

Computer Simulations of Homogeneous Nucleation of Benzene from the Melt

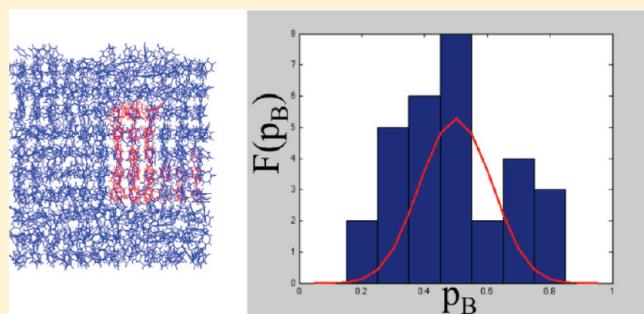
Manas Shah,[†] Erik E. Santiso,^{†,‡} and Bernhardt L. Trout^{*,†}

[†]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[‡]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695, United States

 Supporting Information

ABSTRACT: Nucleation is the key step in crystallization by which the molecules (or atoms or ions) aggregate together, find the right relative orientations, and start to grow to form the final crystal structure. Since nucleation is an activated step involving a large gap in time scales between molecular motions and the nucleation event itself, nucleation must be studied using rare events methods. We employ a technique developed previously in our group known as aimless shooting [Peters, B.; Trout, B. L. *J. Chem. Phys.*, 2006, 125, 054108], which is based on transition path sampling, to generate reactive trajectories between the disordered and ordered phases of benzene. Using the likelihood maximization algorithm, we analyze the aimless shooting trajectories to identify the key order parameters or collective variables to describe the reaction coordinate for the nucleation of benzene from the melt. We find that the local bond orientation and local relative orientation order parameters are the most important collective variables in describing the reaction coordinate for homogeneous nucleation from the melt, as compared to cluster size and space-averaged order parameters. This study also demonstrates the utility of recently developed order parameters for molecular crystals [Santiso, E. E.; Trout, B. L. *J. Chem. Phys.*, 2011, 134, 064109].



1. INTRODUCTION

Crystallization is one of the most important separation and purification techniques employed by the chemical, food, and pharmaceutical industries.^{1,2} Solution crystallization is perhaps the most important unit operation in the pharmaceutical industry, where 90% of the products contain active pharmaceutical ingredients (APIs) and/or excipients in the crystalline state.² Hence it is imperative for these industries to have a precise control on the shape, size, crystal structure, and purity of these crystals as they determine the physical, chemical, mechanical, and bioavailability properties of the final crystal product. Nucleation is the first step in crystallization where an embryo or nucleus of the new phase is created within the metastable phase.^{2–4} Nucleation plays an important role in determining the nature of polymorph (crystal structure), morphology, size, and purity of the final product.^{2,5}

One of the most utilized theoretical frameworks to understand nucleation has been the classical nucleation theory (CNT).^{2,3,6,7} According to CNT, the driving force for nucleation can be described in terms of a competition between two effects: the free energy to form the stable phase, which favors increasing the size of the nucleus, and the unfavorable contribution from the free energy to form the interface between the two phases. Much experimental and theoretical work has been done to underline the shortcomings of CNT to describe the nucleation mechanism.^{2,3,7,6,8–10} One of

the main criticisms of CNT is that the cluster size (mostly assumed to be spherical) is the only relevant order parameter (OP) in the description of nucleation mechanism. Simulation studies by Moroni et al.⁷ have shown that the shape and molecular structure are important physical variables in the description of the nucleation mechanism. The assumption that bulk thermodynamic quantities such as interfacial energy can be applied for events involving few tens or hundreds of atoms or molecules has also shown to be incorrect. Moreover, CNT cannot be used to study multiple pathways possible for nucleation or the multistep nature of the mechanism.^{6,10} More recently, the two-step nucleation model has been proposed for the solution crystallization for small organic molecules, methane hydrates, and proteins, which suggests that density fluctuations are necessary before the orientational ordering of molecules within the cluster can take place.^{2,3,11,12} We need to go far beyond CNT and develop a molecular-based mechanistic approach to understand nucleation.

Despite the obvious importance of nucleation and crystallization in the various fields of science and technology, very little is understood about the actual mechanism itself. Direct observation of nucleation-like events has been experimentally shown

Received: April 16, 2011

Revised: July 19, 2011

Published: July 25, 2011

only for larger molecules, e.g., in protein crystallization¹³ and hard sphere colloidal crystallization.⁸ Apart from these larger molecules, the beginning of nucleation and the overall mechanism are unknown for nucleation in molecular and atomic systems.³ On the other hand, such molecular motions can be observed via computer simulations.¹⁴ The challenge with this approach, however, is that nucleation is a rare event, and such rare events are not easily accessible with standard molecular simulations.⁶ Hence sophisticated computational methodologies are required to gain mechanistic understanding of rare events such as nucleation, and obtain a good estimate of the reaction coordinate. The reaction coordinate is defined as the variable that describes the dynamical progress along the reactive pathway connecting the two states of interest, separated by long time scales or by a large free energy barrier. It can be a function of many collective variables or OPs for the nucleation mechanism. For any reaction or transition, the exact reaction coordinate is the committor probability (p_B), the fraction of trajectory initiated with Boltzmann distributed momenta from a given configuration that commits to the product (B) basin.^{15–18} We are mostly interested in the $p_B = 0.5$ isosurface, which corresponds to the transition state ensemble. OPs only serve as a quantitative metric to characterize the different stable basins, while the reaction coordinate provides information regarding the dynamical mechanisms involved in the transitions.¹⁵ Identifying the correct reaction coordinate will enable molecular level insights into nucleation and will ultimately aid in designing the crystallization processes.

