J|A|C|S
ARTICLES
Published on Web 03/27/2007

Surface-Mediated Nucleation in the Solid-State Polymorph Transformation of Terephthalic Acid

Gregg T. Beckham,[†] Baron Peters,[†,||] Cindy Starbuck,[‡] Narayan Variankaval,[§] and Bernhardt L. Trout*[,†]

Contribution from the Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Chemical Process Development and Commercialization and Process Research, Merck & Co., Inc., Rahway, New Jersey 07065

Received December 6, 2006; E-mail: trout@mit.edu

Abstract: A molecular mechanism for nucleation for the solid-state polymorph transformation of terephthalic acid is presented. New methods recently developed in our group, aimless shooting and likelihood maximization, are employed to construct a model for the reaction coordinate for the two system sizes studied. The reaction coordinate approximation is validated using the committor probability analysis. The transformation proceeds via a localized, elongated nucleus along the crystal edge formed by fluctuations in the supramolecular synthons, suggesting a nucleation and growth mechanism in the macroscopic system.

Introduction

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.[8] 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 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.[21,22] This study focuses on the transformation between forms I and II, both of which pack in triclinic lattices. In a paper

[†] Massachusetts Institute of Technology.
[‡] Chemical Process Development and Commercialization, Merck & Co., Inc.
[§] Process Research, Merck & Co., Inc.
[||] Current Address: CECAM, Ecole Normale Supérieure, 46 Allée d'Italie, 69364 Lyon Cedex 7, France.

(1) Mullin, J. W. *Crystallization*, 4th ed.; Butterworth-Heinemann: Oxford, 2001.
(2) Matsui, M.; Price, G. D. *Nature* **1991**, *351*, 735−737.
(3) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keefe, M.; Yaghi, O. M. *Science* **2002**, *295*, 469−472.
(4) Sawada, K. *Pure Appl. Chem.* **1997**, *69*, 921−928.
(5) Spanos, N.; Koutsoukos, P. G. *J. Cryst. Growth* **1998**, *191*, 783−790.
(6) Irie, M.; Kobatake, S.; Horichi, M. *Science* **2001**, *291*, 1769−1772.
(7) Morris, K. R.; Griesser, U. J.; Eckhardt, C. J.; Stowell, J. G. *Adv. Drug Delivery Rev.* **2001**, *48*, 91−114.
(8) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Adv. Drug Delivery Rev.* **2001**, *48*, 3−26.
(9) Moulton, B.; Zaworotko, M. J. *Chem. Rev.* **2001**, *101*, 1629−1658.
(10) Yu, L. X.; Furness, M. S.; Raw, A.; Outlaw, K. P. W.; Nashed, N. E.; Ramos, E.; Miller, S. P. F.; Adams, R. C.; Fang, F.; Patel, R. M.; Holcombe, F. O., Jr.; Chiu, Y.; Hussain, A. S. *Pharm. Res.* **2003**, *20*, 531−536.
(11) Caira, M. *Top. Curr. Chem.* **1998**, *198*, 163−208.
(12) Morissette, S. L.; Soukasene, S.; Levinson, D.; Cima, M. J.; Almarsson, Ö. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2180−2184.
(13) Mnyukh, Y. *Fundamentals of Solid-State Phase Transitions Ferromagnetism and Ferroelectricity*, 1st ed.; Springer: New York, 1998.
(14) Mnyukh, Y. V. *J. Cryst. Growth* **1996**, *32*, 371−377.
(15) Mnyukh, Y. V. *Mol. Cryst. Liq. Cryst.* **1979**, *52*, 163−200.
(16) Mnyukh, Y. V. *Mol. Cryst. Liq. Cryst.* **1979**, *52*, 201−218.
(17) Cardew, P. T.; Davey, R. J.; Ruddick, A. J. *J Chem. Soc., Faraday Trans.* **1984**, *80*, 659−668.
(18) Tuble, S. C.; Anwar, J.; Gale, J. D. *J. Am. Chem. Soc.* **2004**, *26*, 396−405.
(19) Kaneko, F.; Yano, J.; Tsujiuchi, H.; Tashiro, K.; Suzuki, M. *J. Phys. Chem. B* **1998**, *102*, 327−330.
(20) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125*, 054108.
(21) Bailey, M.; Brown, C. J. *Acta Cryst.* **1967**, *22*, 387−391.
(22) Śledź, M.; Janczak, J.; Kubiak, R. *J. Mol. Struct.* **2001**, *595*, 77−82.

10.1021/ja0687567 CCC: $37.00

that for the first time demonstrated structural polymorph stabilization via twinning, Davey et al.[23,24] 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. On the basis of 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. This work forms the basis and inspiration for this investigation.

