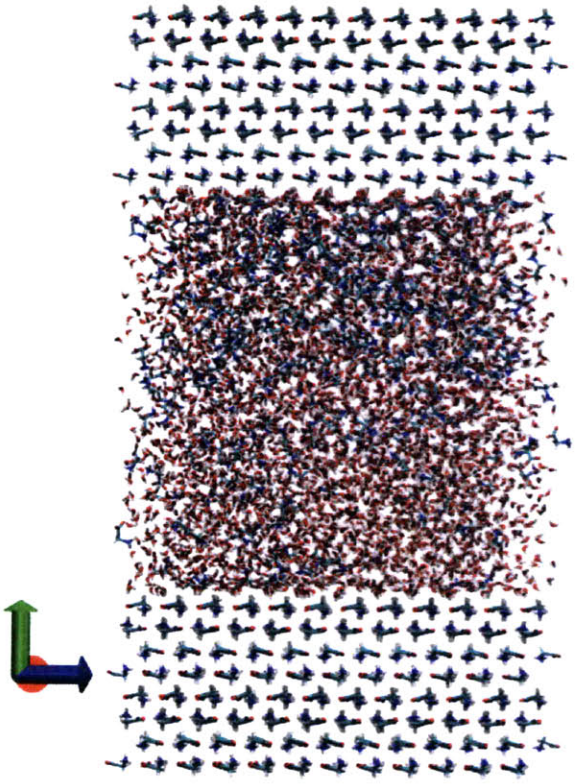


Figure 3-20: Simulation box of the 3.8 mol/L glycine aqueous solution contacting with α glycine (010) face, viewed along the a-axis of the crystal, with 324 glycine molecules and 4000 water molecules in the center solution phase, 800 glycine molecules in the upper crystal phase, and 800 glycine molecules in the lower crystal phase.

increase can probably facilitate the growth of α glycine, it is not large enough to strongly support the link hypothesis or to explain the selective crystallization of α and β glycine since the cyclic dimer fractions of glycine clusters in methanol-water 3:7 mixtures are still higher.

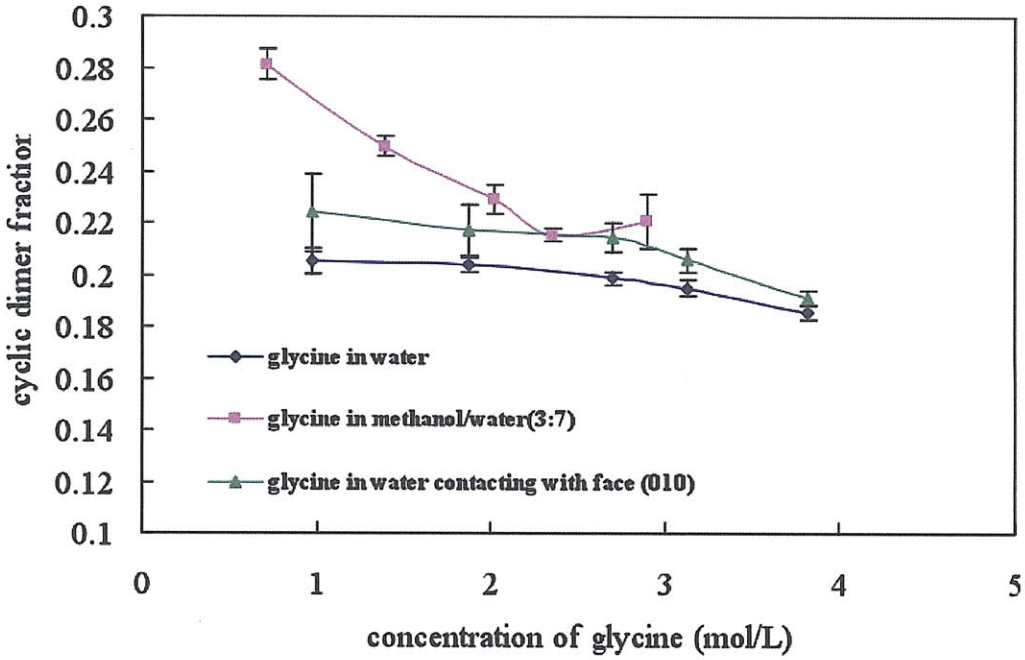


Figure 3-21: Cyclic dimer fractions of the glycine hydrogen bond networks in various solutions at 298 K and 1 atm contacting with the (010) face of α glycine, plotted together with the results without the crystal surface. The distance cutoff for hydrogen bond is (N)H-O(C) < 2.2 $\AA$ and the angle cutoff for hydrogen bond is N-H-O(C) > 140 ^{o}.

3.5.7 Conclusions of the Self-Assembly Study of Glycine

The link hypothesis does not work for the polymorphism in glycine. Glycine clusters formed in methanol-water 3:7 mixtures have higher cyclic dimer composition than those formed in pure water with the same concentration. The glycine open chain dimer is always more stable than the cyclic dimer regardless the presence of methanol. Using two fully solvated monomers as the reference state, cyclic dimers are more stable in the methanol-water mixture than in water. All these suggest that the selective crystallization of α and β glycine from aqueous and alcoholic solutions respectively is not due to the abundance of glycine aggregates in

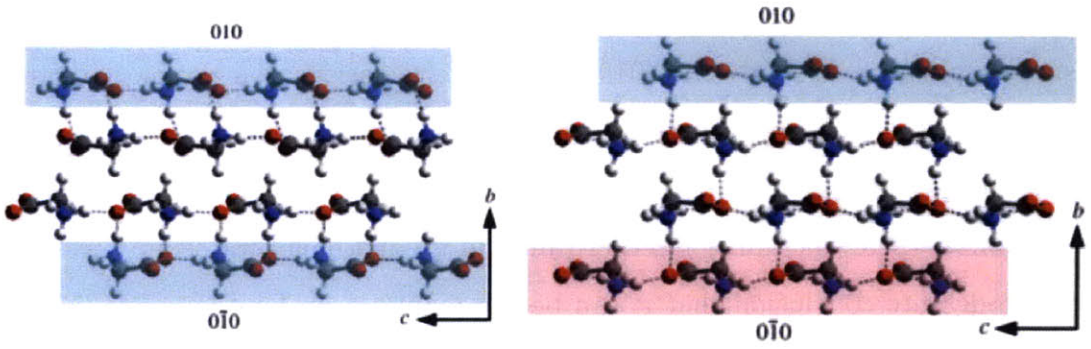
the solution phase with a similar structure to the crystallizing solid form before nucleation happens. The presence of α glycine (010) face, which consists of a double layer of cyclic dimers, induces an increase in the cyclic dimer composition of the glycine clusters in its vicinity. However, the increase, which might accelerate the growth of α glycine, does not significantly offset the cyclic dimer fraction values in the methanol-water mixtures and does not lead to a definite conclusion that it can be the main reason for the crystallization of α glycine.

3.5.8 Other Solvent Mediated Polymorph Selection Mechanisms

3.5.8.1 Methanol Inhibition on the (0±10) Faces of α and β Glycine

α glycine has two crystal surfaces, the symmetric (010) and (0-10) faces, that have exposed C-H bonds. β glycine has the asymmetric (010) and (0-10) faces with C-H bonds exposed on the former and N-H bonds exposed on the latter, as shown in Figure 3-23. Based on this difference in crystal structures, Weissbuch *et al.*[72] suggested that methanol molecules would bind strongly onto both (010) and (0-10) faces of α glycine through the interactions between C-H bond of glycine and the alkyl chain of the alcohol. It would strongly bind onto only the face (010) of β glycine, with C-H bonds exposed, but not the face (0-10) with the N-H bonds exposed. Therefore, they proposed that methanol inhibits the growth of both (010) and (0-10) faces of α glycine but only the (010) face of the β polymorph, but there has been no direct evidence of this hypothesis in literature so far. In this work, we applied molecular dynamics to study the solvent-crystal interactions on both (010) and (0-10) faces of α and β glycine using both pure water and methanol-water 3:7 mixtures with the aim of gaining some molecular level insights into this postulation.

Each simulation cell consists of a solvent-crystal-solvent sandwich with solvent molecules contacting the (010) and (0-10) faces, as shown in Figure 3-23. The α and β glycine crystals placed in the center both consist of 768 molecules from $8a \times 3b \times 8c$ and $8a \times 6b \times 8c$ super cells respectively. The thickness of the crystal in the b direction is about 37 $\mathring{A}$, which is more than twice the cutoff for nonbonded interactions (14 $\mathring{A}$) to ensure the solvent molecules near



(a) (b)

Figure 3-22: The (010) face and (0-10) face of α and β glycine from Weissbuchs previous work.[72] (a) α form exposing C-H bonds to the solution on both faces. (b) β form exposing C-H bonds to the solution on (010) face and N-H bonds to the solution on (0-10) face.

face (010) do not interact with the solvent molecules near face (0-10). The total number of solvent molecules in the system is 2439 for the water simulation and 2000 for methanol-water simulation (30% methanol). The detailed information for all simulation boxes are listed in Table 3.12. Solvent molecules are evenly distributed on both sides of the crystal in the initial configuration. A 2n MD simulation in a NPT ensemble (298 K and 1 atm) with periodic boundary conditions and the particle mesh Ewald summation method was performed to equilibrate the system. The center of mass of the crystal was placed at the origin and images were updated for all molecules around the origin before the simulation and only for solvent molecules during the simulations. To prevent the glycine molecules on crystal surface from diffusing into the solution phase, a harmonic restraint with force constant 8.36 kJ/(mol·$\AA^2$) was again applied to the root mean squared difference of those none hydrogen atoms of the crystal to a reference structure, which was calculated by pre-equilibrating the crystal super cell at 298 K and 1 atm and extracting the average structure. The height of the solvent layer on both sides is at least 20 $\AA$ for all simulations, which is again larger than the 14 $\AA$ cutoff used for nonbonded interactions. After equilibrating the system for 2 ns, another 2 ns production run was harvested to calculate the solvent-crystal interactions, with samples taken every 1 ps. Crystal-solvent interactions were calculated as the pairwise summation of

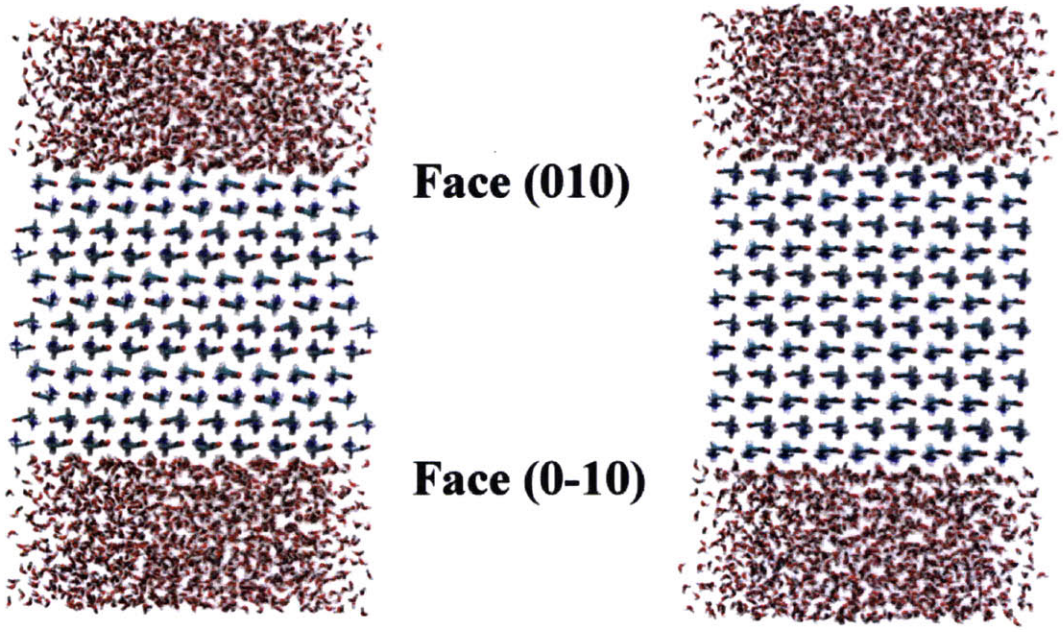


Figure 3-23: Simulation boxes of the (010) and (0-10) faces of α (left) and β glycine (right) contacting with water.

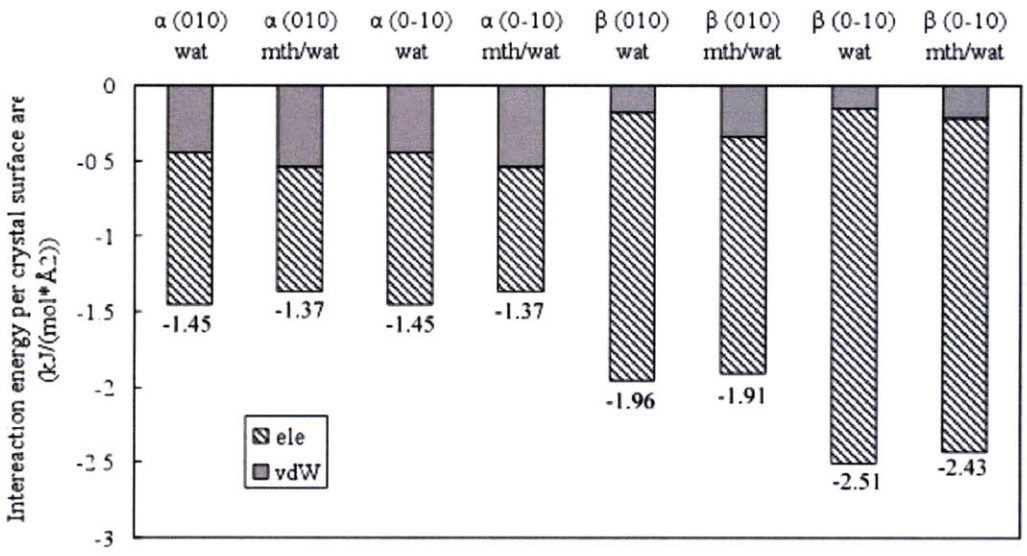
the electrostatic and van der Waals interactions between the crystal and solvent molecules which are less than 14 $\AA$ away (the cutoff for nonbonded interactions), since any solvent molecules farther away than that do not contribute to the interaction energies.

