A Computational Investigation of Nucleation Processes in Organic Crystals

by

Gregg Tyler Beckham

BS Chemical Engineering, Oklahoma State University, 2002
MS Chemical Engineering Practice, Massachusetts Institute of Technology, 2004

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
[February 2008]
October 2007

Signature of Author:_____
Department of Chemical Engineering
October 2, 2007

Certified by:_____
Bernhardt L. Trout, Ph.D.
Associate Professor
Thesis Supervisor

Accepted by:_____
William M. Deen, Ph.D.
Carbon P. Dubbs Professor
Chairman, Committee for Graduate Students

A Computational Investigation of Nucleation Processes in Organic Crystals

Gregg Tyler Beckham

Nucleation processes are ubiquitous in nature and technology. For instance, cloud formation in the atmosphere, the casting of metals, protein crystallization, biomineralization, the production of porous materials, and separation of pharmaceutical compounds from solution are a few examples of relevant nucleation processes. One pathway for nucleation to occur is homogeneous nucleation, in which an embryo of a more stable phase forms within an original metastable medium. Homogeneous nucleation is an activated process, meaning that a free energy barrier must be overcome for the transition to take place, and the height of the free energy barrier determines the rate at which the process will occur. Despite considerable advances in both theoretical and experimental techniques to date, determining nucleation mechanisms for real systems remains a considerable technical challenge. The aim of this thesis is therefore to apply molecular simulation techniques to elucidate nucleation mechanisms in organic crystals. Specifically, the newly developed methods of aimless shooting and likelihood maximization are applied for the first time to study nucleation processes in complex and technically relevant systems.

The first portion of the thesis examines polymorphism, or the ability of a material to pack in different crystal lattices whilst retaining the same chemical composition. Transformation to a more stable polymorph can readily occur in the solid state, which has broad implications in pharmaceutical processing. To date, over 160 mechanisms have been proposed for polymorph transitions in the solid state, but none have been definitively verified. A model compound, terephthalic acid, is chosen for computational studies because it is similar in size to a small molecule therapeutic and exhibits a common bonding motif for organic crystals. Using aimless shooting and likelihood maximization, the mechanism of the solid state polymorph transformation in terephthalic acid is shown to be corner nucleation. The mechanism shows that for a given nucleus size, the interfacial area between the crystalline domains is minimized, thus reducing the unfavorable surface free energy penalty required for nucleation to occur. Furthermore, based on the results presented, it is anticipated that corner nucleation may be a common mechanism for many polymorph transformations in hydrogen bonded crystalline materials.

The second portion of the thesis investigates the mechanism of freezing a subcooled liquid to form a crystal. This phenomenon has widespread application across many technical domains. Similar studies to date on freezing have been limited to model systems, such as Lennard-Jones particles or hard spheres. Benzene is chosen as a model compound. A periodic system is constructed and aimless shooting and likelihood maximization are applied to determine the nature of the critical nucleus. Local order analysis is implemented to distinguish among solid and liquid-like molecules. Preliminary results indicate that the critical nucleus is on the order of 200-300 molecules at 50 K subcooling.

This thesis demonstrates that the complementary molecular simulation techniques of aimless shooting and likelihood maximization offer fundamental insight into nucleation mechanisms in molecular crystals. Knowledge of the mechanism from likelihood maximization is essential for accurate free energies and pathway optimization methods, and it should therefore be applied in computational studies of rare events prior to free energy or rate constant calculations. Moreover, these methods provide quantitative understanding of the important physical variables that determine experimentally observable rates and can further aid in experimental design.

Thesis Supervisor: Bernhardt L. Trout, Associate Professor

Acknowledgements

First, I would like to thank my advisor Professor Bernhardt Trout. He has instilled in me a desire to conduct rigorous work and an appreciation for the search for fundamental understanding as a prerequisite to judicious engineering. In addition, Bernhardt has created a comfortable, good-natured learning environment with friendly, professional support, afforded me a great deal of creative freedom with our research, allowed me to pursue my own interests outside of my thesis project, and has been patient with me in all things. He is an excellent and keenly insightful mentor, and for that I am earnestly grateful.

Thesis committee members Professors Michael Cima, Klavs Jensen, and Sidney Yip provided valuable advice, for which I am appreciative. They challenged me to address the research problems in different and unique ways, and this thesis benefited greatly from their questions and insight.

The initial choice of a project for doctoral work in a new field can be daunting. In our case, we profited greatly from our industrial sponsors in both the Chemical Engineering Research and Development group and the Physical Measurements group at Merck & Co., Inc. Specifically, Michael Thien arranged the initial collaboration and was a fervent supporter of the project throughout. Cindy Starbuck generously and enthusiastically gave her time to the project, especially when I worked at Merck for extended visits. Narayan Variankaval continuously provided helpful suggestions concerning both experimental and computational work. Ivan Lee also deserves special thanks as he volunteered an empty room in his house when I summered in New Jersey to work at Merck in 2003.

Many people at MIT have contributed to the success of this work. Working with the Trout research group has significantly contributed to my overall intellectual experience at MIT. I wish to thank all members past and present for their support and camaraderie during my time at MIT. In particular, Jhih-Wei Chu and Brian Baynes both patiently instructed me in how to use CHARMM and calmed my initial frustrations in molecular simulation. Bin Pan has been a continuous source of helpful discussions and rigorous suggestions as well as a good friend. It has been comforting to have fellow crystal simulators Jie Chen and Srikanth Patala in the group to vent common frustrations with molecular simulation of crystals. Victor Ovchinnikov deserves special thanks as he bore the Promethean duty of bringing NAMD to our group, thus allowing us to answer questions for larger, more complex systems. In addition, he has volunteered substantial amounts of his time to discuss the research presented in this thesis and share his specialized codes, and the thesis has been improved greatly as a result. Gwen Wilcox, Don McGaffigan, and Angelique Scarpa have worked with our group tirelessly to help in all operational aspects. Finally, I am also grateful for generous funding from Merck & Co., Inc., the Singapore-MIT Alliance, and the National Science Foundation via a Graduate Research Fellowship.

One of the greatest friends I made at MIT is Baron Peters. Baron's genuine enthusiasm for conducting research served as an inspiration and model to me. His relentless drive to answer difficult problems led to his pioneering work on aimless shooting and likelihood maximization, which is used throughout and partially developed in this thesis. Baron served as a patient, thoughtful mentor and collaborator, and was always willing to give his time. Spending several weeks working literally side-by-side in Lyon will easily be one of my fondest memories from graduate school. I sincerely thank Baron for his great friendship and

prudent counsel. I wish Baron (and Ban) continued success and happiness, and look forward to any future collaboration in which I might be fortunate enough to participate.

MIT includes many opportunities for professional development beyond the doctoral thesis project. Attending the Practice School was undoubtedly one of the best choices that I made whilst at MIT. Claude Lupis made the experience enjoyable and worthwhile, and for that I am grateful. Counting marshmallows from bowls of Lucky Charms with Mark Sains will remain a fond memory. Belinda Akpa faithfully accompanied me to local bookstores on Sundays. Andy Rees and Iain Burns introduced me to driving on the other side of the road, listening to Tom Waits on rainy evenings, and British-style pub crawls in Hull, all of which was enjoyable and appreciated. In addition, I thank Professor Alan Hatton for providing me with the opportunity to act as the director of the Industrial Immersion program for the Singapore-MIT Alliance in Singapore.

My trajectory to MIT was aided immensely by my admittance to the Oklahoma School of Science and Mathematics. I owe a huge debt of gratitude to the school, and especially to Professors Dorothy Dodd, Jeanette Feldott, Chengde Feng, Yunhua Feng, and Russell Lawson. During undergraduate studies at Oklahoma State University, I was fortunate enough to be rejected as a teaching assistant for the conventional courses in the Chemistry Department. This steered me to work as a teaching assistant in inquiry based chemistry for three years with Dr. Smith Holt. He demonstrated how to apply science and education as complementary tools for social good, and was immensely influential in my decision to attend graduate school.

I am grateful to my mother and father for raising me to be an independent and motivated person and for exposing me to different worldviews by taking me to live abroad as a child. Brenda has been a positive influence as well and for that I thank her. Both my grandfather and father are especially hard working, dedicated people, and have both been excellent role models from my formative days of farm work to the present day as I finish my time at graduate school.

During my time at MIT, I have benefited from many great friendships. Our first year graduate class was an especially fun and diverse group of wonderful people. Kris Wood is easily the most positive person and the best storyteller that I have ever met. He has been both one of my best friends and an inspiration throughout graduate school, and for that I am exceedingly grateful. Jane Rempel, besides being a helpful genius regarding research and an unparalleled proofreader, has also been a great friend. Jane has always been ready for a non-work related diversion over the last five years, especially those involving movie-hopping or afternoon ABP runs. Sanjoy Sircar merits thanks as he was an excellent roommate for four years. Sharon Soong has been a great IM buddy on a daily basis. I was fortunate to sit across the table from the multi-talented Bernat Olle at the first dormitory social event that I attended, and we subsequently spent many evenings together sipping Onix whilst working on homework problems and telling stories. Evenings at Unos with Gwen Wilcox were a definite highlight, as she has been a good friend and a great listener throughout my graduate career; I truly appreciate her support. Shannan O'Shaughnessy and Jake Albrecht were fun-loving partners-in-crime for evenings on the town as well as supportive friends through tough times. Finally, Elizabeth Young has made my last two years at MIT a great joy, as she has been a steadfast, delightful, and goofy (in a good way) companion. My friends made at MIT have my love, respect, and best wishes for the future.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: A COMPUTATIONAL INVESTIGATION OF NUCLEATION PROCESSES IN ORGANIC CRYSTALS

Nucleation processes are ubiquitous in nature and technology. For instance, cloud formation in the atmosphere, the casting of metals, protein crystallization, the formation of inorganic minerals in organisms, the production of porous materials for catalysis and gas absorption, and the separation of high value pharmaceutical compounds from solution are all examples of relevant nucleation processes[1-4]. One pathway by which these processes can occur is homogeneous nucleation, in which an embryo of a more stable phase forms within an original, metastable medium[2]. This type of phase transformation is an activated process, meaning that a free energy barrier must be overcome for the transition to take place, and the height of the free energy barrier determines the rate at which the process will occur. After the cluster is beyond some critical size, the growth of the new phase is spontaneous. In addition, the free energy barrier in nucleation processes depends strongly on temperature, which in turn means that experimental results can be astoundingly difficult to reproduce and theoretical predictions of nucleation rates are rife with uncertainty. As Oxtoby declares "[n]ucleation theory is one of the few areas of science in which agreement of predicted and measured rates to within several orders of magnitude is considered a major success"[3]. It is therefore evident that quantitative understanding of the events leading to nucleation and crystallization is of paramount significance in many areas of science. In light of the need for further understanding of nucleation, this thesis primarily examines two nucleation processes, both in real systems. The first case study elucidates the mechanism of a solid state polymorph transformation in an organic crystal similar in size and motif to organic crystals found in the pharmaceutical industry. The second portion of the thesis examines the mechanism of homogeneous nucleation of an organic crystal from the melt.

To date, experimental approaches for observing random, localized events leading to nucleation and crystallization are limited to relatively large monomeric units such as colloidal particles[5] or to aggregated constituents like large protein nuclei[6], which exist in a range of sizes between 100 nm to 1 μm. While experimental studies of nucleation are relevant for larger length scale processes, current experimental methodologies cannot offer the desired resolution and quantitative understanding for nucleation events that occur on atomic and molecular length scales ranging in magnitude from 1 Å to 10-100 nm. Computational

methods, however, are able to track particles in space and time and therefore are more readily applicable in studying nucleation on a molecular level. Therefore, the approach taken in this thesis is to develop and apply sophisticated computational techniques in order to investigate nucleation events in molecular systems. This approach yields mechanistic information describing the nucleation process in systems of interest. Also, the work described here highlights the development of new computational techniques required to solve problems of this complexity.

1.1 Methodology

A number of studies have proved fruitful in the elucidation of nucleation processes for simple systems. Previous studies can be generally separated into three types of methodological categories:

1) Using well-characterized systems for which the time scale for nucleation is computationally tractable at intermediate to high degrees of subcooling[7-12],

2) Selecting a mechanism *a priori* to force the desired transformation[13-22], or

3) Collecting many unbiased pathways for the transformation[23-28].

The first methodology is limited to obtaining a nucleation event in simple systems at intermediate to large subcooling. For instance, the first publication describing the nucleation of liquid water to hexagonal ice required several microseconds of simulation time to generate a single trajectory at significant subcooling[10]. Figure 1-1 shows snapshots of the trajectory. The "wall clock" time required to perform these microsecond simulations is on the year timescale, and the simulations are not guaranteed to yield useful results *a priori*. The brute force nature of the first approach highlights the limitation of such a methodology for real systems[10]. The second methodology, pioneered by van Duijneveldt and Frenkel[13], often offers valuable insight for simple systems in which the choice of a reaction coordinate, the single variable that quantifies progress along a reaction pathway, is straightforward. However, in systems of increasing complexity, the choice of a reaction coordinate is typically not intuitive, and incorrect reaction coordinates lead to spurious results for quantities of interest such as free energy changes, kinetic rate constants, and the mechanism itself.

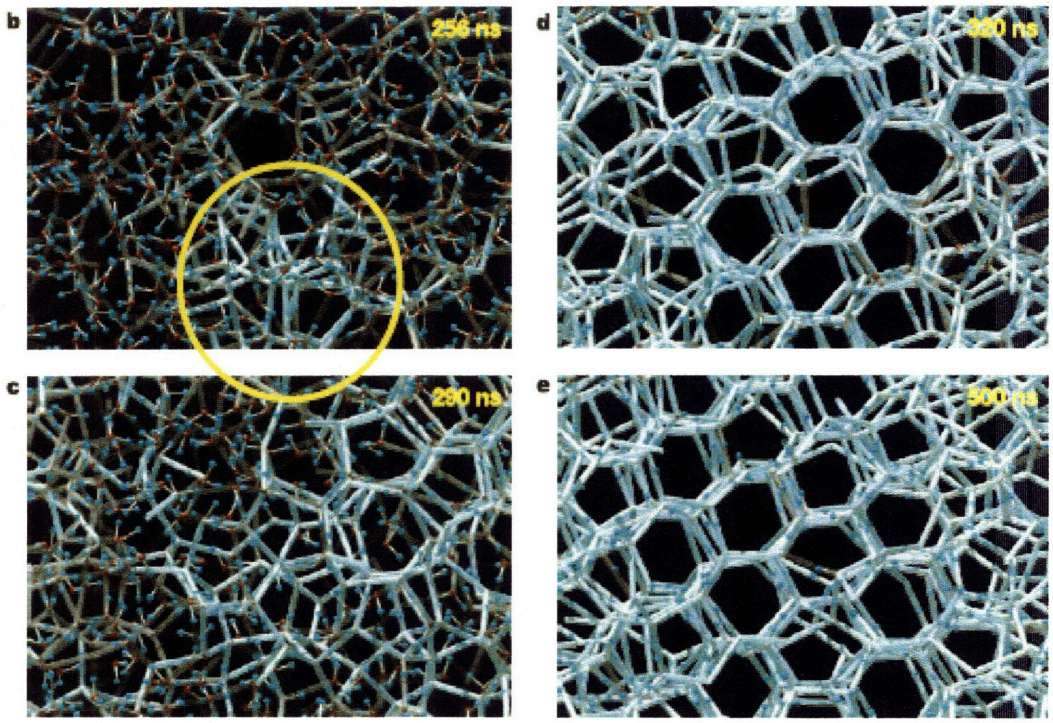


Figure 1-1. Snapshots along a trajectory from liquid water to hexagonal ice, reproduced from Matsumoto *et al.*[10] The trajectory shown was obtained from an MD simulation on the timescale of hundreds of nanoseconds at 230 K. The circled region shows the first sign of order in the system.

The third methodology listed above became readily available with the development of the transition path sampling [TPS]. The TPS method, pioneered by Chandler and colleagues, is an efficient method that allows one to study unbiased, rare events in systems with no previous knowledge of the mechanisms or transition states[29-34]. The method is an iterative, Monte Carlo-like procedure in dynamic space. An initial trajectory connecting the reactant and product basins is generated, which may or may not be physical at the conditions of interest. TPS is initiated from this initial trajectory by perturbing the velocity at a randomly selected point and shooting trajectories in the forward and reverse directions along the new velocity vector. Trajectories are accepted if one end of the trajectory relaxes to the reactant basin and the other reaches the product basin. Trajectories are rejected if both ends connect a single basin. Repetition of the shooting procedure allows one to collect a set of reactive trajectories efficiently by focusing the computational time on the dynamical bottleneck. Figure 1-2 illustrates the TPS shooting algorithm.

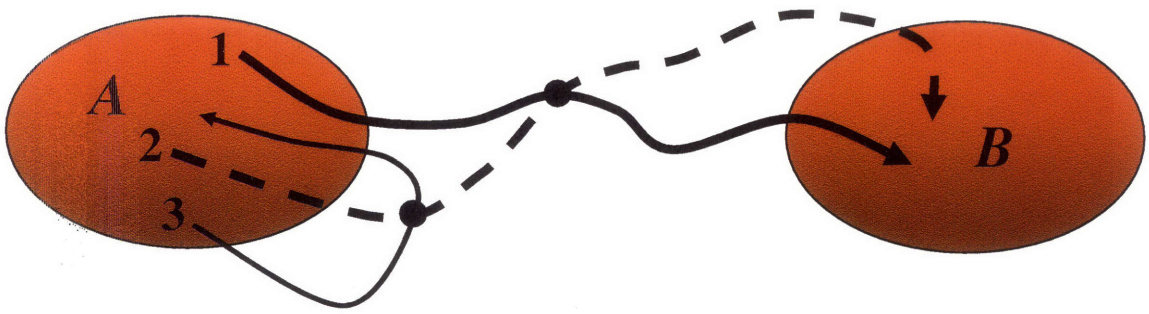


Figure 1-2. Illustration of trial moves in transition path sampling [TPS]. Trajectory 2 is generated by perturbing the momentum slightly at a randomly selected point along trajectory 1, and propagating the forward and reverse directions along the new velocity vector. Since trajectory 2 connects the basins *A* and *B*, it is accepted as a reactive trajectory. Trajectory 3 is generated in the same manner as trajectory 2, and is an example of a rejected trajectory in that it does not connect the basins.

In addition, a variation on the TPS algorithm was developed: aimless shooting[27, 35]. The primary difference in TPS and aimless shooting is that the latter seeds random velocities at the shooting points, rather than perturbing the velocities from the previous trajectory, hence the term "aimless" shooting. This allows faster de-correlation of the transition paths. The other primary difference is that aimless shooting selects only from two shooting points, separated by a pre-specified time displacement whereas TPS randomly selects any point along the previous trajectory. TPS has been applied in nucleation studies such as the freezing of Lennard-Jones particles[23, 25], the formation of small NaCl clusters from solution[24], and nucleation in the Ising model[27, 36]. Other notable demonstrations of TPS across a wide variety of phenomena include the autoionization of water[37], hydrophobic collapse of a polymer[38], cavity formation in boiling water[39], biomolecular isomerization[40], and protein folding[41].

After TPS has been conducted, one can determine the mechanism, or reaction coordinate, from the information contained in the transition path ensemble collected during TPS. Again, the reaction coordinate is the single variable that quantifies progress along a reaction pathway[27]. A good reaction coordinate yields useful insight into the factors which affect experimental reaction rates. From a computational perspective, an accurate reaction coordinate is necessary for free energy or rate constant calculations. Previous TPS studies utilize the committor probability histogram method to determine the reaction coordinate[42-44]. This method is a trial and error procedure that is at least as computationally expensive as TPS itself for each possible reaction coordinate tested. A more efficient method has been developed that uses an informatics approach, likelihood maximization, to extract the best

reaction coordinate from a list of candidate order parameters [OPs][27, 35]. This method is applied after TPS has been conducted. Each OP, or combination of OPs, is assigned a likelihood score, which is used to rank the reaction coordinate approximation. The complementary methods of aimless shooting and likelihood maximization are applied in this thesis to investigate two types of nucleation events in organic crystals.

1.2 Solid state Polymorph Transformation in Terephthalic Acid

A major issue concerning nucleation and crystallization in the pharmaceutical industry is polymorphism, which is defined as the capability of a substance to crystallize into different, but chemically identical, crystalline forms[1]. Polymorphs typically exhibit different physical and chemical properties, which warrant specific design of processes to formulate particular products. An example of widely varying physical properties of polymorphs is illustrated in Figure 1-3, which shows six of the ten polymorphs of 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, known in the literature as ROY[45]. In the pharmaceutical industry, transformation to an undesired polymorph, 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[46]. Oftentimes when compounds are discovered and a process is developed for its production, a more stable, undesired polymorph may appear as in the case with Abbott Laboratories' AIDS drug, Ritonavir. The occurrence of a late appearing polymorph caused setbacks in Ritonavir production as scientists were forced to discover methods to reproduce the desired, metastable polymorph[47]. Cases such as this arise from the exchange between kinetic and thermodynamic stability and can potentially cause extensive and costly problems in development, production, and patent rights.

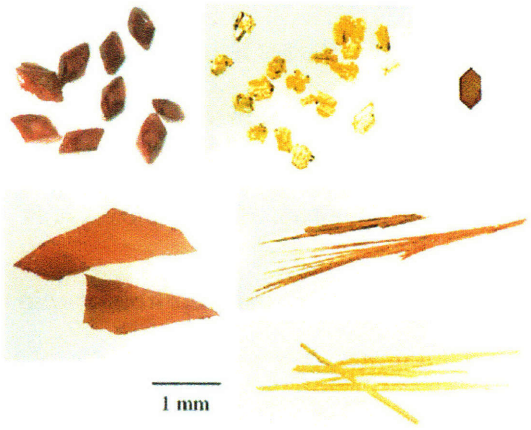


Figure 1-3. Six polymorphs of ROY[45]. The wide variety of physical properties illustrate the diversity that arises from differences in crystal packing arrangements of the same molecule.

Polymorph transformations often occur in the solid state. Mnyukh states that, as of 1998, over 160 mechanisms are reported for solid state polymorph transitions[48]. At present, however, gaining direct molecular level insight to the dynamical events occurring during solid state polymorph transformations is outside of the scope of experimental capabilities, and thus there is no definitive evidence for any particular mechanism. As remarked by Herbstein in an excellent review: "[t]here are relatively few papers about the actual transition directly viewed by microscopic techniques in order to infer the mechanism, and not many about changes in crystal structure...*as the system passes through the transition.*"[49] Two of the most common mechanisms proposed in the literature are 1) nucleation and growth[48, 50-53] and 2) concerted, or martensitic, transformations[54, 55]. As Tuble *et al.* point out, however, there is increasing evidence that transformations thought to be martensitic in nature, actually occur via a nucleation and growth mechanism[48, 49, 54]. An accurate model which describes this process on a molecular level would provide an improved understanding of the mechanism of solid state polymorph transformations in molecular crystals, and could ultimately provide the necessary knowledge for *a priori* selection of processing parameters to promote or prevent transformation from one polymorph to another.

As mentioned previously, determining molecular level pathways in a polymorph transformation is currently difficult from an experimental standpoint due to the inability to directly view morphological changes in crystal structures as a function of time. Molecular simulations, however, provide a convenient approach for the study of polymorph transformations. The objective of the research in this portion of the thesis is to therefore

apply aimless shooting and likelihood maximization[27, 35] to determine the mechanism for the solid state polymorph transformation of a model compound that is of the same approximate size as a small molecule therapeutic: terephthalic acid. Terephthalic acid is chosen as the molecule of study because:

1) it is relatively small and rigid even when completely unconstrained in a dynamic simulation at fully atomistic detail,

2) both of the polymorph crystal structures are known[56],

3) the transformation exhibits interesting macroscopic behavior that remains unexplained in the experimental literature[57, 58], and

4) it represents a ubiquitous motif in organic crystallography with the hydrogen bonded carboxylic acid dimer synthons[59].

1.3 Homogeneous Nucleation of Crystalline Benzene from the Melt

Nucleation of crystalline solids from the melt has been studied with both experimental and computational techniques[5, 10, 12-20, 23, 25, 28, 60]. In a series of three papers, ten Wolde *et al.* described the formation of Lennard-Jones crystals from the liquid state by applying an umbrella potential with bond-orientational order parameters used to convert liquid to solid particles in the Lennard-Jones system[14-16]. This study showed that the crystals typically form as body centered cubic [bcc] clusters and later convert to the more stable face centered cubic [fcc] structure. Moroni, ten Wolde, and Bolhuis later applied TPS to the same system and demonstrated that the critical nuclei can exist as large, more disordered bcc-like lattices as previously found, or as smaller, compact fcc crystals. The earlier publications of ten Wolde *et al.* were certainly pioneering work at the time of publication[14-16], and have been generally applied to many different systems by other researchers[12, 17-21, 61]. However, this later study from Moroni *et al.*, which illustrates that assuming a reaction coordinate yields a biased mechanism, highlights the need for the use of unbiased methods in determining mechanisms in nucleation problems, even for seemingly simple systems. Despite these considerable advances, there still is not a comprehensive mechanistic understanding of nucleation, especially for more complex systems in which the molecules are anisotropic, such as disk shaped molecules like benzene. It is anticipated that the anisotropy of benzene molecules will yield non-spherical critical nuclei. To investigate this, aimless shooting and likelihood

maximization are applied to determine the reaction coordinate for the mechanism of nucleation in benzene. Benzene is chosen because:

1) it is a common model organic molecule and is fairly rigid and relatively small,

2) the OPLS potential describes the liquid and solid states adequately[62, 63],

3) the Form I crystal structure is known[64], and

4) considerable experimental work has been conducted investigate the liquid-solid phase transition[65-70].

1.4 Thesis Outline

Chapter 2 describes the first application of aimless shooting and likelihood maximization to determine the mechanism for two crystal sizes in the solid state polymorph transformation in terephthalic acid. Chapter 3 outlines a methodological alteration to the likelihood maximization algorithm that improves the ability to determine the transition state along the transformation. Chapter 4 discusses the application of these same methods to larger terephthalic acid crystals, and focuses on the importance of local order to the reaction coordinate. Also, the critical nucleus size is examined at the transition state in light of classical nucleation theory. Chapter 5 describes the application of aimless shooting and likelihood maximization to study benzene nucleation from the melt. The preliminary results for benzene nucleation are outlined. A summary of the research and recommendations for future work are outlined in Chapter 6.

1.5 References

1. Mullin, J.W., *Crystallization*. 4th Edition ed. 2001: Butterworth-Heinemann.
2. Debenedetti, P.G., *Metastable liquids: Concepts and principles*. 1st Edition ed. 1996: Princeton University Press.
3. Oxtoby, D.W. *Acc. Chem. Res.*, **31,** p. 91-97 (1998).
4. Hollingsworth, M.D. *Science*, **295,** p. 2410-2413 (2002).
5. Gasser, U., et al. *Science*, **292,** p. 258-262 (2001).
6. Yau, S.T. and P.G. Vekilov. *Nature*, **406,** p. 494-497 (2000).
7. Anwar, J. and P.K. Boateng. *J. Amer. Chem. Soc.*, **120,** p. 9600-9604 (1998).
8. Berim, G.O. and E. Ruckenstein. *J. Chem. Phys.*, **117,** p. 4542-4549 (2002).
9. Berim, G.O. and E. Ruckenstein. *J. Chem. Phys.*, **117,** p. 7732-7737 (2002).
10. Matsumoto, M., S. Saito, and I. Ohmine. **416,** p. 409-413 (2002).
11. Beaucage, P. and N. Mousseau. *Phys. Rev. B*, **71,** p. 094102 (2005).
12. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Chem. Phys.*, **122,** p. 184518 (2005).
13. van Duijneveldt, J.S. and D. Frenkel. *J. Chem. Phys.*, **96,** p. 4655-4668 (1992).
14. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *Faraday Discuss.*, **104,** p. 93-110 (1996).
15. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *J. Chem. Phys.*, **104,** p. 9932-9947 (1996).
16. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *Phys. Rev. Lett.*, **75,** p. 2714-2717 (1995).