Determining pathways of polymorph transformations at the molecular level is difficult from an experimental standpoint, because of 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 similar problems. For example, nucleation has been studied for the freezing of a Lennard-Jones fluid,[25] NaCl nucleation from solution,[26] and the freezing of water.[27] The present study focuses on investigating the molecular level events leading to the transformation in TPA from form I to form II.

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 a 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, and (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 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, a trajectory space Monte Carlo procedure, is a powerful method for sampling the ensemble of transition pathways in complex systems.[28−31] A recent extension of transition path sampling, aimless shooting,[20] 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 (OPs), which describe properties of the system along reactive trajectories. Types of OPs range from bond distances[32] to parameters that measure the coordination numbers and orientations.[26,33,34] For a given system, many OPs will change "adiabatically" along the reactive trajectories. A reaction coordinate is one or more OPs that completely quantify the dynamical progress of the system along the reaction pathway.[28]

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.[20,35−37] In this study, likelihood maximization is applied to screen candidate OPs to determine the OP or set of OPs that best approximates the reaction coordinate.[20]

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.[28,38−40] 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). The paper 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

(23) Davey, R. J.; Maginn, S. J.; Andrews, S. J.; Buckley, A. M.; Cottier, D.; Dempsey, P.; Rout, J. E.; Stanley, D. R.; Taylor, A. *Nature* **1993**, *366*, 248−250.
(24) Davey, R. J.; Maginn, S. J.; Andrews, S. J.; Black, S. N.; Buckley, A. M.; Cottier, D.; Dempsey, P.; Plowman, R.; Rout, J. E.; Stanley, D. R.; Taylor, A. *J. Chem. Soc, Faraday Trans.* **1994**, *90*, 1003−1009.
(25) Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G. *Phys. Rev. Lett.* **2005**, *94*, 235703.
(26) Zahn, D. *Phys. Rev. Lett.* **2004**, *92*, 040801.
(27) Radhakrishnan, R.; Trout, B. L. *J. Am. Chem. Soc.* **2003**, *125*, 7743−7747.
(28) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys.* **2002**, *123*, 1−86.
(29) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964−1977.
(30) Dellago, C.; Bolhuis, P. G.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 9236−9245.
(31) Bolhius, P. G.; Dellago, C.; Chandler, D. *Faraday Discuss.* **1998**, *110*, 421−436.
(32) Chu, J. W.; Brooks, B. R.; Trout, B. L. *J. Am. Chem. Soc.* **2004**, *126*, 16601−16607.
(33) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. *Phys. Rev. B* **1983**, *28*, 784−805.
(34) Radhakrishnan, R.; Trout, B. L. *J. Chem. Phys.* **2002**, *117*, 1786−1796.
(35) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769−6779.
(36) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(37) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106.
(38) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334−350.
(39) Geissler, P. L.; Dellago, C.; Chandler, D. *J. Phys. Chem. B* **1999**, *103*, 3706−3710.
(40) Truhlar, D. G.; Garrett, B. C. *J. Phys. Chem. B* **2000**, *104*, 1069−1072.

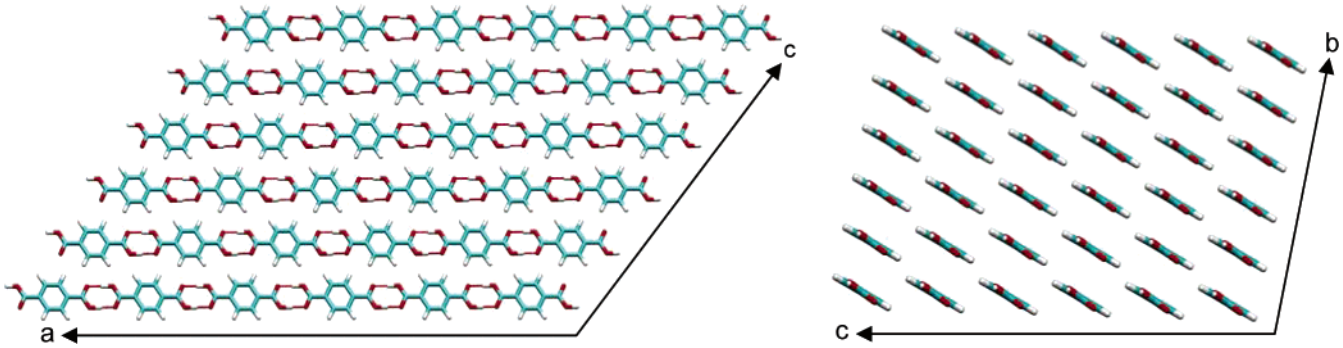

Figure 1. Form I of the 216 molecule (6 × 6 × 6) TPA crystallite from the **b**-axis view (left) and **a**-axis views (right). Hydrogen bonds that form the supramolecular synthons are shown in red-dotted lines between the carboxylic acid groups on each molecule.