The interaction energies normalized based on both the total surface area and the number of glycine molecules in the first crystal surface layer are plotted in Figure 3-24(a) and (b). The contribution from van der Waals interactions and electrostatic interactions are marked out respectively. It is clear that both (010) and (0-10) faces of β glycine interact with solvents (water and methanol-water 3:7 mixture) more strongly than those of the α form and all crystal surfaces interact more strongly with water than with methanol-water 3:7 mixture. The electrostatic interaction is the main contribution in all scenarios. With the presence of methanol, the van der Waals contribution increases slightly, but it does not make up for the loss in the electrostatic interactions. A close look at the β glycine face (010) contacting with the methanol-water solution (Figure 3-25) shows that there are pockets on the surface which

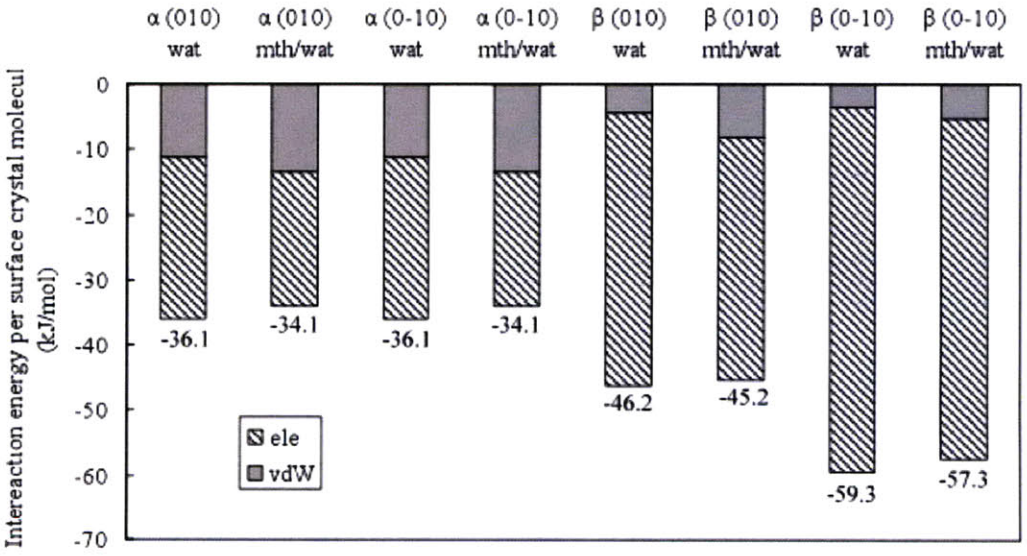
	no. of glycine	thickness. of glycine crystal in b-axis ($\AA$)	no. of water	no. of methanol	thickness of solvent on each size ($\AA$)
α wat	768	37.34	2439	0	19.96
α wat/moh	768	37.34	1400	600	22.51
β wat	768	39.85	2439	0	20.41
β wat/moh	768	39.85	1400	600	23.00

Table 3.12: Details of the simulation boxes of the (010) and (0-10) faces of α and β glycine contacting with water or methanol/water mixture.

are preliminarily occupied by water molecules either due to their smaller size or stronger hydrogen bonding ability, highlighted in those green circles. Methanol molecules near the glycine crystal surface usually have their hydroxyl group pointing toward the surface to form hydrogen bonds, highlighted in the orange circle, instead of forming van der Waals interactions with the exposed C-H bonds. We also see these contacts on the interface between the α glycine (0$\pm$10) faces and the methanol-water solution, which is not shown here for the simplicity. Opposite to the mechanism proposed by Weissbuch, in our simulation study there is no strong binding of methanol molecules onto the (010) and (0-10) faces of α glycine and the (010) face of β glycine (which all have the C-H bonds exposed) and the addition of methanol dilutes the crystal-solvent interactions on all interfaces. Weissbuch also used the fact that the addition of small amount of alcohol gives α crystals with larger (010) and (0-10) faces to support the methanol inhibition on these two faces which leads to their slower growth rates. It is worth to point out that the size of a face in the crystal morphology is decided by the relative growth rates of all faces. It is insufficient to draw conclusions about the methanol inhibition on the b-direction of glycine crystals based on the equilibrium crystal morphology, since the increase in the size of the (010) and (0-10) faces could also be the result of larger increases in the growth rates of other faces after the addition of methanol. Our simulation results at least show that methanol is energetically less favored than water on the (0$\pm$10) faces of both forms.



(a)



(b)

Figure 3-24: Solvent-crystal interaction energies on face (010) and face (0-10) of both α and β glycine for water and methanol-water mixtures calculated using MD simulation at 298 K and 1 atm. (a) normalized based on surface area, 1594.05 $\AA^2$ for α glycine and 1511.84 $\AA^2$ for β glycine (b) normalized based on the number of glycine molecules on crystal surface, 64 for both α and β glycine. The numbers labeled on the plots are the normalized total interaction energies.

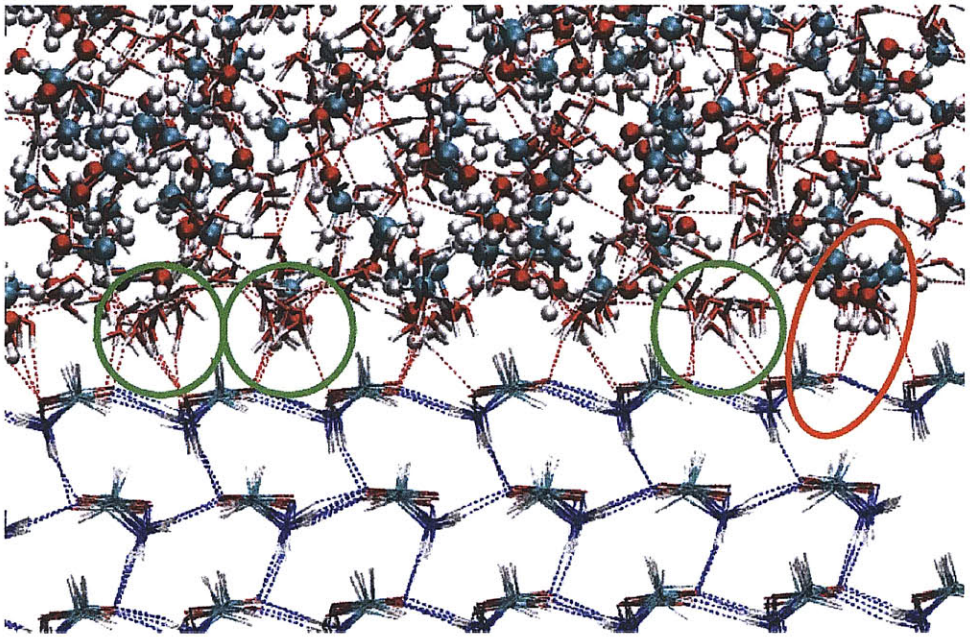


Figure 3-25: A snapshot of the β glycine (010) face interacting with methanol-water mixture, viewed along the c axis of β glycine. Glycine and water molecules are in line representation, and methanol molecules are in ball and stick representation. Hydrogen bonds are in dashed lines. The green circles highlight the water molecules in a pocket on the (010) face. The orange lines highlights the hydrogen bonds formed between methanol molecules and glycine.

3.6 Conclusions

Molecular dynamics is a very useful tool to study the self-assembly process in solution, that has been able to provide us structural information of those solute clusters formed in solutions at atomic level. In the screening of the solvent for potential polymorphs in pharmaceutical crystallization, a wide range of solvents with different chemical and physical properties, like polarity, hydrogen bond donor ability, hydrogen bond acceptor ability, etc., should be investigated to find potential polymorphs. Such an investigation could be performed in *silico*, yielding molecular level insights as a complement to experimental results.

Although the link hypothesis succeeded in explaining the polymorphic outcome of tetrolic acid from different solvents, it failed for the glycine system. Extra caution should be made when drawing conclusions about the potential polymorphic outcome solely from solution studies. To fully understand the selective crystallization of different crystal forms of glycine from pure water, water with alcohol, and water with acid (which gives the γ glycine), we need to study the nucleation process directly using more advanced computational techniques.

Chapter 4

Computer-Aided Rational Selection of Solvents for Improving the Morphology of Needle-like Crystals

4.1 Introduction

Crystals, like ice,[73] can grow into different shapes or morphology, depending on their molecular structures, the packing arrangement in the solid phase, symmetry, as well as the crystallization conditions. Organic molecules used in the chemical, pharmaceutical and food industries commonly exhibit anisotropic structural properties in their crystalline form, which gives rise to needle-like crystals. This type of morphology is usually undesirable in pharmaceutical crystallization because of significant difficulties in downstream particle handling, such as filtration, drying, milling and compaction. Solvents, as demonstrated in many previous studies,[7,8] can affect the morphology of crystals obtained from solutions. However, the selection of the right solvent(s) to minimize the appearance or to reduce the aspect ratio of needle-like crystals is almost always made via a trial and error approach and is often unsuccessful. Recently, computer-aided design has been widely used in various areas, such as drug discovery,[74] protein engineering[75] and process development.[76] In this work, we ap-

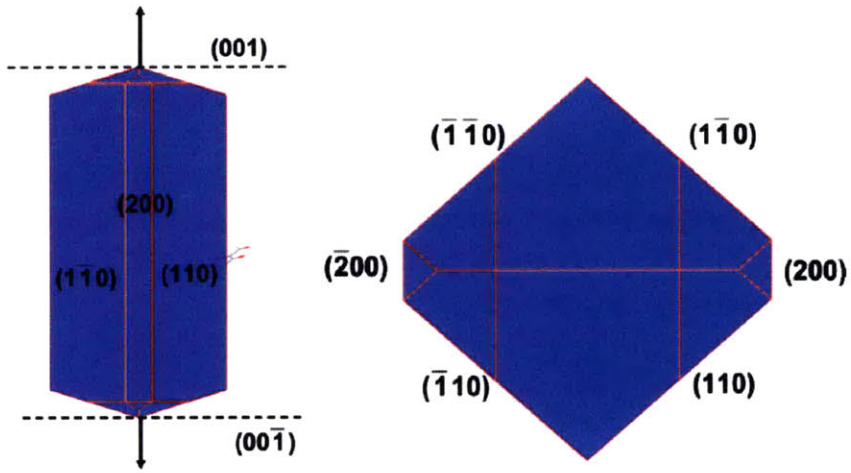
plied this concept to the field of crystal shape engineering and developed a computer-aided solvent selection procedure for improving the morphology of needle-like crystals. The procedure we developed in this work is based on molecular simulations and was is able to provide mechanistic understanding and therein rationale behind the solvent selection.

4.2 Solvent Effect and Model System

The mechanism of morphology modification using different solvents falls into two categories: (1) solvent changing the polymorphic form of the crystal leading to morphology changes and (2) solvent changing the relative growth rates of different crystal faces while preserving the polymorphic form. It is extremely important in the pharmaceutical industry to produce drugs using APIs with consistent and desired polymorphic form. Therefore, one of our most important solvent selection criteria is to preserve the desired polymorphic form of the crystal.

We used 2,6-dihydroxybenzoic acid (DHB) as a model system, which has two polymorphic forms. Form 1 (CCDC refcode LEZJAB01) is metastable,[77] has monoclinic structure and belongs to space group P21/c with a = 5.408 $\mathring{A}$, b = 5.224 $\mathring{A}$, c = 22.986 $\mathring{A}$ and beta = 94.69 ^{o}. Form 2 (CCDC refcode LEZJAB), the stable form,[78] is orthorhombic and belongs to the noncentric space group Pna21 with a = 14.174 $\mathring{A}$, b = 12.132 $\mathring{A}$ and c = 3.8280 $\mathring{A}$. In this work, our efforts have been directed to form 2 crystal only. The reasons are two-fold. Firstly form 2 is the polymorphic form which exhibits a needle-like morphology as shown in Davey's previous study.[33] Secondly it is the most stable form, the more desirable product form in the pharmaceutical industry. These two traits of DHB form 2 render it an ideal example to tackle the common problem in pharmaceutical crystallization: minimizing the needle morphology while preserving the polymorphic form. So far in the literature, no solvent systems have been reported that improve the needle morphology of DHB form 2. In Daveys previous study, needle shaped DHB form 2 crystals were obtained from both toluene and chloroform, with those from chloroform exhibiting the larger aspect ratio. The needle direction aligns with the c axis, bounded by two fast growing faces (001) and (00-1) called needle tips, and surrounded by six slow growing faces (200), (-200), (110), (1-10), (-110)

and (-1-10) called surrounding faces. It is schematically shown in Figure 4-1. Therefore, to improve the morphology and to reduce aspect ratio, a solvent or a solvent mixture which can greatly reduce the growth rates of the needle tips without over sacrificing the growth rates of the surrounding faces is needed.



(a) (b)

Figure 4-1: Morphology of DHB form 2 crystal, predicted using BFDH model from MERCURY and viewed (a) perpendicular to and (b) along the needle direction respectively.

4.3 Morphology Predictions

Many efforts have been dedicated to studying crystal growth and to predicting the equilibrium morphologies of crystals for decades. Wulff construction[79] is the basic theory underlying most morphology prediction models. It states that the total surface free energy of a crystal in equilibrium with its surroundings at constant temperature and pressure would be a minimum for a given volume. This eventually leads to the formula that the ratio of the growth rate of a particular crystal face and its perpendicular distance to the origin stays constant for all crystal faces, as shown in Figure 4-2. Once the relative growth rates of all crystal faces are known, the equilibrium morphology of the crystal can be constructed using Wulff theory. Therefore, the greatest challenge in morphology prediction models is to accurately

predict the relative growth rates of all crystal faces, especially those with low Miller indexes.

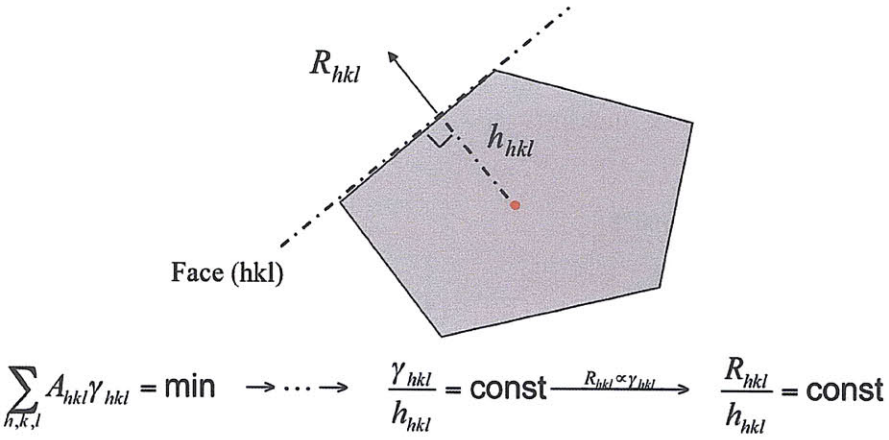


Figure 4-2: Demonstration of Wulff construction. A_{hkl}, γ_{hkl}, R_{hkl} are the area, the specific surface free energy, and the growth rate of face (hkl) respectively. h_{hkl} is the perpendicular distance between the origin and face (hkl).