One of the earliest theoretical efforts in studying nucleation was undertaken by Frenkel and co-workers,^{19–21} who used Monte Carlo (MC) techniques with the umbrella sampling techniques of Torrie and Valleau²² to measure a free energy barrier and the rate of nucleation involved in crystallization of soft repulsive spheres from a supercooled liquid. Their proposed reaction coordinate was the crystalline cluster size, and they used the bond orientational OP to quantify the crystallinity of the system.²³ While umbrella sampling and other biasing methods provide useful insights, a better methodology would not require *a priori* assumption of the reaction coordinate. More recently, a new technique known as transition path sampling (TPS) has been utilized to obtain insights into the mechanism of rare events in a variety of chemical and biological systems.^{16,24–26} Under this framework, an ensemble of unbiased reactive trajectories is obtained between the two basins separated by a large time scale. This methodology has been used to generate nucleation pathways in an Ising model,²⁷ the Lennard-Jones fluid,⁷ and hexagonal ice in water.^{28,29} Zahn and co-workers have used TPS to gain insights into melt crystallization in sodium halides,³⁰ pressure-induced polymorphic transformations,³¹ and the evaporation of water.³² From these studies, it is clear that TPS serves as a powerful tool to obtain unbiased reactive pathways and calculate rates in complex systems.

Recently, a variant of TPS known as the aimless shooting algorithm was developed in our group.^{15,33} Unlike TPS where a new trajectory is obtained by modifying momenta from a previous trajectory, a new trajectory in aimless shooting is obtained by choosing a configuration (near the putative $p_B = 0.5$ isosurface) along the previous trajectory, and drawing fresh momenta from the Boltzmann distribution consistent with the temperature of the system.¹⁵ This leads to a higher degree of decorrelation between successive trajectories than in the original TPS method, as fresh velocities (independent of previous trajectories) are obtained for every trajectory.

Obtaining an estimate of the reaction coordinate for complex systems is quite challenging, but it is also quite necessary for a complete physical understanding of the mechanism. Previously, the method to identify the reaction coordinate was based mostly on guesswork. Recently there have been several approaches to identify reaction coordinates systematically as a function of several OPs, including genetic neural networks (GNNs),³⁴ Bayesian path statistical analysis,²³ and the likelihood maximization technique.^{15,33} We utilize the latter technique to obtain a reaction coordinate as a function of several key OPs or collective variables. Likelihood maximization is especially suited for our application since it only requires the data from aimless shooting (or TPS) simulations.¹⁵ These methods have been utilized to identify key collective variables in several systems such as polymorphic transformations in organic compounds,^{35,36} protein folding,^{37,38} calculation of methane diffusivity in gas hydrates,³⁹ and also acid-catalyzed hydrolysis of peptide bonds.⁴⁰ These methods are especially suitable for studying the nucleation mechanism in small molecule organic crystals, as there is no predetermined reaction coordinate and many OPs can be screened for their validity in describing the nucleation pathway.¹⁵ The role of OPs such as local orientational order and cluster size in describing the mechanism can be better quantified by using such methods.

In this work, we apply aimless shooting and likelihood maximization in order to gain mechanistic insights into the nucleation of organic crystals. We consider the test case of homogeneous nucleation of benzene crystal from the liquid phase and obtain many reactive pathways connecting the solid (Form I crystal) and liquid basin.⁴¹ Benzene is chosen as the molecule of interest because it is a well-studied model organic molecule, both from an experimental^{42–44} and theoretical^{41,45} point of view. We also test the validity of a new set of OPs for molecular organic crystals,⁴⁶ developed recently in our group, for describing the nucleation pathways. In the next section, we provide the details of our system including the key aspects of our aimless shooting simulations and the new set of OPs. In the Results section, we provide our results from aimless shooting statistics and also the likelihood maximization screening of many local OPs. Finally we attempt to provide a physical understanding of our simulation results and compare it with existing theoretical frameworks for melt crystallization, including CNT. In the Discussion section, we present the benefits and limitations of using such methodologies in studying nucleation mechanisms.

2. METHODOLOGY

2.1. System Description. We start with the structure of the benzene Form I crystal (refcode: BENZEN) from the Cambridge Structural Database,⁴⁷ corresponding to the experimental study of Bacon et al.⁴² The Form I crystal of benzene belongs to the space group $Pbca$, and the benzene molecules are arranged on an orthorhombic lattice in the unit cell. The benzene Form I crystal is the stable form at low pressures and temperatures close to room temperature.⁴¹ The values of lattice parameters have been taken from the experimental crystal structure.⁴² A system size of 720 molecules is used for this study, and reactive trajectories connecting solid and liquid states are obtained at 250 K, significantly below the melting temperature of benzene.⁴⁵ The molecular dynamics (MD) simulations were carried out using the CHARMM22 force field^{48,49} in the NAMD software package.⁵⁰