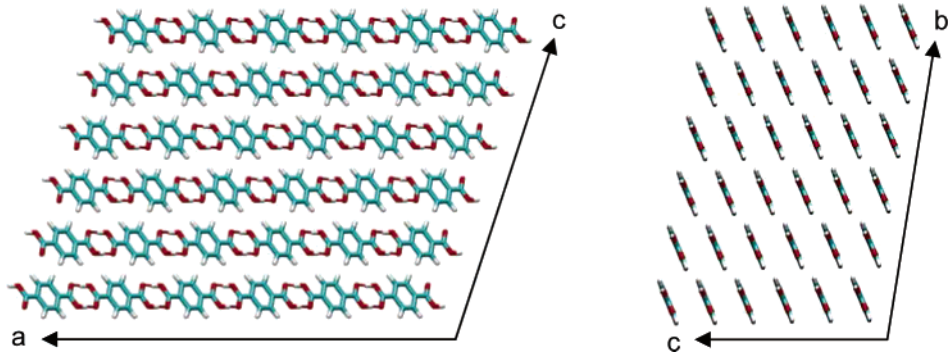

Figure 2. Form II of the 216 molecule (6 × 6 × 6) TPA crystallite from the **b**-axis view (left) and **a**-axis views (right). Hydrogen bonds that form the supramolecular synthons are shown in red-dotted lines between the carboxylic acid groups on each molecule.

conclusions are provided highlighting the physical insight gained into the TPA polymorph transformation.

Methodology

System Description. Adjacent molecules in TPA are connected via 1-dimensional hydrogen-bonded carboxylic acid dimers, termed synthons.[41] These chains, or supramolecular synthons, pack in 2-dimensional sheets. As discussed by Bailey and Brown,[21] 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 6 × 6 × 6 and a 7 × 7 × 7 geometry. These sizes correspond to 216 and 343 molecules, respectively. Figures 1 and 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 6 × 6 × 6 is counted by the number of molecules in the supramolecular synthon, not the crystal unit cells: for example, note that in Figure 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 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.[21] 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[42]) are used either in minimization or molecular dynamics, any minor errors in the hydrogen positions leading to unfavorable configurations are eliminated.

Table 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

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.[18,43,44] A modified CHARMM potential[45] 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.[46] The B3PW91 density functional[47] 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[48,49] (MSK)

(41) Desiraju, G. R. *Angew. Chem., Int. Ed.* **1995**, *34*, 2311−2327.
(42) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(43) Gavezzotti, A.; Fillippini, G. *J. Phys. Chem.* **1994**, *98*, 4831−4837.
(44) Raabe, G. *Z. Naturforsch., A: Phys. Sci.* **2002**, *57a*, 961−966.
(45) MacKerell, A. D.; Wiórkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(46) Frisch, M. J.; et al. *Gaussian 03*, revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.
(47) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

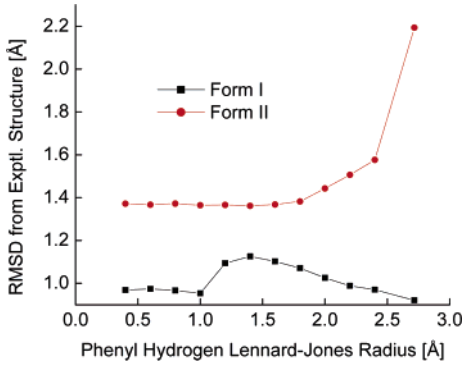

Figure 3. The rmsd as a function of the Lennard-Jones phenyl hydrogen radius for Form I (black square) and Form II (red circle), calculated with the Merz−Singh−Kollman[48,49] partial atomic charge set.

and ChelpG[50] 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 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.

Order Parameters. Over 100 OPs are tested with the likelihood maximization algorithm The selection of trial OPs 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 OPs 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, 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), 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 nonuniformities in the local lattice parameter values, say, for the horizontal layers in Figure 2.

Other OPs 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.

Aimless Shooting. The aimless shooting algorithm, described by Peters and Trout[20], is applied to harvest an ensemble of independent reactive trajectories. As with the transition path sampling method,[28−31] 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 on the basis of 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 plus or minus 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 either to 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[51] 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[45] 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 time step of 1 fs is used with a cutoff for nonbonded 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: $x_{-\Delta t}$, x_0, $x_{+\Delta t}$. We found selecting from two points, $x_{-\Delta t}$, or $x_{+\Delta t}$ is sufficient to sample the transition-state ensemble, and that is what we did in this study.

Trajectories are initiated from points thought to be close to the transition-state region based on 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.

Likelihood Maximization. As described in Peters and Trout,[20] 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:

(48) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.
(49) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.
(50) Breneman, C. M.; Wilberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(51) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690−7697.

$$r(\mathbf{q}) = \alpha_0 + \sum_{k=1}^{m} \alpha_k q_k \quad (1)$$

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:[20]

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

As pointed out by Peters and Trout,[20] 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,

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

over all coefficients and all combinations of OPs determines the best reaction coordinate according to the models of eqs 1 and 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 approximate transition-state isosurface.