In early studies, the primary focus in crystal morphology modeling was the relationship between the internal crystal structure and the external crystal shape. The earliest attempt was the Bravais-Friedel-Donnay-Harker (BFDH) model,[80,81] which relates the crystal habit to the lattice geometry and assumes an inversely proportional relationship between the growth rate and the interplanar distance (Figure 4-3):

$$R_{hkl} \propto \frac{1}{d_{hkl}} \tag{4.1}$$

In spite of its low accuracy, it is still useful for fast identification of the important faces of a grown crystal.

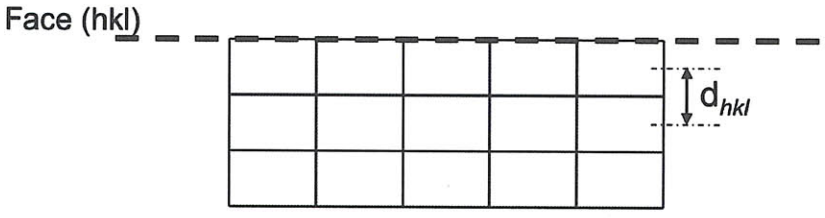


Figure 4-3: Schematic demonstration of the BFDH model.

The attachment energy (AE) model, developed by Hartman and Perdok,[82] started to capture the anisotropic structural information of crystal using quantifiable energies. This model assumes that the growth rate of a surface is proportional to its attachment energy in a vacuum (absolute value), defined as the energy released by adding a growth slice to the existing crystal surface:

$$R_{hkl} \propto \mid E_{hkl}^{vac,att} \mid \text{ and } E_{hkl}^{vac,att} = \frac{1}{2}\left(E_{hkl}^{latt} - E_{hkl}^{sl}\right) \tag{4.2}$$

where $E_{hkl}^{vac,att}$, E_{hkl}^{latt}, and E_{hkl}^{sl} are the attachment energy in a vacuum, lattice energy and slice energy respectively as shown by the top image in Figure 4-4. While overall the predictions have been accurate for sublimation growth crystals, they have not, in general, been accurate for crystals grown from solution, since AE fails to consider the influence of the crystallization solvents and additives. A modified version of this model, the modified attachment energy (MAE) model,[83,84] accounts for the impact of the solvent by reducing the vacuum attachment energy using solvent-crystal interaction energies. In the MAE model, the growth rate of crystal face (hkl) is proportional to the modified attachment energy, which is defined as

$$MAE_{hl} = E_{hkl}^{vac,latt} - \frac{Z_{hkl}^{solvent}}{Z} U_{hkl}^{solution} \tag{4.3}$$

where Z is the number of crystal molecules in a unit cell in the incoming crystal slice, $U_{hkl}^{solution}$ is the specific interaction energy between the crystal face (hkl) and the solvent molecules, and $Z_{hkl}^{solvent}$ is the number of solvent molecules in the volume defined by the unit cell in the incoming crystal slice. This is schematically shown in Figure 4-4. The MAE model was reported to be able to predict the aspirin morphology from aqueous ethanolic solutions better than the AE model.[84] However, this model inherits many of the shortfalls of the original attachment energy model. It is still a thermodynamic model and does not reflect any underlying growth kinetics.

Recently, more detailed kinetic based morphology prediction models have been reported in the literature.[85–87] They require a prior knowledge of the growth mechanism (2D nucle-

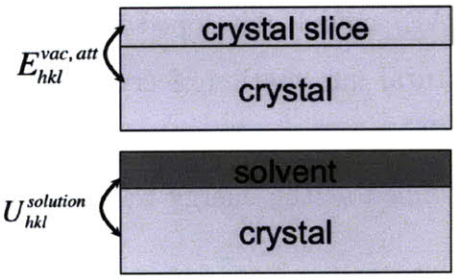


Figure 4-4: Schematic demonstration of the calculation of modified attachment energy.

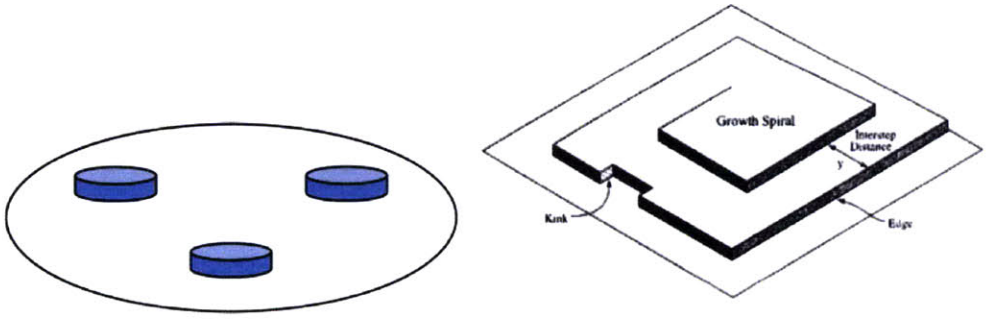
ation or spiral growth) and an empirical estimation of the interfacial free energy. A 2D nucleation mechanism is usually applicable to the crystal faces with medium growth rate, where nucleation on the surface is followed by growth, as shown in Figure 4-5a. This usually happens in the medium supersaturation region of the phase diagram. It adopts the form

$$R_{hkl} \propto \left[n \tan(\pi/n) \frac{(\phi_{hkl}^{edge})^2}{\Delta\mu} \right]^{-1} \tag{4.4}$$

where ϕ_{hkl}^{edge} is the surface energy of edges on face (hkl). The basic idea of the spiral growth mechanism is that dislocations in the crystals are the sources of new steps. A type of dislocation known as a screw dislocation could provide a way for the steps to grow continuously, shown in Figure 4-5b. This mechanism is usually followed by crystal faces with low growth rates, which usually falls in the low supersaturation region of the phase diagram. According to this theory, the crystal growth rate can be estimated as

$$R_{hkl} \propto \frac{d}{\kappa\phi_{hkl}^{edge}} \left[1 + 0.5 \exp(\phi_{hkl}^{kink}/RT) \right]^{-1} \tag{4.5}$$

where ϕ_{hkl}^{edge} and ϕ_{hkl}^{kink} are the surface energies of edges and kinks on face (hkl) respectively. Those models in general are more accurate and are able to capture the impact of solvents and additives. One of the most notable contributions is from Winn and Doherty.[86,88,89] However, those models are unsuitable for needle-like crystals, whose fast growing tips often exhibit roughness and do not follow either growth mechanism.



(a) (b)

Figure 4-5: Kinetic based growth models, (a) 2D nucleation mechanism, (b) spiral growth mechanism.

We employed the modified attachment energy model for the morphology predictions in this work. The reasons are two-folds. First, there are no suitable kinetic based models for the fast growing needle tips. Second, the MAE model is simple and computationally inexpensive, and therefore is more suitable for fast solvent screening. Furthermore, the goal of our morphology prediction is to get a quick estimation of the aspect ratios of the crystals from different solvents with reasonable accuracy, not accurate prediction of all the equilibrium faces and their relative sizes. The modified attachment model with a simple correction for solvent effect can serve the needs in this work, as we demonstrate herein.

Theoretically, the modified attachment energy should be calculated for all possible crystal faces. Practically the list of crystal faces can be narrowed down to those with low Miller index, which, in general, are the main contributors to the equilibrium morphology. In this work, since our goal is to select a solvent(s) to reduce the aspect ratio of the needle-like crystals instead of to predict accurately the size of all faces in the equilibrium morphology, we took advantage of the elongated shape property of needle-like crystals and only focused on the fast growing needle tips and the slow growing surrounding faces parallel to the needle direction to get a quick estimation of the aspect ratio and to speed up the solvent screening process. We started by using the BFDH model from the software MERCURY v2.2 to identify the needle tips and the surrounding faces. After that, MAEs of these faces were calculated using molecular simulations, which will be discussed in detail shortly. Finally the aspect

ratio of the crystal was calculated using the following scheme shown in Figure 4-6.

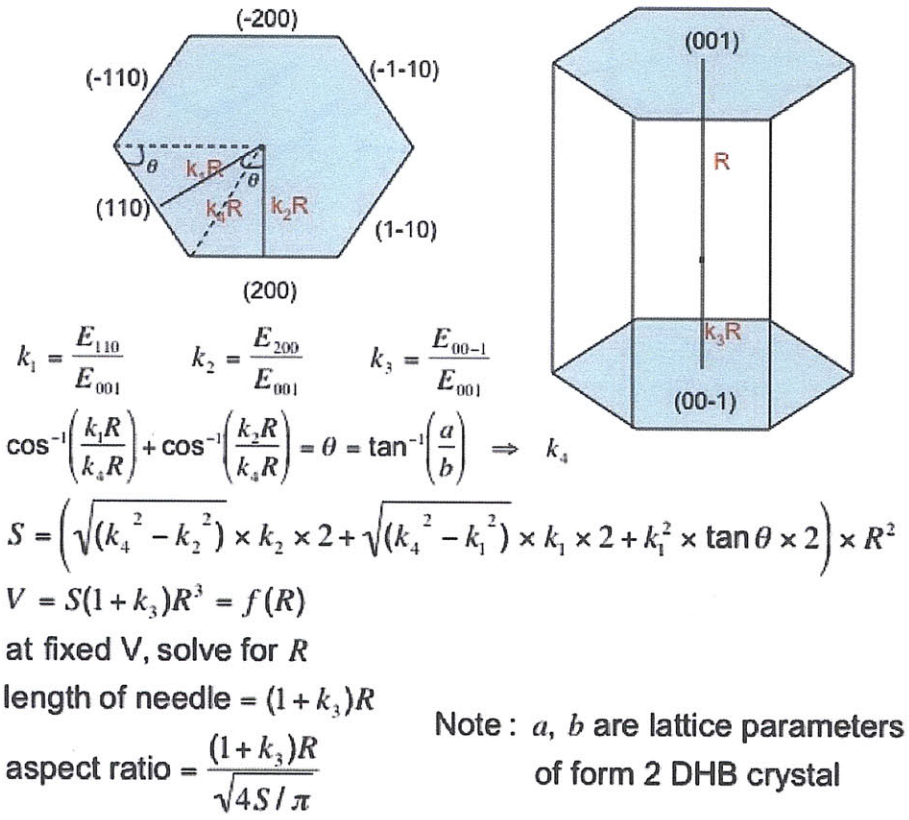


Figure 4-6: A simplified scheme to calculate the aspect ratio of needle-like crystals.

4.4 Force Field and Computational Details

The OPLS-AA force field[14] was used to model DHB and solvent molecules in this study, because it has a good library for organic solvents. The partial charges of DHB are the most sensitive parameters in simulating molecular crystals and were fit specifically for this study. The partial charges of DHB used in this work were derived from *ab initio* calculations using DFT theory with basis set 6-31+g(d,p) and CHELPG analysis. To validate this new charge set for DHB, we performed low temperature (20 K) molecular dynamics simulations for DHB crystal supercells (a $10a \times 10b \times 2c$ supercell for form 1 and a $4a \times 4b \times 12c$ supercell for form 2) in NPT ensembles. Low temperature MD is similar to potential energy minimization

	a	b	c	beta
Form 1	-1.90%	4.29%	-1.40%	0.08%
Form 2	-1.50%	3.39%	0.11%	0

Table 4.1: Percentage changes of lattice parameters of the equilibrium DHB crystal structures compared to the experimental X-ray structures, calculated using CHELPG charges derived from *ab initio* DFT calculations.

method but better in removing bad contacts. After equilibrating the system for 1 ns, the equilibrium crystal structure was calculated as the average of a 2 ns production run with a sampling interval of 1 ps. A comparison of the original OPLS-AA charges for DHB and the rederived charges using the *ab initio* method is shown in Figure 4-7. In general, both forms of DHB crystals have smaller percentage changes in lattice parameters and density when simulated using the rederived charges. The percentage changes in the lattice parameters of the equilibrium structure, compared to the experimental x-ray structures, are all within 5% (as shown in Table 4.1) when the rederived charges were used, which is the typical error in crystal simulations. We did not modify the OPLS-AA force field parameters of the solvent molecules. The ability of these parameters to reproduce the experimental density and heat of vaporization has been carefully tested when they were derived initially. Therefore, we did not perform any further validations. With that, we concluded that this force field was adequate for further simulations.