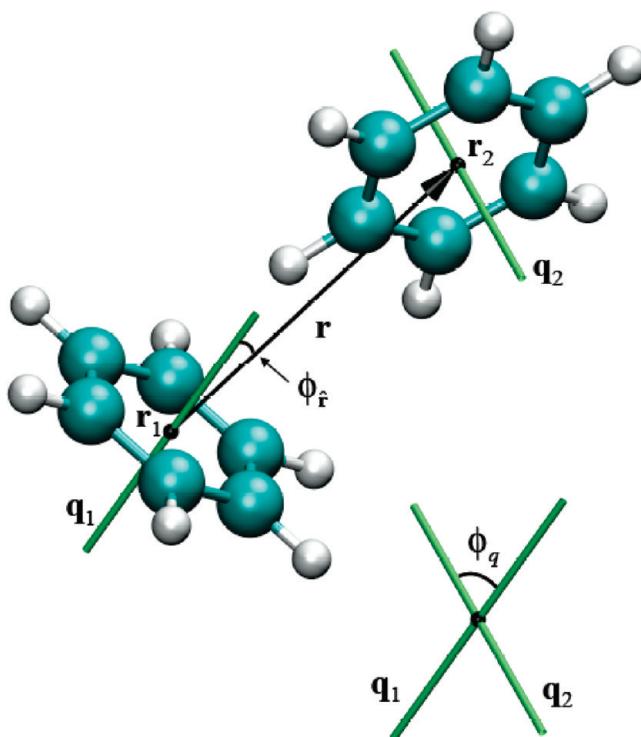


Figure 1. An illustration of the quantities describing the relative configuration of two benzene molecules. The vector \mathbf{r} joins the center of mass of the two molecules. The absolute orientations of the two molecules are represented by the normals to the plane of the molecules (\mathbf{q}_1 and \mathbf{q}_2). The bond orientation is the angle between \mathbf{r} and \mathbf{q}_1 , and the relative orientation is the angle between two absolute orientations.⁴⁶ Reprinted with permission from ref 46. Copyright 2011 American Institute of Physics.

Initial equilibration and energy minimization was carried out at 250 K and 1 bar using the NAMD software package. After equilibration, the system size was $45.13 \text{ \AA} \times 48.32 \text{ \AA} \times 41.84 \text{ \AA}$. The forward and backward MD trajectories for aimless shooting methodology were performed using the NVT ensemble. A time step of 1 fs was used, and a Langevin thermostat with a damping coefficient of 5 ps^{-1} was employed to maintain the temperature at 250 K. Periodic boundary conditions were used, and long-range electrostatics were treated using particle mesh Ewald (PME) summation.

2.2. Order Parameters. OP provide a quantitative metric to distinguish between different stable states. Previous simulation studies involving nucleation and crystallization have primarily utilized lattice parameters,³⁵ self-diffusion coefficients,³⁰ the Lindemann index,^{51,52} and bond orientational OPs^{19,20,23} to distinguish between the liquid and various forms of solid states. However, it is not straightforward to use such OPs to distinguish between various polymorphs of small organic molecules, typically encountered in pharmaceutical applications. Recently, Santiso and Trout⁴⁶ developed a new set of OPs that can be systematically defined for small molecular systems and can also be utilized to distinguish between various polymorphs of a single compound. This methodology requires information that can be easily obtained from a straightforward MD simulation. The description of this set of OPs is based on a generalized pair correlation function that contains all the variables relevant to description of the crystal structure. For any given molecular

crystalline system, this set of OPs accounts for distance order, bond orientational order, relative orientational order, and even internal degrees of freedom such as dihedral angles.⁴⁶ Here we briefly provide the equations and notations for the different OPs used in this study.

In Figure 1, we show two benzene molecules with distance \mathbf{r} between their centers of masses, and having absolute orientations, \mathbf{q}_1 and \mathbf{q}_2 , respectively. For a planar molecule like benzene, the normal to the plane of the molecule is chosen as the absolute orientation of the molecule. The distance OP provides quantification for the various center of mass distances between different molecules arranged on a crystal lattice and is approximated by a Gaussian distribution:

$$\varphi_{\alpha,i}^d = \sum_{j \neq i} \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left[-\frac{(r_{ij} - r_\alpha)^2}{2\sigma_\alpha^2}\right] \quad (1)$$

where the index i denotes the i th molecule, r_{ij} denotes the center of mass distance between the molecules i and j , r_α is the mean center-of-mass distance corresponding to the α -peak, and σ_α is the corresponding standard deviation.

The bond orientation OP measures the orientation of “bonds” joining the centers-of-mass of molecules, while the relative orientation OP measures the orientation of one molecule with respect to the second molecule. The bond orientation between two molecules will be the projection of the bond vector \mathbf{r} joining the centers of mass of two benzene molecules onto the orientation axis of the first molecule and is denoted as ϕ_r (Figure 1). The relative orientation between two molecules is the angle between two orientation axes, which is denoted as ϕ_q (Figure 1). The distributions of bond and relative orientations are modeled using the von Mises distribution.^{46,53} The resulting OPs are given by

$$\varphi_{\alpha,i}^{bo} = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \frac{1}{2\pi I_0(\eta_r^\alpha)} \sum_{j \neq i} \exp\left[-\frac{(r_{ij} - r_\alpha)^2}{2\sigma_\alpha^2}\right] \exp[\eta_r^\alpha \cos 2(\phi_r - \phi_r^\alpha)] \quad (2)$$

$$\varphi_{\alpha,i}^{ro} = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \frac{1}{2\pi I_0(\eta_q^\alpha)} \sum_{j \neq i} \exp\left[-\frac{(r_{ij} - r_\alpha)^2}{2\sigma_\alpha^2}\right] \exp[\eta_q^\alpha \cos 2(\phi_q - \phi_q^\alpha)] \quad (3)$$

where I_0 is the modified Bessel function of the second kind and order 0, and η_r^α and η_q^α are concentration parameters corresponding to the α th peaks in bond orientation and relative orientation, respectively.⁴⁶