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 probability analysis.[28] 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 eq 1.[20,28]

Constructing a p_B histogram requires many configurations from which to shoot. The generation of these configurations is inspired by the BOLAS method.[52] Several random, aimless shooting points are selected close to the predicted transition-state region, as defined by $r(\mathbf{q}) = 0$ in eq 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.[28] 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.

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 $p_B = 0.5$.[28] 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 $\pm 1\%$ 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.

Results

Initial Trajectories. As mentioned previously, for the 216 molecule system, trajectories initiated in the form I basin transform on the order of 1 ns either to form II or to an intermediate basin as measured by the lattice parameters. Figure 4 shows the average lattice parameters for a trajectory of each type: one of which becomes kinetically trapped in the intermediate basin (Figure 4a) and one of which does not (Figure 4b). For reference, the average lattice parameters for a representative trajectory initiated in the form II basin are also shown in Figure 4. Snapshots at germane points during these trajectories are shown in Figures 5 and 6, respectively. As seen in Figures 4b (for the lattice parameters) and 6 (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 5, 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 I orientation. The second snapshot (at 0.60 ns) in Figure 6 also displays this type of interface between the polymorphs, but quickly transforms completely to the form II polymorph.

To confirm the behavior seen in the initial trajectories as illustrated by Figures 4−6, 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 7 shows 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 4−5 kcal/mol, as explained below.

As discussed in the Methodology section, MD umbrella sampling is used to obtain initial configurations for the transformation in the 343 molecule system. Figure 8 shows the results for the potential of mean force (PMF). The PMF curve confirms the behavior observed in the 216 molecule system with a metastable intermediate between the two polymorphs. The free energy barrier to nucleation is approximately 12−13 kcal/mol and the barrier to growth is approximately 3 kcal/mol. Figure 9 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 9.

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,[20] that is, trajectories that reach neither basin in the prespecified 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 con-

(52) Radhakrishnan, R.; Schlick, T. *J. Chem. Phys.* **2004**, *121*, 2436−2444.

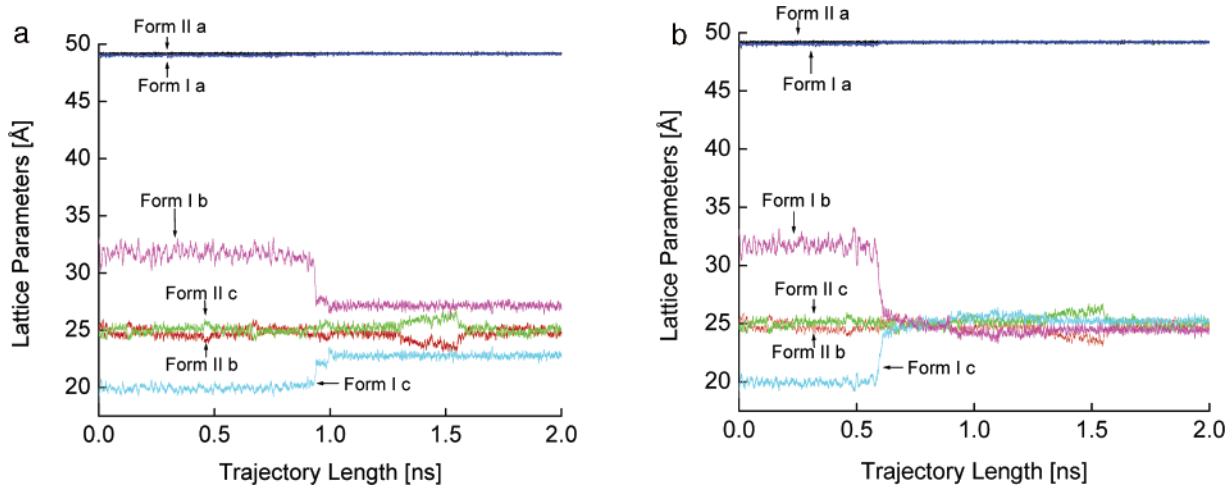

Figure 4. Crystal lattice parameters along a 2 ns trajectory for forms I and II of the 216 molecule system. (a) The change in the lattice parameters at approximately 1.0 ns corresponds to the unbiased nucleation event along the edge of the TPA crystal. The growth phase, however, is not seen in this trajectory 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.