17. Desgranges, C. and J. Delhommelle. *J. Amer. Chem. Soc.*, **128,** p. 15104-15105 (2006).
18. Desgranges, C. and J. Delhommelle. *J. Chem. Phys.*, **126,** p. 054501 (2007).
19. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Amer. Chem. Soc.*, **126,** p. 12286-12287 (2004).
20. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Chem. Phys.*, **122,** p. 104510 (2005).
21. Radhakrishnan, R. and B.L. Trout. *J. Chem. Phys.*, **117,** p. 1786-1796 (2002).
22. Chopra, M., Müller, M., and de Pablo, J.J. *J. Chem. Phys.*, **124,** p. 134102 (2006).
23. Moroni, D., P.R. ten Wolde, and P.G. Bolhuis. *Phys. Rev. Lett.*, **94,** p. 235703 (2005).
24. Zahn, D. *Phys. Rev. Lett.*, **92,** p. 040801 (2004).
25. Trudu, F., D. Donadio, and M. Parrinello. *Phys. Rev. Lett.*, **97,** p. 105701 (2006).
26. Zahn, D. *J. Phys. Chem. B*, **111,** p. 5249-5253 (2007).
27. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
28. Radhakrishnan, R. and B.L. Trout. *J. Amer. Chem. Soc.*, **125,** p. 7743-7747 (2003).
29. Bolhuis, P.G., C. Dellago, and D. Chandler. *Faraday Discuss.*, **110,** p. 421-436 (1998).
30. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **110,** p. 6617-6625 (1999).
31. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **108,** p. 9236-9245 (1998).
32. Dellago, C., et al. *J. Chem. Phys.*, **108,** p. 1964-1977 (1998).
33. Dellago, C., P.G. Bolhuis, and P.G. Geissler. *Adv. Chem. Phys.*, **123,** p. 1-86 (2001).
34. Bolhuis, P.G., et al. *Annu. Rev. Phys. Chem.*, **53,** p. 291-318 (2002).
35. Peters, B., G.T. Beckham, and B.L. Trout. *J. Chem. Phys.*, **127,** p. 1 (2007).
36. Pan, A.C. and D. Chandler. *J. Phys. Chem. B*, **108,** p. 19681-19686 (2004).
37. Geissler, P.G., et al. *Science*, **291,** p. 2121-2124 (2001).
38. ten Wolde, P.R. and D. Chandler. *Proc. Natl. Acad. Sci. U.S.A.*, **99,** p. 6539-6543 (2002).
39. Zahn, D. *Phys. Rev. Lett.*, **93,** p. 227801 (2004).
40. Bolhuis, P.G., C. Dellago, and D. Chandler. *Proc. Natl. Acad. Sci. U.S.A.*, **97,** p. 5877-5882 (2000).
41. Juraszek, J. and P.G. Bolhuis. *Proc. Natl. Acad. Sci. U.S.A.*, **103,** p. 15859-15864 (2006).
42. Du, R., et al. *J. Chem. Phys.*, **108,** p. 334-350 (1998).
43. Geissler, P.G., C. Dellago, and D. Chandler. *J. Phys. Chem. B*, **103,** p. 3706-3710 (1999).
44. Truhlar, D.G. and B.C. Garrett. *J. Phys. Chem. B*, **104,** p. 1069-1072 (2000).
45. Yu, L., et al. *J. Amer. Chem. Soc.*, **122,** p. 585-591 (2000).
46. Caira, M.R. *Topics in Curr. Chem.*, **198,** p. 163-208 (1998).
47. Morissette, S.L., et al. *Proc. Natl. Acad. Sci. U.S.A.*, **100,** p. 2180-2184 (2003).
48. Mnyukh, Y.V., *Fundamentals of solid-state phase transitions ferromagnetism and ferroelectricity*. 1st Edition ed. 1998: 1st Books Library.
49. Herbstein, F.H. *Acta Cryst.*, **B62,** p. 341-383 (2006).
50. Mnyukh, Y.V. *J. Cryst. Growth*, **32,** p. 371-377 (1976).
51. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 163-200 (1979).
52. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 201-218 (1979).
53. Cardew, P.T., R.J. Davey, and A.J. Ruddick. *J. Chem. Soc. - Faraday Trans.*, **80,** p. 659-668 (1984).
54. Tuble, S.C., J. Anwar, and J.D. Gale. *J. Amer. Chem. Soc.*, **126,** p. 396-405 (2004).
55. Kaneko, F., et al. *J. Phys. Chem. B*, **102,** p. 327-330 (1998).
56. Bailey, M. and C.J. Brown. *Acta. Cryst.*, **22,** p. 387-391 (1967).
57. Davey, R.J., et al. *J. Chem. Soc. - Faraday Trans.*, **90,** p. 1003-1009 (1994).
58. Davey, R.J., et al. *Nature*, **366,** p. 248-250 (1993).
59. Desiraju, G.R. *Angew. Chem. Int. Ed.*, **34,** p. 2311-2327 (1995).
60. Valeriani, C., E. Sanz, and D. Frenkel. *J. Chem. Phys.*, **122,** p. 194501 (2005).
61. Shen, V.K. and P.G. Debenedetti. *J. Chem. Phys.*, **111,** p. 3581-3589 (1999).
62. Jorgensen, W.L. and D.L. Severance. *J. Amer. Chem. Soc.*, **112,** p. 4768-4774 (1990).
63. Cacelli, I., et al. *J. Amer. Chem. Soc.*, **126,** p. 14278-14286 (2004).
64. Bacon, G.E., N.A. Curry, and S.A. Wilson. *Proc. Roy. Soc. London*, **279,** p. 98-110 (1954).
65. Bridgman, P.W. *J. Chem. Phys.*, **9,** p. 794-797 (1941).
66. Akella, J. and G.C. Kennedy. *J. Chem. Phys.*, **55,** p. 793-796 (1971).
67. Azreg-Ainou, M. *J. Cryst. Growth*, **299,** p. 195-205 (2007).
68. Azreg-Ainou, M., A. Huseynov, and B. Ibrahimoglu. *J. Chem. Phys.*, **124,** p. 204505 (2006).
69. Craven, C.J., et al. *J. Chem. Phys.*, **98,** p. 8236-8243 (1993).
70. Cansell, F., D. Fabre, and J.P. Petitet. *J. Chem. Phys.*, **99,** p. 7300-7304 (1993).

CHAPTER 2: SURFACE-MEDIATED NUCLEATION IN THE SOLID STATE POLYMORPH TRANSFORMATION OF TEREPHTHALIC ACID

Reproduced in part with permission from G.T. Beckham, B. Peters, C. Starbuck, N. Variankaval, and B.L. Trout, "Surface-Mediated Nucleation in the Solid State Polymorph Transformation of Terephthalic Acid", *J. Amer. Chem. Soc.*, 129, 4714- 4723 (2007).

2.1 Introduction

As described in Chapter 1, polymorphism is the ability of a system to pack into different crystal lattices while retaining the same chemical composition[1]. It is a well known phenomenon with important technical and financial implications in a diverse range of areas in which crystalline materials play a significant role, such as geophysics[2], energy storage[3], biominerals[4, 5], nonlinear optical materials[6], and pharmaceuticals[7-10]. Polymorphs typically exhibit different physical and chemical properties, thus representing an excellent framework for the study of structure-property relationships. The property differences between polymorphs can affect process and product development. The transformation to an undesired polymorph in the pharmaceutical industry, for instance, can lead to different bioavailability in the target organism, which could render the drug ineffective or increase its potency to a dangerous limit[11]. One prominent case is Abbott Laboratories' AIDS drug, Ritonavir: a more stable polymorph appeared after a manufacturing process was developed for its production, leading to significant production setbacks[12].

Polymorph transformations often occur in the solid state. Mnyukh states that, as of 1998, over 160 mechanisms are reported for solid state polymorph transitions[13]. However, presently there is no definitive, verified model of the molecular mechanism of solid state polymorph transformations in molecular crystals. Two of the most common mechanisms proposed in the literature are 1) nucleation and growth[13-17] and 2) concerted, or martensitic, transformations[18, 19]. As Tuble *et al.* point out, however, there is increasing evidence that transformations thought to be martensitic in nature, actually occur via a nucleation and growth mechanism[13, 18]. An accurate model which describes this process on a molecular level would provide an improved understanding of the potentially localized mechanism of solid state polymorph transformations in molecular crystals, and could ultimately provide the necessary knowledge for *a priori* selection of processing parameters to promote or prevent transformation from one polymorph to another.

A specific molecular crystal that exhibits transformations between polymorphs is terephthalic acid, [p-C_6H_4-$(COOH)_2$], which has three known polymorphs, designated as forms I-III[20, 21]. Our work focuses on the transformation between forms I and II, both of which pack in triclinic lattices. In a paper that for the first time demonstrated structural polymorph stabilization via twinning, Davey *et al.*[22, 23] experimentally showed that form II is the more stable polymorph at room temperature and pressure. They also found that the transformation from form I to form II occurs in a temperature range from room temperature to 90 °C. The same authors noted that the transformation is often accompanied by a large release of mechanical energy, sufficient to make the crystal jump during the phase transition. In addition, from their microscopy studies, Davey *et al.*[23] conclude that the overall process appears martensitic in nature. Based on an examination of the crystal structures, the authors propose a chain slide and rotation mechanism for the single crystal transformation. It should be stressed, however, that the intent of the two studies by Davey *et al.* was not to determine the mechanism of transformation, but rather to determine the most stable single-crystal form and then to address the question of stabilization of the metastable form in industrial crystals. Their work forms the basis and inspiration for this investigation.

Determining pathways of polymorph transformations at the molecular level is difficult from an experimental standpoint, due to the difficulty in characterizing localized fluctuations in crystal structures. As a result, the majority of studies concerned with molecular crystalline systems typically characterize the equilibrium properties of polymorphs, rather than definitively ascertaining the molecular level events leading to transformations. Sophisticated molecular simulations, however, provide approaches that could potentially address these processes. These approaches have successfully addressed simpler problems. For example, nucleation has been studied for the freezing of a Lennard-Jones fluid[24], NaCl nucleation from solution[25], and the freezing of water[26]. The present study focuses on investigating the molecular level events leading to the transformation in TPA from form I to form II.

2.2 Overview

To determine the transformation mechanism for any process, it is necessary to determine the reaction coordinate, or the single variable that describes the system along the reaction pathway[20]. Knowledge of the reaction coordinate, and hence the mechanism, can provide essential molecular level insight for judicious engineering of complex systems.

To ascertain the reaction coordinate for any system, one must perform three steps:

1. Sample the configurations in the region in which the transformation of interest occurs.

2. Approximate the reaction coordinate as a function of physically relevant parameters.

3. Validate the reaction coordinate with committor probability analysis.

Standard molecular dynamics (MD) or Monte Carlo methods can readily sample regions of stability. However, when studying transitions of interest, standard methods are not widely applicable because the transition states are typically short lived and infrequently visited. Instead one must utilize techniques designed to sample transition state regions in a more efficient manner (Step 1). Transition path sampling [TPS], a trajectory space Monte Carlo procedure, is a powerful method for sampling the ensemble of transition pathways in complex systems[27-31]. A recent extension of TPS, aimless shooting[32], is especially useful for obtaining reaction coordinates with informatics approaches.

Once adequate sampling of the transition state region is completed, the reaction coordinate is approximated as a function of physically relevant parameters (Step 2). These parameters are termed order parameters (OP's), which describe properties of the system along reactive trajectories. Types of OP's range from bond distances[33] to parameters that measure the coordination numbers and orientations[25, 34, 35]. For a given system, many OP's will change "adiabatically" along the reactive trajectories. A reaction coordinate, however, is one or more OP's that completely quantify the dynamical progress of the system along the reaction pathway[31].

The typical method to determine the most appropriate reaction coordinate until recently was trial and error. Several recently published methods now allow one to determine the reaction coordinate in a more systematic manner[32, 36-38]. In this study, likelihood maximization is applied to screen candidate OP's to determine the OP or set of OP's that best approximates the reaction coordinate[32].

To validate the approximate reaction coordinate, one determines the average probability of reaching the product basin from the transition state region (Step 3). This method is known as committor probability analysis[31, 39-41]. If the true reaction coordinate is known, then firing

trajectories randomly from any configuration on the transition state isosurface should result in an equal number of trajectories that reach the reactant and product basins.

We apply this scheme to the solid state polymorph transformation in terephthalic acid (TPA). This chapter is organized into three sections: first, the details of the methodology are described. This includes the development and screening of the potential, the molecular simulation details, and the construction of the system sizes and shape. Details of the aimless shooting and likelihood maximization algorithms are also provided. Next, results are summarized for the harvesting of initial trajectories for two system sizes and from the likelihood maximization algorithm. Qualitative verification for the models obtained from likelihood maximization is given. Finally, a discussion and conclusions are provided highlighting the physical insight gained into the TPA polymorph transformation.

2.3 Methodology

This section explains the construction of the TPA crystals, and the development and validation of the CHARMM molecular force field. The order parameters that are screened as approximations to the reaction coordinate are described, as well as the computational methods of aimless shooting and likelihood maximization that are applied to the transformation of the TPA crystals. A new method is introduced to harvest independent configurations for verifying the reaction coordinate approximation, and the histogram method is also detailed.

2.3.1 System Description

Adjacent molecules in TPA are connected via 1-dimensional hydrogen-bonded carboxylic acid dimers, termed synthons[42]. This particular motif is ubiquitously found in molecular organic crystals. These chains, or supramolecular synthons, pack in 2-dimensional sheets. As discussed by Bailey and Brown[20], neighboring chains in the form I crystal lie with benzene rings adjacent to the carboxylic acid groups in the next chain; in form II, the carboxylic acid groups from neighboring chains are adjacent to one another. The layers of chains in form I, however, lie with the benzene rings in line, whereas for form II, the benzene rings pack alternately with the carboxylic acid groups.

Two system sizes are studied in full detail: a 6x6x6 and a 7x7x7 geometry. These sizes correspond to 216 and 343 molecules, respectively. Figures 2-1 and 2-2 illustrate the crystal

shapes for the 216 molecule system seen from the **b** and **a** axis views for forms I and II, respectively. The crystal shapes and morphologies approximate the TPA crystals studied by Davey *et al*[23]. It should be noted that 6x6x6 is counted by the number of molecules in the supramolecular synthon, not the crystal unit cells: e.g., note that in Figure 2-1, the chains for both forms are 6 molecules long and there are 6 sets of chains in both chain-perpendicular dimensions. In addition, all molecules are hydrogen bonded through the carboxylic acid group to at least one neighbor. The experimental lattice parameters for the crystals are given in Table 2-1. The lattice vector **a** is equivalent for the two structures as it is the measure of the supramolecular chain length.

The crystal configurations are adopted from structures reported by Bailey and Brown. In the original crystal structure determination, the hydrogen positions are not reported for form II. Hydrogen atoms are therefore added to the molecules with the corresponding bond lengths and angles found in form I. As no constraints (such as SHAKE[43]) are used either in minimization or molecular dynamics, any minor errors in the hydrogen positions leading to unfavorable configurations are eliminated.

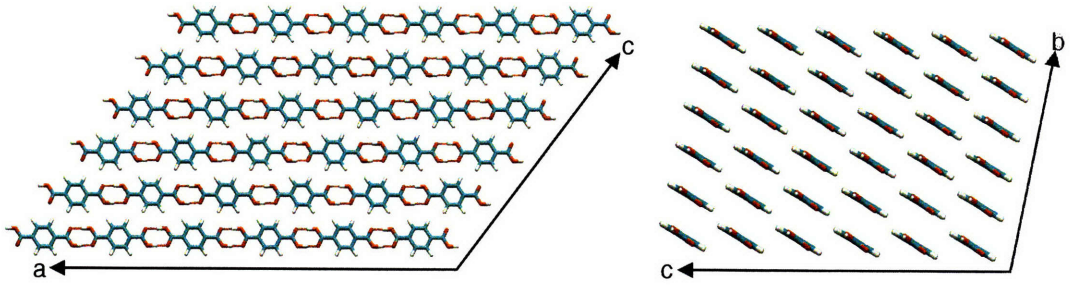


Figure 2-1. Form I of the 216 molecule (6x6x6) TPA crystallite from the b-axis view (left) and a-axis views (right). Hydrogen bonds that form the supramolecular synthons are shown in red-dotted lines between the carboxylic acid groups on each molecule.

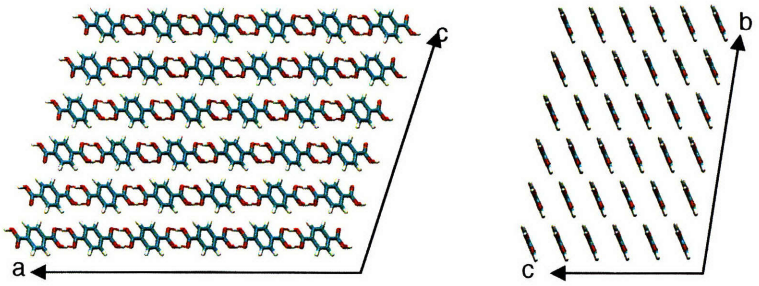


Figure 2-2. Form II of the 216 molecule (6x6x6) TPA crystallite from the b-axis view (left) and a-axis views (right). Hydrogen bonds that form the supramolecular synthons are shown in red-dotted lines between the carboxylic acid groups on each molecule.

Table 2-1. Lattice parameters from the experimental crystal structures.

	Form I	Form II
a	9.54	9.54
b	3.19	5.02
c	6.44	5.34
α	87.25	86.95
β	126.27	104.9
γ	107.36	134.65

2.3.2 Force Field Development

Empirical force fields may not be directly suitable for modeling molecular crystals, especially if the parameters are fit using scenarios that are incongruous with crystal packing[18, 44, 45]. A modified CHARMM potential[46] is thus applied in the simulations. Bond, angle, dihedral, and Lennard-Jones parameters (with the exception of one) are taken from the CHARMM 27 force field library. Since original Lennard-Jones parameters for the phenyl hydrogens are fit to hydration energy, which makes them unsuitable to model a packed crystal, a range of phenyl hydrogen Lennard-Jones radii is tested. In addition, partial atomic charges are extracted from single point energy calculations on a single gas phase TPA molecule with Gaussian 03[47]. The B3PW91 density functional[48] with 6-311G++** basis set is employed. Following the approach of Tuble *et al.*[18], partial atomic charges are calculated with both the Merz-Singh-Kollman[49, 50] (MSK) and ChelpG[51] schemes. Since the intramolecular geometry does not differ significantly between forms I and II, there is little difference in the partial charges calculated between molecules corresponding to the two polymorphs.

As the simulations are performed in vacuum, the potential is screened for the ability to reproduce the experimental form I and form II crystal structures in vacuum, as measured by the root mean square deviation (RMSD) of the minimized crystal from the experimental crystal structures. Figure 2-3 shows the RMSD as a function of the phenyl hydrogen Lennard-Jones radius for the MSK partial charge set. Although the Lennard-Jones radius for the phenyl hydrogens does not significantly affect the RMSD for form II up to 2.0 Å, a value of 1.0 Å is selected as it yields the smallest RMSD for form I. The use of either MSK or ChelpG does not result in significant differences in the RMSD at a fixed phenyl hydrogen Lennard-Jones radius. The charges in the potential are those from the MSK method. Figure 2-4 shows the atom name assignments and Table 2-2 lists the charge for each molecule.

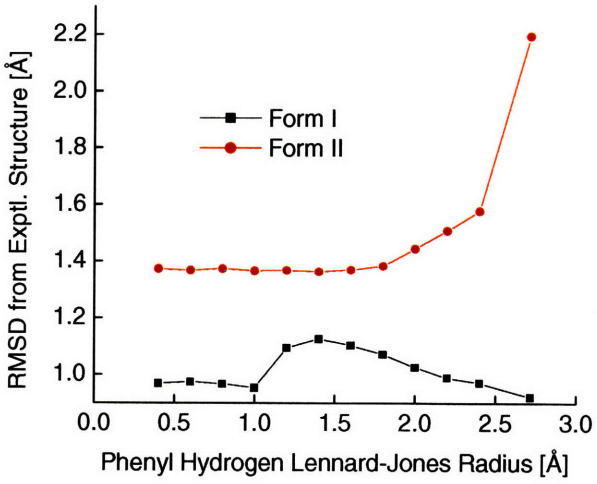


Figure 2-3. RMSD as a function of the Lennard-Jones phenyl hydrogen radius for Form I (■) and Form II (●). Calculated with the Merz-Singh-Kollman[48,49] partial atomic charge set.

Table 2-2. Partial atomic charges from a single point energy calculation on a single gas phase molecule. Partial atomic charges are extracted using the MSK scheme.

Atom name	Partial atomic charge [e]	Atom name	Partial atomic charge [e]
C1	0.0701	C3A	-0.1919
C2	-0.1515	C2A	-0.1541
C3	-0.2009	C1A	0.0823
C4	0.6462	H3A	0.1704
H1	0.4391	H2A	0.1482
H2	0.1463	C4A	0.6307
H3	0.1764	O1A	-0.5582
O1	-0.5568	O2A	-0.5665
O2	-0.5722	H1A	0.4424

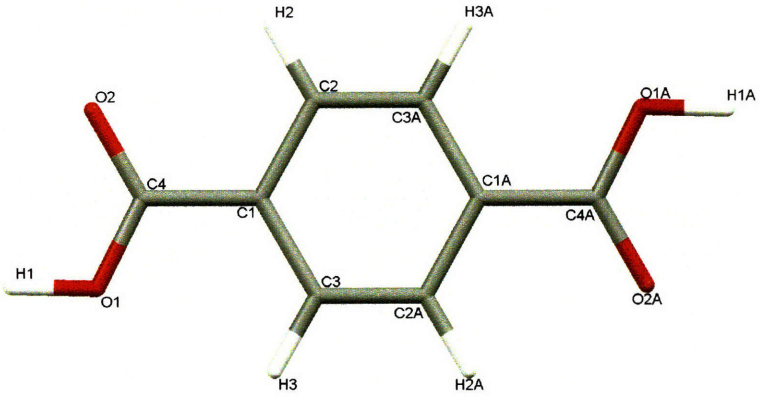


Figure 2-4. Single terephthalic acid molecule with atom labels

2.3.3 Order Parameters

Over 100 OP's are tested in the simulations. The selection of trial OP's is typically an *ad hoc* procedure for which there is no known systematic approach. For a study of polymorph transformations, perhaps the most obvious set of trial OP's is the lattice parameters, which warrant discussion here.

The lattice vectors, **a**, **b**, and **c**, are measured by averaging over the distances of the molecules at the crystal edges. For instance, from Figure 2-2, one measures the **c** lattice parameter by measuring the distance from the molecule at the top left to the top right. This is done for each

molecule pair down the length of the chain. Therefore, for each horizontal layer (as shown in Figure 2-2), one can collect 6 distances for the 216 molecule system and 7 for the 343 molecule system. This procedure is repeated for the second layer and so on until the bottom layer is reached. The value for each lattice parameter is the average over the entire system. Using the measurement method employed in this study, differences, if any exist, in the local lattice parameters through the transformation will aid in the determination of the nature of the transformation. For example, in the case of a nucleation mechanism, a change on one side of the crystal will give rise to large non-uniformities in the local lattice parameter values, say for the horizontal layers in Figure 2-2.

Other OP's are tested, such as the lattice angles, local twisting along the length of the supramolecular chains, supramolecular chain twisting relative to other chains in the crystal, and the variation in the directions of the supramolecular chains.

2.3.4 Aimless Shooting

The aimless shooting algorithm, described by Peters and Trout, is applied to harvest an ensemble of independent reactive trajectories. As with the TPS method[27-31], aimless shooting requires:

1) accurate definitions of the basins of stability and

2) an initial trajectory that connects the stable basins.

Quantitative definitions of the basins are constructed based on the fluctuations of lattice parameters taken from equilibrated MD trajectories kinetically trapped in the respective basins. All basin definitions are taken as the average value of the **b** and **c** lattice vectors ± three standard deviations. The **a** lattice parameter is not included as it does not change significantly between the two forms.

For the 216 molecule system, trajectories initiated in the form I basin are found to transform on the order of 1 ns to either form II or to an intermediate basin "between" the two forms, as measured by the lattice parameters. These trajectories are used for the selection of the initial aimless shooting points. For the 343 molecule system, trajectories initiated in the form I basin are not found to transform either to the intermediate form or to the form II polymorph on the order of several ns. Therefore, MD umbrella sampling is applied following the method

of Kottalam and Case[52] to obtain a set of configurations connecting the stable basins. The ratio of the **b** to **c** lattice vectors is selected as the OP to sample from the form I basin (**b/c** = 1.5-1.7) to the form II basin (**b/c** = 1.0-1.1). A harmonic umbrella potential with a spring constant of 500 kcal/mol is applied with 8 windows along the OP, starting with the form I polymorph. The sampling time for each window is 1.05 ns starting from the endpoint of the previous window and included 50 ps at the start of each new window prior to saving configurations. As in the case of the 216 molecule system, the obtained trajectory is used for the selection of the initial aimless shooting points.

Starting from the well characterized basins, the aimless shooting method is applied. The algorithm contains only one adjustable parameter: time displacement, Δt, to shift along the initial trajectory to generate the new shooting points. The time displacement is carefully selected for optimal efficiency since if Δt is chosen too high, the algorithm will wander too far away from the transition state region leading to a low acceptance rate; if chosen too small, the aimless shooting algorithm will search a smaller amount of shooting point configuration space requiring more trajectories to obtain a good approximation to the reaction coordinate. A time displacement of 1% of the entire reactive trajectory length is found to be adequate to sample the transition state region, giving an acceptance rate in general between 40 and 60%.

Dynamic trajectories are collected in vacuum using the CHARMM package[46] at 300 K in the *NVE* ensemble. For a system this large, temperature fluctuations in the *NVE* ensemble will be on the order of 1% (kinetic energy fluctuations are proportional to 1/root(N), where N = degrees of freedom) and will result in only a small perturbation from the NVT ensemble. A timestep of 1 fs is used with a cutoff for non-bonded interactions of 14 Å. The aimless shooting algorithm originally described in Peters and Trout is amended slightly for this study. The selection of shooting points was originally from 3 points: $\mathbf{x}_{-\Delta t}$, $\mathbf{x}_0$, $\mathbf{x}_{+\Delta t}$. We found selecting from two points, $\mathbf{x}_{-\Delta t}$, or $\mathbf{x}_{+\Delta t}$ is sufficient to sample the transition state ensemble and we therefore used this two-point version of aimless shooting.

Trajectories are initiated from points thought to be close to the transition state region along the initial trajectories. In the interest of efficiency, the trajectory length is set as short as possible. For both the 216 and 343 molecule systems, a total length of 30 ps is sufficient to maintain the level of inconclusive trajectories at or below 15%. A time displacement, Δt, of

300 fs is selected to obtain the desired acceptance rate. Approximately 4000 paths for the 216 molecule system and 3500 paths for the 343 molecule system are collected.

2.3.5 Likelihood Maximization

As described in Peters and Trout[32], the reaction coordinate, r, is modeled as a linear combination of candidate OP's, denoted as $\mathbf{q}$, with α_0 through α_m as adjustable coefficients:

$$r(\mathrm{q}) = \alpha_0 + \sum_{k=1}^{m} \alpha_k q_k \tag{2-1}$$

The reaction coordinate is related to the probability of being on a transition path at some value of the reaction coordinate, r, through the following model[32]:

$$p(\mathrm{TP}|r) = p_0(1 - \tanh^2[r]) \tag{2-2}$$

As pointed out by Peters and Trout[32], this model function exhibits a maximum at r equals zero, which corresponds to the transition state, and decays to zero on either side of the peak. Maximizing the likelihood function,

$$L(\boldsymbol{\alpha}, p_0) = \prod_{k}^{N_{\mathrm{acc}}} p(\mathrm{TP}|r(q_{\mathrm{acc}}^{(k)})) \prod_{k}^{N_{\mathrm{rej}}} (1 - p(\mathrm{TP}|r(q_{\mathrm{rej}}^{(k)}))) \tag{2-3}$$

over all coefficients and all combinations of OP's determines the best reaction coordinate according to the models of Equations 2-1 and 2-2. For the best approximate reaction coordinate, the approximate transition state isosurface can be obtained by maximizing $p(\mathrm{TP}|r)$. This occurs at $r = 0$, so setting $r(\mathbf{q}) = 0$ defines the transition state isosurface.

2.3.6 Reaction Coordinate Validation

As mentioned in the Overview Section, one must determine if the reaction coordinate approximation is correct. This can be done by approximating the probability of reaching the reactant (p_A) or product (p_B) state from the predicted transition state region. This technique is typically referred to in the literature as a committor distribution analysis[31]. One can construct a committor distribution of p_B values (referred to as a p_B histogram) by firing randomly seeded trajectories from the predicted reaction coordinate isosurface, given by setting $r(\mathbf{q}) = 0$ in Equation 2-1[31, 32].

Constructing a p_B histogram requires many configurations from which to shoot. The generation of these configurations is inspired by the BOLAS method[53]. Several random, aimless shooting points are selected close to the predicted transition state region, as defined by $r(\mathbf{q}) = 0$ in Equation 2-1. Very short trajectories are fired randomly from each initial configuration and the endpoints are evaluated to determine if they are within a narrow window on the transition state isosurface. If so, this configuration is accepted and becomes the next shooting point. This process is repeated until an adequate number of configurations is generated from which to shoot reactive trajectories to build a p_B histogram. A collection of points at $r(\mathbf{q})=0$ is thus generated and these points are used to perform the committor probability analysis. It should be noted that BOLAS samples the equilibrium distribution of trajectories within an OP window regardless of the accuracy of the putative reaction coordinate obtained from likelihood maximization. Therefore, the collected points will be extracted from the equilibrium distribution, but if either or both the transition state isosurface and the reaction coordinate are incorrect, the resulting histogram will not be peaked at ½.

To construct the histogram, trajectories are shot from each configuration with a length corresponding to half the length of a reactive trajectory. The endpoints of the trajectories are evaluated and a histogram is constructed of the probability of reaching basin B from the predicted transition state isosurface. The basin definitions for constructing the p_B histograms correspond to the same basin definitions used for the reactant and product basins in the aimless shooting simulations. An adequate approximation to the true reaction coordinate will yield a histogram that is sharply peaked at $p_B = 0.5$[31]. Additionally, one can make a quantitative comparison of the histogram to the binomial distribution, which will have a mean value, $\mu = 0.50$ with a standard deviation, $\sigma = 0.11$.The trajectories for the generation of new configurations are 100 fs long and the endpoint window width at $r = 0$ is constrained within a range of ±1% of the total configuration space sampled, as measured by Δr. For each histogram assembled in this study, approximately 250 shooting points are collected. From each configuration collected, 20 trajectories are shot, corresponding to approximately 5000 trajectories for each histogram. The trajectory length for calculating p_B values is 15 ps, which is half the length of the reactive trajectories in the aimless shooting simulations, again resulting in a low rate of inconclusive paths.