All these OPs are defined for every molecule and every peak for a given crystal structure. We can define the OP for a particular molecule by summing over the different α peaks. For example, we can obtain a single value for bond orientation for a molecule:

$$\theta_i^{bo} = \sum_\alpha \varphi_{i,\alpha}^{bo} \quad (4)$$

Since these OPs are sensitive for a particular polymorph of a compound,⁴⁶ we can imagine that they will be quite high for a crystal and very low for a liquid-like structure. In Figure 2, we plot the number distribution of benzene molecules according to their (a) distance order, (b) bond orientation, and (c) relative orientation OPs values for both liquid and solid molecules. This number

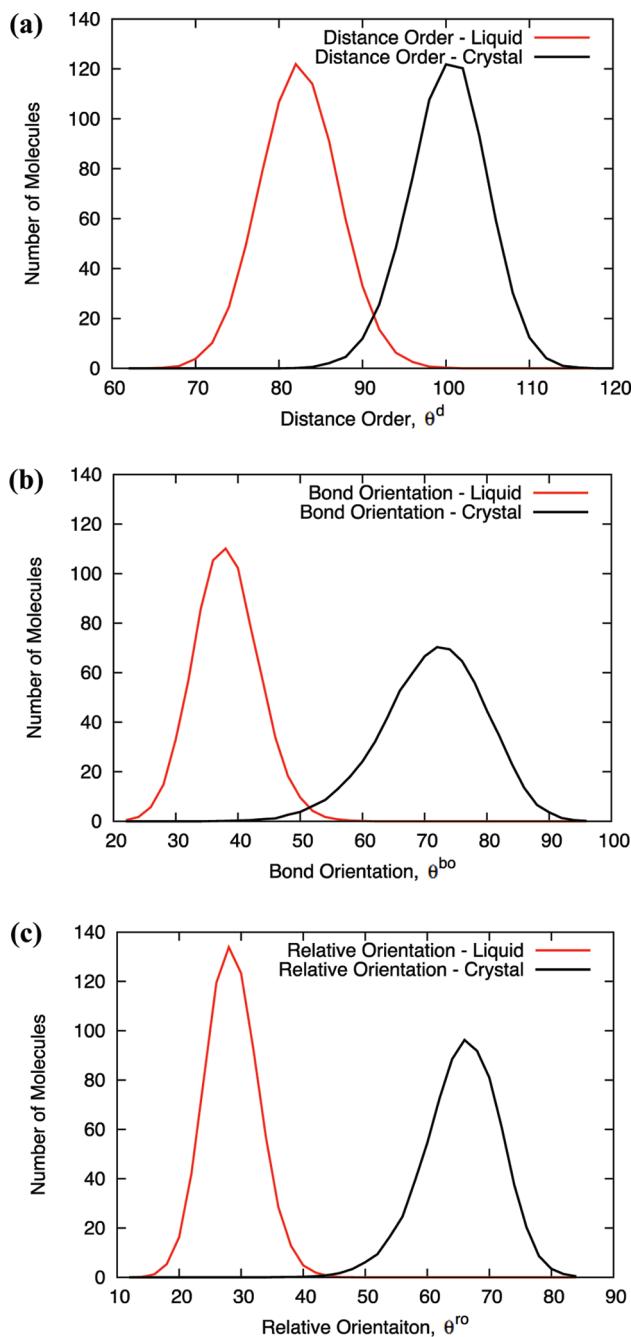


Figure 2. OPs for liquid and solid benzene. The number of molecules of benzene are plotted as a function of three types of OPs: (a) distance order, (b) bond orientation, and (c) relative orientation. These plots are averaged over 500 snapshots each of solid and liquid benzene. It is clear that for all three types of OPs, the liquid and solid states are peaked at different values.

distribution was obtained by averaging over 500 frames of MD simulations of both solid and liquid benzene (1 ns simulations). We can clearly see that for all three different types of OPs, the solid and liquid benzene molecules are peaked at different values of OPs. Thus any of these OPs serve as a good metric to distinguish between solid and liquid benzene. Upon closer look, there is a certain degree of overlap in distance OP space for liquid and solid molecules, as one would expect for a melt (refer to Figure 2a). The degree of overlap is lowest for relative orientation

OPs (Figure 2c). We can imagine that the probability of having the right orientational order (both bond and relative), corresponding to a particular polymorph of a crystal, is extremely low to occur in a liquid system. This seems to indicate that the distinction between solid and liquid benzene molecules is much sharper in bond orientation and relative orientation OPs.⁴⁶

To minimize the number of OPs, the simulation box is divided into cells, and each OP is averaged within the cell and also summed over all the peaks. For example, the bond orientation OP within a cell C is given as

$$\theta_C^{bo} = \frac{1}{N_{c_i \in C}} \sum \sum \varphi_{i,\alpha}^{bo} \quad (5)$$

Finally all these OPs are summed over all the cells in the simulation box to yield three space-averaged parameters. Again, as an illustration, we show this for bond orientation:

$$\Theta^{bo} = \sum_C \theta_C^{bo} \quad (6)$$

To distinguish between solid and liquid basins in our aimless shooting algorithm, we utilize a space-averaged bond orientational OP, Θ^{bo} . In the normalized OP space, if the value of Θ^{bo} is lower than 0.15, it is considered to be liquid. The system is considered to be solid if the value is higher than 0.5. For this particular system, any of these OPs or a combination of them would have sufficed to distinguish between the two states. In the Results section, we show how these OPs vary along a reactive trajectory connecting the solid (Form I) and liquid states of benzene.