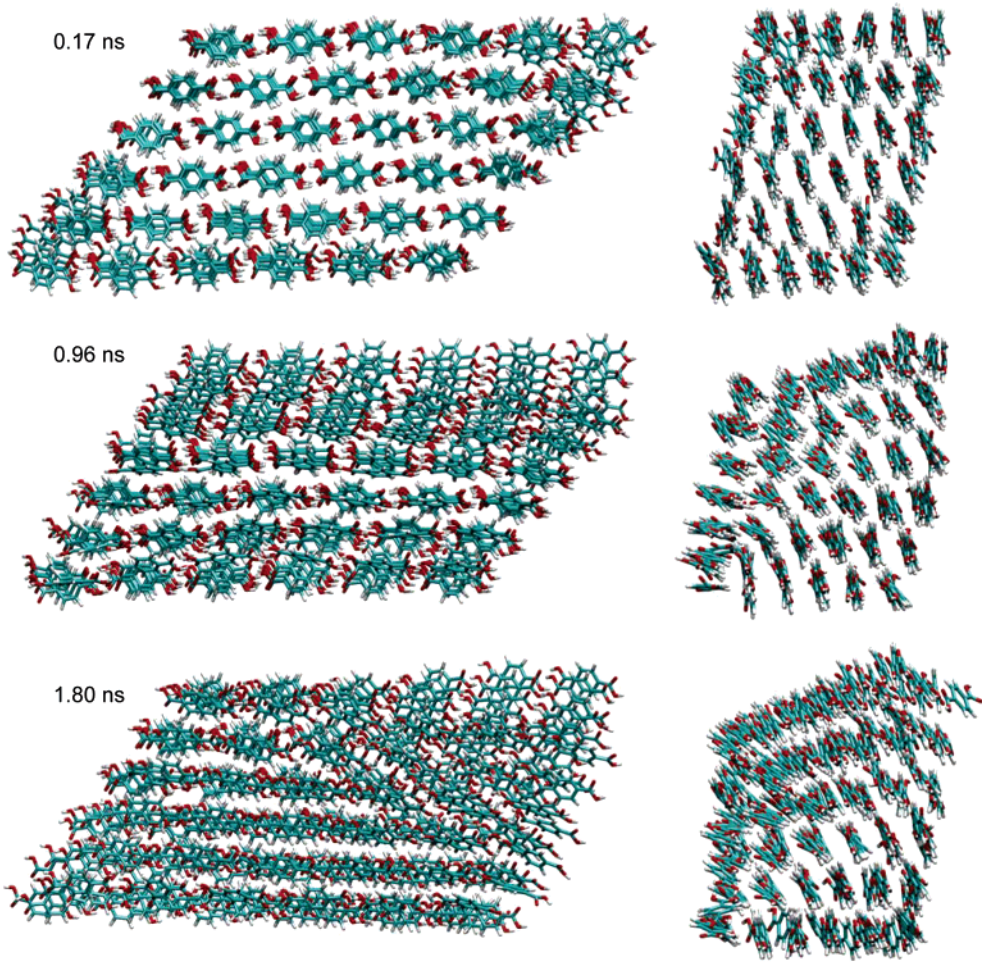

Figure 5. Snapshots from the **b**-axis and **a**-axis view along the initial trajectory for the 216 molecule system. These configurations correspond to the lattice parameters shown in Figure 4a, which becomes kinetically trapped after the nucleation event. From top to bottom, the configurations are taken at 0.17 ns (pretransformation), 0.96 ns (during nucleation), and 1.8 ns (postnucleation, 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.

figurations corresponding to the intermediate form, characteristics of both crystal forms appear. Visualization of the trajectories, as shown in Figures 5, 6, and 9, 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