2.4 Results

This section describes the results obtained for the initial trajectories for the 216 and 343 molecule systems. In addition, details of the aimless shooting and likelihood maximization scheme employed here are provided. The free energy curves are discussed in light of the findings from the initial trajectories.

2.4.1 Initial Trajectories

As mentioned previously, for the 216 molecule system, trajectories initiated in the form I basin transform on the order of 1 ns to either form II or to an intermediate basin as measured by the lattice parameters. Figure 2-5 shows the average lattice parameters for a trajectory of each type: one of which becomes kinetically trapped in the intermediate basin (Figure 2-5a) and one of which does not (Figure 2-5b). For reference, the average lattice parameters for a representative trajectory initiated in the form II basin are also shown on Figure 2-5. Snapshots at germane points during these trajectories are shown in Figures 2-6 and 2-7, respectively. As seen in Figures 2-5b (for the lattice parameters) and 2-7 (for the molecular configurations), excellent agreement is found between the end state of the unbiased trajectory and the form II crystal structure. The second and third snapshots of the kinetically trapped trajectory, shown in Figure 2-6, exhibit characteristics of the form I and form II crystal as most clearly seen in the **a**-axis view. The top left corner of the crystal is the location in which nucleation occurs and the bottom right corner retains the form II orientation. The second snapshot (at 0.60 ns) from Figure 2-7 also displays this type of interface between the polymorphs, but quickly transforms completely to the form II polymorph.

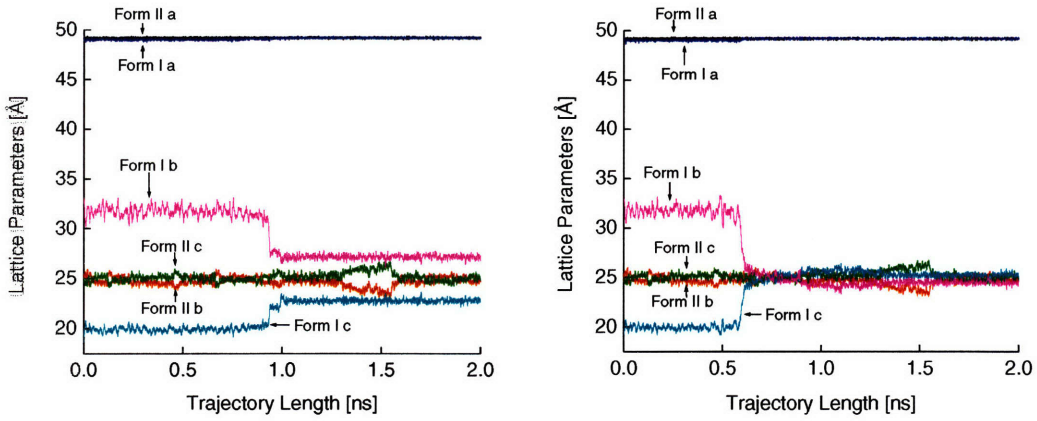


Figure 2-5. Crystal lattice parameters along a 2 ns trajectory for forms I and II of the 216 molecule system. (a) The change in the lattice parameters at approximately 1.0 ns corresponds to the unbiased nucleation event along the edge of the TPA crystal. The growth phase, however, is not seen in this trajectory as the system becomes kinetically trapped between the nucleation and growth events. (b) The change in the lattice parameters at approximately 0.6 ns corresponds to the unbiased form I to form II polymorph transformation. Excellent agreement is seen between the original form II crystal and the lattice vectors after 0.6 ns for the form I trajectory.

To confirm the behavior seen in the initial trajectories as illustrated by Figures 2-4 through 2-6, MD umbrella sampling is applied to the transformation in the 216 molecule system. The **b** lattice parameter is selected as the OP over which to sample. Figure 2-8 shows that the potential of mean force (PMF) curve for this simulation. The free energy exhibits both a barrier to nucleation and a barrier to growth, each approximately 3-4 kcal/mol, as explained below.

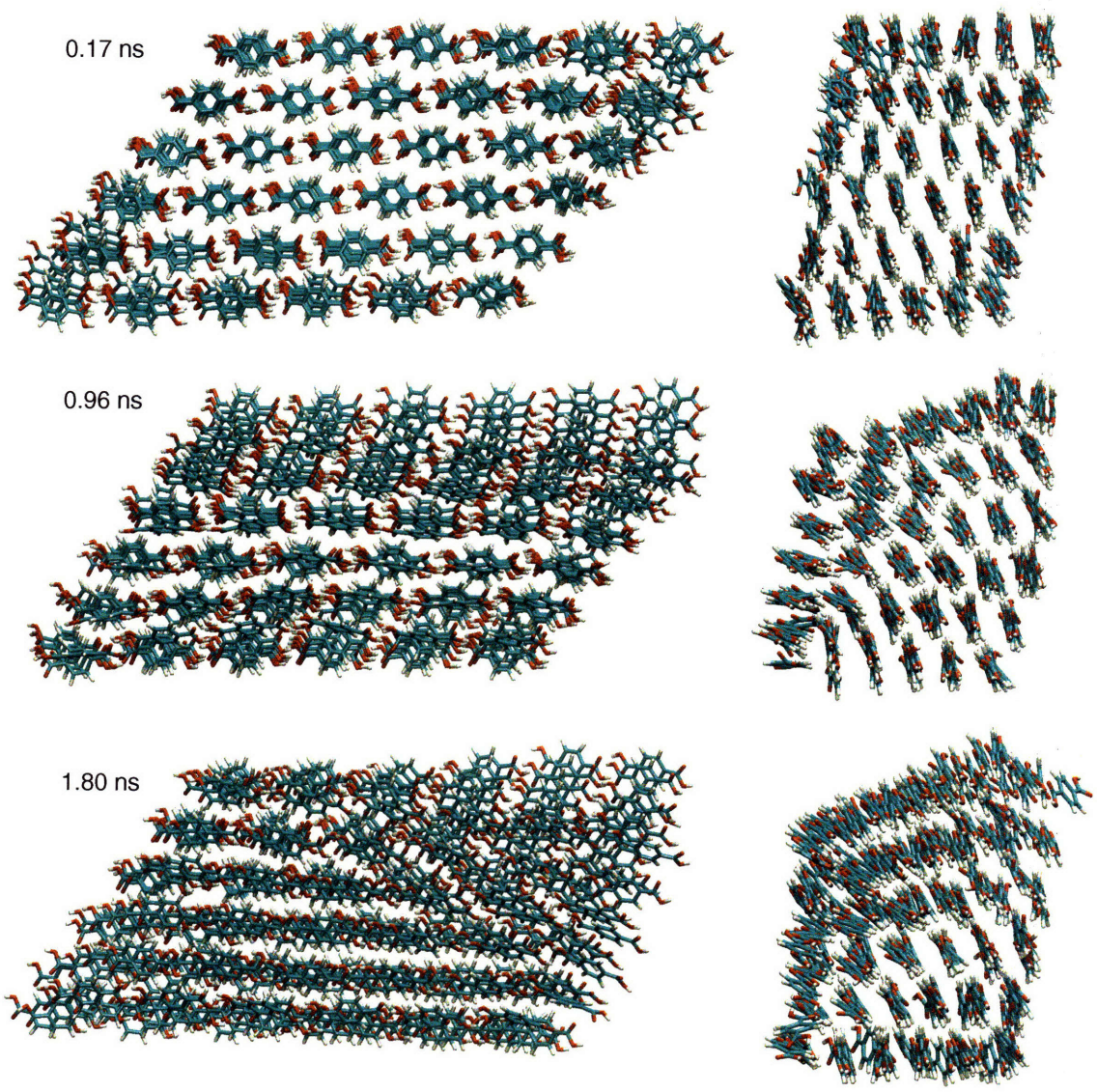


Figure 2-6. Snapshots from the b-axis and a-axis view along the initial trajectory for the 216 molecule system. These configurations correspond to the lattice parameters shown in Figure 4a, which becomes kinetically trapped after the nucleation event. From top to bottom, the configurations are taken at 0.17 ns (pre-transformation), 0.96 ns (during nucleation), and 1.8 ns (post-nucleation, when the crystal is kinetically trapped). Note that the nucleation event takes place at the top left corner (as seen from the a-axis view) and the molecules in the bottom right corner (again, on the a-axis view) retain their original orientation characteristic of form I.

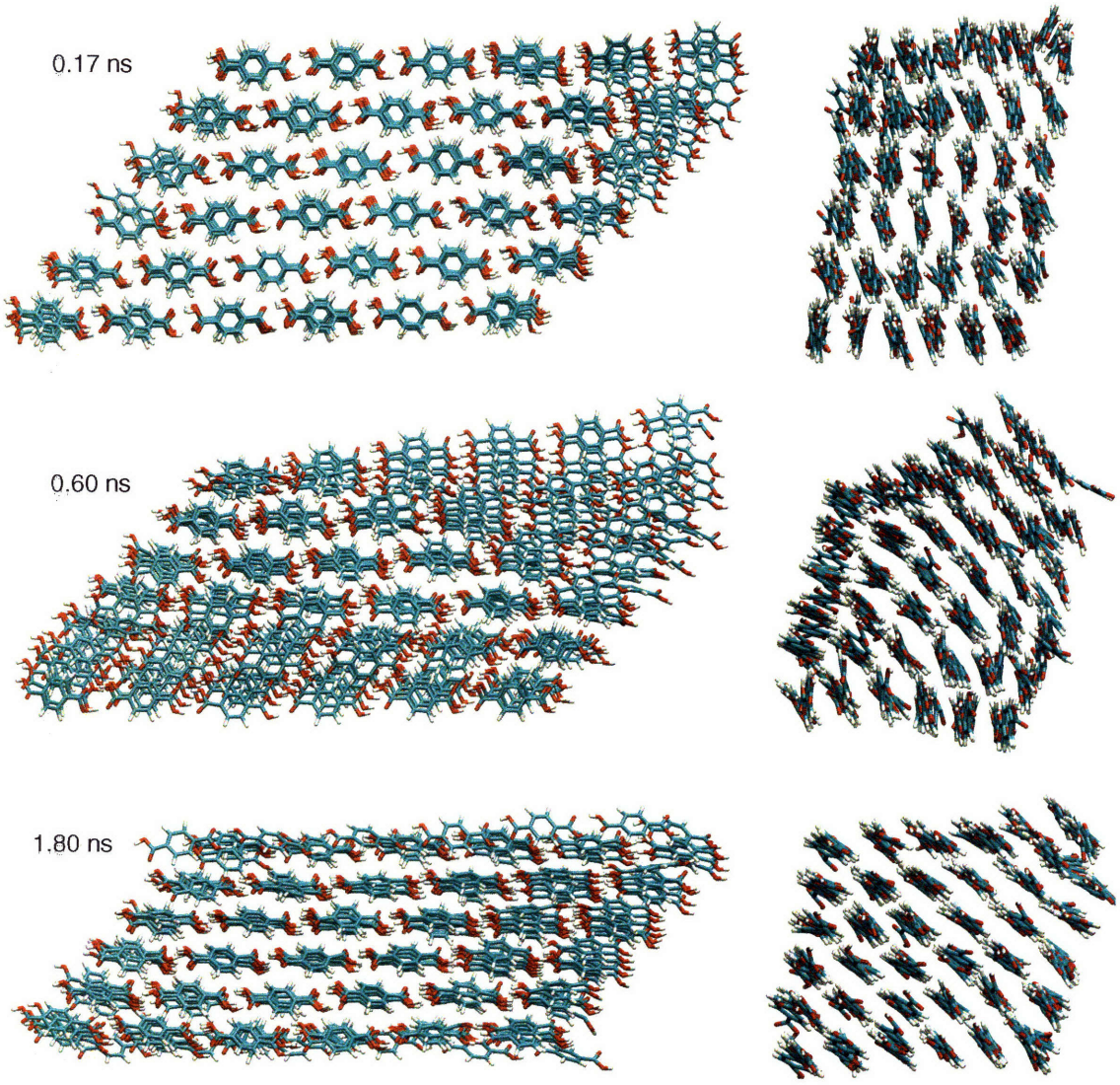


Figure 2-7. Snapshots from the b-axis and a-axis views along the initial trajectory for the 216 molecule system. These configurations correspond to the lattice parameters shown in Figure 4b, which undergoes both nucleation and growth. From top to bottom, the configurations are taken at 0.17 ns (pre-transformation), 0.60 ns (at the onset of nucleation and growth), and at 1.8 ns (in the form II basin). Note that the nucleation event takes place at the top left corner (as seen from the a-axis view) and propagates to the bottom right corner (again, on the a-axis view).

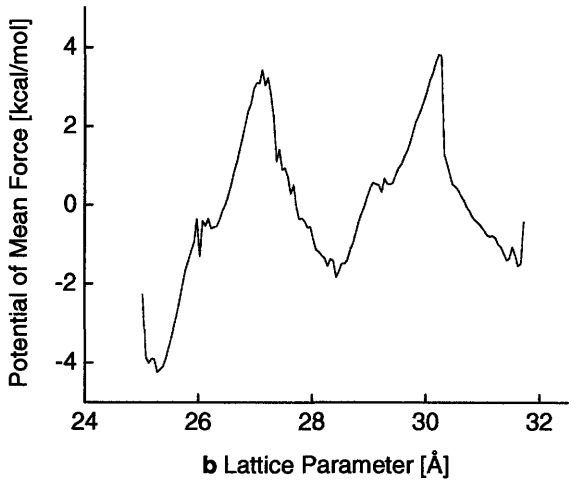


Figure 2-8. Potential of mean force for the 216 molecule TPA system calculated from 8 windows of MD umbrella sampling along the b lattice parameter.

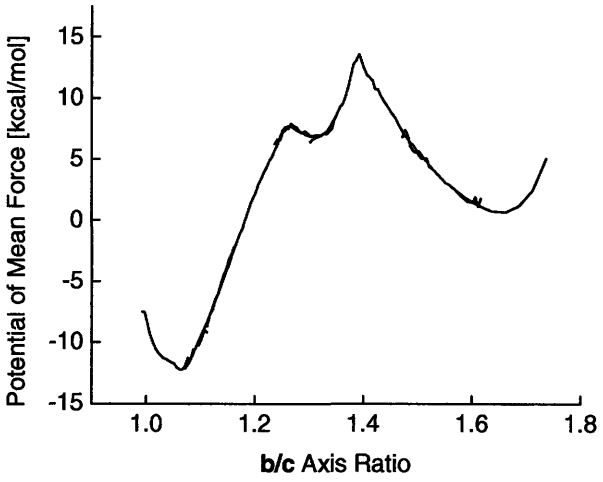


Figure 2-9. Potential of mean force for the 343 molecule TPA system calculated from 8 windows of MD umbrella sampling along the b/c axis ratio.

As discussed in the Methodology section, MD umbrella sampling is used to obtain initial configurations for the transformation in the 343 molecule system and is applied to both system sizes to determine the free energy barrier to nucleation and growth. Figures 2-8 and 2-9 show the results for the potential of mean force (PMF) for the 216 and 343 molecule systems, respectively. The PMF curve for the 343 molecule system confirms the behavior observed in the 216 molecule system with a metastable intermediate between the two polymorphs. Sampling from the form I polymorph to form II (i.e. from right to left on Figures 2-8 and 2-9), the free energy barrier to nucleation is approximately 12-13 kcal/mol

and to growth is approximately 3-4 kcal/mol. Figure 2-10 displays several snapshots of the system along the **b/c** coordinate. Similar behavior is seen in the larger system: a nucleation event occurs at a corner, an interface is formed between sections that are characteristic of either polymorph, and finally a growth phase occurs to complete the transformation. An examination of the final state obtained from MD umbrella sampling exhibits excellent agreement with the form II polymorph, as shown in the last snapshot in Figure 2-10. In both cases, force constants and window centers are chosen to maximize overlap between windows. For the 216 MD umbrella sampling, a force constant of 5 kcal/mol/Å^2 is used with 1.05 ns per window with 50 ps of equilibration, just as in the 343 MD umbrella sampling simulations. In addition, PMF curves in the crystal basins corresponding to the experimental crystal structures are collected without applying a bias.

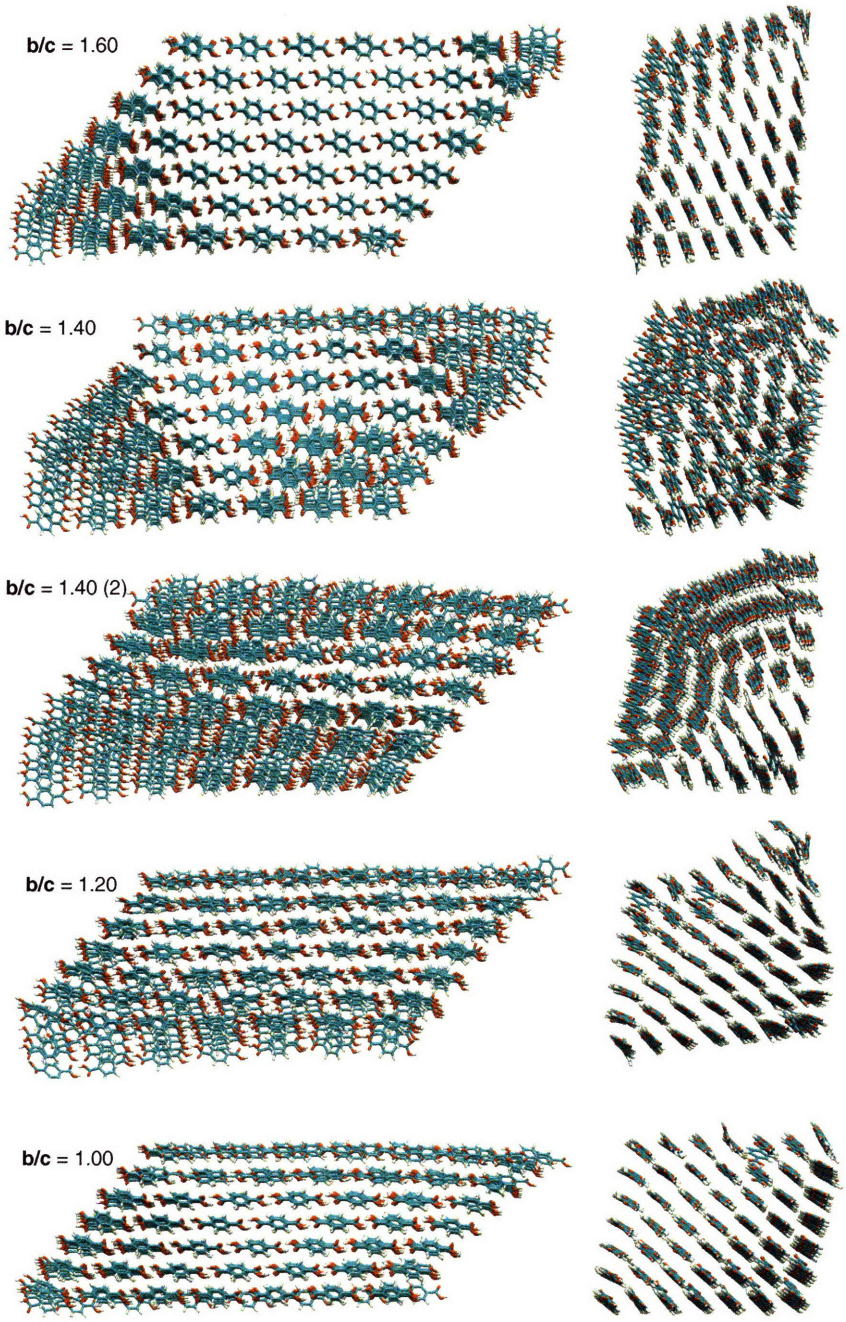


Figure 2-10. Snapshots from the b-axis and a-axis views along the initial trajectory for the 343 molecule system obtained from MD umbrella sampling. These configurations correspond to the PMF curve shown in Figure 8. From top to bottom, the configurations are taken at b/c=1.60 (pre-transformation), b/c = 1.40 (at the onset of nucleation), b/c = 1.40 (as nucleation is occurring at the peak of the PMF curve), b/c = 1.20 (as growth is occurring), and at b/c = 1.00 (in the form II basin). Note that the nucleation event takes place at the top left corner (as seen from the c-axis view) and propagates to the bottom right corner (again, on the c-axis view).

2.4.2 Aimless Shooting

As previously stated, aimless shooting requires accurate definitions of the basins of stability, or the reactant and product. It should be noted that stable intermediates along a proposed reaction coordinate can give rise to a large amount of inconclusive trajectories, i.e. trajectories that reach neither basin in the pre-specified trajectory length. In the case of TPA, the harvesting of initial trajectories between the basins confirms the existence of a metastable intermediate for both system sizes studied. From an examination of the configurations corresponding to the intermediate form, characteristics of both crystal forms appear. Visualization of the trajectories, as shown in Figures 2-6, 2-7 and 2-10, reveals that a nucleation event occurs along the edge of the crystal. Trajectories that become kinetically trapped in the intermediate basin are those in which the nucleation event is not directly followed by propagation throughout the remainder of the crystal, but rather those in which there exists an interface between the two forms. Since the primary interest for this study is to determine the mechanism for nucleation, the intermediate basin is treated as the "product" state. Aimless shooting is therefore conducted between the form I basin and the intermediate state, with the form II basin lumped into the intermediate state. The selected aimless shooting points from the initial trajectories are those that are at a configuration with a **b/c** axis ratio of approximately 1.4. This value corresponds to the first peak in the free energy barrier from the umbrella sampling simulations, which is most likely near the transition state region (in 1 dimension only) for the nucleation event.

2.4.3 Likelihood Maximization

Over 100 OP's are screened to determine the best fit to the reaction coordinate. As mentioned previously, the reaction coordinate is approximated using Equation 2-2. The likelihood maximization results for the 216 and 343 molecule systems are summarized in Table 2-5 for a 1-dimensional model.

From Table 2-5, one can see that the global **b** lattice vector is the most important parameter in the reaction coordinate approximation for the 216 molecule system. For the larger system size studied, however, the ratio of the relevant global lattice vectors receives the highest likelihood score. It should be noted that the two average lattice vectors, **b** and **c**, and the **b/c** ratio are the best three single variable approximations to the reaction coordinate for both system sizes.

Table 2-5. Likelihood maximization approximations to the 1-dimensional reaction coordinate for 4000 aimless shooting paths for the 216 molecule TPA system and 3500 aimless shooting paths for the 343 molecule TPA system[a].

System Size	OPs	p_0	α_0	α_1	Transition State Isosurface (r=0)
216	**b**	0.57	-0.99	2.06	$\mathbf{b}^{\ddagger}$ = 30.195 Å
343	**b/c**	0.73	-1.43	2.83	$\mathbf{b/c}^{\ddagger}$ = 1.433

[a] The $p(\mathrm{TP}|r)$ model given in Equation 2-2 is used to calculate the likelihood function as shown in Equation 3. The corresponding models for the 1-dimensional reaction coordinate approximations are shown for both systems. The OPs in the expression for r are provided on a normalized basis such that $q_i \in [0, 1]$.

2.4.4 Reaction Coordinate Validation

For the 216 molecule system, a histogram is constructed from predicted transition state region for 1 OP: $\mathbf{b}^{\ddagger}$ = 30.195 Å ± 0.014 Å, which corresponds to a p_B window centered at 0.5 with a total width of 0.01. Figure 2-11a shows the p_B histogram for this system. This histogram indicates that the **b** lattice parameter is a good approximation of the reaction coordinate. However, since the histogram is peaked at 0.629, the system is more likely to go to the product basin at this value of $\mathbf{b}^{\ddagger}$ denoting an error in the prediction of the transition state isosurface. The value for $\mathbf{b}^{\ddagger}$ is therefore shifted to 30.400 Å to generate a second histogram with an equivalent number of trajectories, shown in Figure 2-11b. This histogram is sharply peaked near $p_B = 0.5$, which denotes that it is a good approximation to the reaction coordinate and that the value of the transition state isosurface is correct, as desired. The mean value for this histogram is $\mu_h = 0.490$ with a standard deviation, $\sigma_h = 0.144$.

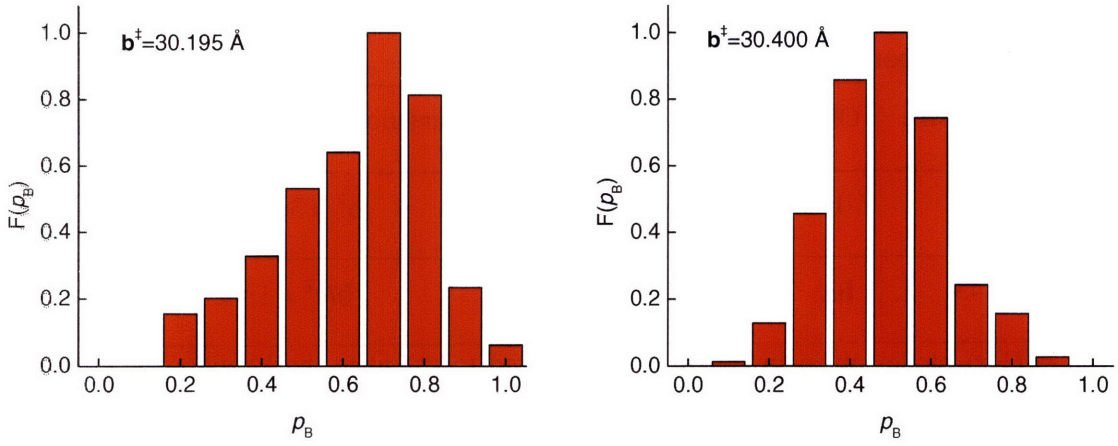


Figure 2-11. (a) Committor probability histogram for 216 molecule system with $b^{\ddagger}$ = 30.195 Å, as predicted from the $p(TP|r)$ model with a 1-D reaction coordinate model. μ_h = 0.629, σ_h = 0.182. (b) Committor probability histogram for 216 molecule system with $b^{\ddagger}$ = 30.400 Å. μ_h = 0.490, σ_h = 0.144.

For the 343 molecule system, a histogram is also constructed from the 1-dimensional reaction coordinate approximation: $b/c^{\ddagger}$ = 1.4327 ± 0.0006, as shown in Figure 2-12. This histogram is broader than the histogram compiled for the smaller system, but nonetheless is peaked at p_B = 0.5, indicating that the 1-dimensional model is qualitatively a valid approximation to the reaction coordinate for this system size. The mean value for this histogram is μ_h = 0.520 with a standard deviation, σ_h = 0.208.

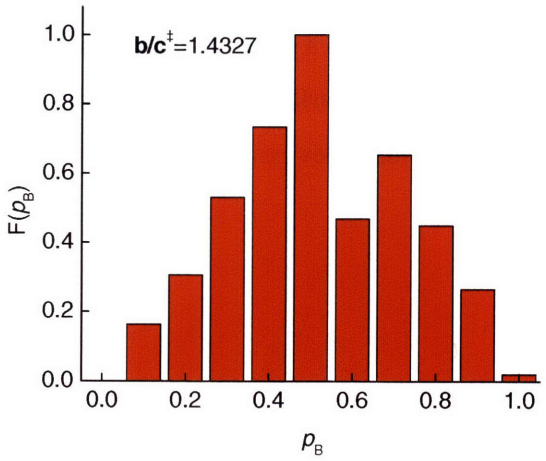


Figure 2-12. Committor probability histogram for 343 molecule system with $b/c^{\ddagger}$ = 1.4333, as predicted from the $p(TP|r)$ model with a 1-D reaction coordinate model. μ_h = 0.520, σ_h = 0.208 .

2.5 Discussion

The unbiased trajectory, as shown in Figures 2-5b and 2-7, simulated with the adapted force field, connects forms I and II for the 216 molecule system with excellent agreement in the lattice parameters. The trajectory that becomes kinetically trapped, as highlighted in Figures 2-5a and 2-7 however, offers the desired snapshot solely of the edgewise nucleation event in the TPA crystal. The fluctuations in the chains at the edge of the crystal, which have the greatest freedom to move, drive the nucleation event for both system sizes studied. This behavior is verified with the p_B histograms shown in Figures 2-11 and 2-12 for both system sizes studied. The second histogram for the 216 molecule system, shown in Figure 2-11b, is constructed based on shifting the value of the transition state isosurface from the original value of 30.195 Å, predicted by likelihood maximization. The inaccuracy in the predicted value may arise from an improper fitting function for likelihood maximization, as will be investigated in the next chapter.

It should be noted that all the trajectories from aimless shooting exhibit a similar transformation mechanism along the same crystal edge, which indicates that the ensemble of trajectories is seemingly equilibrated. In a related matter, as the crystals are symmetric, there is an equivalent crystal edge along which nucleation may occur. However, it was observed from the aimless shooting trials that nucleation only occurs on the edge where nucleation occurred in the initial trajectory. Therefore, the mechanisms, which are most likely identical, are seemingly separated by several kT, making nucleation on the opposite crystal edge not accessible during aimless shooting.