3.3. Aimless Shooting Methodology. The aimless shooting methodology, as described in papers by Peters and Trout,^{15,33} generates an ensemble of reactive trajectories between the two basins. The two conditions necessary for initiating aimless shooting simulations are (i) proper definition of OPs for distinguishing the two basins, and (ii) an initial trajectory connecting the two basins. As described above, we use space-averaged bond orientational OPs⁴⁶ to distinguish between the solid and liquid basins of benzene. The initial trajectory connecting the crystal and liquid is obtained in a slightly unphysical manner by melting the crystal at a higher temperature (600 K) and choosing a configuration along the melting trajectory as the initial shooting point. Within the aimless shooting methodology, shooting points serve as the midpoint of the trajectories and are supposed to sample the transition state ensemble of the reaction (or the $p_B = 0.5$ isosurface).¹⁵ We generate a new trajectory from an old trajectory by performing the following procedure:

- Shift the shooting point along the old trajectory by a small time displacement, $\pm \Delta t$, to obtain a new configuration for the shooting point. This time displacement should be quite small compared to the overall length of one simulation, t_{sim} .
- Obtain new momenta from a Boltzmann distribution corresponding to the simulation temperature and perform MD simulations in the forward ($t_{sim}/2$) and backward ($-t_{sim}/2$) direction to obtain the new trajectories.
- If the new trajectory connects the two basins, then accept it. Otherwise, reject it and repeat steps (i)–(iii).

The choice of the parameters, time displacement (Δt), and actual length of trajectory (t_{sim}) are critical for the efficiency of the aimless shooting simulations. We would ideally prefer to have a very short t_{sim} in order to save computational time by increasing

the acceptance rate, but this value is mainly governed by the underlying dynamics of the transition. Earlier studies used a value of 1% for the ratio $\Delta t/t_{\text{sim}}$ for higher efficiency and a good acceptance rate of trajectories.^{15,35} However, upon trying various combinations for these aimless shooting parameters for this particular system, we found that the highest acceptance rate was obtained for the case of $\Delta t = 2$ ps and $t_{\text{sim}} = 0.8$ ns. This leads to an acceptance ratio of 32%, a value close to previous studies.^{36,40,54} However, the rate of inconclusive trajectories is quite high for this particular system, and we discuss this issue of sampling reactive pathways for the nucleation mechanism later in the Discussion section.

2.4. Likelihood Maximization. As described in earlier studies,^{15,33,35–37,40,55} likelihood maximization is a methodology to obtain an approximate reaction coordinate, r , as a function of various relevant OPs. We use a simple linear model to describe r as a function of the various OPs, denoted by \mathbf{q} , where the various α_k ($k = 0–m$) act as coefficients to be determined using an optimization procedure.

$$r(\mathbf{q}) = \alpha_0 + \sum_{k=1}^m \alpha_k \quad (7)$$

The choice of a linear model is for convenience only, and, in principle, any function can be utilized. The model function for the committer probability, $p_B(r)$, was chosen to be^{33,35,40}

$$p_B(r) = \frac{1}{2} [1 + \tanh(r)] \quad (8)$$

It is not necessary to chose this particular model function, but any model for $p_B(r)$ must satisfy the conditions for values of committer probability at transition state, reactant, and the product phases: $p_B(r=0) = 0.5$, $p_B(r=-\infty) = 0$, and $p_B(r=\infty) = 1$. This model of committer probability is used to maximize the likelihood function only using outcomes of forward half trajectories described in the following equation:¹⁵

$$L(\alpha) = \prod_{k=1}^B p_B(r(\mathbf{q}^{(k)})) \prod_{k=1}^{\neq B} (1 - p_B(r(\mathbf{q}^{(k)}))) \quad (9)$$

The values of OPs ($\mathbf{q}^{(k)}$) correspond to all the shooting points of conclusive trajectories obtained from the aimless shooting methodology. The log likelihood $l(\alpha) = \ln L(\alpha)$ function is maximized using an optimization algorithm such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method⁵⁶ to obtain the optimal coefficients, α_k . The particular value of r as a function of different collective variables (\mathbf{q}), for which the log likelihood is maximized, is considered to be the approximate reaction coordinate. The likelihood maximization screening is carried out for all possible combinations of increasing numbers of collective variables until the Bayesian information criterion (BIC) shows that adding variables does not improve the model.¹⁵

2.5. Reaction Coordinate Validation. This step is performed after the likelihood maximization calculations to validate the model for the transition state ($r=0$) obtained. On the basis of the definition of committer probability, $p_B(r)$, for a true transition state, the value of p_B should peak at 0.5.¹⁶ From the transition state isosurface ($r=0$) obtained using the likelihood maximization method, many trajectories are shot with random initial velocities, conforming to the Boltzmann distribution. If the histogram of p_B values obtained from these trajectories peaks at 0.5, then the isosurface is good representation of the true transition state.¹⁶ Several new shooting points are needed for construction of a p_B histogram. New shooting points were obtained by em-