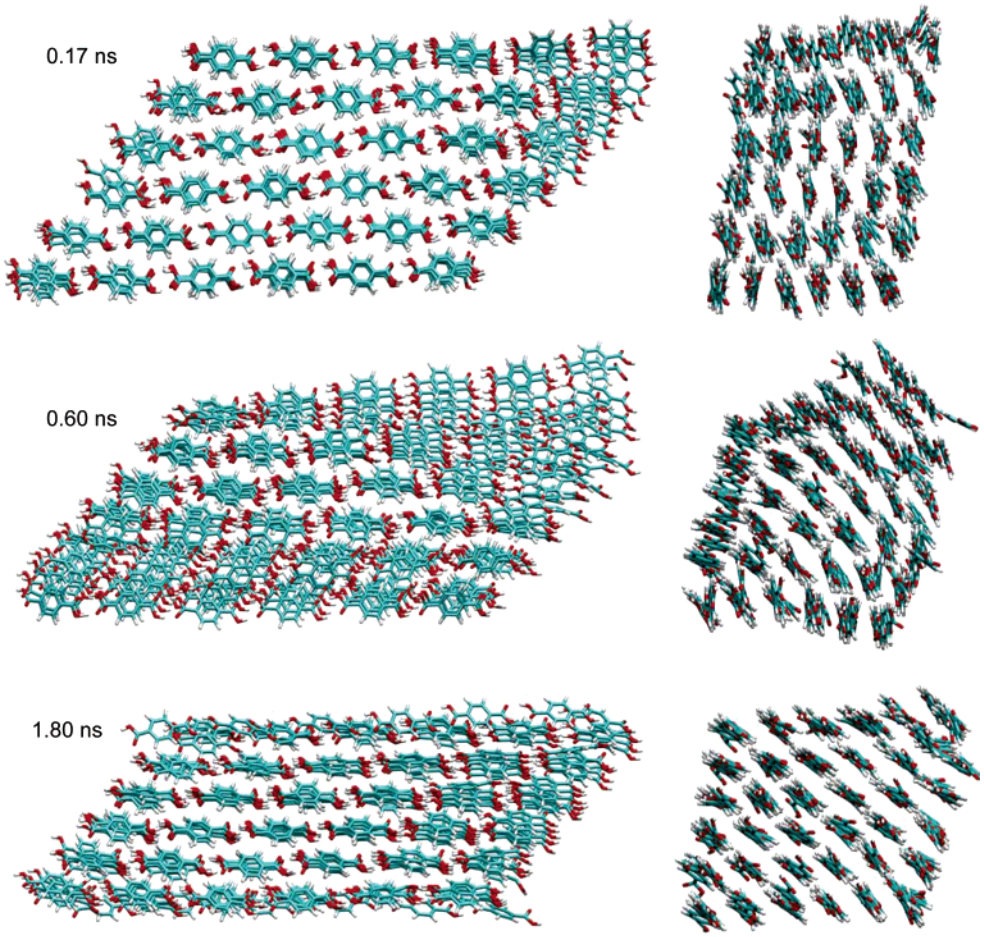

Figure 6. Snapshots from the **b**-axis and **a**-axis views along the initial trajectory for the 216 molecule system. These configurations correspond to the lattice parameters shown in Figure 4b, which undergoes both nucleation and growth. From top to bottom, the configurations are taken at 0.17 ns (pretransformation), 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).

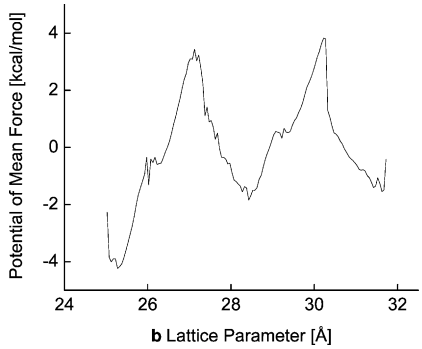

Figure 7. Potential of mean force for the 216 molecule TPA system calculated from 8 windows of MD umbrella sampling along the **b** lattice parameter.

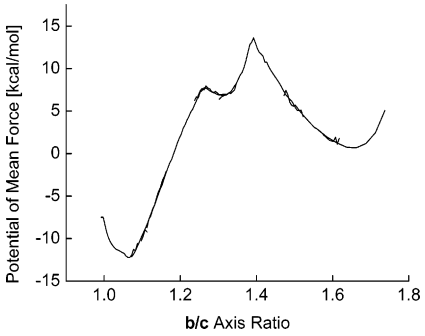

Figure 8. Potential of mean force for the 343 molecule TPA system calculated from 8 windows of MD umbrella sampling along the **b/c** axis ratio.

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.

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

From Table 2, one can see that the average **b** lattice vector is the most important parameter in the reaction coordinate approximation for the 216 molecule system. For the larger