As mentioned in the Methodology Section, it is shown how the OP for the lattice parameters is condensed into either a single global average lattice parameter or separated by layers of the supramolecular synthons. Although the reaction coordinate is captured with the global average lattice parameter (or the ratio of the global lattice parameters), in the case of a nucleation mechanism, non-uniformity should be apparent in the local lattice parameters measured on each supramolecular synthon layer. For example, Figure 2-13 shows this behavior for the **c** lattice vector for the 216 molecule system. Figure 2-13 corresponds to the trajectory that undergoes the full polymorph transformation as shown in Figures 2-5 and 2-7. The outside layers (1 and 6) exhibit significant non-uniformity as do the next two layers (2 and 5), but to a lesser extent. This provides quantitative evidence that the mechanism in the

TPA system is nucleation and growth. It should be noted that if the crystal were significantly larger, a global average lattice parameter that spans the entire crystal may not be the best approximation to the reaction coordinate, as will be investigated in a subsequent chapter. This is because the nucleation, even as seen in these system sizes, is a localized event in that it occurs along a specific edge of the crystal. Because of the sizes used in this study, the global lattice parameter is still able to capture the local changes on a single edge. However, the aimless shooting and likelihood maximization technique can be applied to study the polymorph transformations for any system size using any OP, as desired and deemed computationally feasible.

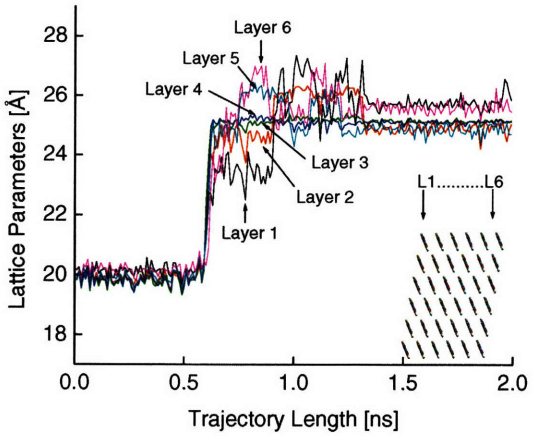


Figure 2-13. Change in the local c lattice parameter for each layer in the polymorph transformation of TPA. The differences in the local OP denote that there is a localized event occurring during the transition, namely edgewise nucleation. This corresponds to the transformation seen in Figure 5b, in which both the nucleation and growth events occur. The inset shows the local lattice parameter measurements from Layer 1 to Layer 6 along the crystal.

The shape of the nucleus, formed by synthons along the edge of the crystal for the TPA system, is starkly different from the nucleus that behaves as a mean field as given by classical nucleation theory[54]. This prediction of elongated nuclei is qualitatively reasonable as the hydrogen bonds formed in the direction of the supramolecular synthons are much stronger than the van der Waals interactions in the other two directions.

Comparison of Figures 2-8 and 2-9 reveals that the free energy barrier depends on the length of the supramolecular synthon. The 12-13 kcal/mol free energy barrier in the 343 molecule system is an inaccessible barrier to surmount in a direct MD simulation, as is demonstrated. The free energy barrier for this type of nucleation mechanism is expected to scale for a perfect

crystal as the transformation for the 2 system sizes studied requires the deformation of the entire edge of the crystal. If it were computationally feasible to simulate much larger TPA crystals at present, such that the surface area to volume ratio more closely matches the experimental value, there will most likely be a synthon length at which the hydrogen bonds would be broken during the nucleation event. However, if this edgewise nucleation occurs along a very long supramolecular chain and the barrier to the growth process remains small (order of a few kcal/mol, as seen in both Figures 2-8 and 2-9), the growth process would result in a large, sudden energy change, which can explain the "jumping" behavior seen in the experiments conducted by Davey *et al.*. An order of magnitude analysis from the present simulations is possible to bolster this hypothesis. For example, in the 216 molecule system, the major lattice parameter expansion occurs over 0.5 nm in approximately 30 ps, which corresponds to a transformation velocity of approximately 8 m/s. The total possible kinetic energy for this system is therefore on the order of 0.3 kcal/mol. As shown in Figure 2-13, the free energy barrier is on the order of 3-4 kcal/mol, which is greater than the estimated kinetic energy, so if only a small fraction of the energy is released as kinetic energy, the jumping should be observed. This order of magnitude estimate illustrates that the nucleation mechanism along the predicted crystal edge with fast growth kinetics is sufficient to explain the observed jumping behavior of the crystal.

It is also evident from both trajectories that an interface forms between the two polymorphs during the nucleation event. This observation reaffirms that this transformation initiates via localized, surface mediated nucleation. As previously mentioned, however, a distinct interface is not observed in the crystal during the phase transition as reported by Davey *et al.*, even though the simulations at this small scale clearly show a phase boundary. This absence of a visible boundary between the two polymorphs forms as reported by Davey *et al.* could be explained by the extremely fast growth process, which is not directly treated in this study. A plausible explanation for the disparity between experiment and simulation is most likely that this is a nucleation controlled process and the growth phase follows the nucleation rapidly as confirmed by the free energy barriers measured in this work.

2.6 Summary and Conclusions

In this study, the initiation event leading to the transformation in the solid state polymorph transformation of terephthalic acid is shown to be nucleation. Specifically, the polymorph

transformation from form I to form II proceeds via a surface mediated nucleation mechanism due to the freedom of movement of the supramolecular chains on the surface. This mechanism is confirmed for two system sizes: a 216 molecule system and a 343 molecule system.

The technique of likelihood maximization reveals that the global average lattice parameters can be used to approximate the reaction coordinate. This is qualitatively verified with committor distribution analyses. A trace of the localized lattice parameters on each layer of the crystal along the transformation show significant non-uniformity, again, providing evidence for a nucleation mechanism. Also, the free energy barrier is observed to scale with the length of the edge of the crystal on which nucleation occurs. It is proposed that in a nucleation-controlled event, that this mechanism could still exhibit the "jumping" behavior observed by Davey *et al.*

To our knowledge, this study represents the first successful mechanistic investigation of solid state polymorph transformations in molecular crystals, leading to an enhanced understanding of nucleation processes in complex systems at the molecular level. Furthermore, the methods employed in this study allow the unbiased distinction between transformation mechanisms in molecular systems. This study also illustrates the applicability of aimless shooting and likelihood maximization to efficiently and systematically hone in on the most important collective variables needed to approximate the reaction coordinate.

2.7 References

1. Mullin, J.W., *Crystallization*. 4th Edition ed. 2001: Butterworth-Heinemann.
2. Matsui, M. and G.D. Price. *Nature*, **351,** p. 735-737 (1991).
3. Eddaoudi, M., et al. *Science*, **295,** p. 469-472 (2002).
4. Sawada, K. *Pure Appl. Chem.*, **69,** p. (1997).
5. Spanos, N. and P.G. Koutsoukos. *J. Cryst. Growth*, **191,** p. (1998).
6. Irie, M., S. Kobatake, and M. Horichi. *Science*, **291,** p. 1769-1772 (2001).
7. Vippagunta, S.R., H.G. Brittain, and D.J.W. Grant. *Adv. Drug Deliv. Reviews*, **48,** p. 3-26 (2001).
8. Morris, K.R., et al. *Adv. Drug Deliv. Reviews*, **48,** p. 91-114 (2001).
9. Moulton, B. and M.J. Zaworotko. *Chem. Reviews*, **101,** p. 1629-1658 (2001).
10. Yu, L.X., et al. *Pharm. Research*, **20,** p. 531-536 (2003).
11. Caira, M.R. *Topics in Curr. Chem.*, **198,** p. 163-208 (1998).
12. Morissette, S.L., et al. *Proc. Natl. Acad. Sci. U.S.A.*, **100,** p. 2180-2184 (2003).
13. Mnyukh, Y.V., *Fundamentals of solid-state phase transitions ferromagnetism and ferroelectricity*. 1st Edition ed. 1998: 1st Books Library.
14. Mnyukh, Y.V. *J. Cryst. Growth*, **32,** p. 371-377 (1976).
15. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 163-200 (1979).
16. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 201-218 (1979).
17. Cardew, P.T., R.J. Davey, and A.J. Ruddick. *J. Chem. Soc. - Faraday Trans.*, **80,** p. 659-668 (1984).
18. Tuble, S.C., J. Anwar, and J.D. Gale. *J. Amer. Chem. Soc.*, **126,** p. 396-405 (2004).
19. Kaneko, F., et al. *J. Phys. Chem. B*, **102,** p. 327-330 (1998).

20. Bailey, M. and C.J. Brown. *Acta. Cryst.*, **22,** p. 387-391 (1967).
21. Śledź, M., J. Janczak, and R. Kubiak. *J. Molec. Struct.*, **595,** p. 77-82 (2001).
22. Davey, R.J., et al. *J. Chem. Soc. - Faraday Trans.*, **90,** p. 1003-1009 (1994).
23. Davey, R.J., et al. *Nature*, **366,** p. 248-250 (1993).
24. Anwar, J. and P.K. Boateng. *J. Amer. Chem. Soc.*, **120,** p. 9600-9604 (1998).
25. Zahn, D. *Phys. Rev. Lett.*, **92,** p. 040801 (2004).
26. Radhakrishnan, R. and B.L. Trout. *J. Amer. Chem. Soc.*, **125,** p. 7743-7747 (2003).
27. Bolhuis, P.G., C. Dellago, and D. Chandler. *Faraday Discuss.*, **110,** p. 421-436 (1998).
28. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **110,** p. 6617-6625 (1999).
29. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **108,** p. 9236-9245 (1998).
30. Dellago, C., et al. *J. Chem. Phys.*, **108,** p. 1964-1977 (1998).
31. Dellago, C., P.G. Bolhuis, and P.G. Geissler. *Adv. Chem. Phys.*, **123,** p. 1-86 (2001).
32. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
33. Chu, J.W., B.R. Brooks, and B.L. Trout. *J. Amer. Chem. Soc.*, **126,** p. 16601-16607 (2004).
34. Steinhardt, P.J., D.R. Nelson, and M. Ronchetti. *Phys. Rev. B*, **28,** p. 784-805 (1983).
35. Radhakrishnan, R. and B.L. Trout. *J. Chem. Phys.*, **117,** p. 1786-1796 (2002).
36. Ma, A. and A.R. Dinner. *J. Phys. Chem. B*, **109,** p. 6769-6779 (2005).
37. Best, R.B. and G. Hummer. *Proc. Natl. Acad. Sci. U.S.A.*, **102,** p. 6732-6737 (2005).
38. Maragliano, L., et al. *J. Chem. Phys.*, **125,** p. 024106 (2006).
39. Du, R., et al. *J. Chem. Phys.*, **108,** p. 334-350 (1998).
40. Truhlar, D.G. and B.C. Garrett. *J. Phys. Chem. B*, **104,** p. 1069-1072 (2000).
41. Geissler, P.G., C. Dellago, and D. Chandler. *J. Phys. Chem. B*, **103,** p. 3706-3710 (1999).
42. Desiraju, G.R. *Angew. Chem. Int. Ed.*, **34,** p. 2311-2327 (1995).
43. Ryckaert, J., G. Ciccotti, and H. Berendsen. *J. Comput. Phys.*, **23,** p. (1977).
44. Gavezzotti, A. and G. Fillippini. *J. Phys. Chem. B*, **98,** p. 4831-4837 (1994).
45. Raabe, G. Z. *Naturforsch*, **57a,** p. 961-966 (2002).
46. MacKerell, A.D., J. Wiórkiewicz-Kuczera, and M. Karplus. *J. Amer. Chem. Soc.*, **117,** p. 11946-11975 (1995).
47. M.J. Frisch, e.a., *Gaussian 03, revision b.05*. 2003: Pittsburgh PA.
48. Becke, A.D. *J. Chem. Phys.*, **98,** p. 5648-5652 (1993).
49. Singh, U.C. and P.A. Kollman. *J. Comput. Chem.*, **5,** p. 129-145 (1984).
50. Besler, B.H., K.M. Merz, and P.A. Kollman. *J. Comput. Chem.*, **11,** p. 431-439 (1990).
51. Breneman, C.M. and K.B. Wilberg. *J. Comput. Chem.*, **11,** p. 361-373 (1990).
52. Kottalam, J. and D.A. Case. *J. Amer. Chem. Soc.*, **110,** p. 7690-7697 (1988).
53. Radhakrishnan, R. and T. Schlick. *J. Chem. Phys.*, **121,** p. 2436-2444 (2004).
54. Turnbull, D. and J.C. Fisher. *J. Chem. Phys.*, **17,** p. 71-73 (1949).

CHAPTER 3: EXTENSIONS TO THE LIKELIHOOD MAXIMIZATION APPROACH APPLIED TO THE SOLID STATE POLYMORPH TRANSFORMATION IN TEREPHTHALIC ACID

Reproduced in part with permission from:

1. G.T. Beckham, B. Peters, C. Starbuck, N. Variankaval, and B.L. Trout, "Surface-Mediated Nucleation in the Solid State Polymorph Transformation of Terephthalic Acid", *J. Amer. Chem. Soc.*, 129, 4714- 4723 (2007)
2. B. Peters, G.T. Beckham, and B.L. Trout, "Extensions to the Likelihood Maximization Approach for Finding Reaction Coordinates", *J. Chem. Phys.*, 127, 034109 (2007).

3.1 Introduction

Chapter 2 outlined the initial work conducted on the solid state polymorph transformation in terephthalic acid. The polymorph transformation is shown to proceed via fluctuations of supramolecular chains at the surface of the crystal. The reaction coordinate is determined with likelihood maximization and verified with the committor histogram method. The research described in Chapter 2 represents the first application of aimless shooting and likelihood maximization to a real system. This chapter briefly describes an extension to both the aimless shooting and likelihood maximization routines that is motivated by the work described in Chapter 2.

The aimless shooting briefly described in Chapter 2 does not follow the original algorithm described by Peters and Trout[1]. It does, however, obey detailed balance and is simpler to implement in prepackaged molecular simulation codes such as CHARMM[2] and NAMD[3]. The methodological change to the aimless shooting algorithm is described more fully in this chapter.

As shown in Figure 2-11, the transition state isosurface as predicted by the $p(\mathrm{TP}|r)$ model is not correct for the 6x6x6 terephthalic acid crystal (216 molecules). The histogram for the reaction coordinate from the $p(\mathrm{TP}|r)$ model is shifted towards to the product basin with a mean value $\mu_h = 0.63$. Originally, the value for the transition state isosurface was arbitrarily shifted from 30.195 Å to 30.400 Å, and the resulting histogram is sharply peaked at $\mu_h = 0.490$. To amend the issue of locating the transition state isosurface with likelihood maximization, a new half trajectory likelihood based on a new model function is developed

and applied to the original aimless shooting data for the 6x6x6 crystal. This chapter briefly describes this new model and the application to the 6x6x6 terephthalic acid system.

3.2 Two Point Version of Aimless Shooting

Let the duration of a reactive trajectory be t and choose a small $\Delta t << t$. The application described in Chapter 2, for instance, used $\Delta t = 0.01t$. Like other transition path sampling algorithms, aimless shooting requires an initial reactive trajectory. However, it should be noted that because aimless shooting reinitializes the velocity of the system at the shooting point, velocities are not required. Many methods exist for the generation of an initial trajectory[4-9]. Shift the initial trajectory in time so that the configuration along the trajectory at $t = 0$, $\mathbf{x}_0$, has a committor probability near ½, $p_B(\mathbf{x}_0) \approx ½$. In practice, a shift that gives the starting configuration $\mathbf{x}_0$ with a finite probability of reaching both the product and reactant states is acceptable for aimless shooting. Once an initial trajectory has been obtained, aimless shooting is performed by iterating these three steps:

1) Select a shooting point on the old trajectory, $x_{-\Delta t}$ or $x_{\Delta t}$, with a probability ½ for each position.

2) Draw new momenta from the Boltzmann distribution and dynamically propagate the system from the position and the time at the selected shooting point to $\pm t/2$.

3) Accept the new trajectory if it joins the reactant and product states, *A* and *B*.

Each shooting point should be saved with the forward and backward outcomes (A or B), regardless of whether the shooting point results in a reactive or nonreactive trajectory. This simple scheme samples the true transition path ensemble if the dynamics preserve the equilibrium ensemble. Collective variables to be tested as reaction coordinates must be computed at the shooting points from aimless shooting. Thus, the shooting points should be saved with the trajectory outcomes, which allows additional collective variables to be tested using the original aimless shooting data if the originally screened variables are insufficient.

3.3 Half Trajectory Likelihood

The aimless shooting and likelihood maximization scheme is widely applicable in determining reaction coordinates as a function of one or several pertinent variables. The original paper describing this methodology[1] describes two likelihood maximization routines

both of which use the forward and backward likelihood trajectories to compute the likelihood score. The more general scheme described uses the probability of being on a transition path at some point in configuration space, **x**, denoted $p(\mathrm{TP}|\mathbf{x})$.

$$p(\mathrm{TP}|r) = p_0(1-\tanh^2[r]) \tag{3-1}$$

The likelihood function for Equation 3-1 is then,

$$L_{\mathrm{TP}}(\boldsymbol{\alpha}, p_0) = \prod_k^{N_{\mathrm{acc}}} p(\mathrm{TP}|r(q_{\mathrm{acc}}^{(k)})) \prod_k^{N_{\mathrm{rej}}} (1-p(\mathrm{TP}|r(q_{\mathrm{rej}}^{(k)}))) \tag{3-2}$$

However, the originally described $p(\mathrm{TP}|r)$ model function to fit the reaction coordinate is stationary at the transition state isosurface, given by r=0 as shown in Figure 3-1a. Therefore the likelihood L_{TP} is not sufficiently sensitive to the exact location of the transition state along the reaction coordinate, r. The results described in the previous chapter suggest that maximization of L_{TP} tends to shift the reaction coordinate to fit the tails of $p(\mathrm{TP}|\mathbf{x})$ at the expense of the transition state region. To correct this, a scheme is proposed that is based on forward-trajectory outcomes that directly models $p_{\mathrm{B}}(\mathrm{x})$ regardless of whether the dynamics are diffusive or ballistic, as given by Equation 3-3.

$$p_B(r) = \tfrac{1}{2}[1+\tanh(r)] \tag{3-3}$$

This function is shown in Figure 3-1b. The committor probability can be directly approximated by maximizing the likelihood:

$$L = \prod_{x_k \to B} p_B(r(x_k)) \prod_{x_k \to A} (1-p_B(r(x_k)) \tag{3-4}$$

The notation $\mathbf{x_k} \to B$ (or A) indicates a product over all shooting points ($\mathbf{x_k}$) that result in trajectories that go to state B (or A) in forward time. Equation 3-4 discards half of each trajectory, but when the dynamics are not purely diffusive, the forward and backward outcomes are correlated. Therefore, Equation 3-4 discards an amount of information that is at most half of the available information. Peters, Beckham, and Trout describe this in significantly more detail and show the model functions as applied to the Muller potential[10].

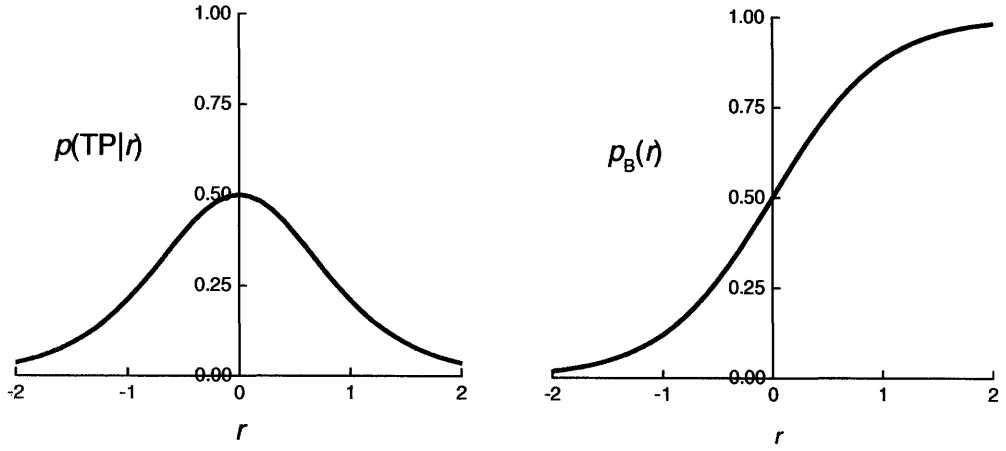


Figure 3-1. Model functions for the likelihood maximization algorithm. (a) $p(TP|r)$ as given in Expression 3-1. The function is stationary at the transition state surface given by $r = 0$. (b) $p_B(r)$ as given by Expression 3-3. This model function is used in the half trajectory likelihood formulation of likelihood maximization.

3.4 Application of Half-Trajectory Likelihood Maximization to Terephthalic Acid

As described in Chapter 2, aimless shooting and likelihood maximization are applied to the polymorph transformation in a 6x6x6 crystal of terephthalic acid. The original $p(TP|r)$ function given by Equation 3-1 is used as the model function for likelihood maximization. The same aimless shooting data for the 6x6x6 system are used as input to the half-trajectory likelihood maximization scheme using Expressions 3-3 and 3-4. The application of the half trajectory likelihood maximization also identifies the lattice vector **b** as the best reaction coordinate. The coefficients for the reaction coordinate from both models is given in Table 3-1 for the form $r = \alpha_0 + \alpha_1\mathbf{b}$. The values are reported as $1/\alpha_1$, the lattice parameter distance over which the system changes from reactant to product, and $-\alpha_0/\alpha_1=\mathbf{b}^{\ddagger}$, the value of the lattice parameter at the predicted transition state isosurface.

Table 3-1. Likelihood maximization results for the 1-dimensional reaction coordinate for 4000 aimless shooting paths for the 6x6x6 terephthalic acid system for both the $p(\mathrm{TP}|r)$ and $p_B(r)$ models.

Model	Transition State Surface (r=0)	p_0	$1/\alpha_1$ (Å)	$-\alpha_0/\alpha_1$ (Å)
$p(\mathrm{TP}\|r)$	$\mathbf{b}^{\ddagger}$ = 30.195 Å	0.57	-0.99	2.06
$p_B(r)$	$\mathbf{b}^{\ddagger}$ = 30.240 Å	--	-1.43	2.83

Committor probability distributions are compiled for both model functions and shown in Figure 3-2[11-14]. Figure 3-2a shows that Expressions 3-1 and 3-2 yield a good reaction coordinate in that it is peaked, but the transition state surface is shifted to the product side of the true transition state surface. In Chapter 2, the location of the transition state surface was arbitrarily shifted towards the product basin, which yields a histogram sharply peaked around ½ as shown in Figure 2-11. Figure 3-2b shows that maximizing the half-trajectory likelihood given by Equations 3-3 and 3-4 more accurately predicts the transition state surface location than the $p(\mathrm{TP}|r)$ model. Table 3-2 lists the mean and variance from the histograms and the intrinsic mean and variance of the committor probability distributions from the two likelihood scores[14]. As intended, the likelihood maximization scheme based on the $p_B(r)$ model shifts the transition state towards the reactant basin and yields a histogram peaked near p_B = ½.

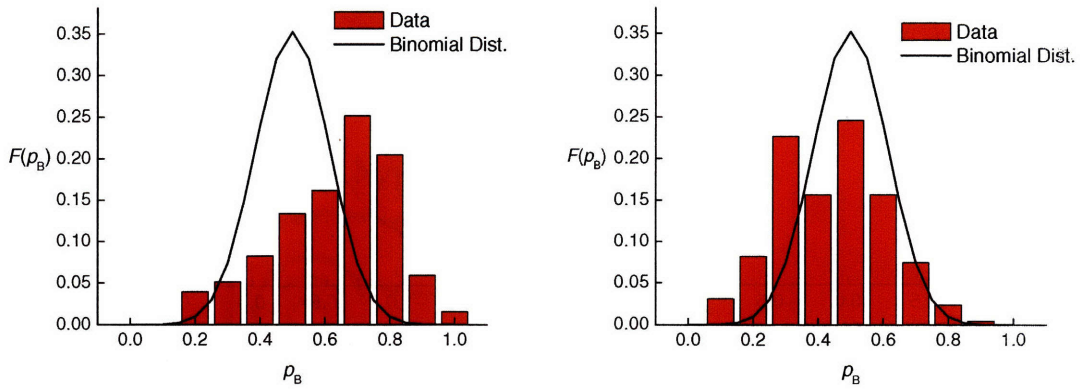


Figure 3-2. (a) Committor probability histogram for 216 molecule system with $\mathbf{b}^{\ddagger}$ = 30.195 Å, as predicted from the $p(\mathrm{TP}|r)$ model with a 1-D reaction coordinate model. (b) Committor probability histogram for 216 molecule system with $\mathbf{b}^{\ddagger}$ = 30.240 Å as predicted from the $p_B(r)$ model with a 1-D reaction coordinate model.

Table 3-2. Mean (μ) and standard deviations (σ) of the intrinsic committor probability distribution for transition state surfaces from the two likelihood maximization schemes. The half trajectory likelihood gives a transition state surface ($b^{\ddagger}$ = 30.240 Å) that is closer to the true transition state surface.

Model	Transition State Surface (r=0)	Histogram		Intrinsic	
		μ_h	σ_h	μ	σ
$p(\mathrm{TP}\|r)$	$b^{\ddagger}$ = 30.195 Å	0.629	0.182	0.629	0.148
$p_B(r)$	$b^{\ddagger}$ = 30.240 Å	0.441	0.164	0.441	0.118

Figure 3-3 shows the $p(\mathrm{TP}|r)$ and $p_B(r)$ models and the aimless shooting data plotted as a function of the reaction coordinate, r. Figure 3-3 is prepared by binning the data points according to the respective reaction coordinates in increments of 1/12th of the range of reaction coordinates sampled. The error bars are binomial standard deviations from the number of data points and the mean in each bin: $[\mu(1-\mu)/n]^{1/2}$ from the binomial standard deviation.

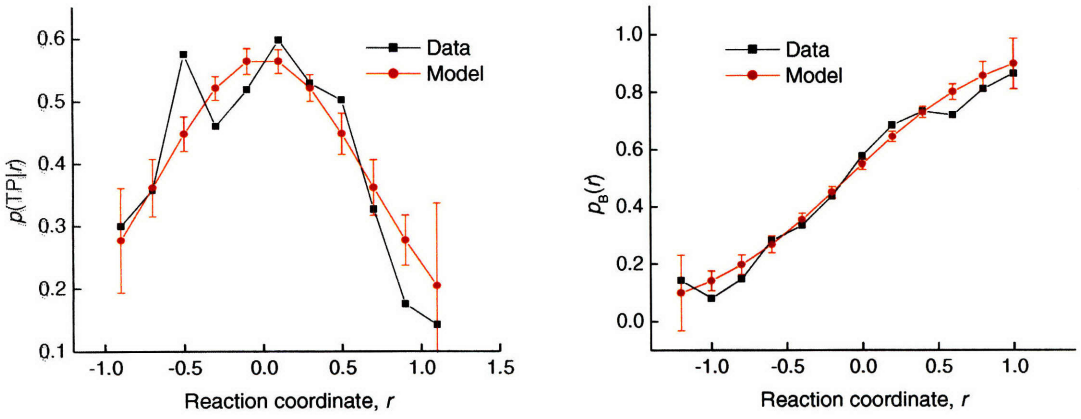


Figure 3-3. Aimless shooting data plotted as a function of the reaction coordinate r and compared to (a) optimized $p(\mathrm{TP}|r)$ function and (b) optimized $p_B(r)$ function. Note that the error bars appear on the model, not the data. The error bars show how far shooting point data should deviate from the probabilities $p(\mathrm{TP}|r)$ and $p_B(r)$ for a perfect reaction coordinate model.

The $p_B(r)$ model shown in Figure 3-3b appears to capture the transition state region more effectively than the $p(\mathrm{TP}|r)$ model. Confidence intervals on the reaction coordinate parameters can be derived to compare the reliability of these two approaches. Figure 3-4 shows the distribution of $\mathbf{b}^\ddagger$ estimates from bootstrapping with 4000 aimless shooting points per bootstrap sample, which is the same size as the original data set. The distribution of $\mathbf{b}^\ddagger$ was created by generating 10,000 estimates. The half trajectory likelihood based on $p_B(r)$ generates a narrower distribution of $\mathbf{b}^\ddagger$ estimates than the likelihood function based on $p(\mathrm{TP}|r)$. Interpolation of the data in Table 3-2 suggests that the surface $\mathbf{b}^\ddagger = 30.22$ Å would give a committor probability distribution centered to $\mu = ½$. That value is indicated by the arrow on the **b**-axis in Figure 3-4. The median $\mathbf{b}^\ddagger$ estimate from the half trajectory likelihood is only slightly close to 30.22 Å, but the typical deviation from 30.22 Å is much smaller because of the narrower distribution.