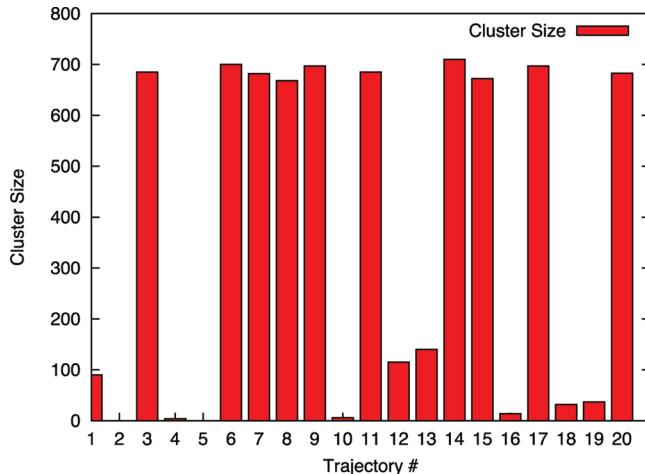


Figure 3. Cluster size (N_c) for 20 different outcomes for trajectories initiated from a putative shooting point.

ploying the BOLAS algorithm as described in previous works.^{39,57} Finally, we evaluate the mean and variance of the histograms by following the procedure outlined in a recent article.⁵⁸

3. RESULTS

3.1. Initial Trajectory. The equilibrium configuration of Form I crystal is obtained after running a MD simulation in the NPT ensemble at 250 K, as described in the above section. One of the two criteria for initiating aimless shooting simulations is to have an initial trajectory connecting the two basins. Clearly it will be almost impossible to obtain a trajectory connecting solid and liquid phase at 250 K by solely running MD simulations, because of the activation barrier associated with having an interface between solid and liquid phases at a temperature below the melting point. In order to circumvent this issue, the following protocol is utilized: (i) melt the crystal at a very high temperature using NVT ensemble (600 K for this system size), (ii) chose a configuration along this melting trajectory as a first guess for the shooting point, (iii) shoot a few MD trajectories from this guessed shooting point and observe whether both the basins (solid and liquid) can be accessed, and (iv) if the trajectories from this guessed shooting point cannot reach both solid and liquid basins, then a new configuration along the melting trajectory is selected and steps (ii)–(iv) are repeated. Ideally we would like to start our simulations with an initial shooting point very close to the $p_B \sim 0.5$ isosurface (transition state ensemble) for efficient sampling by aimless shooting simulations. However, in practice, a configuration that has a finite probability of reaching both reactant and product basins is quite acceptable for initiating aimless shooting simulations.³³

Figure 3 shows the outcome, in terms of cluster size, of a few trajectories initiated from a putative “transition state” as our guess for the first shooting point. We calculate cluster size (N_c) based on bond orientational OPs, following the same procedure of Frenkel and co-workers.²¹ We evaluate the largest connected crystalline cluster (N_c) for the final image of the MD trajectory. A molecule is considered to be “crystalline” if the value of normalized bond orientation (θ_i^{bo}) is greater than 0.5. Then a cluster analysis is carried out to determine the largest connected cluster of “crystalline” molecules in the system. This involves finding all the crystalline molecules and then checking for nearest neighbors

for crystallinity. Admittedly, there is some ambiguity in this particular definition of crystalline cluster, and we will discuss this issue in more detail later on. However, for the current purpose of obtaining a configuration close to the transition state, such a metric would suffice. For the purpose of finding a feasible shooting point, we need to obtain states that have very small crystalline cluster sizes (liquid-like) as well as states with very large crystalline cluster size (solid-like). In Figure 3, the values of cluster size for the outcomes of 20 MD simulation trajectories initiated from a putative shooting point are shown. The initial velocities are chosen at random from the Boltzmann distribution consistent with the temperature. We observed that states with very low values of N_c (~ 0), as well as states with very high values of N_c (~ 700) are obtained. Thus our initial guess for shooting point

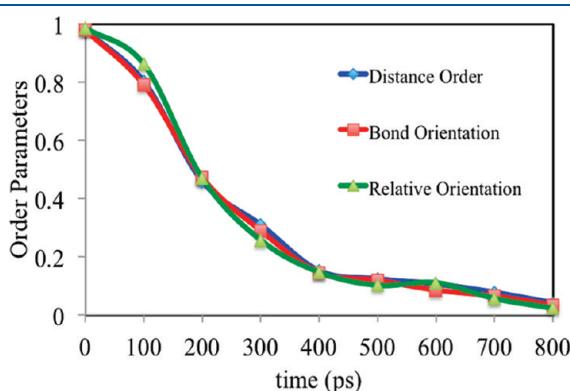


Figure 4. Normalized space averaged OPs as a function of time for a given reactive trajectory.

serves to obtain trajectories to sample both solid and liquid states and can thus serve as an adequate starting point to sample the transition state and obtain several reactive trajectories. Starting with this particular initial configuration, we perform the aimless shooting simulations. Before we discuss the results for aimless shooting statistics, it is worthwhile to see the variation of OPs (space-averaged and local) and crystalline cluster size along a reactive trajectory connecting solid and liquid states.