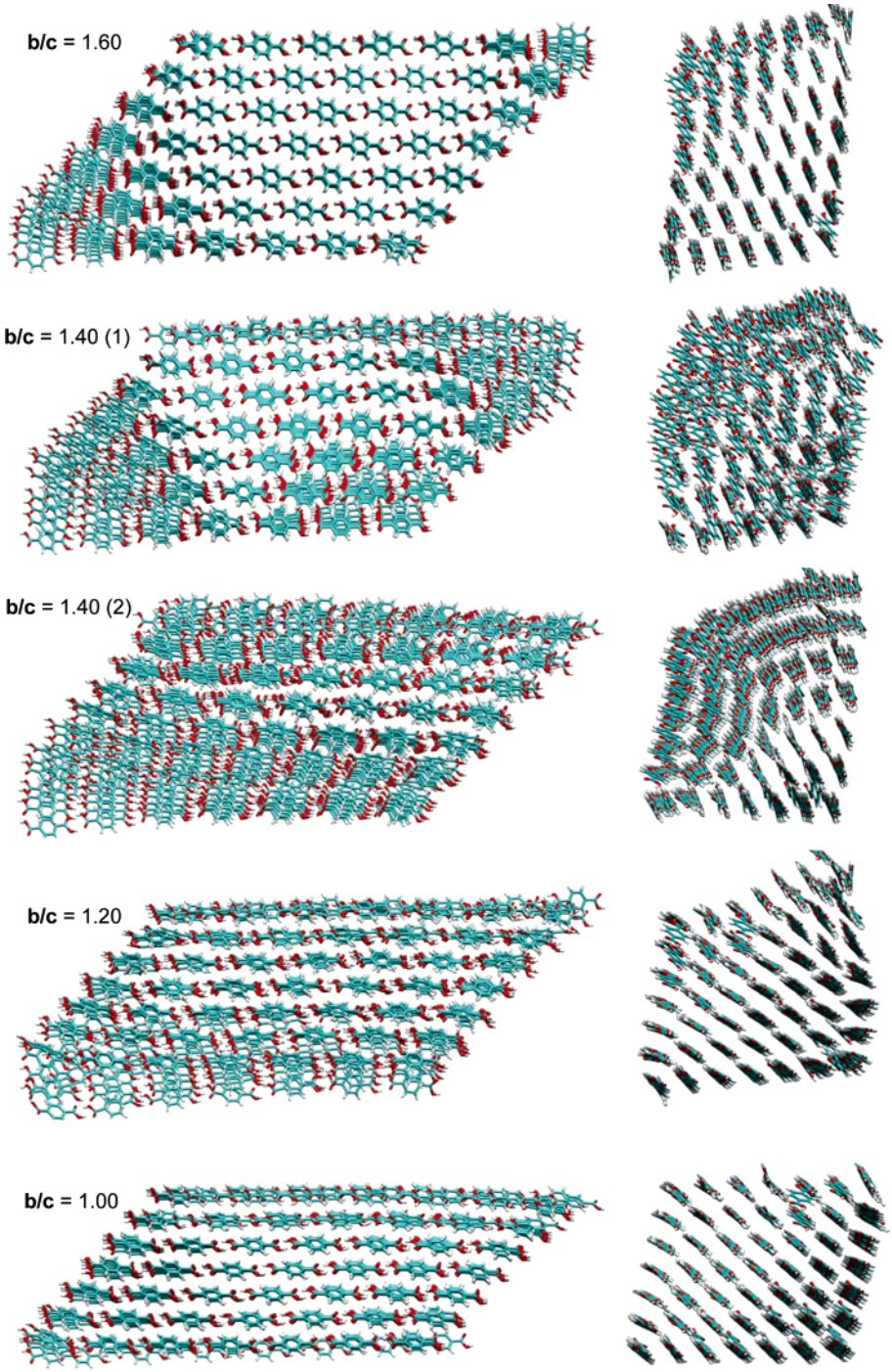

Figure 9. Snapshots from the **b**-axis and **a**-axis views along the initial trajectory for the 343 molecule system obtained from MD umbrella sampling. These configurations correspond to the PMF curve shown in Figure 8. From top to bottom, the configurations are taken at **b**/**c** = 1.60 (pretransformation), **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).

system size studied, however, the ratio of the relevant 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.

Reaction Coordinate Validation. For the 216 molecule system, a histogram is constructed from the predicted transition-state region for 1 OP: $\mathbf{b}^{\ddagger} = 30.195 \pm 0.014$ Å, which corresponds to a p_B window centered at 0.5 with a total width of 0.01. Figure 10a 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, possibly arising from an inadequate amount of statistics. 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 10b. This histogram is sharply peaked near $p_B = 0.5$, which denotes that it is a good approximation to the reaction coordinate and that

Table 2. Likelihood Maximization Results 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	$\mathbf{b}/\mathbf{c}^{\ddagger} = 1.433$

[a] The $p(\mathrm{TP}|r)$ model given in eq 2 is used to calculate the likelihood function as shown in eq 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 $q_i \in [0, 1]$.

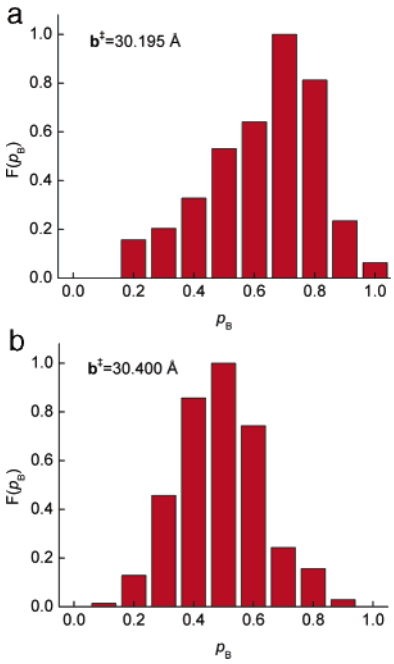

Figure 10. (a) Committor probability histogram for 216 molecule system with $\mathbf{b}^{\ddagger} = 30.195$ Å, as predicted from the $p(\mathrm{TP}|r)$ model with a 1-D reaction coordinate model: $\mu_h = 0.629$, $\sigma_h = 0.182$. (b) Committor probability histogram for 216 molecule system with $\mathbf{b}^{\ddagger} = 30.400$ Å: $\mu_h = 0.490$, $\sigma_h = 0.144$.