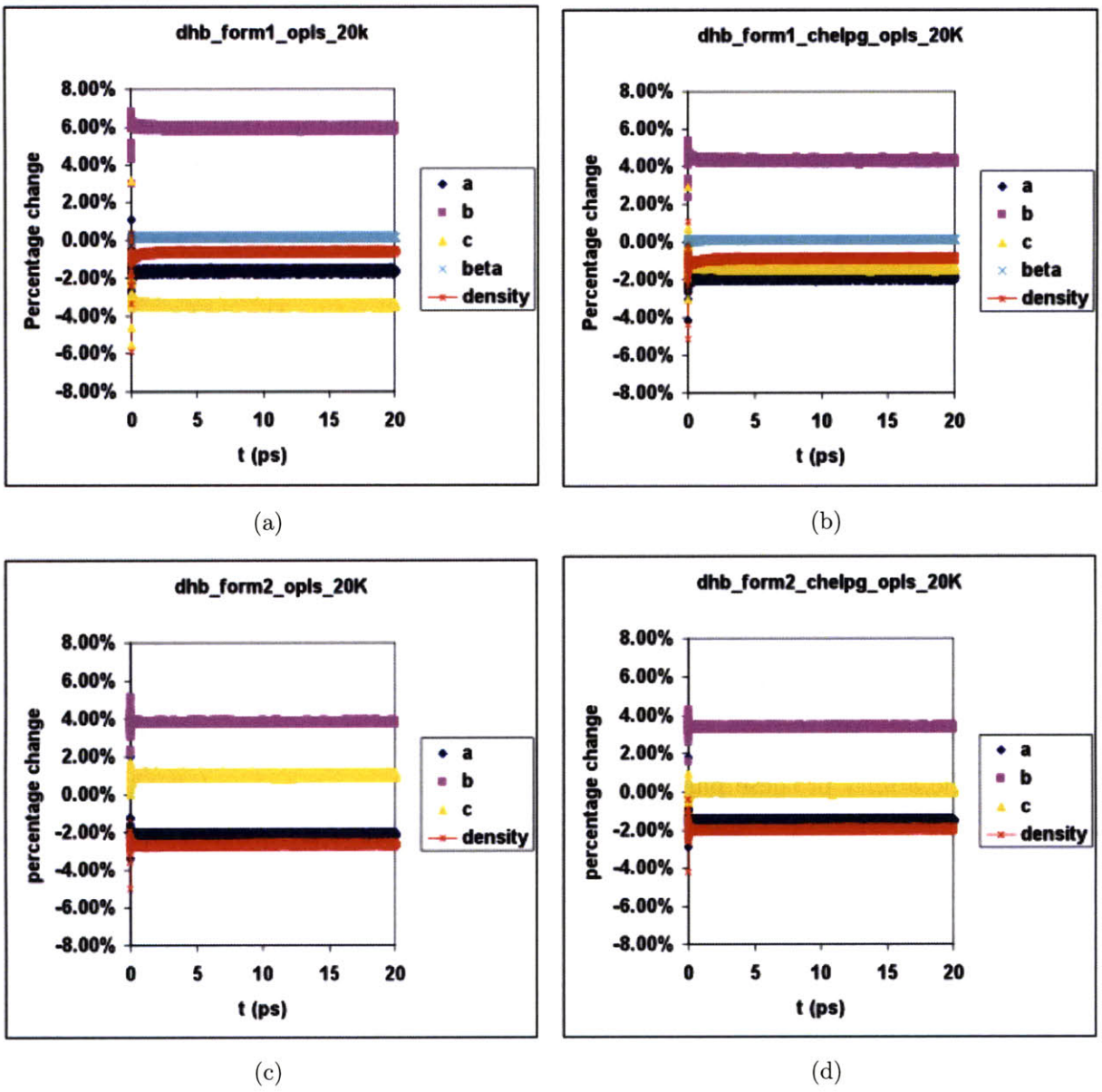


Figure 4-7: Percentage changes of the lattice parameters and the density of the DHB crystals in MD simulations at 20 K and 1 atm using different partial charges. (a) Form 1 crystal simulated using original OPLS-AA charges, (b) form 1 crystal simulated using rederived CHELPG charges, (c) form 2 crystal simulated using original OPLS-AA charges, and (d) form 2 crystal simulated using rederived CHELPG charges. All bonded parameters and LJ parameters are taken from OPLS-AA force field.

	supercell size in MD	total no. of molecules	size of bulk crystal	size of crystal slice
face (001)	$4a \times 3b \times 8c$	384	$4a \times 3b \times 7c$	$4a \times 3b \times 1c$
face (200)	$3a \times 4b \times 12c$	576	$2.5a \times 4b \times 12c$	$0.5a \times 4b \times 12c$
face (110)	$4a \times 3d \times 12c$ [a]	576	$4a \times 3d \times 11c$	$4a \times 3d \times 1c$

Table 4.2: System details of vacuum attachment energy calculation.
[a] d-axis is pointing to the (-110) direction and is of unit length $\sqrt{a^2 + b^2}$.

To calculate the attachment energy, a supercell of DHB form 2 was built and relaxed using molecular dynamics at 298 K and 1 atm with periodic boundary conditions. The particle mesh Ewald summation method was used to correct for the electrostatic interactions. After equilibrating the system for 2 ns, the equilibrium crystal structure was calculated as the average of 4 ns sampling. Then, the attachment energies $E_{hkl}^{vac,att}$ for faces (001), (200) and (110) were calculated using the equilibrium crystal structure. $E_{hkl}^{vac,att}$ for faces (00-1), (-200), (1-10), (-110) and (-1-10) were obtained using crystal symmetry. Here, we will take $E_{001}^{vac,att}$ as an example to explain the details. First, we defined a crystal layer parallel to the (001) face and of a height of one unit length of cell parameter c as the crystal slice and the rest crystal as the crystal bulk (shown in Figure 4-8). $E_{hkl}^{vac,att}$ was calculated by a pairwise summation of the interaction energies between the DHB molecules in the crystal slice and the DHB molecules in the crystal bulk and normalized by the number of molecules in the crystal slice. Since the cutoff distance for non-bonded interactions was set as 14.0 $\mathring{A}$, the crystal bulk should have a height of at least 14.0 $\mathring{A}$. The details of all MD boxes used in the attachment energy calculation are listed in Table 4.2.

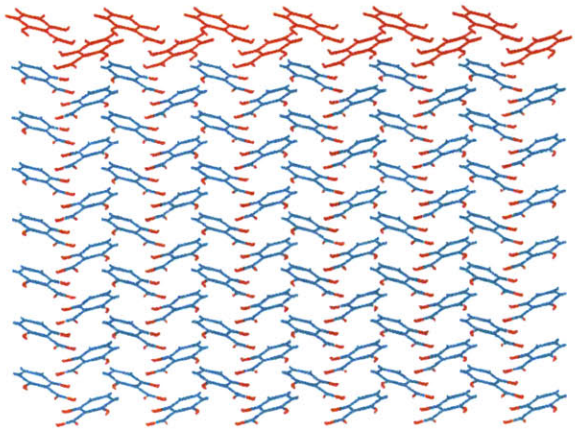


Figure 4-8: Demonstration of the calculation of vacuum attachment energy. The crystal face studied is face (001). Molecules in red representation are molecules in the crystal slice, and the rest are in the crystal bulk.

To calculate the interaction energies between the solvent molecules and the crystal surface, $U_{hkl}^{solution}$, a solvent-crystal-solvent sandwich was set up as the simulation box (shown in Figure 4-9), with the crystal face of interest contacting the solvent molecules. We performed molecular dynamics of this simulation box in the NPT ensemble (298 K and 1 atm) with 3D periodic boundary conditions. With this special setup, we modeled a 2D infinite crystal-solution interface. To prevent the surface crystal molecules diffusing into the solution phase, a harmonic constraint with force constant 8.36 kJ/(mol·$\AA^2$) was applied to the root mean squared difference of the actual crystal structure in the simulation. The equilibrium crystal structure obtained in the attachment energy study was used as the reference structure. An assumption underlying this is that the DHB molecules at the interface adopt the same structure as the bulk crystal. All solvent molecules were able to move freely in the simulation. After equilibrating the system for 2 ns, the interfacial configurations were sampled for 4 ns with 1 ps intervals. Solvent molecules within one crystal slice thickness above the crystal surface were selected as the solvent layer and used to calculate the solvent-crystal interaction energies. Finally, the interaction energies were normalized by the number of crystal molecules in the crystal slice, defined above in the attachment energy calculations. Details of all simulation boxes used in $U_{hkl}^{solution}$ calculations are listed in Appendix A . Our approach,

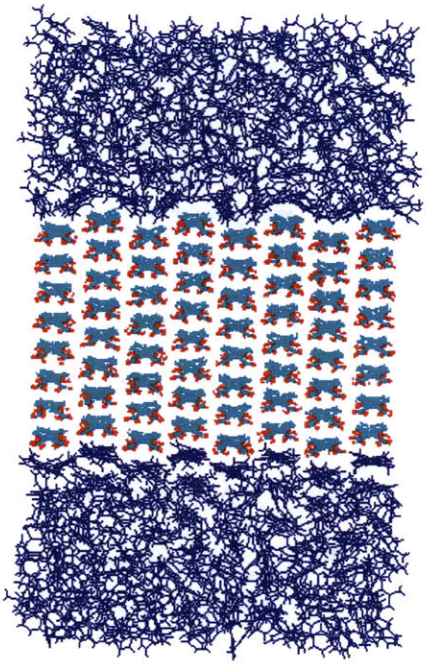


Figure 4-9: Demonstration of the calculation of crystal-solvent interaction energy. DHB face (001) and face (00-1) are shown here. Molecules in blue representation are toluene molecules, and the rest are DHB molecules.

different from the probe method in literature,[84] took into account the different configurations of the solvent molecules in the solvent layer and gave ensemble averaged MAEs. There are another two advantages of using molecular dynamics to calculate MAEs:

1. When dealing with solvent mixtures, the MD approach uses local concentrations near the crystal surface instead of the bulk values.

2. It can track the movement of each atom and can provide microscopic information about the interactions on the interface.

4.5 Results and Discussion

4.5.1 Single Solvents

We studied five solvents: toluene, chloroform, ethanol, diethyl ether and acetone. These solvents possess very different properties, such as polarity, aromaticity and hydrogen bond donor/acceptor ability. Toluene is a non-polar solvent with an aromatic ring, while chloroform is slightly more polar and is a weak hydrogen bond donor as noted in our polymorphism study of tetrolic acid. Ethanol is a common solvent in pharmaceutical crystallization, whose hydroxyl group can act as both a hydrogen bond donor and an acceptor. Diethyl ether and acetone are only hydrogen bond acceptors. It is worth pointing out that we started by using acetone as the representative of hydrogen bond acceptors but it was found to] react with DHB in our experimental validations later on. The reaction is shown in Appendix B. As a consequence, we switched to diethyl ether, which is similar to acetone as it is solely a hydrogen bond acceptor but it does not react with DHB.

Absolute values of the modified attachment energies (MAE) of the two needle tips and the six surrounding faces of DHB form 2 were calculated for five solvents: toluene, chloroform, ethanol, diethyl ether and acetone. Results are listed in Table 4.3. Note that the (200) and (-200) faces have the same MAE values due to the crystal symmetry, thus only face (200) is listed in the table. This is the same case for faces (110), (-1-10), (1-10) and (-110). The larger the MAE is, the faster the face grows. Therefore, an effective solvent to improve the morphology of DHB should have larger MAEs for the (001) and (00-1) faces and, at the same time, smaller MAEs for the (200) and (110) faces. The DHB crystals grown from chloroform are predicted to have larger aspect ratios than those grown from toluene, which agree with Daveys experimental observations. In all solvents studied, the needle tips have larger MAEs than the surrounding faces, which suggested that DHB has a very high propensity to crystallize with a needle-like morphology. Interestingly, although the (001) and (00-1) faces have the same attachment energy values, they have different MAEs. This implies they have asymmetric behavior when interacting with solvent molecules. In toluene,

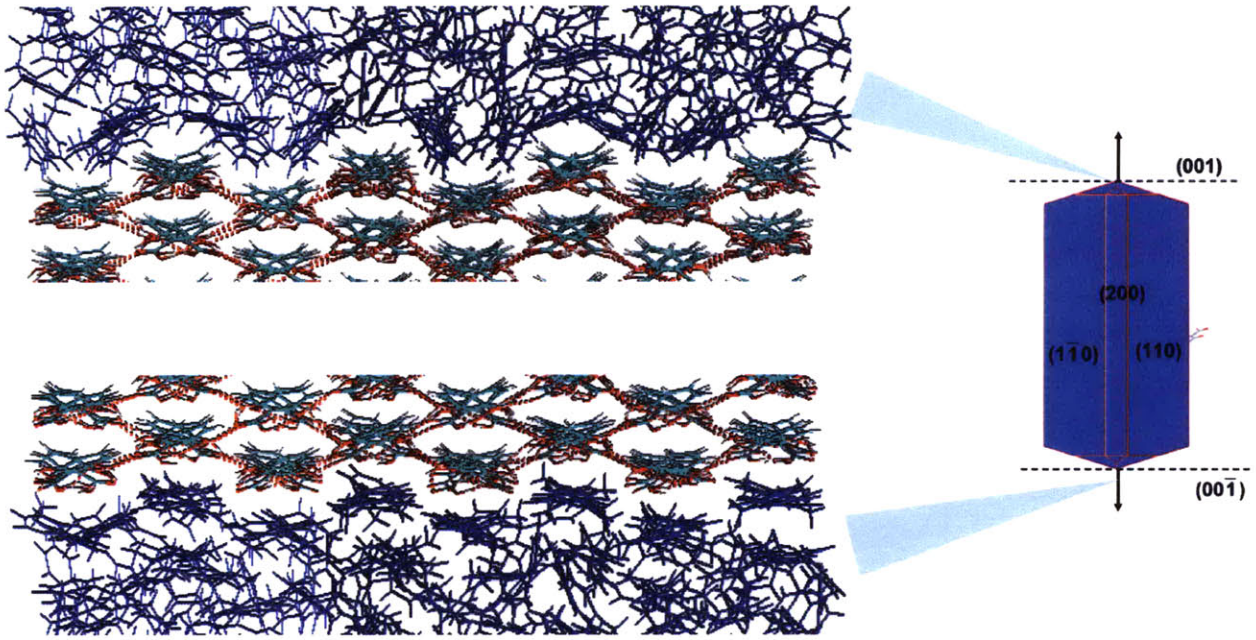
the (00-1) face has smaller MAE than the (001) face and therefore grows slower. A close look at the crystal-toluene interface (Figure 4-10) shows that toluene molecules stack onto the (00-1) face through π-π interactions of the aromatic rings, but are distributed more randomly near the face (001). However, it is the opposite for diethyl ether. Diethyl ether molecules, acting as hydrogen bond acceptors, form hydrogen bonds with the (001) face, but not with the (00-1) face. This leads to stronger solvent-crystal interactions on the former one and therefore inhibition of its growth. Ethanol gives the smallest sum of MAEs for the (001) and (00-1) faces, which means that it inhibits the growth of both needle tips to the largest extent. Ethanol also gives the smallest MAEs for the surrounding faces suggesting strong ethanol-crystal interactions formed on those faces as well. The net effect is that the predicted aspect ratio of the crystal from ethanol is the largest among all the four solvents. A snapshot of the solvent-crystal interfaces on the surrounding faces shows ethanol molecules form hydrogen bonds with the (110) face by acting as both donors and acceptors and form hydrogen bonds with the (200) face by acting mainly as donors (Figure 4-11). In other words, although the hydrogen bonding ability of ethanol effectively inhibits the growth of both needle tips, its donor ability over inhibits the growth of the surrounding faces. This is further confirmed by examining the diethyl ether-crystal interactions on the (200) and (110) faces. Diethyl ether, as an acceptor only, does not form hydrogen bonds with the (200) face and forms only half of the number of hydrogen bonds with the (110) face compared to ethanol. Therefore, even though it is less effective in inhibiting the growth of needle tips compared to ethanol, it does not over inhibit the growth of the surrounding faces and therefore still gives a smaller aspect ratio. The MAE values from acetone are still calculated using molecular dynamics, although acetone reacts with DHB and cannot be used as the crystallization solvent in practice. In addition to the awareness of the limitation of this procedure in selecting the right solvents, there is something else learned from the study of acetone. Although acetone is a hydrogen bond acceptor solely, it is a much stronger acceptor than diethyl ether. It over-inhibits the growth of the surrounding the (200) and (110) faces and thus gives a larger aspect ratio value than diethyl ether. Thus, insights gained from the single solvent studies are:

1. No single solvent studied here gives a better DHB morphology than toluene, which are still long thin needles.

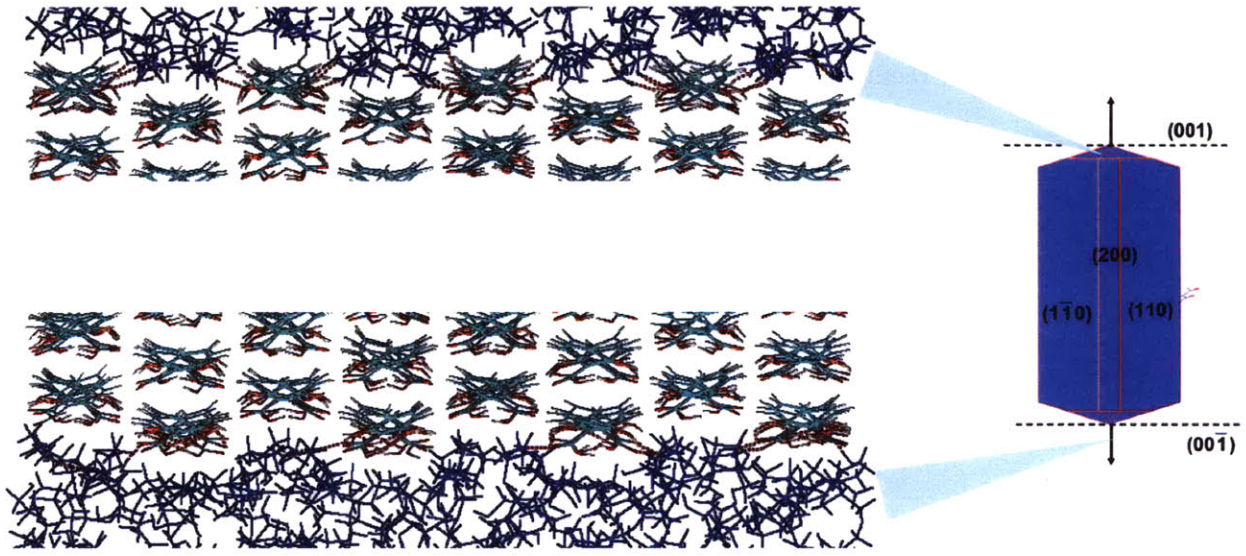
2. Hydroxyl groups inhibit the growth of all faces and does not serve the purpose of reducing the aspect ratio.

3. Aromatic rings effectively inhibit the growth of the (00-1) face and a weak hydrogen bond acceptor, like the ether oxygen, effectively inhibits the growth of the (001) face without over-inhibiting the growth of the surrounding faces.

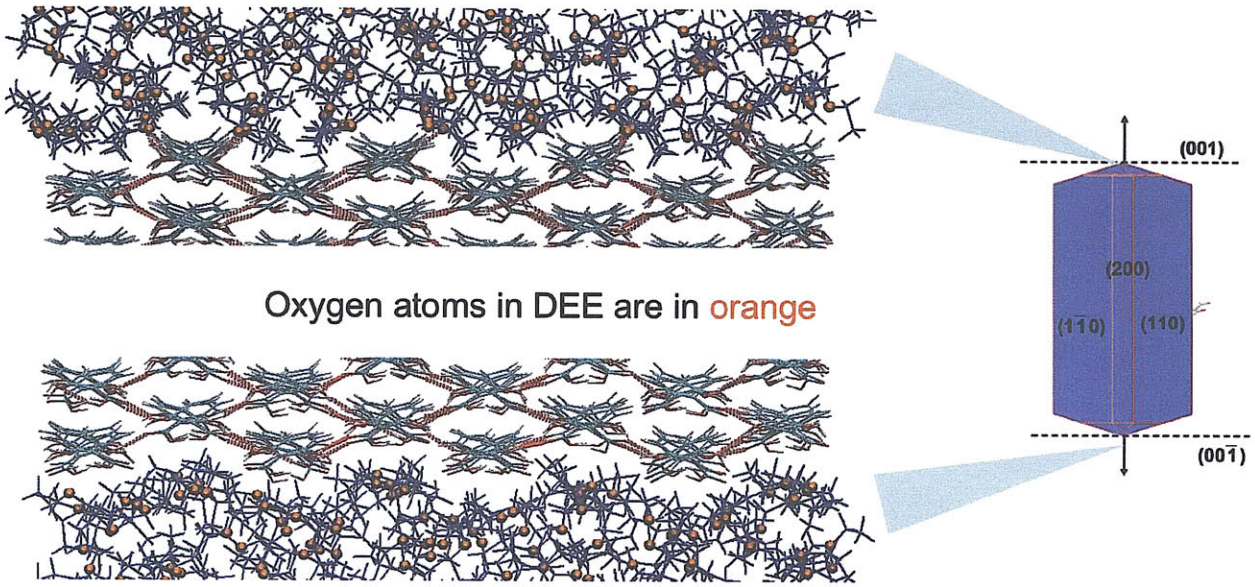
This naturally led to our next study on solvent mixtures.



(a) DHB form 2 face (00±1)-ethanol. Ethanol forms hbonds with both face (001) and face (00-1).

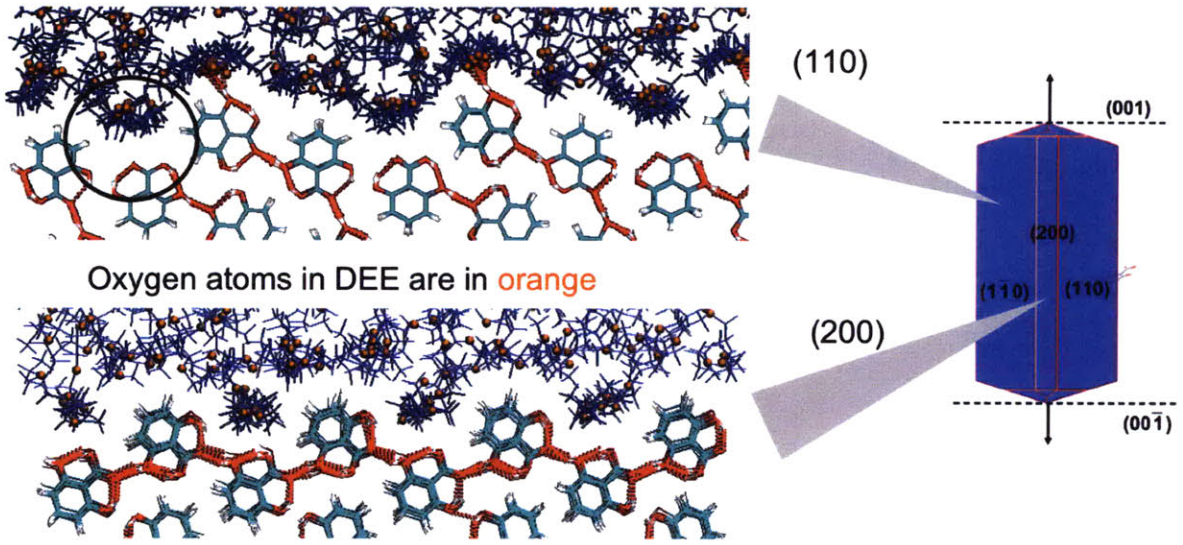


(b) Ethanol forms hydrogen bonds with both face (001) and face (00-1).

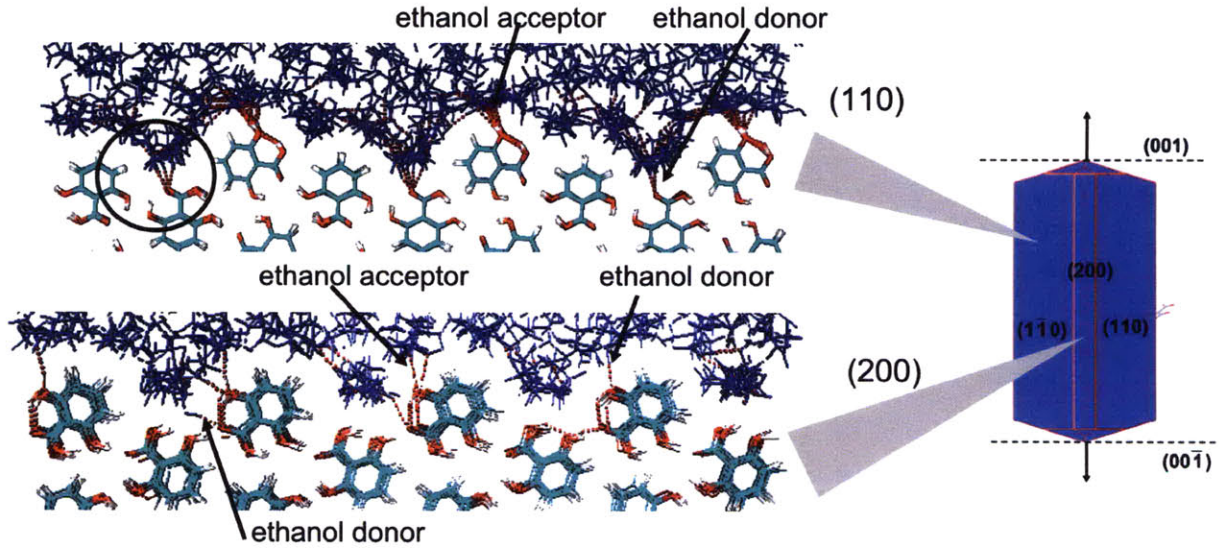


(c) Diethyl ether forms hydrogen bonds with face (001) only.

Figure 4-10: Snapshots of the crystal-solvent interfaces near DHB form 2 crystal face (00±1). (a)Toluene as the solvent, (b) ethanol as the solvent, and (c) diethyl ether as the solvent.



(a) Diethyl ether is only a hbond acceptor and forms fewer hydrogen bonds with surrounding faces.



(b) Ethanol is both a hydrogen bond donor and an acceptor and over inhibits the growth of surrounding faces.

Figure 4-11: Snapshots of the crystal-solvent interfaces near DHB form 2 crystal face (110) and face (200). (a) diethyl ether as the solvent, (b) ethanol as the solvent. The black circles highlight the hydrogen bonding difference between diethyl ether and ethanol.

	Toluene (TLN)	Chloroform	Ethanol	Diethylether (DEE)	Acetone
MAE(001)	-33.27 ± 0.11	-34.42 ± 0.09	-20.97 ± 0.11	-27.48 ± 0.08	-22.93 ± 0.09
MAE(00-1)	-26.84 ± 0.07	-29.52 ± 0.11	-27.61 ± 0.20	-34.57 ± 0.07	-36.74 ± 0.11
MAE(200)	-15.81 ± 0.06	-13.82 ± 0.10	-13.27 ± 0.15	-17.21 ± 0.05	-14.52 ± 0.05
MAE(110)	-20.84 ± 0.03	-18.81 ± 0.03	-11.76 ± 0.08	-18.59 ± 0.03	-15.65 ± 0.05
Aspect Ratio	1.34	1.59	1.75	1.47	1.68

Table 4.3: Modified attachment energies and predicted aspect ratios of DHB from single solvents.
Note: MAEs are in the unit of kJ/mol. MAE(200) = MAE(-200). MAE(110) = MAE(-1-10) = MAE(1-10) = MAE(-110).

4.5.2 Solvent Systems of the desired functional groups

In the single solvent studies, we identified the important functional groups for improving the aspect ratio of DHB form 2 crystals: the ether group (hydrogen bond acceptor) and the aromatic ring. Following that, we studied two types of solvent systems: (1) a mixture of toluene and diethyl ether and (2) a molecule with all the required functional group, e.g. anisole (as shown in Figure 4-12). We compared both the effectiveness and the economy of these two types of solvent systems. A series of toluene/diethyl ether mixtures with different mole ratios were also studied to find the optimal solvent composition.

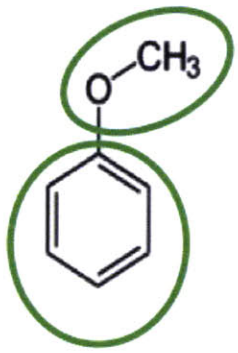


Figure 4-12: Structure of anisole. The circles highlight the required functional groups.

Modified attachment energies of the two needle tips and the six surrounding faces were again calculated using molecular simulation and the results are listed in Table 4.4. The predicted aspect ratio of DHB crystals from the toluene/diethyl ether mixture with the mole ratio 4:1 is the smallest among all the mixtures studied here and is smaller than those from either toluene or diethyl ether alone. The system 1 out performed the system 2. One possible reason is that the aromatic ring adjacent to the ether oxygen in anisole contributes additional steric hindrance when it forms hydrogen bonds with face (001), which weakens the strength of hydrogen bond formed and dampens the inhibition effect. This is supported by the slightly higher MAE (001) value for anisole than that of the TLN/DEE (4:1) mixture. Moreover, anisole is more expensive than both toluene and diethyl ether. Therefore, system 1 is both more effective and economically viable than system 2 in improving the morphology of DHB form 2 crystals. An interesting observation is that the optimal mixture requires a

diethyl ether mole faction of less than 0.5, as shown in Table 4.4. This is probably due to the fact that diethyl ether can still form hydrogen bonds with face (110) and it has potential to over inhibit the growth of face (110) if its concentration in the mixture is too high. One last thing we want to point out here is that the linear relationship between the growth rate and the MAE used in this morphology prediction model tends to underestimate the aspect ratio of needle-like crystal, which might have led to the very small differences in the predicted aspect ratios from different solvents, as listed in Table 4.4. For example, on the average, the calculated MAEs in this study are around -20kJ/mol. Doubling the MAE value for any particular face while keeping the MAEs of the other faces fixed would result in a one-fold increase in its growth rate and also a one-fold increase in its distance away from the origin on a strictly proportional basis. However, a 20 kJ/mol increase in the absolute value of the MAE would cause a ~3000 fold increase ($e^{\frac{20kJ/mol}{RT}}$ with T = 298 K) in its distance away from the origin on an exponential basis. The valuable information from the MAE, however, is the trend and the relative aspect ratio values. Our predictions suggest that the TLN/DEE (4:1) mixture is very promising in improving the morphology of the needle-like DHB form 2 crystal.