3.2. Reactive Trajectories. A reactive trajectory is one that connects two metastable basins separated by a free energy barrier or by large time scales. In Figure 4, we show the variation of the molecular OPs along one such reactive trajectory connecting the crystal and liquid basins. We plot normalized values of space-averaged distance order, bond orientation and relative orientation as a function of time along the reactive trajectory. We can clearly see that these molecular OPs vary from a very high value (~ 1) at the crystal basin (at 0 ps) and decrease to a very low value (~ 0) at the liquid basin (at 800 ps). On the crystal side, almost all the molecules have very high values of OPs indicated by a high value for the space-averaged OPs. We also observe a strong degree of correlation between the three different types of OPs, indicating that any one could be used to categorize the phase as liquid or solid. At the shooting point of this particular trajectory, we observe that, globally, the state shares more similarity with the liquid phase rather than the solid phase.

A better quantification of the degree of crystallinity can be obtained by looking at local OPs at a molecular level. In Figure 5, we show key snapshots from the same reactive trajectory going from the crystal state to the liquid state. The molecules are color coded according to their bond orientation OP values (θ_i^{bo}): blue for high values of bond orientation OP, red for very low values,

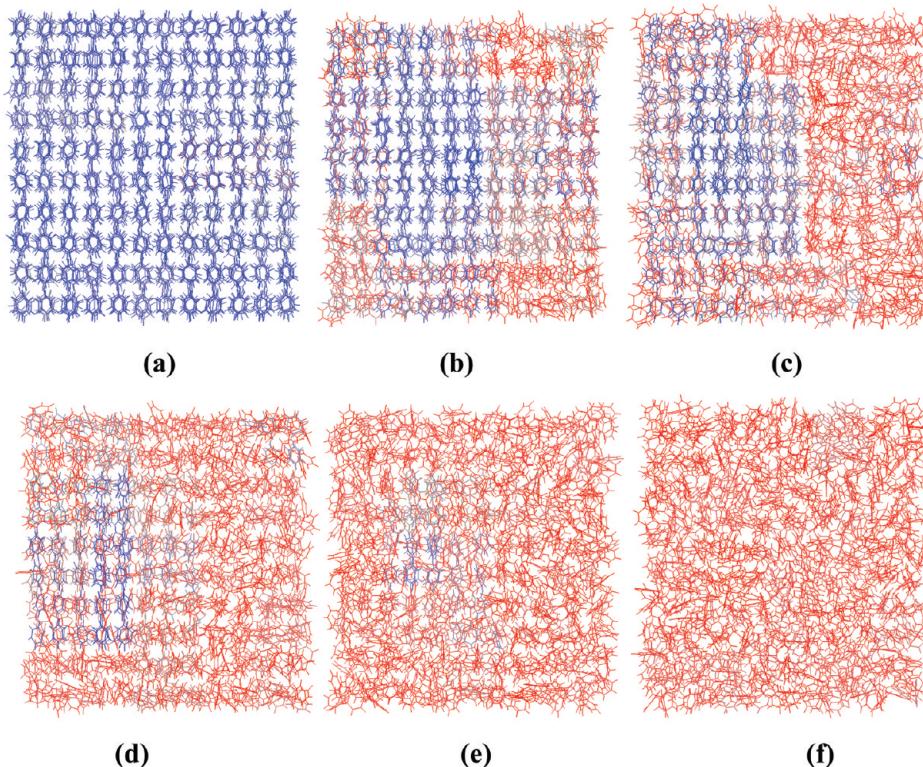


Figure 5. Snapshots of configuration of benzene molecules corresponding to the reactive trajectory in Figure 4. The various snapshots are for (a) 0 ps (crystal basin), (b) 100 ps, (c) 300 ps, (d) 350 ps, (e) 400 ps (shooting point), and (f) 800 ps (liquid basin). The molecules are colored according to their bond orientation OP values: blue - high value of OP; red - low value of OP; gray - intermediate values.

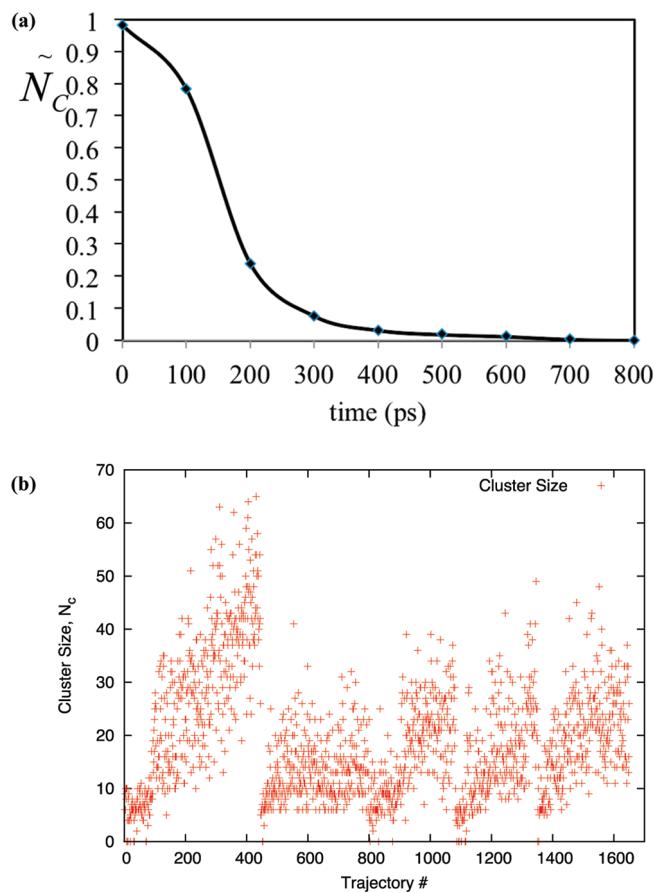


Figure 6. (a) Normalized cluster size (\tilde{N}_C) calculated on the basis of bond orientation OP for the reactive trajectory discussed in Figures 4 and 5. (b) Values of N_C across all trajectories.