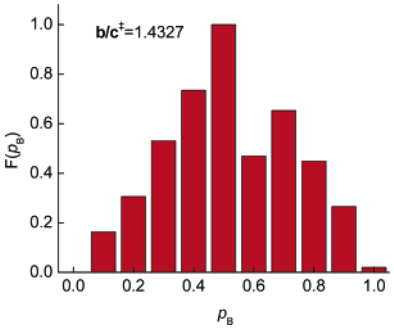

Figure 11. Committor probability histogram for 343 molecule system with $\mathbf{b}/\mathbf{c}^{\ddagger} = 1.4333$, as predicted from the $p(\mathrm{TP}|r)$ model with a 1-D reaction coordinate model: $\mu_h = 0.520$, $\sigma_h = 0.208$.

the value of the transition-state isosurface is correct, as desired. The mean value for this histogram is $\mu = 0.490$ with a standard deviation $\sigma = 0.144$.

For the 343 molecule system, a histogram is also constructed from the 1-dimensional reaction coordinate approximation: $\mathbf{b}/\mathbf{c}^{\ddagger} = 1.4327 \pm 0.0006$, as shown in Figure 11. 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 $\mu = 0.520$ with a standard deviation $\sigma = 0.208$.

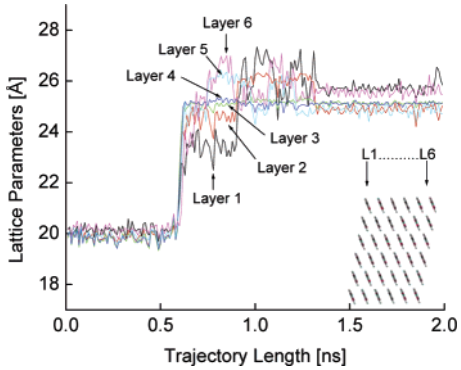

Figure 12. Change in the local **c** lattice parameter for each layer in the polymorph transformation of TPA. The differences in the local OP denote that there is a localized event occurring during the transition, namely edgewise nucleation. This corresponds to the transformation seen in Figure 4b, in 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.

Discussion

The unbiased trajectory, as shown in Figures 4b and 6, 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 4a and 5 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 10 and 11 for both system sizes studied. The second histogram for the 216 molecule system, shown in Figure 10b, is constructed on the basis of 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 inadequate number of trajectories sampled with aimless shooting.

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.

In the Methodology section it is shown how the OP for the lattice parameters is condensed into a single, average lattice parameter. Although the reaction coordinate is captured with the single, average lattice parameter (or the ratio of the average lattice parameters), in the case of a nucleation mechanism, nonuniformity should be apparent in the local lattice parameters measured on each supramolecular synthon layer. For example, Figure 12 shows this behavior for the **c** lattice vector for the

216 molecule system. Figure 12 corresponds to the trajectory that undergoes the full polymorph transformation as shown in Figures 5 and 7. The outside layers (1 and 6) exhibit significant nonuniformity 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 single, average lattice parameter that spans the entire crystal may not be the best approximation to the reaction coordinate. 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 single 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.

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.[53] 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 7 and 8 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 7 and 8), 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.[23] 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 7, the free energy barrier is on the order of 4−5 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.,[23] 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.[23,24] 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.

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 owing 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 shows that the average lattice parameters can be used to approximate the reaction coordinate. This is qualitatively verified with committor probability analyses. A trace of the localized lattice parameters on each layer of the crystal along the transformation shows significant nonuniformity, 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.[23]

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.[54] 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.

Acknowledgment. We thank Merck & Co., Inc. for their financial support of this study. In addition, G.T.B. thanks the National Science Foundation for a Graduate Research Fellowship. We especially thank Dr. Ning Shan for insightful discussions and Prof. Roger Davey for describing the phase transformations from an experimental standpoint. We are grateful to Dr. Michael Thien, Dr. Ivan Santos, Dr. Robert Wenslow, Dr. Lou Crocker, and many others at Merck & Co., Inc. for their support. In addition, we thank Jane Rempel at MIT for both technical and editorial contributions to this manuscript.

Supporting Information Available: Complete ref 46. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0687567

(53) Turnbull, D.; Fisher, J. C. *J. Chem. Phys.* **1949**, *17*, 71−73.
(54) It should be noted that after this manuscript was accepted, a study by Anwar, J.; Tuble, S. C.; Kendrick, J. *J. Am. Chem. Soc.* **2007**, *9*, 2542−2547 was published which shows concerted molecular displacements in a periodic system of DL-norleucine crystals using molecular dynamics techniques.