	TLN	DEE	DEE/TLN (1:9)	DEE/TLN (2:8)	DEE/TLN (3:7)	DEE/TLN (5:5)	anisole
MAE(001)	-33.27 ± 0.11	-27.48 ± 0.08	-31.26 ± 0.14	-29.65 ± 0.10	-29.05 ± 0.11	-28.21±0.13	-30.34 ± 0.09
MAE(00-1)	-26.84 ± 0.07	-34.57 ± 0.07	-26.90 ± 0.09	-27.25 ± 0.11	-28.15 ± 0.09	-28.72±0.10	-27.25 ± 0.10
MAE(200)	-15.81 ± 0.06	-17.21 ± 0.05	-15.94 ± 0.05	-15.95 ± 0.05	-16.05 ± 0.05	-16.33±0.06	-14.27 ± 0.05
MAE(110)	-20.84 ± 0.03	-18.59 ± 0.03	-19.98 ± 0.04	-19.48 ± 0.05	-19.25 ± 0.07	-18.88±0.05	-18.52 ± 0.04
Aspect Ratio	1.34	1.47	1.33	1.33	1.35	1.35	1.44

Table 4.4: Modified attachment energies and predicted aspect ratios of DHB from solvent systems with desired functional groups.
Note: MAEs are in the unit of kJ/mol.

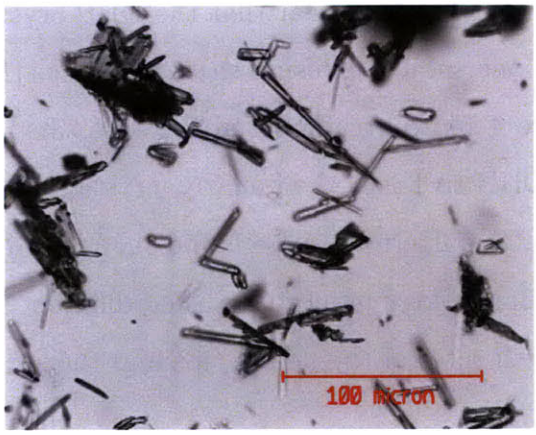
4.6 Experimental Validation

Slurry aging is a commonly used method for modifying crystal morphology, which allows crystals to grow into the equilibrium shape. For our experimental validation of the computer predictions DHB slurries were prepared using DHB form 2 crystals and various solvents including toluene, diethyl ether, ethanol, TLN/DEE (4:1) mixture and anisole. All the chemicals used in this work were purchased from Sigma-Aldrich. The commercial DHB form 2 crystal is of purity $\geq$ 98% and all the solvents are of purity $\geq$ 99.9%. DHB slurries with excess of solids were stirred at 30 ^oC for 3 hours using magnetic stirrer at 800 rpm and then kept static at room temperature (23 ^oC) for 7 days. Crystal samples before and after aging in the slurries were prepared for analysis by filtration of the slurry. The commercial DHB crystals were also used in the morphology study. All crystal samples were visualized using an OLYMPUS BXS1optical microscope with an Infinite 1 camera to study the morphology. A 50 fold magnification was used. All images were captured and analyzed in the Infinity Analyze software.

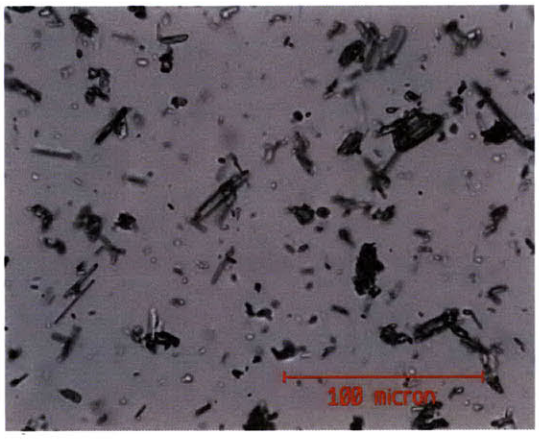
As shown in Figure 4-13a, the commercial DHB crystals exhibit a long thin needle-like morphology. Before aging in toluene, diethyl ether and the TLN/DEE (4:1) mixture (Figure 4-13b, d, f, h and j), the DHB crystals are either in long thin needles or fine particles as a result of breakage of needles during stirring. There is also significant amount of agglomerates formed by the fine particles. The morphology of DHB crystal does not change significantly after aging in toluene (Figure 4-13c), The solubilities of DHB form 2 in toluene, diethyl ether and TLN/DEE 4:1 mixture at 30 ^oC and 23 ^oC can be found in Appendix C. All DHB crystals grow into long needles after aging in diethyl ether, ethanol and anisole (Figure 4-13e, i and k). There is an obvious improvement in the morphology of DHB form 2 crystals obtained from the TLN/DEE 4:1 mixture (Figure 4-13g), the solvent mixture recommended by the computer predictions. The crystals show significantly reduced aspect ratios.

Power X-ray diffraction (PXRD) patterns of the DHB crystals aged in the slurries were measured using PANalytical X'pert Pro with X'pert data collector software to check the polymorphic form. PXRD patterns of DHB form 1, form 2 and hydrate were simulated using

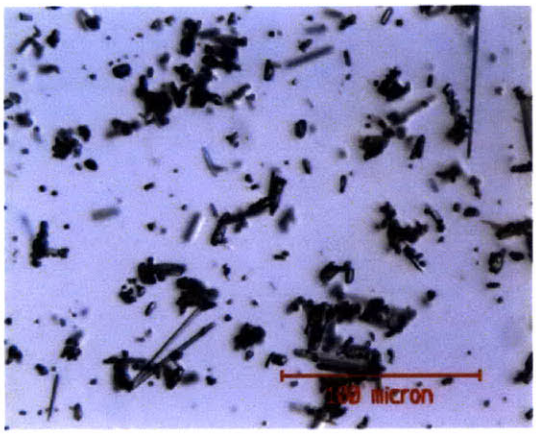
the software MERCURY v2.2 as well with the cif files obtained from the Cambridge Crystallographic Data Centre (CCDC). It is clear that the DHB crystals aged in the TLN/DEE (4:1) mixture matches the one simulated using the crystalline structure of the polymorphic form 2, as shown in Figure 4-14. Slurry aging also transformed the small amount of form 1 present in commercial DHB to form 2. Similarly, crystals harvested from other slurries are all form 2 (whose PXRD patterns can be found in Appendix D). This confirms that the change in the crystal morphology induced by using different solvents is the result of the change in the relative growth rates of the needle tips and the surrounding faces rather than the polymorphic form.



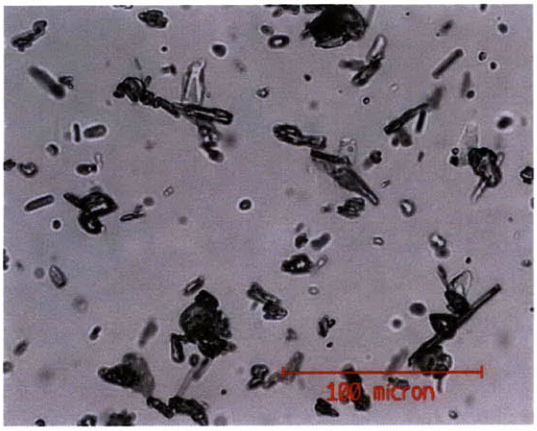
(a)



(b)

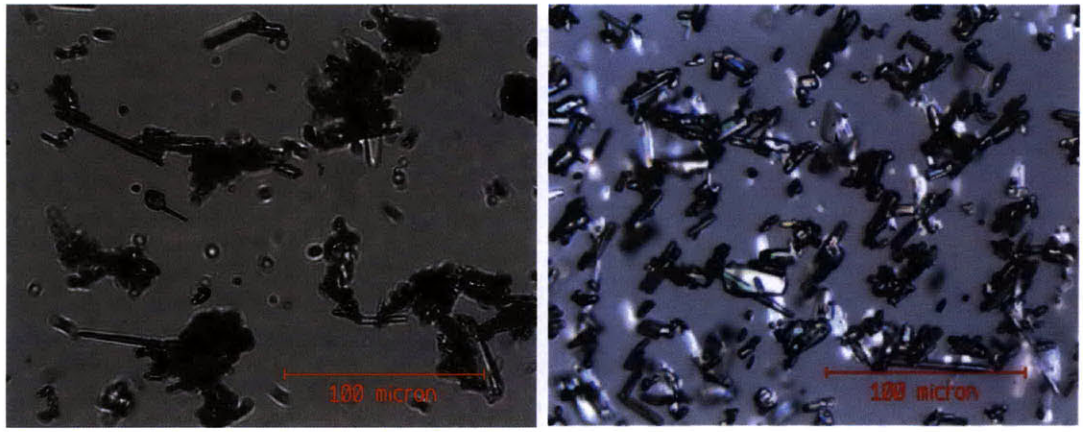


(c)

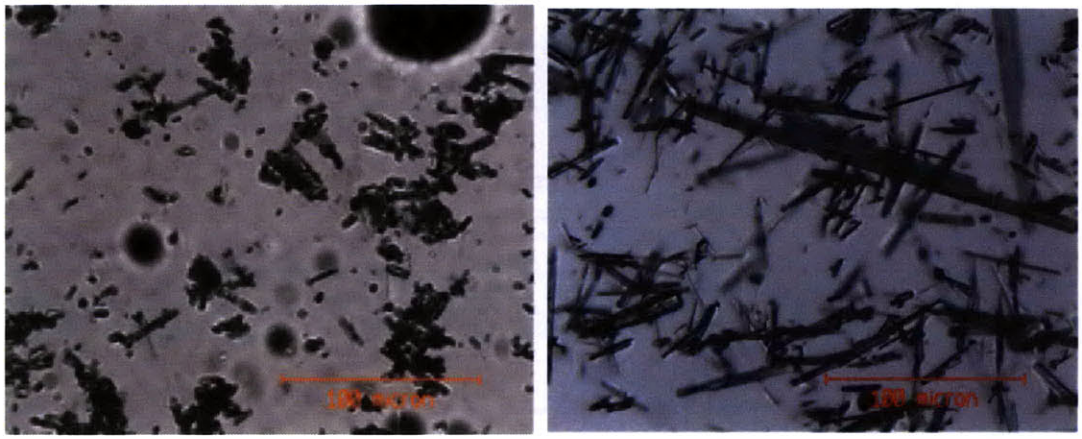


(d)

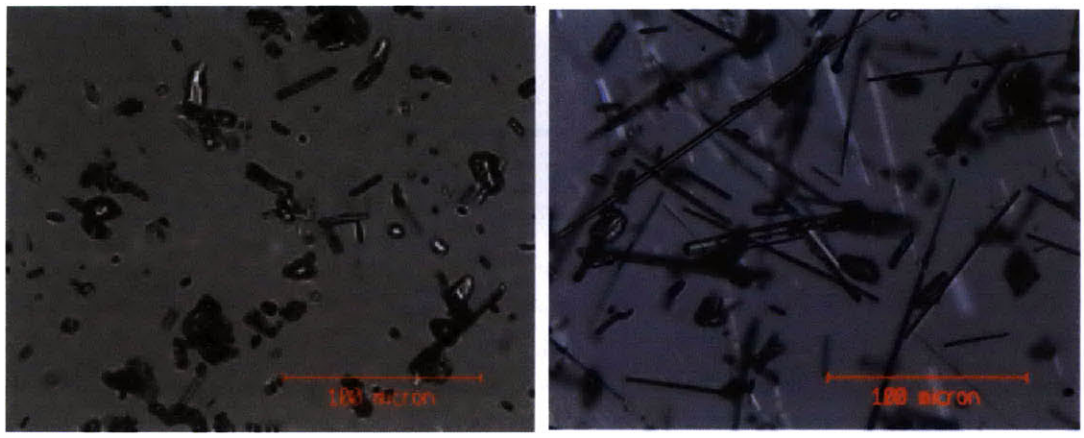
(e)



(f) (g)



(h) (i)



(j) (k)

Figure 4-13: DHB form 2 crystals aged in different solvents. (a) commercial, (b) before aging in toluene, (c) after aging in toluene, (d) before aging in diethyl ether, (e) after aging in diethyl ether, (f) before aging in TLN/DEE (4:1) mixture, (g) after aging in TLN/DEE (4:1) mixture, (h) before aging in ethanol, (i) after aging in ethanol, (j) before aging in anisole and (k) after aging in anisole.

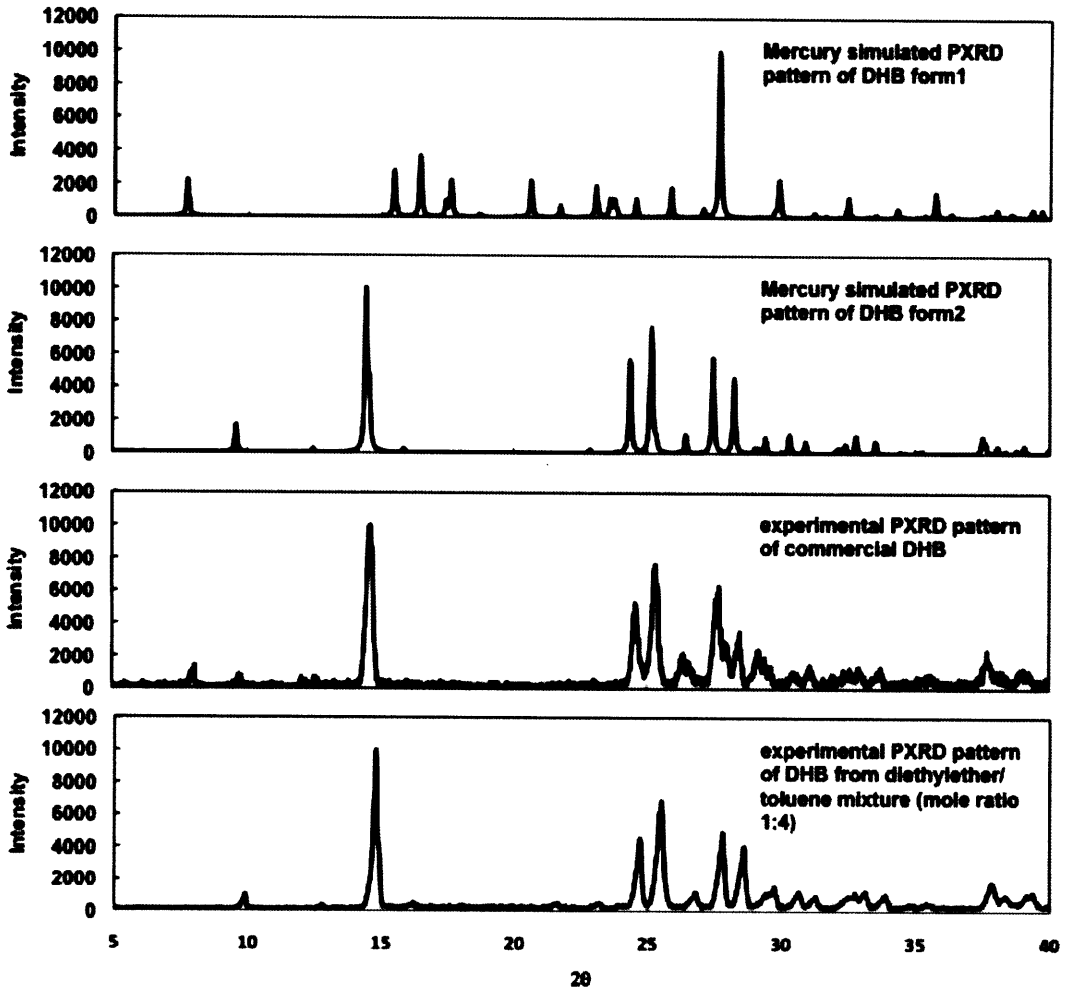


Figure 4-14: PXRD pattern of DHB crystals. From the top to the bottom: MERCURY simulated PXRD pattern for DHB form 1 crystal, MERCURY simulated PXRD pattern for DHB form 2 crystal, experimental PXRD pattern of commercial DHB form 2 crystal, and experimental PXRD pattern for DHB crystals aged in toluene/diethyl ether (4:1) mixture.