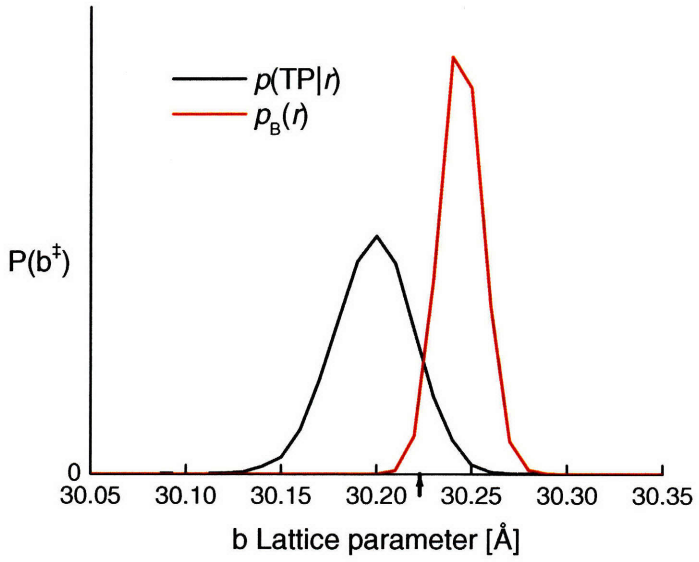


Figure 3-4. Distributions of the maximum likelihood estimate from $\mathbf{b}^\ddagger$ from 10000 bootstrap samples with 4000 trajectories in each sample. The red curve is from the half-trajectory likelihood of Equation 3-4, and the black curve is from the likelihood obtained via Equation 3-2. The value of $\mathbf{b}^\ddagger$ indicated by the arrow is obtained by interpolating the location of m = ½ from the histogram results at $\mathbf{b}^\ddagger$ = 30.240 Å and 30.195 Å.

Figure 3-5 shows how the $\mathbf{b}^\ddagger$ estimates from the two likelihood approaches vary with the number of trajectories in the likelihood function. The likelihood based on $p_B(r)$ consistently provides a narrower distribution of $\mathbf{b}^\ddagger$ than the likelihood based on $p(\mathrm{TP}|r)$. Based on the

terephthalic acid example, the $\mathbf{b}^{\ddagger}$ estimates from the likelihood based on $p_B(r)$ are also more accurate. Interestingly, when many trajectories are included in the likelihood, the variance in the distribution of $\mathbf{b}^{\ddagger}$ estimates increases linearly with the inverse square root of the number of trajectories even though the trajectories are interdependent parts of the likelihood function.

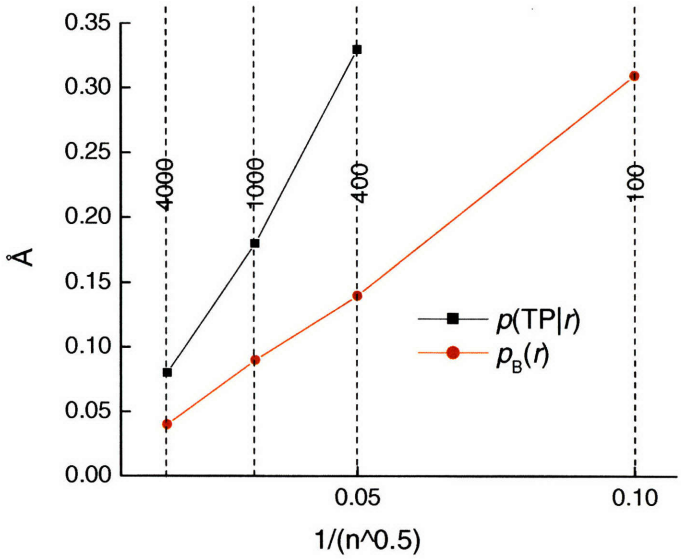


Figure 3-5. Width of a 95% confidence interval from bootstrap sampling for the transition state value of the reaction coordinate b (in Å) as a function of the number of aimless shooting trajectories used in the two likelihood functions, *n*. The width from L_{TP} from $p(TP|r)$ is shown in black and the width from the half trajectory likelihood from $p_B(r)$ is shown in red. The half trajectory likelihood consistently provides a smaller interval that is also more accurate as evidenced by the histograms in Figure 3-2.

3.5 Summary

The half-trajectory likelihood scheme based on the commitment probability, $p_B(r)$, is shown to predict the transition state isosurface more accurately than the transition path probability, $p(TP|r)$, when applied to the 6x6x6 terephthalic acid system. In addition, a simplified version of aimless shooting is presented to facilitate real applications and more easily implement in molecular simulation packages. See Peters *et al.* for more detail[10].

3.6 References

1. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
2. MacKerell, A.D., J. Wiórkiewicz-Kuczera, and M. Karplus. *J. Amer. Chem. Soc.*, **117,** p. 11946-11975 (1995).
3. Phillips, J.C., et al. *J. Comput. Chem.*, **26,** p. 1781-1802 (2005).
4. van Erp, T.S. and P.G. Bolhuis. *J. Comp. Phys.*, **205,** p. 157-181 (2005).
5. Allen, R.J., D. Frenkel, and P.R. ten Wolde. *J. Chem. Phys.*, **124,** p. 194111 (2006).
6. Voter, A.F. *Phys. Rev. Lett.*, **78,** p. 3908-3911 (1997).

7. Kottalam, J. and D.A. Case. *J. Amer. Chem. Soc.*, **110,** p. 7690-7697 (1988).
8. Schlitter, J., et al. *Mol. Sim.*, **10,** p. 291-308 (1993).
9. Hu, J., A. Ma, and A.R. Dinner. *J. Chem. Phys.*, **125,** p. 114101 (2006).
10. Peters, B., G.T. Beckham, and B.L. Trout. *J. Chem. Phys.*, **127,** p. 1 (2007).
11. Du, R., et al. *J. Chem. Phys.*, **108,** p. 334-350 (1998).
12. Geissler, P.G., C. Dellago, and D. Chandler. *J. Phys. Chem. B*, **103,** p. 3706-3710 (1999).
13. Truhlar, D.G. and B.C. Garrett. *J. Phys. Chem. B*, **104,** p. 1069-1072 (2000).
14. Peters, B. *J. Chem. Phys.*, **125,** p. 241101 (2006).

CHAPTER 4: IDENTIFICATION OF A NON-INTUITIVE REACTION COORDINATE IN THE POLYMORPH TRANSFORMATION OF TEREPHTHALIC ACID VIA LIKELIHOOD MAXIMIZATION

Reproduced in part with permission from G.T. Beckham, B. Peters, and B.L. Trout, "Identification of a Non-intuitive Reaction Coordinate in the Polymorph Transformation of Terephthalic Acid via Likelihood Maximization", submitted, 2007.

4.1 Introduction

A technologically important case of nucleation is the inter-conversion between different crystal forms, or polymorphs[1, 2]. Polymorphic systems are an ideal framework for the study of structure-function relationships as polymorphs typically exhibit different physical and chemical properties. In the pharmaceutical industry, understanding and controlling polymorphism is often of great importance as different crystal forms lead to variations in both processing and in pharmacokinetics. A particularly high profile case of polymorphism in the pharmaceutical industry was highlighted when Abbott Laboratories' AIDS drug, Ritonavir converted to a more stable polymorph in the manufacturing phase, causing significant production setbacks and financial losses[3]. Further interest in polymorphism has stemmed from recent attempts to patent specific polymorphs[4]. Thus, understanding nucleation mechanisms leading to transformation between polymorphs would be of substantial benefit to controlling polymorphism, and would provide complementary kinetic information to the thermodynamic problem of polymorph prediction[5-9]. Solid state polymorph transformations in molecular crystals have received significant attention from the experimental community[2, 10-18]. To date, approximately 160 mechanisms have been proposed for solid state polymorph transformations in molecular crystals[10]. At present, however, gaining direct molecular level insight to the dynamical events occurring during solid state polymorph transformations is outside of the scope of experimental capabilities, and thus there is little definitive evidence for any particular mechanism. As pointed out by Herbstein in an excellent review: "[t]here are relatively few papers about the actual transition directly viewed by microscopic techniques in order to infer the mechanism, and not many about changes in crystal structure...*as the system passes through the transition.*"[18]

Molecular simulation approaches alternatively offer the ability to track individual molecules in space and time, and can be utilized to study crystallization processes such as solid state polymorph transformations. To date, many computational approaches have proved fruitful for

investigating nucleation processes for a wide range of complexity from model systems[19-30] to atomic and molecular crystals[31-39]. Several past studies either utilized biasing methods, which assume a mechanism *a priori* and force the system from one state to another[29, 30, 32, 36, 40], or alternatively, run many trajectories at conditions where the metastable state is highly unstable so that the "rare" event occurs on accessible timescales[20, 31, 37]. While these methods provide valuable insight, a more ideal methodology would examine nucleation without assuming a mechanism *a priori* and simulate conditions comparable to experimental supersaturations and subcooling. For rough, high dimensional free energy landscapes, which are inherent in nucleation processes, transition path sampling [TPS][41-44] enables the collection of unbiased pathways that connect stable end states, such as two polymorphs, by efficiently focusing the computational time on the dynamical bottleneck. Several research groups have successfully utilized the TPS method to study crystal nucleation[23, 25-27, 34, 35, 39]. In addition, recent developments have led to methods that systematically determine the mechanism, or reaction coordinate[26, 45-47]. The Genetic Neural Network approach of Ma and Dinner[46] and the subsequently developed likelihood maximization technique of Peters *et al.*[26, 47], both use an informatics approach to approximate the reaction coordinate as a function of physically relevant variables. These two methods have recently been compared, and it was shown that likelihood maximization requires less trajectory information to obtain an accurate reaction coordinate[47]. Accurate reaction coordinates are important because they enable calculations of kinetically meaningful free energy barriers[48-54] and accurate rate constants[55-61].

Chapter 2 presented a computational study that illustrated the likelihood maximization approach by elucidating the mechanism of the solid-state polymorph transformation in terephthalic acid[39]. Terephthalic acid was chosen as the model system for in-depth computational studies because the crystal structures of the Form I and Form II polymorphs are known[62], the polymorph transformation exhibits a reluctance to nucleate unless massively subcooled or mechanically perturbed and extremely fast growth kinetics in an experimental setting[12, 63]. Also, the molecules are relatively small and fairly rigid even when completely unconstrained in a molecular dynamics (MD) simulation, and its structure represents an ubiquitous motif in organic crystals such as those found in the pharmaceutical industry with a carboxylic acid dimer synthon that forms long molecular chains[64]. In the previous study, two system sizes were investigated: a 6x6x6 (216 molecules) and a 7x7x7 (343 molecules) crystal, both with free boundaries. The system undergoes a nucleation step, and the free

energy barrier to nucleation increases as a function of system size. The mechanism as determined with likelihood maximization shows that fluctuations of the supramolecular synthon chains on the surface of the crystals, which have the most freedom to move in vacuum, drive the polymorph transformation. This mechanism, or reaction coordinate, can be approximated in these two small systems by changes in the global **b** lattice parameter and the global **b/c** lattice parameter ratio, respectively. The term "global" in this sense denotes an average over the entire crystal, which is measured on the nanometer length scale.

While global order parameters [OPs] serve as a viable approximation to the reaction coordinate for the 6x6x6 and 7x7x7 crystals, one can imagine that on a substantially larger length scale, an average lattice parameter measured over the entire crystal will not be able to capture the fluctuations on a specific side or corner of the crystal. Therefore, we propose that these same global OPs will not be sufficient reaction coordinates in larger systems[39]. Instead, we hypothesize that for larger crystals localized fluctuations will lead to nucleation events. The goal of this study is to investigate whether local OPs become important for larger TPA crystal sizes, and if so, to determine the nature of the important local OPs. Several types of local coordinates, such as new, local **b/c** lattice parameters and the nucleus size of the Form II polymorph are tested as possible reaction coordinates.

4.2 Methodology

The molecular force field developed and validated in Chapter 2 is used for this study. Two system sizes are examined in this study: an 8x8x8 terephthalic acid crystal and a 10x10x10 crystal, which corresponds to 512 and 1000 molecules, respectively. For the 8x8x8 system, the same scheme as described in Chapter 2 is employed here:

1) an initial trajectory is collected;

2) aimless shooting is used to harvest reactive trajectories between the reactant and product;

3) likelihood maximization is applied to determine the reaction coordinate;

4) a histogram test is used to validate the reaction coordinate approximation; and

5) the free energy barrier is determined using the correct reaction coordinate.

Only steps 1-3 are conducted for the 10x10x10 system due to the significant increase in computational expense.

4.2.1 Obtaining Initial Trajectories

Initial trajectories for the 7x7x7 crystal (343 molecules) were previously obtained using MD umbrella sampling over the global **b/c** lattice parameter ratio. The application of MD umbrella sampling over the global **b/c** lattice parameter ratio, however, does not yield a set of configurations connecting the Form I and Form II crystal structures for the 8x8x8 crystal. This preliminary result suggests that the global **b/c** lattice parameter is not a viable reaction coordinate for the 8x8x8 system. Therefore, MD umbrella sampling over a newly defined local **b/c** lattice parameter ratio is applied between minimized crystals from Form I to Form II with CHARMM[65]. Figure 4-1 shows the outline of the new local **b/c** lattice parameter ratio that is used for both the Form I and Form II crystals. This corner is chosen because it is crystallographically equivalent to the corner that undergoes significant structural changes during the nucleation step in the smaller systems.

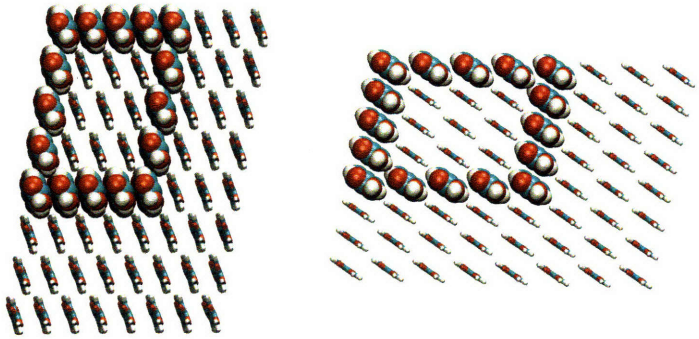


Figure 4-1. Form I and Form II terephthalic acid crystals from the a-axis view with the outline of the local b/c lattice parameter ratio used to obtain the initial trajectory for input to the aimless shooting method. This particular local OP was chosen because the same crystallographically equivalent corner in the 6x6x6 and 7x7x7 molecule systems undergoes significant changes during the nucleation event.

The initial trajectory for the 10x10x10 is generated using targeted molecular dynamics (TMD) in CHARMM[66]. These two different methods, umbrella sampling and TMD, are applied to harvest initial trajectories for the 8x8x8 and 10x10x10 crystals, respectively, to examine the effect of the biasing method on the reaction coordinate obtained with likelihood maximization. To obtain an initial trajectory with TMD that retains some physical relevance,

the Δ(RMSD) per time step is set to 2.5×10^{-6} Å/fs. The total trajectory length is therefore 2.3 ns with the initial RMSD of approximately 5.8 Å between the minimized end states.

4.2.2 Aimless Shooting

Aimless shooting pathways are collected with the two-point version of aimless shooting described in detail in Chapter 3. All MD trajectories for aimless shooting are obtained using the NAMD simulation package[67] in the NVE ensemble at 300 K. A time step of 1 fs is used with a cutoff for nonbonded interactions of 14 Å. The slight differences in the nonbonded energy switching functions between CHARMM and NAMD do not cause noticeable differences in the relevant quantities such as average lattice parameters, most likely because of the generous nonbonded energy cutoff distance of 14 Å. The total trajectory length for aimless shooting is 60 ps, which maintains the fraction of inconclusive trajectories well below 1%. A time displacement, Δt, of 600 fs is applied to achieve the desired acceptance rate. The first 100 aimless shooting trajectories are discarded to remove any effects from the initial, biased trajectories. Following dynamic equilibration, 1500 paths are collected for the 8x8x8 system and 500 paths are collected for the 10x10x10 system for use in likelihood maximization. The basin definitions are based on the global **b/c** lattice parameter. It should be noted that basin definitions do not need to be representative of the reaction coordinate, but must quantitatively distinguish the stable basins. An endpoint with **b/c** greater than 1.55 is characterized as Form I and those with **b/c** less than 1.40 are assigned to the post nucleation basin. These basin definitions are based on MD trajectories in the respective basins, as shown in Figure 2-5.

4.2.3 Likelihood Maximization

The newly developed half-trajectory likelihood maximization algorithm[47] is applied to the aimless shooting results to extract an approximation to the reaction coordinate. The reaction coordinate, r, is modeled as a linear combination of candidate OPs denoted as **q**, with α_0 through α_m as adjustable coefficients given by:

$$r(q) = \alpha_0 + \sum_{k=1}^{m} \alpha_k q_k \tag{4-1}$$

As described in Chapter 3, the half-trajectory likelihood method yields a more accurate approximation to the transition state isosurface and applies whether the dynamics are ballistic

or diffusive. In this formulation of likelihood maximization, the reaction coordinate is related to probability of committing to basin B as given by:

$$p_B(r) = \frac{1}{2}[1 + \tanh(r)] \tag{4-2}$$

The corresponding likelihood function is given as:

$$L = \prod_{x_k \to B} p_B(r(x_k)) \prod_{x_k \to A} (1 - p_B(r(x_k)) \tag{4-3}$$

Expression 4-3 is maximized over all coefficients and possible combinations of candidate OPs, yielding a likelihood score for each potential reaction coordinate. Each possible reaction coordinate receives a likelihood score and the Bayesian information criterion (BIC) is calculated to determine the statistical significance of different models.

Several hundred OPs are tested for the increased system sizes. All possible combinations of the local **b/c** axis ratios are screened as candidate variables such as the one shown in Figure 4-1. Also, the global lattice vectors, local lattice angles, local twisting along supramolecular chains, and directional variation of the supramolecular chains relative to one another are screened.

Classical nucleation theory assumes that the nucleus size is the reaction coordinate[68]. Therefore, a cluster size analysis is performed to identify nuclei of the more stable Form II crystal based on the unit cell environment around each molecule. To do this, the **b** and **c** lattice parameters around each molecule are calculated. The **b** lattice parameter is calculated by measuring the distance to the adjacent molecules along the **b** axis, and the **c** lattice parameter is measured similarly. If the **b/c** ratio for a molecule is greater than 1.55, the molecules are labeled as Form I, and if the **b/c** ratio is less than 1.15, it is labeled as Form II. This local classification is referred to as a "unit cell OP". These unit cell OPs are then used to quantify the cluster size for both polymorphs. Molecules with lattice parameters in between those values are tagged as interface molecules.

4.2.4 Committor Probability Distribution

The histogram test is applied to quantify the error in the reaction coordinate from likelihood maximization[69-72]. The histogram should be centered at ½, implying that points on the $r = 0$

surface are transition states with $p_B = ½$. For the 8x8x8 crystal, 20 trajectories of 30 ps each are fired randomly from independent configurations generated by a method inspired by the equilibrium path sampling method[53], as described in our previous study. The same simulation package (NAMD) and conditions are used as for the aimless shooting trajectories. The mean and standard deviation of the resulting committor distribution is calculated and compared to the intrinsic standard deviation as described by Peters[72]. This approach yields a quantitative measure of the reaction coordinate error.

4.2.5 Free Energy Calculations

After validating the reaction coordinate with the histogram test, the free energy is calculated with MD umbrella sampling for the 8x8x8 crystal. Windows are 1.05 ns long with 50 ps of equilibration at the start of each reaction coordinate window. The weighted histogram analysis method is applied to construct the free energy curve[50, 73].

4.3 Results

Figure 4-2 shows configurations along a reactive aimless shooting trajectory for the 8x8x8 crystal. The frame shown in Figure 4-2c is the shooting point. The snapshots appear qualitatively similar to the mechanism previously obtained for the 6x6x6 and 7x7x7 crystals shown in Chapter 2 as well as the 10x10x10 crystal in this chapter. Figure 4-3 shows the global **b/c** values at the endpoints of the aimless shooting trajectories for both the 8x8x8 and 10x10x10 systems. The distributions of global **b/c** values are split into two regions in the ranges of **b/c** < 1.4 and **b/c** > 1.55, indicating that the basin definitions are correct and that the trajectory length is sufficient. In addition, there is a slightly bimodal distribution for the product basin (**b/c** < 1.4), which is more pronounced in the 8x8x8 crystal. The majority of the trajectories that commit to the Form II basin reach a global **b/c** value of 1.0. As found for both the 6x6x6 and 7x7x7 crystals previously[39], some of the trajectories that end in the Form II basin stop at values of **b/c** between 1.3 and 1.4 indicating that there may be a metastable intermediate in the free energy landscape, as measured by this OP.

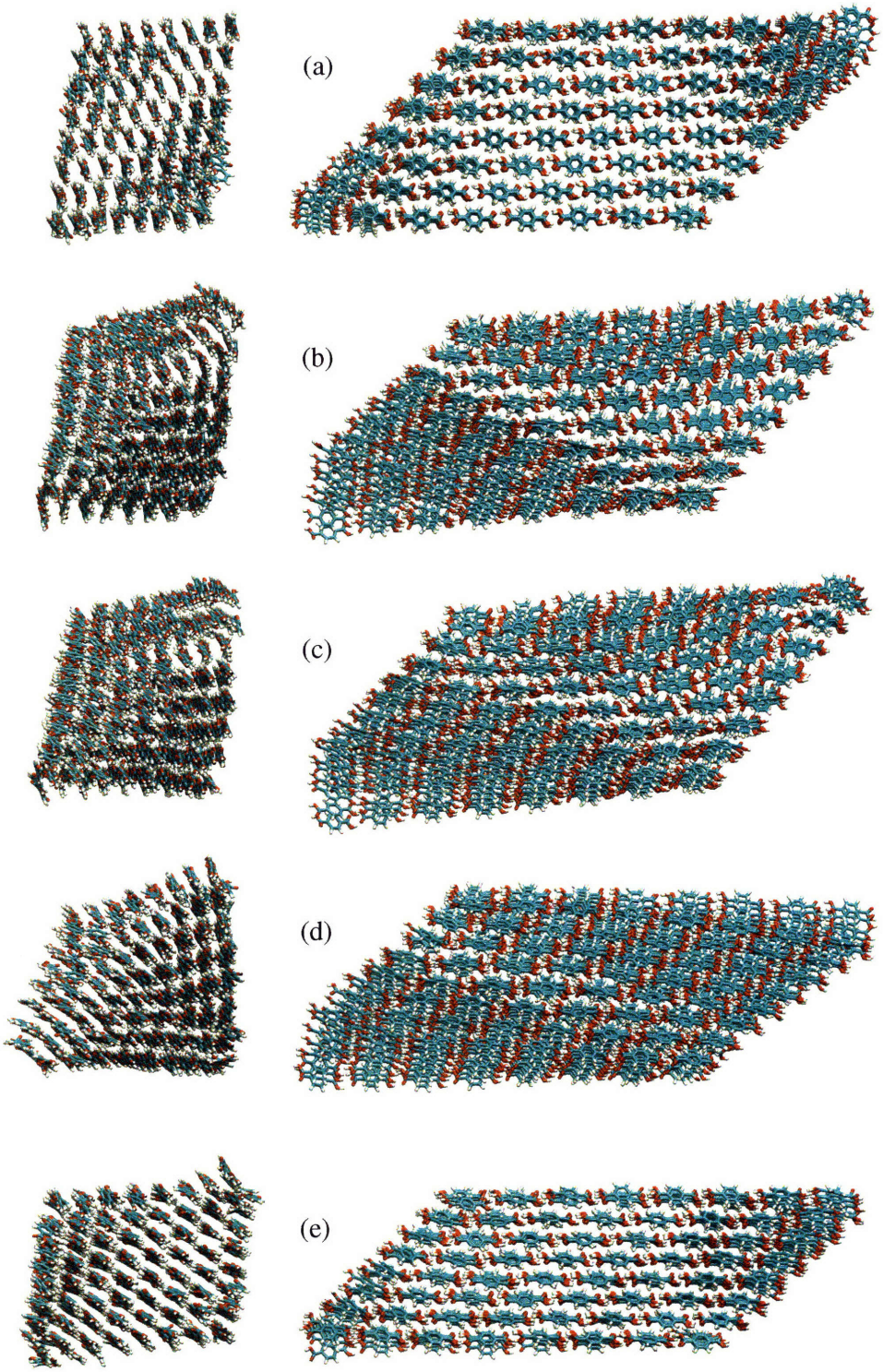


Figure 4-2. Snapshots taken from a reactive trajectory from Form I to Form II. The configuration shown in frame (c) is the shooting point in aimless shooting. Configurations shown in frame (a) and (e) are in the Form I and Form II basin, respectively.

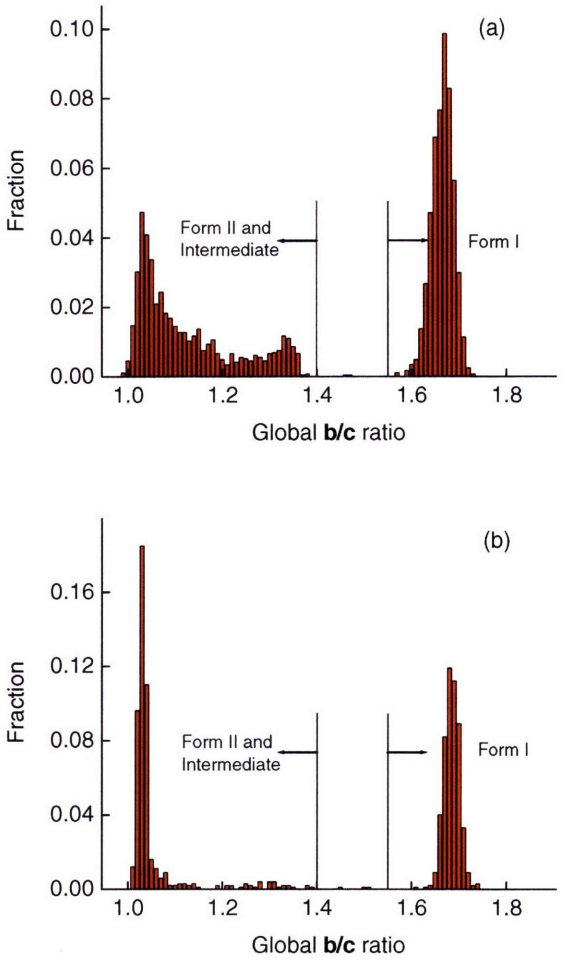


Figure 4-3. Endpoints of aimless shooting simulations as a function of the global b/c axis ratio for the (a) 8x8x8 and (b) 10x10x10 crystal. The global b/c lattice parameter ratio is used for the quantitative basin definitions in the aimless shooting procedure.

Several results from likelihood maximization are presented for comparison in Table 4-1. The likelihood maximization results for a 1-dimensional reaction coordinate in the 8x8x8 system are extracted from 1500 aimless shooting trajectories. The BIC for these data is 3.7. Three OPs are shown for the 1-dimensional case: the nucleus size for the Form II cluster, the global **b/c** lattice parameter for comparison to results in our previous study, and one of the local **b/c** OPs for a specific corner of the crystal. Both the Form II nucleus size and the global **b/c** axis ratio are inadequate 1-dimensional approximations to the reaction coordinate when compared to the local OP. The molecules used to measure the local **b/c** OP are highlighted in Figure 4-4 from two **a**-axis views for a configuration near the transition state isosurface. The reaction coordinate is 6 layers in both the **b** and **c** dimensions, and is located in a different corner than

the OP used to obtain an initial trajectory. The reason for this will be examined in the Discussion.

Approximately 500 trajectories are collected for the 10x10x10 crystal. The optimal reaction coordinate for this system is also a local **b/c** lattice parameter ratio. In this case, it is 5 layers in the **b** and **c** directions. The reaction coordinate is in a crystallographically equivalent corner to the reaction coordinate found for the 8x8x8 system.

Likelihood maximization is also re-applied to 1000 aimless shooting trajectories for the 7x7x7 system from the previous study[39]. The Form II cluster size is screened as a potential reaction coordinate. Table 4-1 shows that the global **b/c** axis ratio remains the optimal reaction coordinate, and that the average Form II cluster size is approximately 105 molecules.

Table 4-1. Likelihood maximization results for the reaction coordinate in the 8x8x8 and 10x10x10 crystals. In addition, the likelihood scores are included for the 7x7x7 optimal reaction coordinate from the previous study[39] along with the Form II nucleus size for comparison.

System Size	OP	Likelihood Score	Transition State Isosurface ($r = 0$)
8x8x8 [512 molecules] [(a)]	Form II nucleus size	-751.35	84 molecules
	Global **b/c**	-714.73	1.50
	b/c for corner 4, depth = 6 layers	-685.55	1.39
10x10x10 [1000 molecules] [(b)]	Form II nucleus size	-308.57	81 molecules
	Global **b/c**	-298.98	1.48
	b/c for corner 4, depth = 5 layers	-255.88	1.35
7x7x7 [343 molecules] [(c)]	Form II nucleus size	-570.41	105 molecules
	Global **b/c**[39]	-503.82	1.43

(a) BIC = 3.1
(b) BIC = 3.5
(c) BIC = 3.6

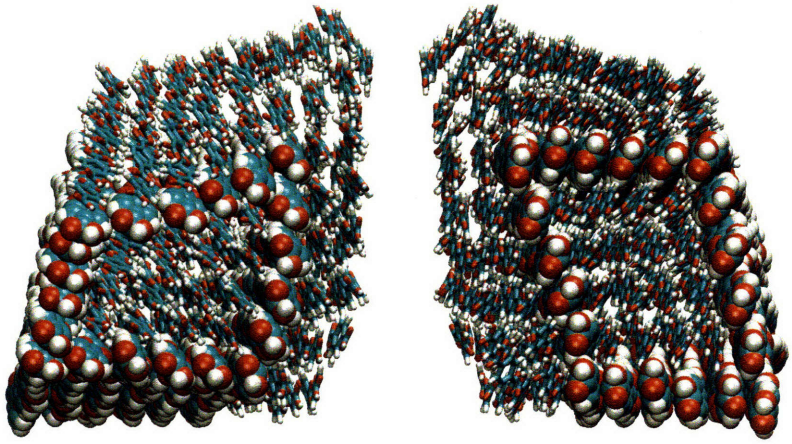


Figure 4-4. Snapshots from the a-axis view of the 8x8x8 system near the transition state. The molecules in space-fill format outline the local b/c axis ratio that is the best reaction coordinate approximation as determined by likelihood maximization.