4.7 Conclusions

We successfully demonstrated the use of computational modeling to select rationally a solvent system that improves the aspect ratio of needle-like crystals. A diethyl ether/toluene (mole ratio 1:4) mixture, proposed rationally by our modeling approach, improved the morphology of the needle-like DHB form 2 crystals. The modified attachment energy and the simplification of focusing only on the needle tips and the surrounding faces were sufficiently accurate to provide a very useful tool for fast solvent screening. It allowed us to understand the interactions at the crystal-solvent interfaces at atomic level and select solvents based on those important functional groups identified using molecular simulations. We have thus developed a rational solvent selection procedure which goes far beyond the typical trial and error approach. This approach is generally applicable for the improvement of other needle-like crystals, as schematically depicted in Figure 4-15.

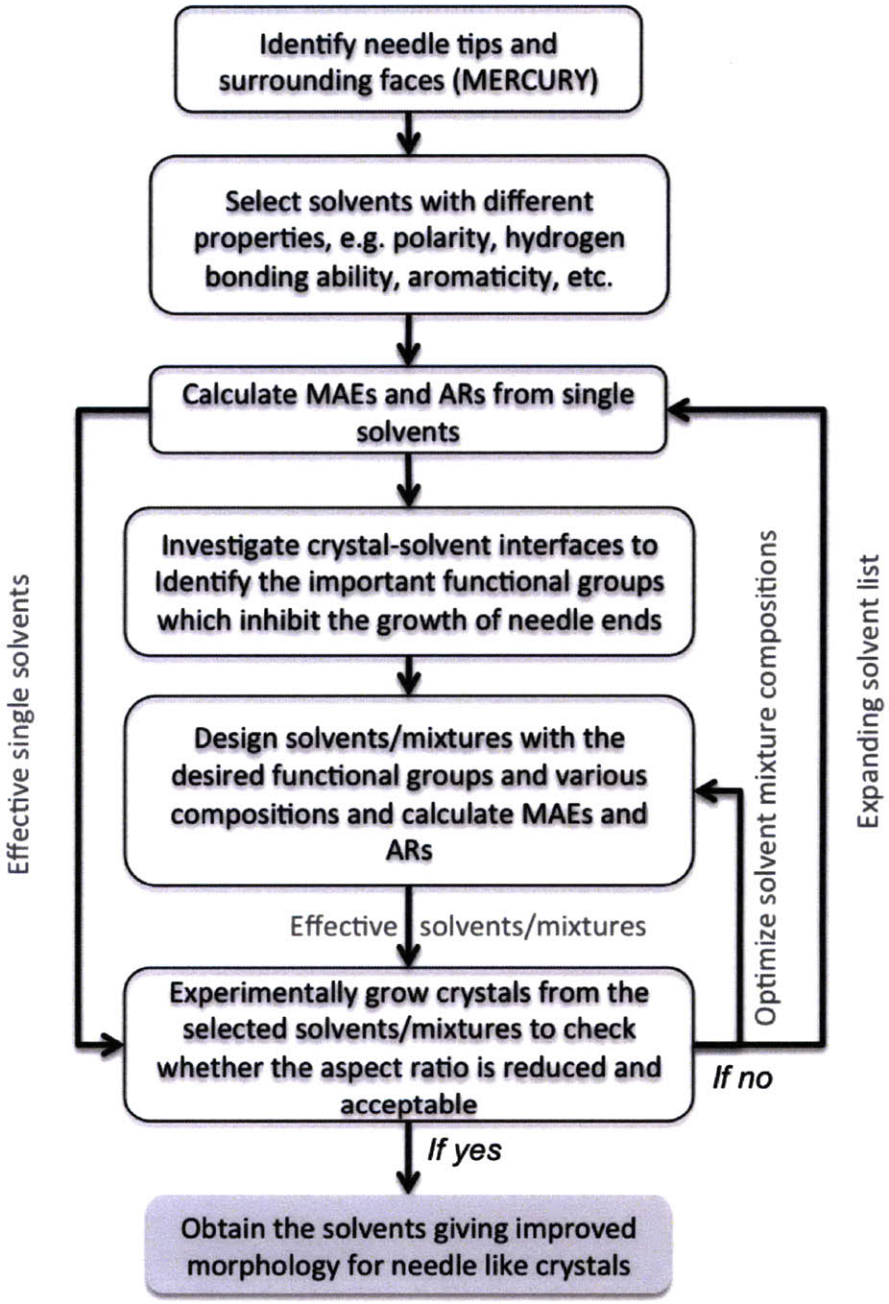


Figure 4-15: A flowchart of the computer-aided solvent selection procedure for improving the morphology of needle-like crystals.

Chapter 5

Conclusions and Future Work

In this chapter we present some of the salient conclusions of the thesis and possible future directions that the present work could be extended in.

5.1 Molecular Modeling in Understanding Crystal Polymorphism

Molecular modeling is a very useful tool in studying the self-assembly process in solution, which is able to provide structural information on those solute clusters formed in solutions at the atomic level. In the screening of solvent for potential polymorphs in pharmaceutical crystallization, a wide range of solvents with different chemical and physical properties, such as polarity, hydrogen bond donor ability, hydrogen bond acceptor ability, etc., should be investigated to find potential polymorphs. Such an investigation could be performed in *silico*, yielding molecular level insights as a complement to experimental results.

Although link hypothesis succeeded in explaining the polymorphic outcome of tetrolic acid crystallizations from different solvents, it failed in the case of glycine. Extra caution should be used when drawing conclusions about the potential polymorphic outcome solely from solution studies. The applicability of the link hypothesis would become clearer once it has been applied to study more polymorphic systems. Besides that, other parameters

linking the solution chemistry to the solid state polymorphic outcome should be explored. It is possible that the cyclic dimer fraction is not the right parameter in the case of glycine which might have been the cause in missing the link.

The mechanism of the selective crystallization of α and β-glycine from pure and alcoholic aqueous solutions is still not clear. To fully understand the selective crystallization of the different crystal forms of glycine from pure water, water with alcohol, and water with acid (which gives the γ glycine), direct study of the nucleation process is needed, which requires the development of more advanced computational techniques for efficient sampling of rare events.

5.2 Molecular Modeling in Understanding Crystal Morphology

We successfully demonstrated the use of computational modeling to rationally select a solvent system for improving the aspect ratio of needle-like crystals. A diethyl ether/toluene (mole ratio 1:4) mixture improved the morphology of the needle-like DHB form 2 crystal, which was predicted by molecular modeling and verified by bench-top experiments. The modified attachment energy and the simplification of focusing only on the needle tips and the surrounding faces, though low in accuracy, is a very useful tool for fast solvent screening. Moreover, we were able to understand the interactions at the crystal-solvent interfaces at atomic level and select solvents based on those important functional groups identified using molecular simulations. This made the solvent selection procedure developed in this work no longer a pure trial and error approach. We were able to introduce some rational design elements. Finally, this approach is also transferrable for the improvement of other needle-like crystals.

In this work, we only studied single solvents and solvent mixtures. It would be interesting to extend the solvent list to additives which are usually present in very small concentrations to check whether the model developed in this work can predict the effect of small amount

of additives. We also pointed out at the end of the solvent mixture studies that the linear relationship between the modified attachment energy and the growth rate tends to underestimate the aspect ratios of needle-like crystals. One future direction would be to modify this linear relationship to a functional form which can better predict the real aspect ratio values. Ultimately, it would be of great importance to understand the growth mechanism of needle tips and develop a kinetic-based morphology prediction model for needle-like crystals, which can capture the impact of both solvent/additive and supersaturation.

5.3 Molecular Modeling in Other Areas of Crystallization

There are many other areas of crystallization where molecular modeling could make a contribution, like predicting the solubility of a novel drug compound, selecting a surface for heterogeneous nucleation, designing cocrystals, etc. With the development of more accurate force fields to describe the interactions between molecules, more powerful super computers and more advanced computational methods, molecular modeling will be used to study larger and more complicated problems in crystallization.

Appendix A

Crystal-Solvent Interaction Energy Calculations for DHB

The Details of the simulation boxes used in the calculation of crystal-solvent interaction energy, $U_{hkl}^{solution}$, are listed in the Table A.1.

crystal face	crystal size	no. of crystal molecules	no. of solvent molecules								
			TLN	chloro-form	ethanol	DEE	D/T [a] (1:9)	D/T (1:4)	D/T (3:7)	D/T (5:5)	anisole
(00±1)	$4a \times 3b \times 8c$	384	672	930	1110	584	58/522	116/464	174/406	290/290	672
(±200)	$3a \times 4b \times 12c$	576	852	1214	1280	666	66/594	113/532	198/462	330/330	852
(±1±10)	$4a \times 3d \times 12c$ [b]	576	520	864	1104	518	51/459	103/412	153/357	255/255	520

Table A.1: System details of crystal-solvent interaction energy calculations for DHB.
[a] D/T: DEE/TLN
[b] d-axis is pointing to the (-110) direction and is of unit length $\sqrt{a^2 + b^2}$.

Appendix B

Reaction between 2,6-Dihydroxybenzoic Acid and Acetone

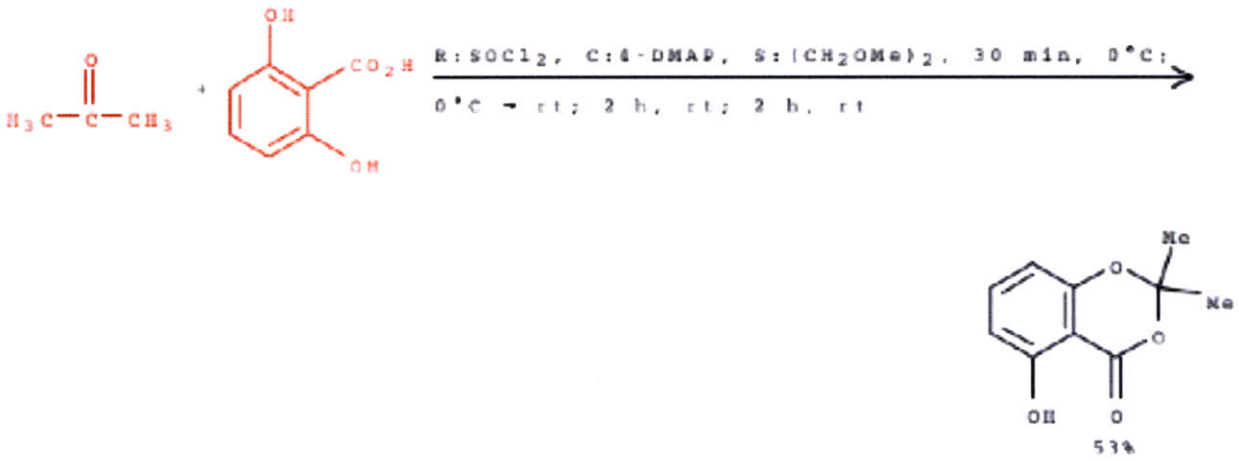


Figure B-1: Reaction between 2,6-dihydroxybenzoic acid and acetone.[90]
Note: Reactants: 2, Reagents: 1, Catalysts: 1, Solvents: 1, Steps: 1, Stages: 1

Appendix C

Solubility of DHB form 2

Solubilities of DHB form 2 in toluene, diethyl either and TLN/DEE 4:1 mixture at both 30 ^oC and 23 ^oC were measured using the evaporation method. DHB slurries with excess of solids were stirred at the temperature of interest for 3 hours using magnetic stirrer at 800 rpm and then kept static for 12 hours. Three samples of each saturated solution were taken from the top clear solution for evaporation. Results are listed in Table C.1

	toluene	diethyl ether	TLN/DEE (4:1)
23^oC	0.0006 ± 0.0001	0.0782 ± 0.0002	0.5296 ± 0.0026
30^oC	0.0010 ± 0.0001	0.0867 ± 0.0005	0.5869 ± 0.0005

Table C.1: Solubility of DHB form 2 in toluene, diethyl ether and TLN/DEE 4:1 mixture. Note: All solubilities are of the unit g DHB/g solvent. Molecular weight of toluene and diethyl ether are 92.14 and 74.12 g/mol respectively.

Appendix D

PXRD Patterns of DHB Crystals

PXRD patterns of DHB crystals aged in toluene, diethyl ether, anisole, chloroform, and ethanol are shown in Figure D-1. All crystals harvested are of polymorphic form 2.