The committor probability histogram for the local **b/c** OP for the 8x8x8 crystal is shown in Figure 4-5. It is peaked at $\mu_h = 0.588$ and the standard deviation is $\sigma_h = 0.127$ The intrinsic mean and standard deviation are calculated as described by Peters[72], and are $\mu = 0.588$ and $\sigma = 0.108$, respectively. For reference, the distribution given by the intrinsic parameters is plotted along with the histogram with the assumed functional form of the normal distribution. The relative uncertainty in the intrinsic mean and standard deviation are $\Delta\mu/\mu = 0.0215$ and $\Delta\sigma/\sigma = 0.199$, respectively[72]. This quantitatively demonstrates that the local **b/c** OP is a viable approximation to the reaction coordinate.

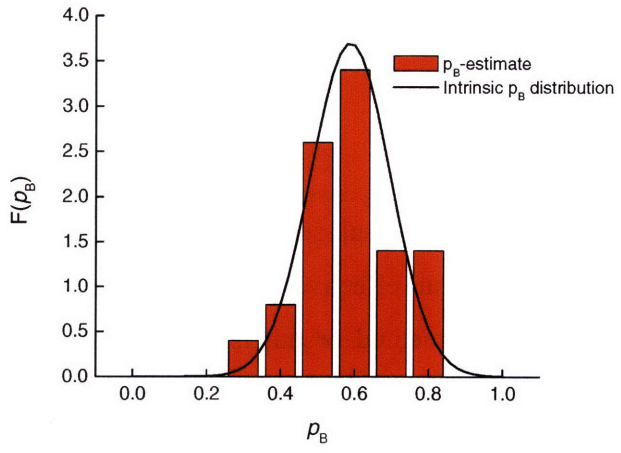


Figure 4-5. Normalized committor probability histogram for the 8x8x8 system with a local $b/c^{\ddagger} = 1.387$, as predicted from the $p_B(r)$ model with a 1-dimensional reaction coordinate. $\mu_h = 0.588$, $\sigma_h = 0.127$. The intrinsic distribution with $\mu = 0.588$ and $\sigma = 0.108$, is plotted for comparison.

Figure 4-6 shows the free energy along the local **b/c** OP in the 8x8x8 crystal found in likelihood maximization. As the system transforms from the Form I polymorph to Form II, the system overcomes a free energy barrier to nucleation, and passes through a metastable intermediate prior to the growth phase. In addition, as in our previous study, the barrier to nucleation scales with system size, whereas the barrier to growth remains relatively small at 2-3 kcal/mol. The nucleation barrier for the 6x6x6 crystal and for the 7x7x7 crystal is approximately 4 kcal/mol and 12 kcal/mol, respectively[39]. For the 8x8x8 system, the free energy barrier to nucleation is approximately 22 kcal/mol.

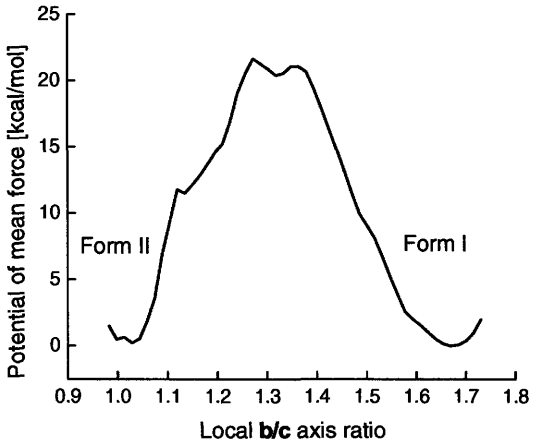


Figure 4-6. Potential of mean force calculated from MD umbrella sampling as a function of the local b/c reaction coordinate from likelihood maximization for the 8x8x8 system. The weighted histogram analysis method[50, 73] is applied to combine the umbrella sampling windows.

4.4 Discussion

The discussion section is divided into two sub-sections: the first sub-section highlights the utility of likelihood maximization as an effective method to identify non-intuitive reaction coordinates. It is stressed that this method or others with similar intentions[46] should be applied to systematically determine reaction coordinates in complex systems. Also, the implications of the results are briefly discussed in the broader context of solid state polymorph transformations in hydrogen bonded crystalline materials.

4.4.1 Implications for Theory

Our results indicate that the global **b/c** lattice parameter is not a good approximation to the reaction coordinate relative to the local **b/c** OPs for the larger systems, even though it is a natural first choice for a reaction coordinate when studying this crystal transformation in vacuum. In searching for an initial trajectory, for instance, MD umbrella sampling over this global **b/c** axis ratio does not yield a viable pathway for the 8x8x8 crystal.

It is initially surprising that local **b/c** OP in the corner of the crystal, as shown in Figure 4-4, is the reaction coordinate, rather than an OP similar to the one chosen to obtain an initial trajectory. This reaction coordinate is surprising because the corner chosen to obtain the initial trajectory appears to undergo significant structural changes near the transition state region. To investigate this finding further, it is useful to label the Form II cluster according to the unit cell OPs for configurations that the best reaction coordinate identified as transition states. Figure 4-7 shows an **a**-axis view of the transition state configuration with the molecules in blue for the Form II unit cell OP and in green for the Form I unit cell OP. The red wireframe molecules have local environments that are not indicative of either the Form I or Form II polymorph. This figure clearly shows that the Form II cluster exists in the corner of the crystal identified by the local **b/c** reaction coordinate. Although the cluster size is not the best reaction coordinate identified by likelihood maximization, it is still useful to illustrate that the intuitive Form II cluster size parameter coincides with the local reaction coordinate identified by likelihood maximization. To further verify that the Form II cluster size exhibits correlation with the reaction coordinate, the nucleus size is shown in Figure 4-9 for the 8x8x8 system as a function of the best reaction coordinate according to likelihood maximization. As expected, the Form II nucleus size increases with r, which is concomitant with the product basin and the nucleus size decreases closer to the reactant basin. The discovery of a reaction coordinate better than the intuitive nucleus size coordinate highlights the value of likelihood maximization as a tool to extract non-intuitive reaction coordinates from trajectory data.

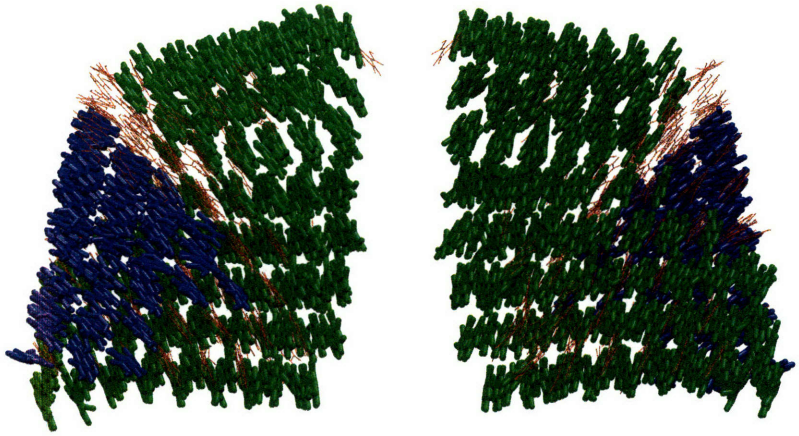


Figure 4-7. Snapshots from the a-axis view of the 8x8x8 system near the transition state. The molecules in blue are the molecules in the Form II lattice, whereas the green molecules are in the Form I lattice. The wireframe molecules in red are classified as being in neither basin, and form the interface between the two polymorphs.

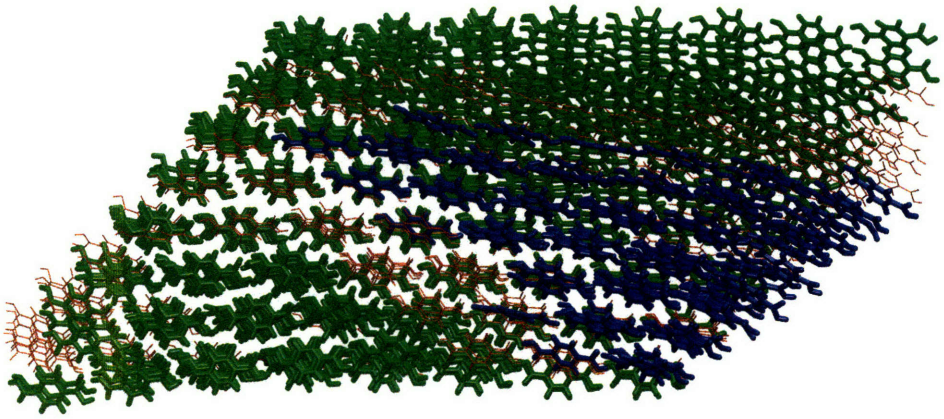


Figure 4-8. Snapshot from the b-axis view of the 8x8x8 system near the transition state. The molecules in blue are the molecules in the Form II lattice, whereas the green molecules are in the Form I lattice. The wireframe molecules in red are classified as neither crystal, and form the interface between the two polymorphs.

In addition, as shown in Table 4-1, the nucleus size for the transition states for the 7x7x7, 8x8x8, and 10x10x10 crystals are approximately equivalent. The values range from 81 molecules to 105 molecules. The observation of three crystal sizes with similar mechanisms and critical nuclei of similar sizes suggests that simulations of larger TPA crystals may continue to transform via corner nucleation.

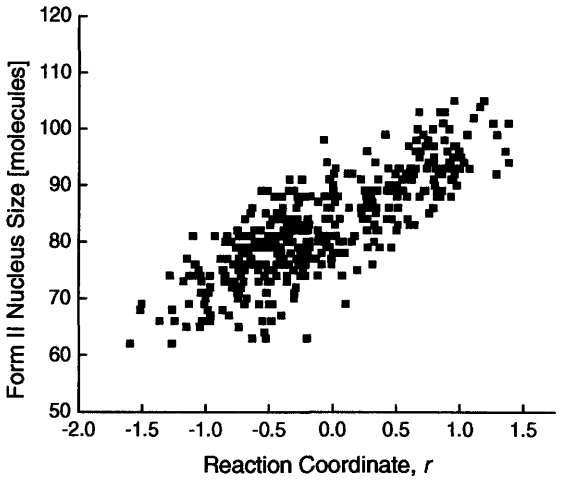


Figure 4-9. Form II nucleus size as a function of the local b/c reaction coordinate for the 8x8x8 crystal. The product basin is associated with large, positive *r* and the reactant basin corresponds to large, negative *r*. Therefore, as expected, the more stable Form II nucleus decreases in size as the system approaches the reactant, or Form I crystal, and increases as the system nears the product, or Form II crystal. This plot demonstrates that even though the nucleus size is not the optimal reaction coordinate, it remains a useful illustration to explain the location of the reaction coordinate determined from likelihood maximization.

4.4.2 Significance for solid state polymorph transformations

For a given nucleus size, corner nucleation leads to the smallest possible interfacial area between the two domains. In the interior of a perfect crystal, however, the interfacial area for a nucleus of the Form II crystal would be much larger at an equivalent nucleus size, and would therefore have a substantially higher free energy barrier to nucleation. Unless the crystal is very large such that the probability of observing a bulk nucleation event is many orders of magnitude higher, it is anticipated that corner nucleation will be a common phenomenon in hydrogen bonded molecular crystals. In summary, corner nucleation is a pathway by which the surface free energy penalty can be minimized in a nucleation event. Although not directly addressed here, nucleation events may also initiate from internal defects within the bulk of the crystal. The number of these internal defects will also scale with the crystal size which may also favor other nucleation mechanisms for large crystals. Investigation of these alternative nucleation mechanisms would be a valuable complement to

this study. Additionally, simulations of larger system sizes are needed to fully characterize the growth process.

4.5 Conclusions

This study explores size effects on the corner nucleation mechanism for the solid state polymorph transformation in TPA from Form I to Form II. Likelihood maximization shows that changes in local order rather than global order become more important as the system size increases. Committor probability analysis verified these reaction coordinates. The free energy curve along the reaction coordinate exhibits a large nucleation barrier and a negligible barrier to growth. These results suggest an explanation for large subcooling or mechanical perturbation required to initiate the transformation, as well as the almost instantaneous growth kinetics seen in the experiments[12, 63].

To verify that the mechanism is a corner nucleation process, it is demonstrated that a nucleus of the more stable Form II polymorph forms at a corner of the crystal. A cluster analysis shows that the average size of the Form II nucleus at the transition state is approximately 80 to 100 molecules, and that an interface exists between the two polymorphs. In addition, it is shown that although the number of molecules in the Form II nucleus is not the best reaction coordinate for the system sizes explored, a definite correlation exists between the actual reaction coordinate and nucleus size.

The corner nucleation mechanism for the solid state polymorph transformation in TPA yields the smallest possible interfacial area between the two polymorphs at a given nucleus size, which reduces the surface free energy penalty required to form a new polymorphic domain. Based on these results, corner nucleation may be a common polymorph transition mechanism in crystals of hydrogen bonded materials.

This study also illustrates the value of likelihood maximization as an effective method for mechanistic studies of rare events in complex systems. Likelihood maximization is appropriate to any problem to which transition path sampling can be applied, which spans many areas of scientific and technical interest. Specifically, likelihood maximization as demonstrated here identifies a non-intuitive reaction coordinate, which etches out the corner around the Form II nucleus, from over 500 candidate order parameters. The recognition of a reaction coordinate that is not directly intuitive highlights the merits of likelihood

maximization. Because incorrect reaction coordinates can yield misleading results for quantities of interest, aimless shooting and likelihood maximization are effective ways to obtain correct reaction coordinates as input for free energy calculations[48-54] and pathway optimizations[45, 74] that require *a priori* specification of the reaction coordinate or reaction coordinates components.

4.6 References

1. Mullin, J.W., *Crystallization*. 4th Edition ed. 2001: Butterworth-Heinemann.
2. Vippagunta, S.R., H.G. Brittain, and D.J.W. Grant. *Adv. Drug Deliv. Reviews*, **48,** p. 3-26 (2001).
3. Morissette, S.L., et al. *Proc. Natl. Acad. Sci. U.S.A.*, **100,** p. 2180-2184 (2003).
4. Thayer, A.M. *C&EN*, **85,** p. 31-34 (2007).
5. Leusen, F.J.J. *J. Cryst. Growth*, **166,** p. 900-903 (1996).
6. Lommerse, J.P.M., et al. *Acta. Cryst.*, **B56,** p. 697-714 (2000).
7. Price, S.L. *Adv. Drug Deliv. Reviews*, **56,** p. 301-319 (2004).
8. Abraham, N.L. and M.I.J. Probert. *Phys. Rev. B*, **73,** p. 224104 (2006).
9. Raiteri, P., R. Martonak, and M. Parrinello. *Angew. Chem. Int. Ed.*, **44,** p. 3769-3773 (2005).
10. Mnyukh, Y.V., *Fundamentals of solid-state phase transitions ferromagnetism and ferroelectricity*. 1st Edition ed. 1998: 1st Books Library.
11. Cardew, P.T., R.J. Davey, and A.J. Ruddick. *J. Chem. Soc. - Faraday Trans.*, **80,** p. 659-668 (1984).
12. Davey, R.J., et al. *Nature*, **366,** p. 248-250 (1993).
13. Morris, K.R., et al. *Adv. Drug Deliv. Reviews*, **48,** p. 91-114 (2001).
14. Davey, R.J. and J. Garside, *From molecules to crystallizers*. Oxford chemistry primers. 2001: Oxford University Press.
15. Zhang, G.G.Z., et al. *Journal of Pharmaceutical Sciences*, **91,** p. 1089-1100 (2002).
16. Yamanobe, M., H. Takiyama, and M. Matsuoka. *J. Chem. Eng. Japan*, **24,** p. 247-254 (2002).
17. Kim, Y.S., H.C. Paskow, and R.W. Rousseau. *Crystal Growth & Design*, **5,** p. 1623-1632 (2005).
18. Herbstein, F.H. *Acta Cryst.*, **B62,** p. 341-383 (2006).
19. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *Faraday Discuss.*, **104,** p. 93-110 (1996).
20. Anwar, J. and P.K. Boateng. *J. Amer. Chem. Soc.*, **120,** p. 9600-9604 (1998).
21. Auer, S. and D. Frenkel. *Nature*, **409,** p. 1020-1023 (2001).
22. Berim, G.O. and E. Ruckenstein. *J. Chem. Phys.*, **117,** p. 7732-7737 (2002).
23. Moroni, D., P.R. ten Wolde, and P.G. Bolhuis. *Phys. Rev. Lett.*, **94,** p. 235703 (2005).
24. Brendel, K., G.T. Barkema, and H. van Beijeren. *Phys. Rev. E.*, **71,** p. 031601 (2005).
25. Pan, A.C., et al. *J. Phys. Chem. B*, **110,** p. 3692-3696 (2006).
26. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
27. Trudu, F., D. Donadio, and M. Parrinello. *Phys. Rev. Lett.*, **97,** p. 105701 (2006).
28. Punnathanam, S. and P.A. Monson. *J. Chem. Phys.*, **125,** p. 024508 (2006).
29. Desgranges, C. and J. Delhommelle. *J. Amer. Chem. Soc.*, **128,** p. 15104-15105 (2006).
30. Desgranges, C. and J. Delhommelle. *J. Chem. Phys.*, **126,** p. 054501 (2007).
31. Matsumoto, M., S. Saito, and I. Ohmine. **416,** p. 409-413 (2002).
32. Radhakrishnan, R. and B.L. Trout. *J. Chem. Phys.*, **117,** p. 1786-1796 (2002).
33. Radhakrishnan, R. and B.L. Trout. *J. Amer. Chem. Soc.*, **125,** p. 7743-7747 (2003).
34. Zahn, D. *Phys. Rev. Lett.*, **92,** p. 040801 (2004).
35. Zahn, D. *J. Phys. Chem. B*, **111,** p. 5249-5253 (2007).
36. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Amer. Chem. Soc.*, **126,** p. 12286-12287 (2004).
37. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Chem. Phys.*, **122,** p. 184518 (2005).
38. Beaucage, P. and N. Mousseau. *Phys. Rev. B*, **71,** p. 094102 (2005).
39. Beckham, G.T., et al. *J. Amer. Chem. Soc.*, **129,** p. 4714-4724 (2007).
40. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *J. Chem. Phys.*, **104,** p. 9932-9947 (1996).
41. Bolhuis, P.G., C. Dellago, and D. Chandler. *Faraday Discuss.*, **110,** p. 421-436 (1998).
42. Dellago, C., et al. *J. Chem. Phys.*, **108,** p. 1964-1977 (1998).
43. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **110,** p. 6617-6625 (1999).
44. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **108,** p. 9236-9245 (1998).
45. Maragliano, L., et al. *J. Chem. Phys.*, **125,** p. 024106 (2006).

46. Ma, A. and A.R. Dinner. *J. Phys. Chem. B*, **109,** p. 6769-6779 (2005).
47. Peters, B., G.T. Beckham, and B.L. Trout. *J. Chem. Phys.*, **127,** p. 1 (2007).
48. Torrie, G.M. and J.P. Valleau. *J. Comput. Phys.*, **23,** p. 187-199 (1977).
49. Kottalam, J. and D.A. Case. *J. Amer. Chem. Soc.*, **110,** p. 7690-7697 (1988).
50. Kumar, S., et al. *J. Comput. Chem.*, **13,** p. 1011-1021 (1992).
51. Kong, X. and C.L. Brooks. *J. Chem. Phys.*, **106,** p. 2414-2423 (1996).
52. Laio, A. and M. Parrinello. *Proc. Natl. Acad. Sci. U.S.A.*, **99,** p. 12562-12566 (2002).
53. Radhakrishnan, R. and T. Schlick. *J. Chem. Phys.*, **121,** p. 2436-2444 (2004).
54. Ensing, B., et al. *Acc. Chem. Res.*, **39,** p. 73-81 (2006).
55. Garrett, B.C. and D.G. Truhlar. *J. Amer. Chem. Soc.*, **101,** p. 4534-4548 (1979).
56. Frenkel, D. and B. Smit, *Understanding molecular simulations: From algorithms to applications*. 2nd ed. Computational science series. Vol. 1. 2002.
57. Peters, B., A.T. Bell, and A. Chakraborty. *J. Chem. Phys.*, **121,** p. 4453-4460 (2004).
58. Moroni, D., P.G. Bolhuis, and T.S. van Erp. *J. Chem. Phys.*, **120,** p. 4055-4065 (2004).
59. van Erp, T.S. and P.G. Bolhuis. *J. Comp. Phys.*, **205,** p. 157-181 (2005).
60. Best, R.B. and G. Hummer. *Proc. Natl. Acad. Sci. U.S.A.*, **102,** p. 6732-6737 (2005).
61. Allen, R.J., D. Frenkel, and P.R. ten Wolde. *J. Chem. Phys.*, **124,** p. 194111 (2006).
62. Bailey, M. and C.J. Brown. *Acta. Cryst.*, **22,** p. 387-391 (1967).
63. Davey, R.J., et al. *J. Chem. Soc. - Faraday Trans.*, **90,** p. 1003-1009 (1994).
64. Desiraju, G.R. *Angew. Chem. Int. Ed.*, **34,** p. 2311-2327 (1995).
65. MacKerell, A.D., J. Wiórkiewicz-Kuczera, and M. Karplus. *J. Amer. Chem. Soc.*, **117,** p. 11946-11975 (1995).
66. Schlitter, J., et al. *Mol. Sim.*, **10,** p. 291-308 (1993).
67. Phillips, J.C., et al. *J. Comput. Chem.*, **26,** p. 1781-1802 (2005).
68. Turnbull, D. and J.C. Fisher. *J. Chem. Phys.*, **17,** p. 71-73 (1949).
69. Du, R., et al. *J. Chem. Phys.*, **108,** p. 334-350 (1998).
70. Geissler, P.G., C. Dellago, and D. Chandler. *J. Phys. Chem. B*, **103,** p. 3706-3710 (1999).
71. Truhlar, D.G. and B.C. Garrett. *J. Phys. Chem. B*, **104,** p. 1069-1072 (2000).
72. Peters, B. *J. Chem. Phys.*, **125,** p. 241101 (2006).
73. Grossfield, A., *An implementation of wham: The weighted histogram analysis method*. 2003.
74. Miller, T.F., E. Vanden-Eijnden, and D. Chandler. *Proc. Natl. Acad. Sci. U.S.A.*, **104,** p. 14559-14564 (2007).

CHAPTER 5: HOMOGENEOUS NUCLEATION OF BENZENE FROM THE MELT

5.1 Introduction

Benzene is a ubiquitous functional group in many naturally occurring and industrially important chemicals and has thus received considerable experimental and theoretical attention[1-23]. Benzene exhibits multiple phases, and the phase diagram has recently been evaluated with the metadynamics approach of Raiteri *et al.*[15] as reproduced in Figure 5-1. It undergoes multiple high pressure order-order phase transitions, polymerizes at high pressure and temperature, and nucleates from the liquid to the Form I crystal from atmospheric pressure up to approximately 2 GPa.

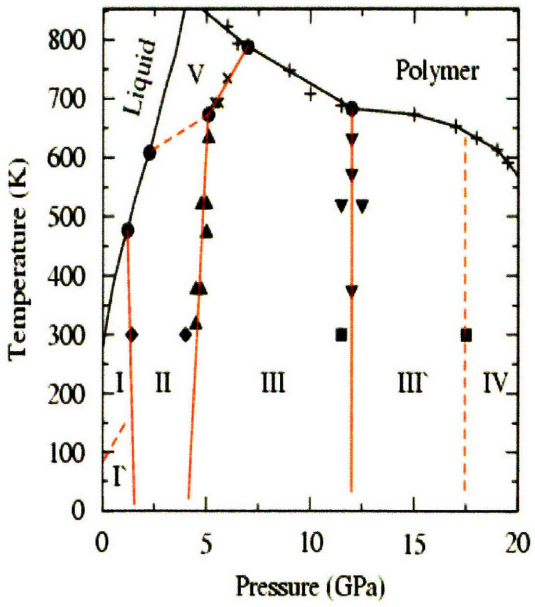


Figure 5-1. Proposed phase diagram of benzene elucidated by metadynamics, reproduced from Raiteri *et al.*[15] The labeled points are from references 4 and 7.

This chapter examines the mechanism of homogeneous nucleation from liquid benzene to the Form I crystal. Nucleation of crystalline solids from the melt is of general scientific and technical interest[24-28], and has been studied extensively with molecular simulation techniques[29-42]. As discussed in Chapter 1, previous studies to date have focused on model systems or small molecules such as, nitrogen[35, 37], carbon dioxide[36], or water[40, 41]. Transition path sampling [TPS][43-48] allows for the collection of many unbiased reactive pathways, and has been applied to melt nucleation in model systems[38, 39]. Application of the TPS method to molecular systems should yield further insight into the liquid to solid phase transition in cases

when then molecules are anisotropic in both shape and intermolecular connectivity. It is anticipated that the anisotropy of benzene will yield critical nuclei that are anisotropic in shape. Benzene is chosen as the system of interest because:

1) it is a common model organic molecule and is fairly rigid and relatively small,

2) the OPLS potential describes the liquid and solid states adequately[5, 12]

3) the Form I crystal structure is known[2], and

4) considerable experimental work has been conducted with regard to the liquid-solid phase transition[1, 3, 7, 8, 18, 20].

This chapter is organized as follows: previous studies of homogeneous nucleation from the melt are reviewed. The methodology, including the potential and the labeling of the solid and liquid particles, is provided and order parameters [OPs] screened in likelihood maximization are included. Preliminary results are presented as to the mechanism of the nucleation event.

5.2 Background

Classical nucleation theory [CNT] serves as a useful framework for understanding homogeneous nucleation processes[49]. The basis of CNT is that thermal fluctuations give rise to the appearance of small nuclei of a more stable phase and occasionally produce a long chain of favorable energetic fluctuations, thereby creating a nucleus exceeding the critical size. Although this more stable phase has favorable lower free energy, there is a free energy penalty associated with the creation of an interface. The free energy of the more stable phase is therefore the sum of a favorable, negative volume term and an unfavorable, positive surface term. For a spherical nucleus, the free energy change, ΔG, is given as:

$$\Delta G = -\frac{4}{3}\pi R^3 \Delta G_v + 4\pi R^2 \gamma \qquad (5\text{-}1)$$

where R = radius of the nucleus, ΔG_V = bulk free energy difference per unit volume between the metastable and more stable phases, and γ= surface free energy of the more stable phase per unit area. The maximum of the free energy expression in Equation 5-1 corresponds to the critical size of the nucleus, $R_{critical}$:

$$R_{\text{critical}} = \frac{2\gamma}{\Delta G_v} \tag{5-2}$$

Combining Equations 5-1 and 5-2 produces an expression for the maximum free energy barrier of nucleation, $\Delta G_{\text{critical}}$,

$$\Delta G_{\text{critical}} = \frac{4}{3}\pi\gamma \cdot R_{\text{critical}}^2 \tag{5-3}$$

The assumed reaction coordinate in CNT is the size of a solid nucleus, which is assumed to behave as a mean field and possess properties concomitant with that of the bulk solid. However, CNT has been shown to fail in quantitatively describing both the mechanism and rate of nucleation[24, 42, 50, 51]. It is therefore desirable to apply methods to understand the limitations of CNT and gain insight into homogeneous nucleation in real systems. As mentioned in Chapter 1, experimental techniques still lack the ability to track random molecular level events in a homogeneous medium except for systems where the elementary particles range from hundreds of nanometers to micrometers in size, such as colloids[52] or aggregated proteins on a pre-designed nucleation surface[53]. Computational techniques, however, have demonstrated that mechanistic studies are possible on nucleation processes from the melt at intermediate to large subcooling[29-31, 33-39, 42, 54, 55].

An approach to study crystal nucleation was originally developed by van Duijneveldt and Frenkel based on non-Boltzmann sampling over an *a priori* assumed reaction coordinate[42]. In a series of three papers, ten Wolde *et al.* applied the approach of van Duijneveldt and Frenkel[42] to describe the formation of Lennard-Jones crystals from the melt[29-31]. The chosen reaction coordinate is a set of bond-orientational OPs[56] used to convert the liquid particles to solid particles. The work of ten Wolde *et al.* demonstrated that the crystals typically form first as body centered cubic [bcc] clusters and then convert to the more stable face centered cubic [fcc] structure. In a later study, Moroni, ten Wolde, and Bolhuis applied transition path sampling [TPS] to the same system and demonstrated that the critical nuclei can exist as large, more disordered bcc-like nuclei as previously found, or as smaller, compact fcc nuclei[38]. The three earlier publications of ten Wolde *et al.* were most certainly pioneering investigations at the time of publication[29-31], and have been generally and successfully applied to many systems by other groups[33-37, 57, 58]. However, this later study from Moroni *et al.*, made possible by the development of TPS, highlights the point that assuming a reaction

coordinate *a priori* does not guarantee comprehensive and accurate results[38]. The study from Moroni *et al.* emphasizes the need for the use of unbiased methods in determining mechanisms in nucleation problems, even for seemingly simple transformations such as the freezing of Lennard-Jones particles.