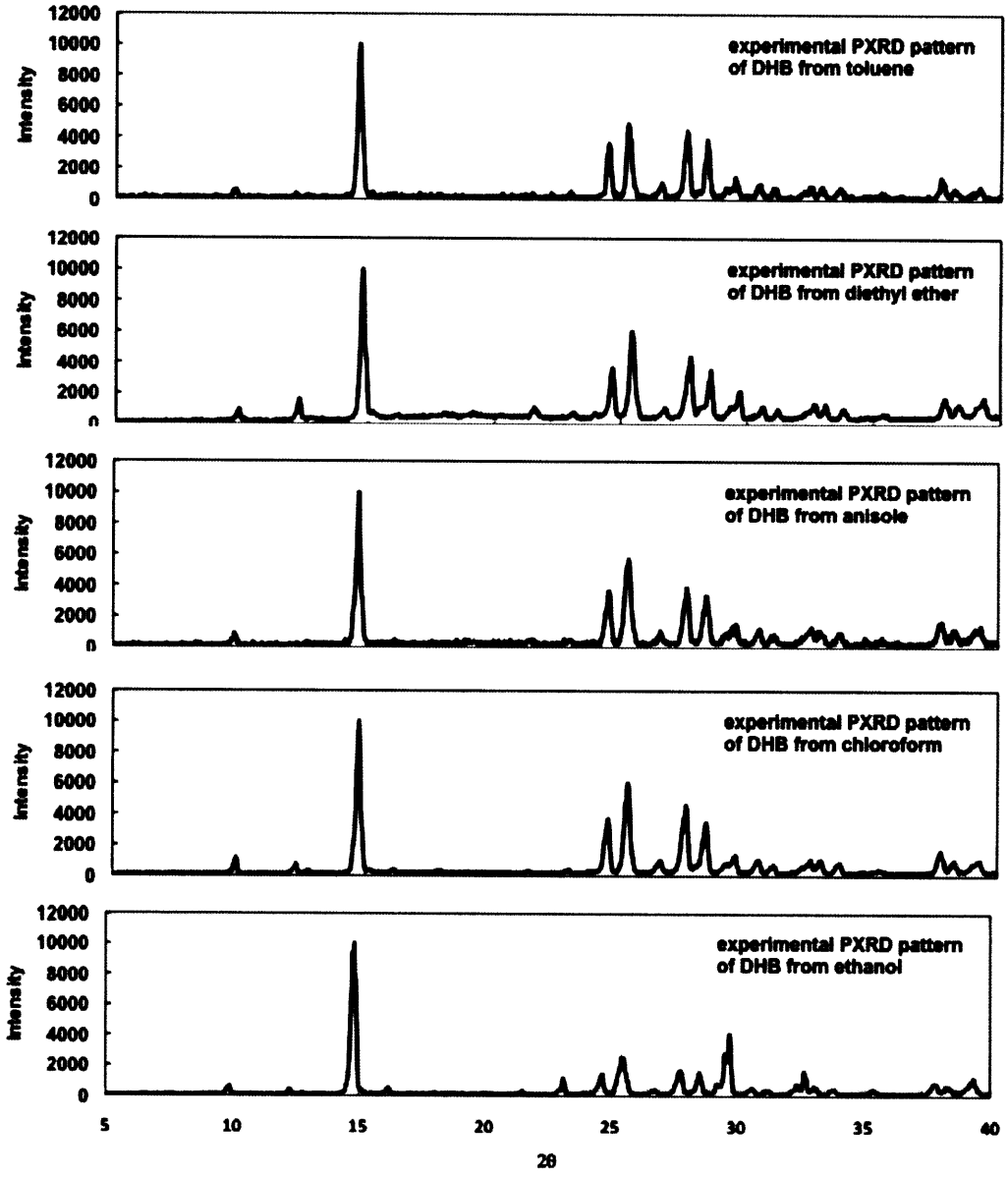


Figure D-1: PXRD patterns of DHB crystals aged in toluene, diethyl ether, anisole, chloroform, and ethanol.

Bibliography

(1) Fisher, J.; D., T. *Journal Of Chemical Physics* **1949**, *17*, 71–73.

(2) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Advanced Drug Delivery Reviews* **2001**, *48*, 3–26.

(3) Hunter, C. A. *Angewandte Chemie International Edition* **2004**, *43*, 5310–5324.

(4) Afoakwa, E. O.; Paterson, A.; Fowler, M. *Trends in Food Science & Technology* **2007**, *18*, 290–298.

(5) Fujiwara, M.; Nagy, Z. K.; Chew, J. W.; Braatz, R. D. *Journal of Process Control* **2005**, *15*, 493–504.

(6) Myerson, A. S. *Handbook of Industrial Crystallization*, 2nd ed.; Butterworth-Heinemann, 2002.

(7) Cano, H.; Gabas, N.; Canselier, J. P. *Journal of Crystal Growth* **2001**, *224*, 335–341.

(8) Davey, R. J.; Mullin, J. W.; Whiting, M. J. L. *Journal of Crystal Growth* **1982**, *58*, 304–312.

(9) Lovette, M. A.; Browning, A. R.; Griffin, D. W.; Sizemore, J. P.; Snyder, R. C.; Doherty, M. F. *Industrial & Engineering Chemistry Research* **2008**, *47*, 9812–9833.

(10) Price, S. L. *Physical Chemistry Chemical Physics* **2008**, *10*, 1996–2009.

(11) Myerson, A. S. *Moluecular Modeling: Applications in Crystallization*; Cambridge University Press, 1999.

(12) Klamt, A. *Journal of Physical Chemistry* **1995**, *99*, 2224–2235.

(13) Ouvrard, C.; Price, S. L. *Crystal Growth & Design* **2004**, *4*, 1119–1127.

(14) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.

(15) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal Of Computational Chemistry* **1983**, *4*, 187–217.

(16) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.

(17) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *Journal Of Physical Chemistry A* **1999**, *103*, 3596–3607.

(18) Sun, H. *Journal of Physical Chemistry B* **1998**, *102*, 7338–7364.

(19) Sun, H.; Ren, P.; Fried, J. R. *Computational and Theoretical Polymer Science* **1998**, *8*, 229–246.

(20) Ponder, J. W.; Case, D. A. *Protein Simulations* **2003**, *66*, 27–+.

(21) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2006**, *1764*, 1647–1676.

(22) Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227–249.

(23) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press, 2001; Vol. 1.

(24) Mullin, J. *Crystallization, 4th Edition*; Butterworth-Heinemann, 2001.

(25) Yu, L.; Stephenson, G. A.; Mitchell, C. A.; Bunnell, C. A.; Snorek, S. V.; Bowyer, J. J.; Borchardt, T. B.; Stowell, J. G.; Byrn, S. R. *Journal Of The American Chemical Society* **2000**, *122*, 585–591.

(26) Caira, M. R. Crystalline polymorphism of organic compounds. In *Design Of Organic Solids*; 1998; Vol. 198, pp 163–208.

(27) Kempf, D. J. et al. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 2484–2488.

(28) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharmaceutical Research* **2001**, *18*, 859–866.

(29) Law, D.; Krill, S. L.; Schmitt, E. A.; Fort, J. J.; Qiu, Y. H.; Wang, W. L.; Porter, W. R. *Journal of Pharmaceutical Sciences* **2001**, *90*, 1015–1025.

(30) Ahn, S. Y.; Guo, F.; Kariuki, B. M.; Harris, K. D. M. *Journal Of The American Chemical Society* **2006**, *128*, 8441–8452.

(31) Scholl, J.; Bonalumi, D.; Vicum, L.; Mazzotti, M.; Muller, M. *Crystal Growth & Design* **2006**, *6*, 881–891.

(32) Parveen, S.; Davey, R. J.; Dent, G.; Pritchard, R. G. *Chemical Communications* **2005**, 1531–1533.

(33) Davey, R. J.; Blagden, N.; Righini, S.; Alison, H.; Quayle, M. J.; Fuller, S. *Crystal Growth & Design* **2001**, *1*, 59–65.

(34) Blagden, N.; Davey, R. J. *Crystal Growth & Design* **2003**, *3*, 873–885.

(35) Ostwald, W. *Physical Chemistry* **1897**, *22*, 289–330.

(36) Davey, R. J.; Blagden, N.; Righini, S.; Alison, H.; Ferrari, E. S. *Journal Of Physical Chemistry B* **2002**, *106*, 1954–1959.

(37) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; WILEY, 2000.

(38) Desiraju, G. R. *Angewandte Chemie International Edition* **1995**, *34*, 2311–2327.

(39) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition path sampling. In *Advances In Chemical Physics*; 2002; Vol. 123, pp 1–78.

(40) Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G. *Physical Review Letters* **2005**, *94*, 235703.

(41) Beckham, G. T.; Peters, B.; Starbuck, C.; Variankaval, N.; Trout, B. L. *Journal of the American Chemical Society* **2007**, *129*, 4714–4723.

(42) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. *Journal Of Physical Chemistry B* **2006**, *110*, 3323–3329.

(43) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S. *Crystal Growth & Design* **2006**, *6*, 1788–1796.

(44) Towler, C. S.; Taylor, L. S. *Crystal Growth & Design* **2007**, *7*, 633–638.

(45) Chattopadhyay, S.; Erdemir, D.; Evans, J. M. B.; Ilavsky, J.; Amenitsch, H.; Segre, C. U.; Myerson, A. S. *Crystal Growth & Design* **2005**, *5*, 523–527.

(46) Erdemir, D.; Chattopadhyay, S.; Guo, L.; Ilavsky, J.; Amenitsch, H.; Segre, C. U.; Myerson, A. S. *Physical Review Letters* **2007**, *99*, 115702.

(47) Hamad, S.; Hughes, C. E.; Catlow, C. R. A.; Harris, K. D. M. *Journal of Physical Chemistry B* **2008**, *112*, 7280–7288.

(48) Huang, J.; Stringfellow, T. C.; Yu, L. *Journal of the American Chemical Society* **2008**, *130*, 13973–13980.

(49) Gavezzotti, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P. *Chemistry-A European Journal* **1997**, *3*, 893–899.

(50) Gavezzotti, A. *Chemistry-A European Journal* **1999**, *5*, 567–576.

(51) Leiserowitz, L. *Acta Crystallographica Section B* **1976**, *32*, 775–802.

(52) Leiserowitz, L.; Benghiat, V. *Journal of the Chemical Society Perkin Transactions 2* **1972**, 1763 – 1768.

(53) `http://www.clippercontrols.com/info/dielectric_constants.html`.

(54) Abraham, M. H.; Platts, J. A. *Journal of Organic Chemistry* **2001**, *66*, 3484–3491.

(55) Darden, T.; Perera, L.; Li, L. P.; Pedersen, L. *Structure with Folding & Design* **1999**, *7*, R55–R60.

(56) Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Oxford, 1987.

(57) `http://chemfinder.cambridgesoft.com`.

(58) Majer, V.; Svoboda, V. *Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Blackwell Scientific Publications: Oxford, 1985.

(59) Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187–199.

(60) Torrie, G. M.; Valleau, J. P. *Journal of Chemical Physics* **1977**, *66*, 1402–1408.

(61) Kottalam, J.; Case, D. A. *Journal of American Chemistry Society.* **1988**, *110*, 7690–7697.

(62) Peters, B.; Trout, B. L. *Journal Of Chemical Physics* **2006**, *125*, 054108.

(63) Peters, B.; Beckham, G. T.; Trout, B. L. *Journal Chemical Physics* **2007**, *127*, 034109.

(64) Perlovich, G. L.; Hansen, L. K.; Bauer-Brandl, A. *Journal of Thermal Analysis and Calorimetry* **2001**, *66*, 699–715.

(65) Albrecht, G.; Corey, R. B. *Journal of the American Chemical Society* **1939**, *61*, 1087–1103.

(66) Iitaka, Y. *Acta Crystallographica* **1960**, *13*, 35–45.

(67) Iitaka, Y. *Acta Crystallographica* **1961**, *14*, 1–10.

(68) Myerson, A. S.; Lo, P. Y. *Journal of Crystal Growth* **1990**, *99*, 1048–1052.

(69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926–935.

(70) Price, S. L.; Hamad, S.; Torrisi, A.; Karamertzanis, P. G.; Leslie, M.; Catlow, C. R. A. *Molecular Simulation* **2006**, *32*, 985–997.

(71) MacKerell, A. D. `http://mackerell.umaryland.edu/Empirical_FF_Dev.html`.

(72) Weissbuch, I.; Torbeev, V. Y.; Leiserowitz, L.; Lahav, M. *Angewandte Chemie-International Edition* **2005**, *44*, 3226–3229.

(73) Hallett, J.; Mason, B. J. *Nature* **1958**, *181*, 467–469.

(74) Kapetanovic, I. M. *Chemico-Biological Interactions* **2008**, *171*, 165–176.

(75) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 11937–11942.

(76) Karunanithi, A. T.; Achenie, L. E. K.; Gani, R. *Chemical Engineering Science* **2006**, *61*, 1247–1260.

(77) Macgillivray, L. R.; Zaworotko, M. J. *Journal of Chemical Crystallography* **1994**, *24*, 703–705.

(78) Gdaniec, M.; Gilski, M.; Denisov, G. S. *Acta Crystallographica Section C-Crystal Structure Communications* **1994**, *50*, 1622–1626.

(79) Wulff, G. *Zeitschrift Fur Kristallographie Und Mineralogie* **1901**, *34*, 449–530.

(80) Docherty, R.; Clydesdale, G.; Roberts, K. J.; Bennema, P. *Journal of Physics D-Applied Physics* **1991**, *24*, 89–99.

(81) Donnay, J. D. H.; Harker, D. *American Mineralogist* **1937**, *22*, 446–467.

(82) Hartman, P.; Perdok, W. G. *Acta Crystallographica* **1955**, *8*, 49–52.

(83) Lu, J. J.; Ulrich, J. *Crystal Research and Technology* **2003**, *38*, 63–73.

(84) Hammond, R. B.; Pencheva, K.; Ramachandran, V.; Roberts, K. J. *Crystal Growth & Design* **2007**, *7*, 1571–1574.

(85) Liu, X. Y.; Boek, E. S.; Briels, W. J.; Bennema, P. *Nature* **1995**, *374*, 342–345.

(86) Winn, D.; Doherty, M. F. *AIChE Journal* **1998**, *44*, 2501–2514.

(87) Piana, S.; Reyhani, M.; Gale, J. D. *Nature* **2005**, *438*, 70–73.

(88) Winn, D.; Doherty, M. F. *AIChE Journal* **2000**, *46*, 1348–1367.

(89) Winn, D.; Doherty, M. F. *Chemical Engineering Science* **2002**, *57*, 1805–1813.

(90) Rountree, J. S. S.; Murphy, P. V. *Organic Letters* **2009**, *11*, 871–874.