To the knowledge of the author, TPS has not been applied to nucleation problems in complex systems in which the molecules are anisotropic, such as disk shaped molecules like benzene. It is anticipated that the anisotropy of benzene will yield critical nuclei that do not behave as a sphere, as assumed in CNT. To investigate the nucleation mechanism, a similar approach as Moroni *et al.* is taken [38]. In this case, aimless shooting and likelihood maximization are applied to determine the mechanism of nucleation in benzene from a subcooled liquid to the Form I crystal.

5.3 Methodology

This section describes the simulation methods in detail. The overall approach and a discussion of the physical conditions and system size are first provided. The force field is introduced and validated with equilibrium MD simulations. A labeling system is described in detail that enables discernment between liquid and crystalline molecules, based on previous work of ten Wolde *et al.*[29-31] The method used to obtain the initial trajectory is discussed and quantitative basin definitions are given. A revised version of aimless shooting is described, which is implemented for computational efficiency, and relevant parameters used in the aimless shooting simulations are provided. The OPs screened in likelihood maximization are introduced as well.

5.3.1 General Approach

The overall approach to determine the nucleation mechanism is similar to that described in Chapters 2 and 4 for the solid state polymorph transformation in terephthalic acid. Because the goal of this study is solely to examine nucleation, the reactant basin is classified as the liquid state and the product basin is classified as the liquid state and a post-critical nucleus of the Form I crystal. Because nucleation and growth are two different processes, the growth of the crystal throughout the cell is not considered in aimless shooting. Additionally, as will be shown, studying crystal growth with the entire simulation cell incurs significant computational cost. The overall methodology is as follows:

1) An initial shooting point for aimless shooting is collected from a melting trajectory that connects the reactant and product basins,

2) Aimless shooting is applied at a fixed temperature and density (i.e. the *NVT* ensemble) to collect several hundred trajectories,

3) Likelihood maximization is utilized to extract the best reaction coordinate approximation for the nucleation event.

The physical conditions and the system size for the simulation are intimately related, and therefore merit comment. The temperature and pressure chosen for the simulation will determine the size of the critical nucleus. At deeper subcooling for example, the critical nucleus will be smaller. It is therefore in the interest of efficiency to choose a temperature at some intermediate subcooling (ΔT >10 K for instance) rather than within several degrees K of coexistence. For the selection of system size, it is conventional to simulate a system 10-50x the size of the critical nucleus to avoid finite size effects[39, 42]. Selecting a size 10-50x the critical nucleus requires having some *a priori* knowledge of the approximate size of the critical nucleus. The critical nucleus size is therefore roughly approximated for the temperature and density chosen in a manner similar to locating putative transition states and quantifying basin definitions as described in Dellago *et al.*[48] The chosen temperature is 250 K and an average density corresponding to 0.1 GPa is used. Experimentally, this temperature corresponds to a subcooling of 50 K[3]. It will be shown that at 250 K, the critical nucleus size is around several hundred molecules. Therefore, a system size of 6912 molecules is used for all the simulations described.

5.3.2 Force Field

The OPLS force field is applied to benzene[5]. This potential, developed by Jorgensen *et al.*[5], is convenient for application in pre-packaged codes such as CHARMM[59] and NAMD[60] because it is fit to compatible functional forms. It should be noted that a more accurate potential for benzene has been developed, but the non-conventional functional form used severely limits widespread applicability[12]. The OPLS potential is parameterized to reproduce experimental quantities such as liquid density and heat of vaporization, and it has also been shown to reasonably reproduce the Form I crystal phase[12]. The intermolecular parameters for

the OPLS benzene force field are listed in Table 5-1. The intramolecular force field parameters are taken from the CHARMM force field library[59].

Table 5-1. Intermolecular force field parameters for benzene from the OPLS potential[5]

Atom	σ[Å]	ε[kcal/mol]	q [e]
C	3.55	0.07	-0.115
H	2.42	0.03	0.115

5.3.3 Equilibration

To determine the average crystal density at various temperatures and pressures of interest, equilibrium MD simulations are conducted in the *NPT* ensemble. NAMD is used to collect all MD trajectories for this study with a 1.0 fs time step[60]. The cutoff for nonbonded interactions is 14 Å, and the switching function is activated at 12 Å. Partice mesh Ewald summation is applied to account for long range electrostatic interactions with 6^{th} order interpolation[61]. The simulations conducted in the *NPT* ensemble utilize Langevin temperature control with a damping coefficient of 5 ps^{-1}, and the Nose-Hoover barostat is applied with a piston period of 200 fs and a damping time scale of 100 fs[61].

Three 400 ps equilibrium *NPT* trajectories are initiated in the crystal phase to equilibrate the system volume and determine the effect of temperature and pressure on density. The deviation from the experimental lattice parameters is monitored. Table 5-2 shows the average values of the lattice parameters for the three cases in the *NPT* ensemble. As shown, the lattice parameters deviate less than 8% for all cases. As mentioned previously, the crystallization simulations for aimless shooting are conducted at 250 K and 0.1 GPa.

Table 5-2. Lattice parameters from the experimental crystal structure and *NPT* MD simulations at various conditions

Lattice Parameter	Crystal Structure[2]	250 K, 0.1 GPa	300 K, 0.1 GPa	250 K, 10^{-4} GPa
a [Å]	88.776	86.378 ± 0.129	88.003 ± 0.158	87.644 ± 0.147
b [Å]	113.22	115.748 ± 0.092	116.724 ± 0.113	116.641 ± 0.096
c [Å]	81.336	87.434 ± 0.142	87.715 ± 0.160	87.806 ± 0.139

5.3.4 Obtaining an Initial Shooting Point for Aimless Shooting Simulations

To collect an initial trajectory connecting the Form I crystal and the liquid, the following approach is taken:

1) A crystal cell is equilibrated in the *NPT* ensemble at 250 K and 0.1 GPa.

2) The volume is constrained at the average value and the system is heated to 400 K. The trajectory is run at 400 K until the cell is a disordered liquid.

3) Randomly seeded trajectories are initiated from configurations along the melting trajectory to locate a putative transition state.

These randomly seeded trajectories are shot from several points with different starting cluster sizes. The cluster size is determined at the end of the 100 ps trajectories to find a putative transition state. Figure 5-1 shows the results 10 trajectories shot from a point with 205 solid molecules. The results are plotted as $N_{cluster}$ with $N_o = 205$ molecules as shown by the dashed line.

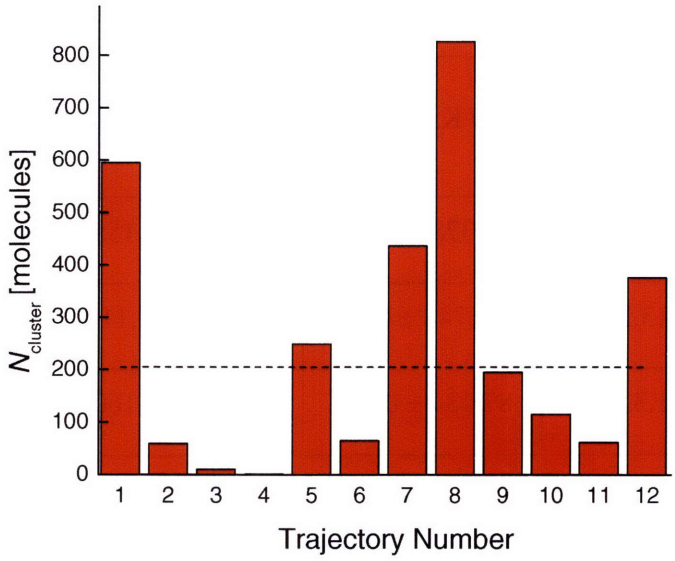


Figure 5-1. Number of solid molecules in a cluster at the end of 12 randomly seeded 100 ps MD trajectories. These trajectories are collected from a point along the initial melting trajectory to search for a putative transition state from which to initiate aimless shooting. The starting configuration contains 205 molecules, as indicated by the dashed line.

5.3.5 Mean Square Displacement

The mean square displacement is monitored in the disordered state to ensure that the disordered state is a liquid rather than a glass-like state in which the molecules have no ability to diffuse in a translational manner. Also the mean square displacement is monitored in the solid state to ascertain if the crystal is stable at these temperatures. Four equilibrium simulations are conducted at 250 K and 300 K, both at 0.1 GPa, for the liquid and solid states. The liquid is prepared by superheating the crystal to 500 K and melting in a 1 ns simulation, then cooled down to the temperature of interest. An equilibration period is conducted at the subcooled temperature of interest of 100 ps. Figure 5-2 depicts the mean square displacement for each of these simulations. As shown, the liquid systems exhibit translational diffusion at both 300 and 250 K, and the crystals are stable over the timescale of the simulations.

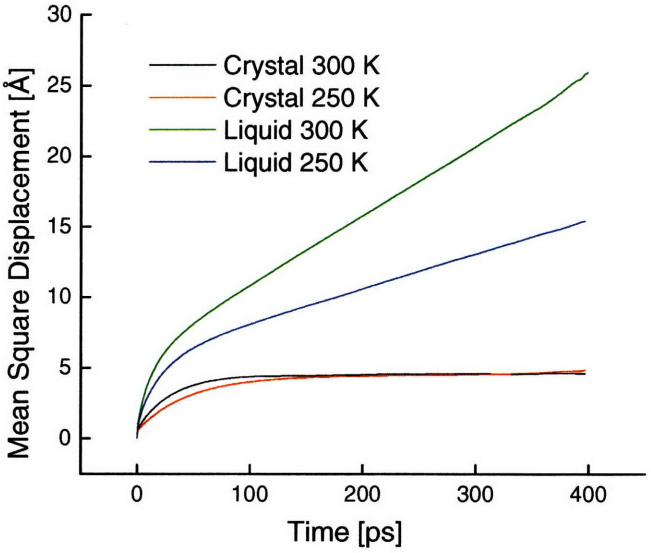


Figure 5-2. Mean square displacement results for the Form I benzene crystal and liquid at 250K and 300 K and 0.1 GPa. These results indicate that benzene can remain in the liquid state at 250 K (rather than a glassy state), and that the crystals are stable over a long time scale at both 300 and 250 K at 0.1 GPa.

5.3.6 Molecule Labeling System

To properly examine the disorder-order transformation, which is hypothesized to occur via nucleation on a local length scale, it is essential to discern between solid and liquid particles. The approach of ten Wolde *et al.* is adopted here to label solid-like and liquid-like molecules during the course of the simulation[29-31]. The bond-orientational order parameters of Steinhardt *et al.* are used to monitor the crystallinity of individual molecules[56]. These local paramters are defined as:

$$\bar{q}_{\mathrm{lm}}(i) \equiv \frac{1}{N_{\mathrm{b}}(i)} \sum_{j=1}^{N_{\mathrm{b}}(i)} Y_{\mathrm{lm}}(\hat{\mathbf{r}}_{ij}) \tag{5-1}$$

Where $Y_{\mathrm{lm}}(\hat{\mathbf{r}}_{ij})$ are the spherical harmonic functions, $\hat{\mathbf{r}}_{ij}$ is a unit vector in the direction of the "bonds" between molecule i and its neighbor molecule j summed over all $N_{\mathrm{b}}(i)$ neighbors. The 6$^{\mathrm{th}}$ order bond orientational order parameter, which has been shown to be a viable OP in other studies, is selected as the specific spherical harmonic function to apply. To each molecule i, a local (2x6+1)-dimensional vector $\mathbf{q}_6(i)$ with components calculated as:

$$\tilde{q}_{6m}(i) \equiv \frac{\bar{q}_{6m}(i)}{\left[\sum_{m=-6}^{m=6} \left|\bar{q}_{6m}(i)\right|^2\right]^{1/2}} \tag{5-2}$$

The dot product of the resulting $\mathbf{q}_6$ vectors for each molecule i and neighbor j is defined as:

$$\mathbf{q}_6(i) \cdot \mathbf{q}_6(j) \equiv \sum_{m=-6}^{m=6} \tilde{q}_{6m}(i) \tilde{q}_{6m}(j)^* \tag{5-3}$$

Ordered molecules, which will vibrate around a lattice position, should have a non-zero value for the dot product given in Equation 5-3, whereas liquid molecules should have a wide distribution of values centered close to zero. Therefore, two molecules are considered to be "connected" if $\mathbf{q}_6(i) \cdot \mathbf{q}_6(j)$ exceeds a specific value. As ten Wolde *et al.* mention in their studies on Lennard-Jones nucleation, at any given instance, a liquid molecule can also be labeled crystalline at any given instance because it can attain a coherent, but fleeting, bond orientationally ordered state. Therefore, the second criterion of ten Wolde *et al.* is added in this study as well: a molecule is considered crystalline if the number of connections exceeds a certain value, i.e. if the dot product given in Equation 5-3 exceeds the threshold value for a minimum number of neighbors.

Equilibrium trajectories are harvested for both the crystal and liquid to determine:

1) the proper cutoff distance to use for the bond orientational OPs;

2) the threshold values for the $\mathbf{q}_6(i) \cdot \mathbf{q}_6(j)$ expression; and

3) the necessary number of connections to label a particle as crystalline.

Figure 5-3 shows the center of mass radial distribution functions (RDF) for equilibrated Form I benzene crystals and the liquid at 250 K and 300 K. The RDF for the crystalline systems is identical and the liquid shows only slight deviation. Figure 5-4 shows the center of mass locations of the nearest neighbors taken from a frame in the equilibrium simulation. From the crystal RDFs, the cutoff distance to search for nearest neighbors is set to 8.5 Å.

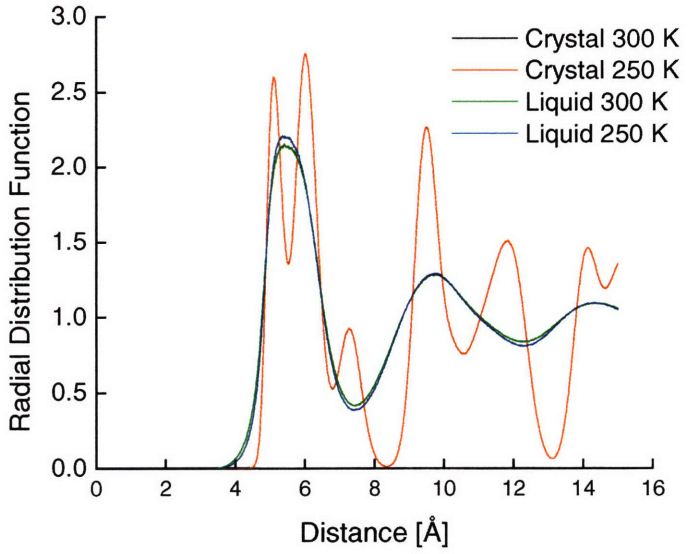


Figure 5-3. Center of mass radial distribution functions for the Form I benzene crystal and liquid at 250K and 300 K at 0.1 GPa. The first coordination shell for crystalline benzene exists out to approximately 8.5 Å and this distance is therefore used to search for nearest neighbors when calculating the bond orientational OPs. The radial distribution functions for the crystals at 250 and 300 K at 0.1 GPa are identical.

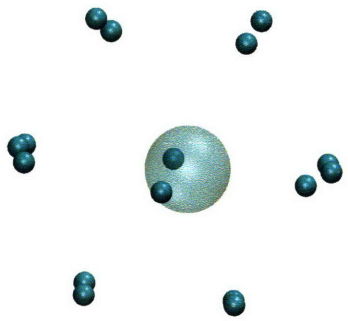


Figure 5-4. Center of mass positions for the 16 nearest neighbors of a single benzene molecule. This snapshot is taken from an equilibrium simulation at 250 K and 0.1 GPa. The neighbors are shown as small beads and the center molecule is shown in transparent spacefill.

Once the cutoff distance for the $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ expression is set, the distributions for crystal and liquid can be calculated. Figure 5-5 displays the distribution function for the $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ results. These distributions are collected from 20 independent configurations from the crystal and liquid simulations at 250 K, and the results from the 300 K simulation are

indistinguishable. The distributions suggest that the $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ cutoff is 0.08, which is used in the molecule labeling.

As mentioned, ten Wolde *et al.* also introduced a second criterion in the labeling process related to the number of connections per particle[29-31]. Figure 5-6 shows the number of connections per particle for both a crystal and a liquid simulation. These results are collected from the same simulations used to generate the $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ distributions, shown in Figure 5-5. The connection threshold is set to 5 connections. Therefore a molecule i with 5 nearest neighbors having $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)>=0.08$ will be labeled as crystalline, and all other molecules will be labeled as liquid.

In addition, a cluster analysis is performed to determine the cluster sizes for any configuration. The algorithm determines which particles in the simulation cell are crystalline, then searches over nearest neighbors to etch out the cluster size. This is prepared to characterize the basins for aimless shooting, as will be discussed.

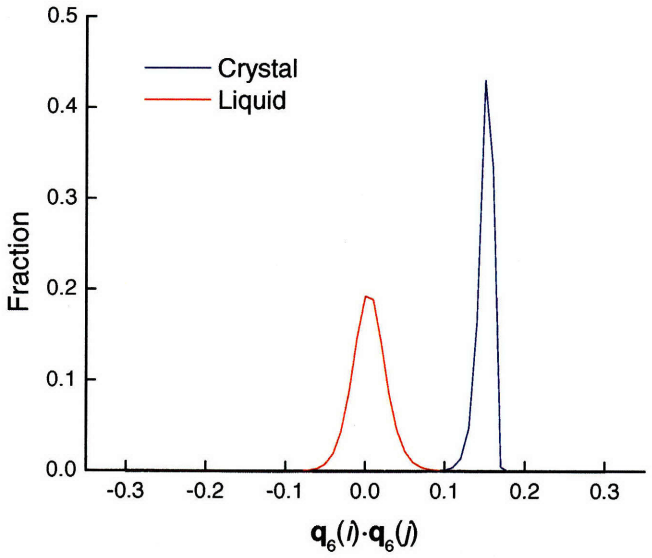


Figure 5-5. Distribution of $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ function given by Equation 5-3 for all neighboring particles i and j in an equilibrated crystal and liquid at 250 K and 0.1 GPa. These distributions are averaged over 20 independent frames.

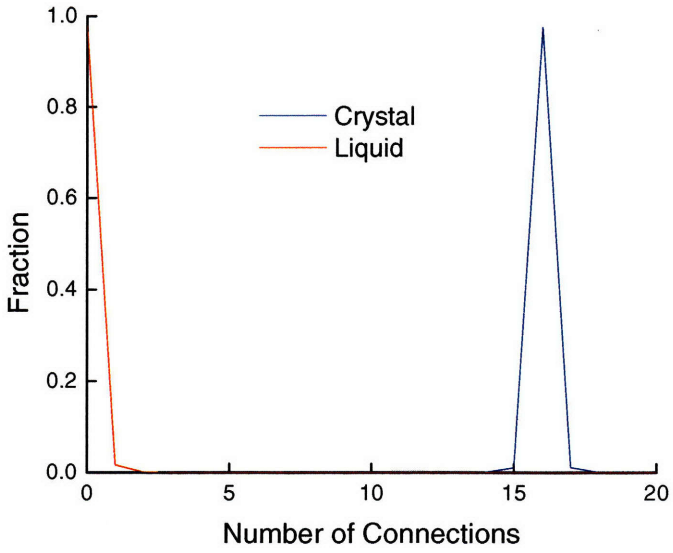


Figure 5-6. Distributions of the number of nearest neighbor connections per particle in an equilibrated crystal and liquid at 250 K and 0.1 GPa. These distributions are averaged over 20 independent frames.

5.3.7 Aimless Shooting

Aimless shooting requires basin definitions and an initial shooting point. The collection of the initial aimless shooting point is described previously. The basin definitions are based on the number of solid benzene molecules in the cluster. The reactant basin is defined as a configuration with less than 150 molecules in the cluster, and the product is defined as a configuration with greater than 450 molecules in the nucleus.

As shown in the Figure 5-1, the trajectory length required to commit to the basin is in some cases greater than 100 ps. For example, points 5 and 9 do not commit to either basin in a 100 ps MD trajectory. Therefore, the total length of the aimless shooting trajectories is set to 600 ps to maintain the frequency of inconclusive trajectories below 10%. The trajectory length of 600 ps makes the aimless shooting simulations for benzene nucleation computationally very expensive relative to the aimless shooting simulations described in Chapters 2 and 4.

5.3.8 Likelihood Maximization

Since the aimless shooting simulations are extremely computationally expensive, likelihood maximization will be conducted after the collection of several hundred trajectories. To date,

approximately 90 trajectories have been collected. As described in Chapter 3, the reaction coordinate will be approximated with the half trajectory likelihood using Equations 3-3 and 3-4. The OPs screened include

1) nucleus size,

2) nucleus surface area as measured by the solvent accessible surface area,

3) nucleus shape as measured by the principal moments of inertia,

4) cluster crystallinity as measured by the Steinhardt OPs[56] (Q_4, Q_6, Q_8, W_4, W_6, W_8),

5) global crystallinity as measured by the Steinhardt OPs[56] (Q_4, Q_6, Q_8, W_4, W_6, W_8),

6) nucleus shape as measured by the moments of inertia,

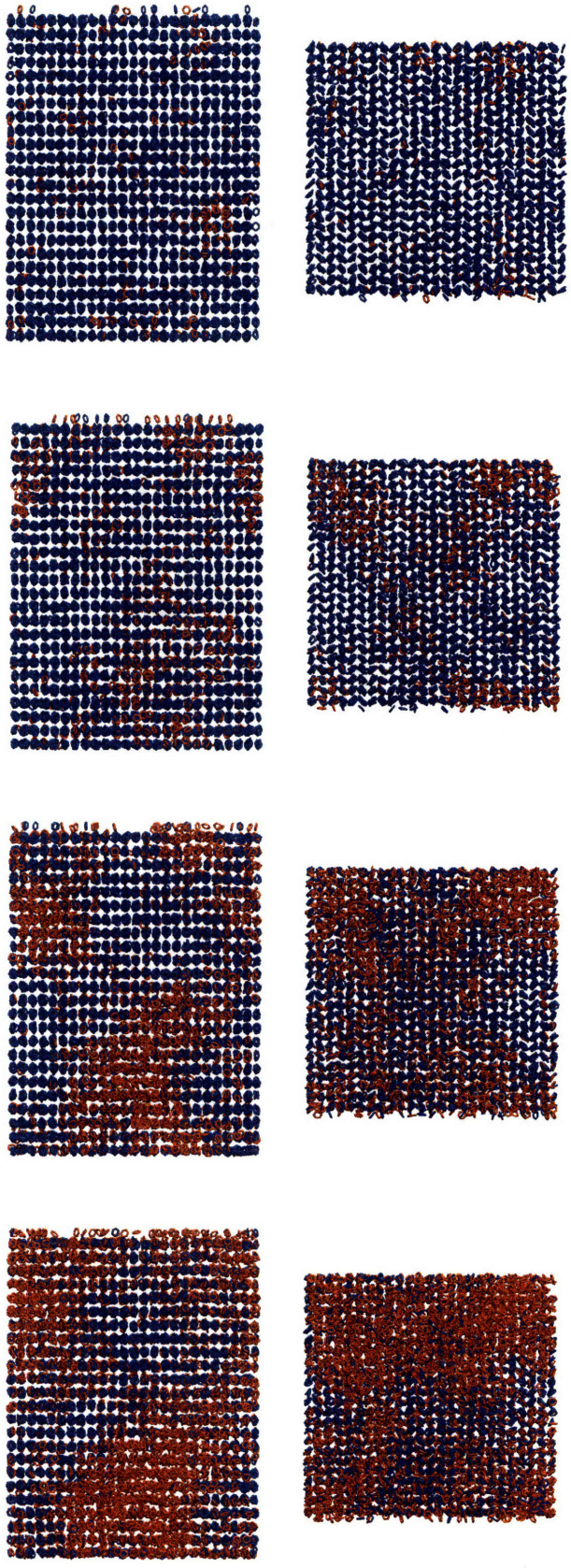
7) energetic OPs such as the potential energy of the cluster surface molecules.

The nucleus surface area is measured with the solvent accessible surface area algorithm in CHARMM[59]. The nucleus shape is measured by the same method utilized by Pool and Bolhuis[62]. The cluster crystallinity is measured by locating the nearest neighbor connections between all of the molecules in the cluster, and the global crystallinity is measured by applying the same method to the entire simulation cell[29-31, 42, 56].

5.4 Results

5.4.1 Initial Trajectory

Seven snapshots of the initial trajectory are shown in Figure 5-7. As described previously, the trajectory goes from a crystal state equilibrated at 250 K and is superheated to 400 K until the cell is a disordered liquid. The point used for aimless shooting is similar to the penultimate snapshot shown in Figure 5-7 with a solid cluster with 320 molecules. The last configuration shown in Figure 5-7 is composed entirely of liquid benzene molecules.



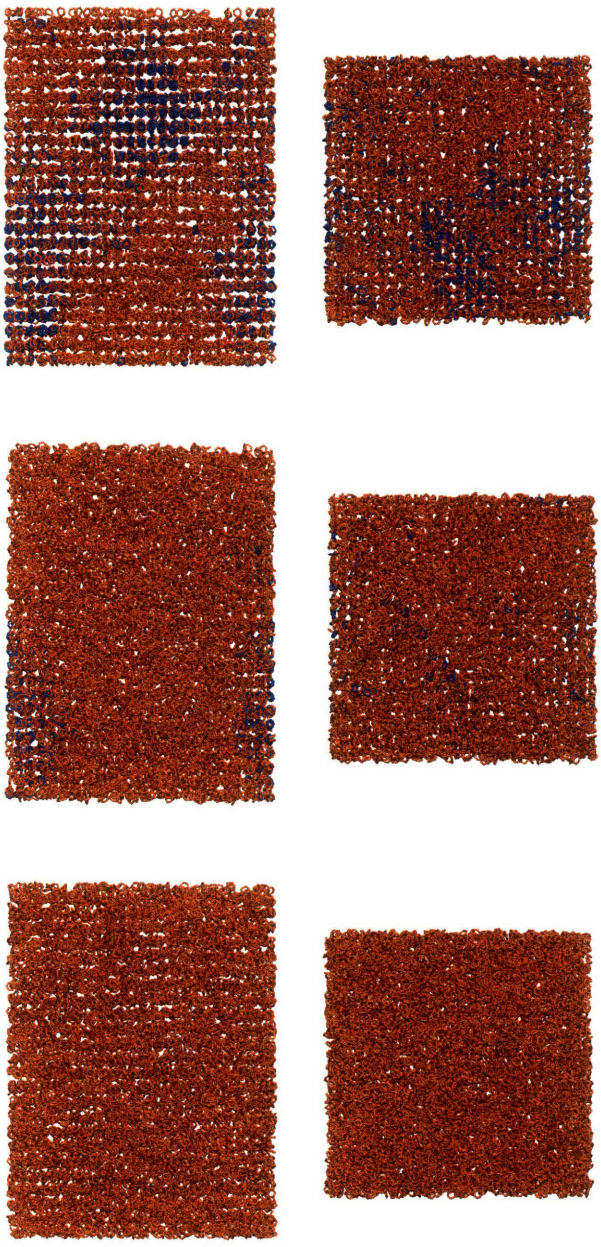


Figure 5-7. Snapshots from the a-axis and c-axis views along the initial melting trajectory at 400 K. The liquid molecules are red and the solid molecules are blue. The respective views are shown in a 1:1 scale. The penultimate snapshot is similar to the configurations used for input to aimless shooting. The last snapshot is completely liquid.

5.4.2 Aimless Shooting and Likelihood Maximization

Aimless shooting simulations are ongoing. The nucleus size, as obtained from the 80 aimless shooting trajectories obtained thus far is between 200 and 300 molecules in size. An example

of a nucleus from a successful aimless shooting point is shown in Figure 5-8. As shown, the nucleus, labeled with blue surface shading, is far from spherical.

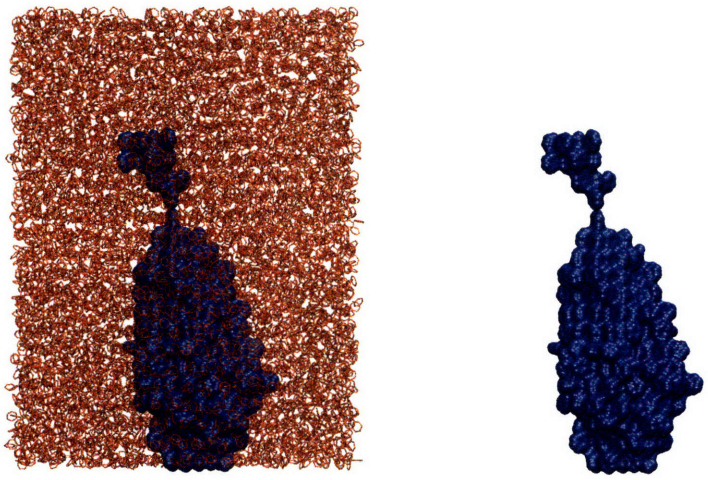


Figure 5-8. Form I crystal nucleus from a reactive aimless shooting trajectory. The nucleus molecules are labeled with blue surface shading and the liquid molecules are shown in red wireframe.

5.4.3 Characterization of Growth Time

The time for crystal growth throughout the entire cell is not studied in aimless shooting as the trajectory length would make the simulations computationally intractable. However, to characterize the growth time, several endpoints from aimless shooting trajectories that nucleate are collected, and 2 ns trajectories are shot. This is done to characterize the time for growth throughout the entire crystal cell. Figure 5-9 shows configurations from the **b**-axis view of an example of a trajectory for which the entire crystal cell crystallizes in 2 ns. Again, solid molecules are labeled in blue and liquid molecules in red.

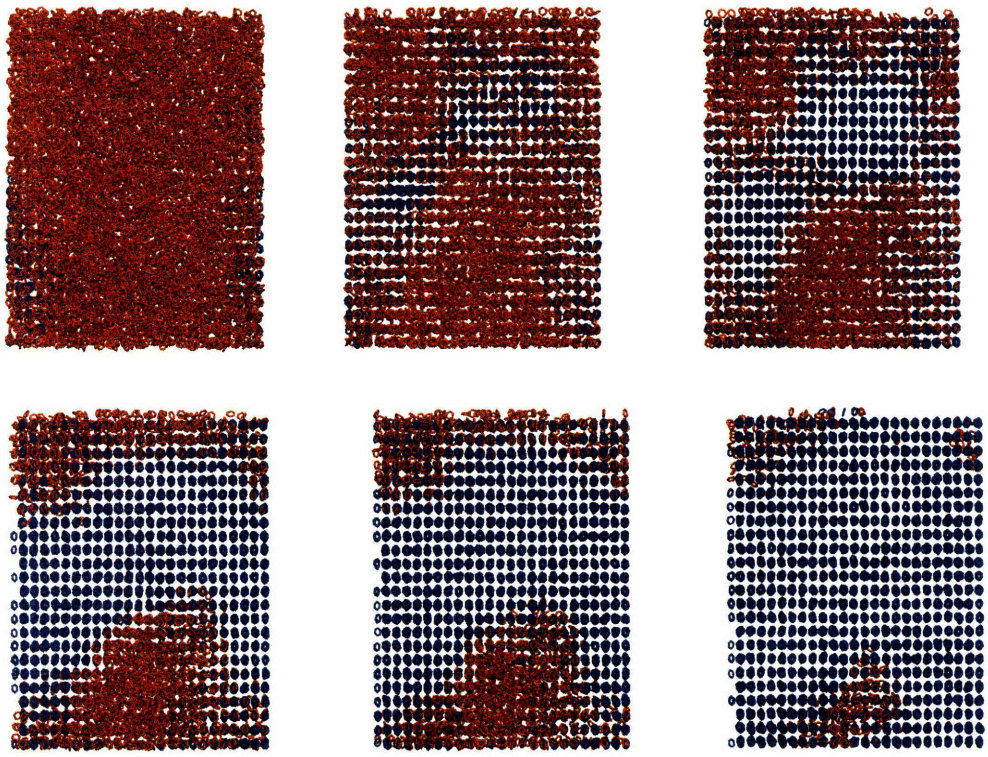


Figure 5-9. Snapshots from the b-axis views along the growth trajectory at 250 K. Solid molecules are shown in blue and liquid molecules are labeled red.

5.5 Discussion

For very large systems in which the commitment time is long, TPS becomes incredibly computationally expensive. A single aimless shooting trajectory for benzene is 1.2 ns with a system size of 84000 atoms in periodic boundaries. For a statistically relevant collection of trajectories, this amount to simulation times up to hundreds of ns. For very large systems such as benzene in which the basin commitment time is long, TPS becomes incredibly computationally expensive. This study therefore highlights the need for further methodological developments to sample diffusive transition paths in a manner that allows the extraction of reaction coordinates. The transition interface sampling [TIS] approach of van Erp *et al.* is based on calculating fluxes through pre-defined interfaces along a transition to extract rate constants[63, 64]. TIS is based on similar principles as TPS and can therefore easily be implemented in pre-packaged molecular simulation codes. However, unlike TPS (and aimless shooting), TIS assumes a pre-defined reaction coordinate. It is speculated that the choice of the reaction coordinate does not lead to large error[64], but this has yet to be verified.

In addition, there is no systematic method to extract the reaction coordinate from TIS results. Assuming that van Erp and Bolhuis are correct in stating that the choice of reaction coordinate is not important in the interface definition, merging of likelihood maximization and TIS could yield a methodology to simultaneously calculate rate constants and elucidate mechanisms in diffusive systems.

5.6 Summary and Conclusions

This chapter describes the study of the homogeneous nucleation from the subcooled melt of benzene to the Form I crystal. Aimless shooting is applied in the *NVT* ensemble to the liquid-solid transition at approximately 50 K subcooling at a density corresponding to 0.1 GPa. Preliminary results indicate that the critical nucleus size is approximately 200-300 molecules at these physical conditions.

Likelihood maximization will be applied after several hundred aimless shooting trajectories are collected. Nucleus size, shape, and surface area will be screened as approximations to the reaction coordinate. To our knowledge, the liquid to solid phase transformation has only been definitively elucidated for a model system[38, 39] and for water in an *a postieri* manner[40]. Thus, the crystallization mechanism in benzene represents a nucleation process of unprecedented molecular complexity. This study also demonstrates that aimless shooting and likelihood maximization are effective methods to elucidate nucleation mechanisms in complex systems.

5.7 References

1. Bridgman, P.W. *J. Chem. Phys.*, **9,** p. 794-797 (1941).
2. Bacon, G.E., N.A. Curry, and S.A. Wilson. *Proc. Roy. Soc. London*, **279,** p. 98-110 (1954).
3. Akella, J. and G.C. Kennedy. *J. Chem. Phys.*, **55,** p. 793-796 (1971).
4. Thiery, M.M. and J.M. Leger. *J. Chem. Phys.*, **89,** p. 4255-4271 (1988).
5. Jorgensen, W.L. and D.L. Severance. *J. Amer. Chem. Soc.*, **112,** p. 4768-4774 (1990).
6. Thiery, M.M., J.M. Besson, and J.L. Bribes. *J. Chem. Phys.*, **96,** p. 2633-2654 (1992).
7. Cansell, F., D. Fabre, and J.P. Petitet. *J. Chem. Phys.*, **99,** p. 7300-7304 (1993).
8. Craven, C.J., et al. *J. Chem. Phys.*, **98,** p. 8236-8243 (1993).
9. Craven, C.J., P.D. Hatton, and G.S. Pawley. *J. Chem. Phys.*, **98,** p. 8244-8255 (1993).
10. Thiery, M.M. and C. Rerat. *J. Chem. Phys.*, **104,** p. 9079-9089 (1996).
11. van Eijck, B.P., et al. *Acta. Cryst.*, **B54,** p. 291-299 (1998).
12. Cacelli, I., et al. *J. Amer. Chem. Soc.*, **126,** p. 14278-14286 (2004).
13. Sinnokrot, M.O. and C.D. Sherrill. *J. Phys. Chem. A*, **108,** p. 10200-10207 (2004).
14. Macias, A.T. and A.D. MacKerell. *J. Comput. Chem.*, **26,** p. 1452-1463 (2005).
15. Raiteri, P., R. Martonak, and M. Parrinello. *Angew. Chem. Int. Ed.*, **44,** p. 3769-3773 (2005).
16. Sinnokrot, M.O. and C.D. Sherrill. *J. Phys. Chem. A*, **110,** p. 10656-10668 (2006).
17. Budzianowski, A. and A. Katrusiak. *Acta. Cryst.*, **B62,** p. 94-101 (2006).
18. Azreg-Ainou, M. *J. Cryst. Growth*, **299,** p. 195-205 (2007).
19. Kearley, G.J., M.R. Johnson, and J. Tomkinson. *J. Chem. Phys.*, **124,** p. 044514 (2006).
20. Azreg-Ainou, M., A. Huseynov, and B. Ibrahimoglu. *J. Chem. Phys.*, **124,** p. 204505 (2006).

21. Amovilli, C., et al. *Theor. Chem. Acc.*, **117,** p. 885-901 (2007).
22. Hunter, C.A., et al. *J. Chem. Soc. - Perkin Trans.*, **2,** p. 651-669 (2001).
23. Jennings, W.B., B.M. Farrell, and J.F. Malone. *Acc. Chem. Res.*, **34,** p. 885-894 (2001).
24. Debenedetti, P.G., *Metastable liquids: Concepts and principles*. 1st Edition ed. 1996: Princeton University Press.
25. Mullin, J.W., *Crystallization*. 4th Edition ed. 2001: Butterworth-Heinemann.
26. Sastry, S. *Nature*, **416,** p. 376-377 (2002).
27. Debenedetti, P.G. *AIChE J.*, **51,** p. 2391-2395 (2005).
28. Debenedetti, P.G. *Nature*, **441,** p. 168-169 (2006).
29. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *Faraday Discuss.*, **104,** p. 93-110 (1996).
30. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *J. Chem. Phys.*, **104,** p. 9932-9947 (1996).
31. ten Wolde, P.R., M.J. Ruiz-Montero, and D. Frenkel. *Phys. Rev. Lett.*, **75,** p. 2714-2717 (1995).
32. Valeriani, C., E. Sanz, and D. Frenkel. *J. Chem. Phys.*, **122,** p. 194501 (2005).
33. Desgranges, C. and J. Delhommelle. *J. Amer. Chem. Soc.*, **128,** p. 15104-15105 (2006).
34. Desgranges, C. and J. Delhommelle. *J. Chem. Phys.*, **126,** p. 054501 (2007).
35. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Amer. Chem. Soc.*, **126,** p. 12286-12287 (2004).
36. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Chem. Phys.*, **122,** p. 184518 (2005).
37. Leyssale, J.M., J. Delhommelle, and C. Millot. *J. Chem. Phys.*, **122,** p. 104510 (2005).
38. Moroni, D., P.R. ten Wolde, and P.G. Bolhuis. *Phys. Rev. Lett.*, **94,** p. 235703 (2005).
39. Trudu, F., D. Donadio, and M. Parrinello. *Phys. Rev. Lett.*, **97,** p. 105701 (2006).
40. Radhakrishnan, R. and B.L. Trout. *J. Amer. Chem. Soc.*, **125,** p. 7743-7747 (2003).
41. Matsumoto, M., S. Saito, and I. Ohmine. **416,** p. 409-413 (2002).
42. van Duijneveldt, J.S. and D. Frenkel. *J. Chem. Phys.*, **96,** p. 4655-4668 (1992).
43. Bolhuis, P.G., et al. *Annu. Rev. Phys. Chem.*, **53,** p. 291-318 (2002).
44. Bolhuis, P.G., C. Dellago, and D. Chandler. *Faraday Discuss.*, **110,** p. 421-436 (1998).
45. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **110,** p. 6617-6625 (1999).
46. Dellago, C., P.G. Bolhuis, and D. Chandler. *J. Chem. Phys.*, **108,** p. 9236-9245 (1998).
47. Dellago, C., et al. *J. Chem. Phys.*, **108,** p. 1964-1977 (1998).
48. Dellago, C., P.G. Bolhuis, and P.G. Geissler. *Adv. Chem. Phys.*, **123,** p. 1-86 (2001).
49. Turnbull, D. and J.C. Fisher. *J. Chem. Phys.*, **17,** p. 71-73 (1949).
50. Oxtoby, D.W. *Acc. Chem. Res.*, **31,** p. 91-97 (1998).
51. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
52. Gasser, U., et al. *Science*, **292,** p. 258-262 (2001).
53. Yau, S.T. and P.G. Vekilov. *Nature*, **406,** p. 494-497 (2000).
54. Anwar, J. and P.K. Boateng. *J. Amer. Chem. Soc.*, **120,** p. 9600-9604 (1998).
55. Auer, S. and D. Frenkel. *Nature*, **409,** p. 1020-1023 (2001).
56. Steinhardt, P.J., D.R. Nelson, and M. Ronchetti. *Phys. Rev. B*, **28,** p. 784-805 (1983).
57. Shen, V.K. and P.G. Debenedetti. *J. Chem. Phys.*, **111,** p. 3581-3589 (1999).
58. Radhakrishnan, R. and B.L. Trout. *J. Chem. Phys.*, **117,** p. 1786-1796 (2002).
59. MacKerell, A.D., J. Wiórkiewicz-Kuczera, and M. Karplus. *J. Amer. Chem. Soc.*, **117,** p. 11946-11975 (1995).
60. Phillips, J.C., et al. *J. Comput. Chem.*, **26,** p. 1781-1802 (2005).
61. Frenkel, D. and B. Smit, *Understanding molecular simulations: From algorithms to applications*. 2nd ed. Computational science series. Vol. 1. 2002.
62. Pool, R. and P.G. Bolhuis. *J. Chem. Phys.*, **126,** p. 244703 (2007).
63. Moroni, D., P.G. Bolhuis, and T.S. van Erp. *J. Chem. Phys.*, **120,** p. 4055-4065 (2004).
64. van Erp, T.S. and P.G. Bolhuis. *J. Comp. Phys.*, **205,** p. 157-181 (2005).

CHAPTER 6: SUMMARY AND FUTURE WORK

6.1 Summary

The primary objective of this thesis is to elucidate nucleation mechanisms in molecular systems. Newly developed molecular simulation techniques are applied that allow one to systematically identify mechanisms in an unbiased manner. This approach represents a significant methodological improvement over using biased methods in which a mechanism is assumed *a priori*[1] and using trial and error methods to obtain the reaction coordinate[2-5]. Two specific problems were investigated: a solid state polymorph transformation in a hydrogen bonded molecular crystal, terephthalic acid, and the nucleation of liquid benzene to the Form I crystal.

Chapters 2-4 describe the mechanism of the solid state polymorph transformation in terephthalic acid. This particular application is motivated by the occurrence of solid state polymorph transformations in the pharmaceutical industry[6], and therefore guided the selection of the model compound. Terephthalic acid, which is bonded via a common supramolecular synthon found in small molecule therapeutics[7], is shown to undergo a nucleation event initiated at the corner of the crystal. For smaller system sizes, the global average lattice parameters can be used to approximate the reaction coordinate[8]. It is shown that the global lattice parameters serve as the reaction coordinate because for very small crystals, the fluctuations on the surface that drive the transformation give rise to large fluctuations in the global lattice parameters.

As the system size is increased, the mechanism becomes more localized, as shown for the 8x8x8 and 10x10x10 systems in Chapter 4. In all cases, the barrier to nucleation scales with the length of the supramolecular synthons, but the barrier to growth is negligible. This suggests an explanation for the experimental observations: first for the significant reluctance to undergo the transformation unless mechanically perturbed or massively subcooled, and second for the extremely fast growth kinetics once the transformation begins[9, 10]. The nucleus of the more stable Form II polymorph is also identified. A cluster analysis shows that the average size of the Form II nucleus at the transition state is approximately 80 molecules, and a distinct interface exists between the two polymorphs. This corner nucleation event yields the smallest possible interfacial area between the two polymorphs at a given nucleus size.

This observation provides evidence that solid state polymorph transformations will occur where the free energy penalty associated with creating a surface is minimized. In light of classical nucleation theory, undergoing the mechanism that minimizes the free energy penalty associated with the creation of a surface justifies a corner nucleation event. To our knowledge, this study represents the first successful mechanistic investigation of solid state polymorph transformations in molecular crystals, leading to an enhanced understanding of nucleation processes in complex systems at the molecular level. Further study is warranted to examine other possible pathways of nucleation in solid state polymorph transformations, which can eventually provide a comprehensive picture of nucleation in the solid state.

Chapter 5 examines the homogeneous nucleation event in a subcooled melt of benzene to the Form I crystal. Aimless shooting is applied in the NVT ensemble to the liquid-solid transition at approximately 50 K subcooling at a density corresponding to 0.1 GPa. Preliminary results indicate that the critical nucleus size is approximately 200 to 300 molecules at these physical conditions. Likelihood maximization will be applied after several hundred aimless shooting trajectories are collected. Nucleus size, shape, and surface area will be screened as approximations to the reaction coordinate. To our knowledge, the liquid to solid phase transformation has only been definitively elucidated for a model system[11, 12] and for water in an *a postieri* manner[13]. Thus, the crystallization mechanism in benzene represents a nucleation process of unprecedented molecular complexity.

From a methodological standpoint, this thesis represents the first two applications of aimless shooting and likelihood maximization as effective, complementary methods to identify the reaction coordinate of rare events in complex systems. Knowledge of a good reaction coordinate can highlight the factors that influence experimental reaction rates. These methods should also be implemented prior to either free energy or rate constant calculations, as spurious reaction coordinates will yield misleading results for quantities of interest[14, 15].

6.2 Future Work

As discussed in Chapter 1 and demonstrated in Chapters 2-5, computational approaches provide valuable mechanistic insight into the nucleation process that experiments cannot probe. Future computational work should be considered in two parts:

1) applications of the current methodology to different systems, and

2) methodological advances that are needed to expand the ability of simulations to probe questions regarding nucleation.

Four applications are proposed that merit future study and two methodological needs for computational nucleation studies are discussed.

6.2.1 Applications of Aimless Shooting and Likelihood Maximization to Nucleation Processes

1. *Solid state polymorph transformation in terephthalic acid*: A natural extension of Chapters 2 and 4 involve further study of terephthalic acid. One advantage of using this system is that an effective potential has been developed, as described in Chapter 2. This study could be two-fold: first, one can investigate the size of the critical nucleus in systems much larger than those examined here. As mentioned in the conclusions of Chapter 4, one should scale up by a factor of at least 5-10x or more to determine if the mechanism remains the same and if so, if the corner nucleus size is stable at approximately 80 to 100 molecules. As described in Chapter 4, it is anticipated that this will be the case. The second portion of the study could examine the polymorph transformation in full periodic boundary conditions [PBCs]. The mechanism in PBCs would presumably be different than with free surfaces. Targeted molecular dynamics could be applied to obtain a putative transition state for aimless shooting simulations. Free energy calculations should be performed to directly compare the two mechanisms. It is hypothesized that the free energy barrier would be much higher in a perfect bulk crystal, and therefore this approach would provide a definitive comparison to the results obtained with the systems in vacuum.

   Further extension, both in free boundaries and in the bulk crystal, would involve the introduction of defects and their influence on the mechanism and free energy barrier. Although there are many types of defects possible in molecular crystals, this problem would be of significant further interest to industry.

2. *Solid state polymorph transformation in other systems*: Another set of project involve the application of similar methods to other systems of interest. One possible molecule to investigate is DL-norleucine. Two studies have been published by Anwar *et al.* concerning DL-norleucine that argue that solid state polymorph transformations are martensitic, although the authors concede that there is still some debate surrounding this

issue[16, 17]. To support their hypothesis, Anwar *et al.* developed a potential for crystalline DL-norleucine[16] and presented MD simulations that show entire crystal layers displaced simultaneously in PBCs[17]. The authors argue that their study represents interface advancement in solid state polymorph transformations. However, "interface advancement" seems to be crystal growth in this case. The two studies by Anwar *et al.* compared to the work presented in this thesis highlight the need for further study of both nucleation and growth processes for these types of transformations, and the need for a clear delineation between nucleation and growth. Specifically, one portion of a future project would involve investigating the nucleation event in DL-norleucine with free boundaries or defects, while the second project would involve studying the growth process of terephthalic acid in the hope of reconciling the views presented in the two different approaches.

Another type of study could examine solid state polymorph transformations with different packing motifs and transformations in which hydrogen bonds are broken and reformed. The hydrogen bonded supramolecular chains in terephthalic acid are shown to remain intact during the polymorph transformation. However, many crystals undergo transformations in which the intermolecular connectivity is perturbed more significantly based on the crystal structures. The α–β polymorph transformation in resorcinol, for instance, requires that intermolecular hydrogen bonds be broken[18]. It is expected that the mechanism will involve both significant intermolecular and intramolecular rearrangement. Other examples of molecules that exhibit this behavior and have received significant attention in the literature are oxalic acid[19] and ROY[20]. Herbstein outlines a large set of molecular systems that exhibit polymorph transformations in the solid state, and provides an excellent summary of current experimental knowledge for each system[21].

3. *Secondary and heterogeneous nucleation*: It is common practice in the pharmaceutical industry to use seeding techniques to induce secondary nucleation[22]. Also, crystallization induced by impurities or surfaces in vessels is a common pathway for nucleation to occur. Frenkel *et al.* have reported two studies on heterogeneous nucleation in hard sphere colloids[23, 24], but substantial questions still remain, especially in systems that exhibit polymorphism. The recent work of Yu *et al.* demonstrates that cross nucleation can readily occur in molecular systems[25-27]. Determining the mechanism of cross nucleation

or heterogeneous nucleation in a molecular system would yield insight on how to control nucleation of specific polymorphs.

4. *Pressure driven nucleation in molecular solids*: Crystals in the pharmaceutical industry are subjected to significant mechanical and thermal stress during processing[22, 28]. As shown for single crystals at metastable conditions, mechanical perturbations can induce transformation[9, 10, 29-32]. Molecular simulation packages allow the application of pressure in an anisotropic manner when PBCs are applied. This allows for the study of pressure driven transformations in the bulk. However, recent advances by Dellago and coworkers have enabled researchers to investigate the effect of pressure on systems with free surfaces[33, 34]. The method is based on the use of an ideal gas pressure bath as a thermostat and the pressure medium. The pressure of the ideal gas particles can be adjusted by adjusting the number of particles. Implementation of this method into molecular simulation packages will allow for the study of pressure induced transformations in molecular crystals at surfaces and defects.

6.2.2 Methodological Improvements for Nucleation Studies

1. *Potential development for crystallization simulations*: The force field used to describe energetic interactions is a key limitation in many crystal simulations[8, 16, 35]. Tuble *et al.* presented a method for developing force fields for crystal simulations[16]. Their method involves screening sets of partial charges derived from quantum calculations with a dielectric continuum. For the Lennard-Jones intermolecular interactions, parameters are screened from conventional force field libraries. This approach was slightly amended for terephthalic acid in Chapter 2 and provided a viable force field. However, many organic crystals have highly flexible intramolecular degrees of freedom, and would therefore require another level of parameter screening. For instance, the α and β polymorphs of malonic acid have significant torsion angle differences in the carboxylic acid groups.[36, 37] Development of a general technique to screen force fields for crystal simulations would be an incredibly valuable tool for researchers interested in solving problems related to nucleation and crystallization. It is recommended that several case studies be included in a study to highlight different problems that may arise in the potential development, such as substantially dissimilar intramolecular conformations, different supramolecular interconnectivity for the same molecule, etc.

2. *Implementation and development of methods to study diffusive crystal nucleation problems*: The nucleation of a solute from a supersaturated solution is an important pathway by which products are separated from solvent in industry[28]. Experimental approaches to study nucleation from solution suffer from similar disadvantages as melt nucleation. However, a significant challenge to computational approaches also exists: specifically, the diffusive nature of the transition incurs significant computational expense, as shown in Chapter 5. For a transition path sampling [TPS] approach, this makes the commitment time overly long and severely limits the number of trajectories that can be collected.

   Several methods have been developed to examine these types of transitions, with varying success. The transition interface sampling [TIS] approach of van Erp *et al.* is based on calculating fluxes through pre-defined interfaces along a transition to extract rate constants[38]. TIS is based on similar principles as TPS and can therefore easily be implemented in pre-packaged molecular simulation codes. However, unlike TPS (and aimless shooting), TIS assumes some pre-defined reaction coordinate. It is speculated that the choice of the reaction coordinate does not lead to large error[38], but this has yet to be verified. In addition, there is no systematic method to extract the reaction coordinate from TIS results. Assuming that van Erp and Bolhuis are correct in stating that the choice of reaction coordinate is not important in the interface definition[38], merging of likelihood maximization[14, 15] and TIS could yield a methodology to simultaneously calculate rate constants and elucidate mechanisms in diffusive systems.

   The minimum free energy path [MFEP] method, developed by Vanden-Eijnden *et al.* is an elegant method to determine the free energy in a complex process along a path of maximum likelihood[39, 40]. The MFEP method differs from likelihood maximization in that it uses all possible collective variables as the reaction coordinate rather than using the transition state ensemble to extract the most important variables to approximate the reaction coordinate. However, one advantage of the MFEP method over aimless shooting for certain cases is that the implementation might be more computationally efficient. Also, the transition state configurations at the top of the free energy barrier can be characterized *a postieri.*

6.3 References

1. van Duijneveldt, J.S. and D. Frenkel. *J. Chem. Phys.*, **96,** p. 4655-4668 (1992).
2. Du, R., et al. *J. Chem. Phys.*, **108,** p. 334-350 (1998).
3. Geissler, P.G., C. Dellago, and D. Chandler. *J. Phys. Chem. B*, **103,** p. 3706-3710 (1999).
4. Dellago, C., P.G. Bolhuis, and P.G. Geissler. *Adv. Chem. Phys.*, **123,** p. 1-86 (2001).
5. Quaytman, S.L. and S.D. Schwartz. *Proc. Natl. Acad. Sci. U.S.A.*, **104,** p. 12253-12258 (2007).
6. Caira, M.R. *Topics in Curr. Chem.*, **198,** p. 163-208 (1998).
7. Desiraju, G.R. *Angew. Chem. Int. Ed.*, **34,** p. 2311-2327 (1995).
8. Beckham, G.T., et al. *J. Amer. Chem. Soc.*, **129,** p. 4714-4724 (2007).
9. Davey, R.J., et al. *J. Chem. Soc. - Faraday Trans.*, **90,** p. 1003-1009 (1994).
10. Davey, R.J., et al. *Nature*, **366,** p. 248-250 (1993).
11. Moroni, D., P.R. ten Wolde, and P.G. Bolhuis. *Phys. Rev. Lett.*, **94,** p. 235703 (2005).
12. Trudu, F., D. Donadio, and M. Parrinello. *Phys. Rev. Lett.*, **97,** p. 105701 (2006).
13. Radhakrishnan, R. and B.L. Trout. *J. Amer. Chem. Soc.*, **125,** p. 7743-7747 (2003).
14. Peters, B., G.T. Beckham, and B.L. Trout. *J. Chem. Phys.*, **127,** p. 1 (2007).
15. Peters, B. and B.L. Trout. *J. Chem. Phys.*, **125,** p. 054108 (2006).
16. Tuble, S.C., J. Anwar, and J.D. Gale. *J. Amer. Chem. Soc.*, **126,** p. 396-405 (2004).
17. Anwar, J., S.C. Tuble, and J. Kendrick. *J. Amer. Chem. Soc.*, **129,** p. 2542-2547 (2007).
18. Yoshino, M., et al. *J. Phys. Chem. A*, **103,** p. 2775-2783 (1999).
19. Derissen, J.L. and P.H. Smit. *Acta Cryst.*, **B30,** p. 2240-2242 (1974).
20. Yu, L., et al. *J. Amer. Chem. Soc.*, **122,** p. 585-591 (2000).
21. Herbstein, F.H. *Acta Cryst.*, **B62,** p. 341-383 (2006).
22. Mullin, J.W., *Crystallization*. 4th Edition ed. 2001: Butterworth-Heinemann.
23. Auer, S. and D. Frenkel. *Phys. Rev. Lett.*, **91,** p. 015703 (2003).
24. Cacciuto, A., S. Auer, and D. Frenkel. *Nature*, **428,** p. 404-406 (2004).
25. Yu, L. *J. Amer. Chem. Soc.*, **125,** p. 6380-6381 (2003).
26. Chen, S.A., H.M. Xi, and L. Yu. *J. Amer. Chem. Soc.*, **127,** p. 17439-17444 (2005).
27. Tao, J. and L. Yu. *J. Phys. Chem. B*, **110,** p. 7098-7101 (2006).
28. Davey, R.J. and J. Garside, *From molecules to crystallizers*. Oxford chemistry primers. 2001: Oxford University Press.
29. Mnyukh, Y.V. *J. Cryst. Growth*, **32,** p. 371-377 (1976).
30. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 163-200 (1979).
31. Mnyukh, Y.V. *Molec. Cryst. and Liquid Cryst.*, **52,** p. 201-218 (1979).
32. Mnyukh, Y.V., *Fundamentals of solid-state phase transitions ferromagnetism and ferroelectricity*. 1st Edition ed. 1998: 1st Books Library.
33. Grünwald, M. and C. Dellago. *Mol. Phys.*, **104,** p. 3709-3715 (2006).
34. Grünwald, M., E. Rabini, and C. Dellago. *Phys. Rev. Lett.*, **96,** p. 255701 (2006).
35. Gavezzotti, A. and G. Fillippini. *J. Phys. Chem. B*, **98,** p. 4831-4837 (1994).
36. Fukai, M., T. Matsuo, and H. Suga. *Thermoch. Acta*, **183,** p. 215-243 (1991).
37. Ganguly, S., et al. *Chem. Phys. Lett.*, **69,** p. 227-229 (1980).
38. van Erp, T.S. and P.G. Bolhuis. *J. Comp. Phys.*, **205,** p. 157-181 (2005).
39. Maragliano, L., et al. *J. Chem. Phys.*, **125,** p. 024106 (2006).
40. Miller, T.F., E. Vanden-Eijnden, and D. Chandler. *Proc. Natl. Acad. Sci. U.S.A.*, **104,** p. 14559-14564 (2007).