and gray for intermediate values. In the first image (Figure 5a), as expected, all the molecules have very high values of bond orientation OPs. Along the reactive trajectory (Figure 5b–f), we clearly observe that the number of molecules with high value of bond orientational OP decrease gradually as one approaches the liquid basin (Figure 5f). In the last image, we do not observe any molecule with a high value of bond orientational OP (Figure 5f). For the case of the shooting point (400 ps, Figure 5e), we see a very tiny cluster of molecules that are colored gray/blue at the left-center part of the system surrounded by a sea of disordered molecules. Starting from that tiny cluster, we see it growing in size in images d and c, eventually leading up to a fully formed crystal in image a. On the other side, the same cluster dissolves, and we obtain a liquid phase with hardly any molecules possessing any degree of orientational order. It should be noted that we would obtain similar looking snapshots if the molecules were colored according to their distance order or relative orientation OP values. We quantify the size of this cluster as previously described in the “Initial Trajectory” subsection. We calculate \tilde{N}_C , which measures the size of the crystalline cluster, as a function of time elapsed along this reactive trajectory. In Figure 6a, we plot the value of normalized \tilde{N}_C for the same reactive trajectory shown in Figures 4 and 5. As expected, we see the value to be close to 1 near the crystal side of the trajectory, whereas the value drops to 0 at the liquid end. The more intriguing aspect is that the size of crystalline cluster at the shooting point is really small, comparable to that of the liquid state. A small crystalline cluster

Table 1. Aimless Shooting Statistics

backward	forward	# trajectories
liquid	liquid	280
liquid	crystal	134
crystal	liquid	143
crystal	crystal	317

indicates that our original choice of system size was justified in studying nucleation events in the benzene melt.

In Figure 6b, we plot the crystalline cluster size, N_C for the entire collection of shooting points obtained from aimless shooting simulations. The range of values for N_C extends from very few to up to 70 molecules. The average value of N_C for all the shooting points is around 32 molecules. Again a small cluster size compared to the overall system size (720 molecules) provides a justification for its use in this study. Another key thing to note from Figure 6b is the range of values of cluster size (and correspondingly the molecular OPs) that shooting points can take. This implies that the shooting points do not comprise a single fixed configuration, but comprise an ensemble of configurations with a range of OPs and crystalline cluster sizes. We have used a slightly arbitrary criterion here to define the cluster size. However, upon looking at Figures 2 and 4, using $\theta_i^{bo} > 0.5$ seems like a reasonable criteria to label a molecule as “crystalline”. If we had chosen a different criterion (say, based on relative orientation), the statistics in Figure 6b would change, but the qualitative features would remain the same. We will explore the validity of choosing cluster size as a viable candidate for the reaction coordinate in the following sections.

3.3. Aimless Shooting Statistics. We generate 1650 aimless shooting trajectories out of which 874 were conclusive, i.e., both forward and backward trajectories committed to a particular basin. A relatively low (53%) percentage for conclusive trajectories probably implies that the mechanism is highly diffusive,⁵⁸ and the trajectories spend lot of time sampling the region around the shooting point before committing to a particular basin. One way to alleviate this issue is by increasing the trajectory length.⁵⁸ However, increasing the trajectory length to 1 and 1.2 ns did not improve the rate of conclusive trajectories. Another possible solution is perhaps using less stringent criteria for definition of the basins. If we consider only conclusive trajectories, we obtain an acceptance rate of 32%, which is a reasonable number for such types of path sampling simulations.^{36,54} We present the statistics of conclusive trajectories in Table 1. It displays the statistics of outcomes of forward and backward trajectories, where both trajectories end either in liquid or solid basin. We obtain four cases out of which only the second and third types correspond to reactive trajectories. We would like to remind the readers that there is no bias for either backward or forward trajectory to reach a particular basin under the aimless shooting algorithm. Thus, if we are allowed to collect a large number of reactive trajectories, we should observe equal sampling of both the basins by backward and forward trajectories. For our particular system, we see that both the liquid and solid basins are sampled at a similar frequency, with a marginally higher occurrence for observation of the solid state (refer to Table 1). This partial preference for solid state can also be observed in Figure 6b where we observe that the cluster size for shooting points tends to increase, implying that shooting points drift toward the solid basin. This was also observed in a previous study⁴⁰ where the individual numbers for trajectories committing to reactant and product basins were

Table 2. 1-OP Likelihood Scores^a

number of OPs	OP	log likelihood ($\ln(L)$)
1	averaged distance order	-777.1
1	averaged bond orientation	-769.2
1	averaged relative orientation	-774.7
1	cluster size	-794.7

^a BIC = 3.39.

even more skewed. It is not very surprising that shooting points will tend to drift toward the basin with lower free energy, the solid basin in this particular case of melt crystallization. However, because of the self-correcting nature of aimless shooting algorithm, the shooting points drift back to the transition state region as depicted in Figure 6b. When a shooting point drifts toward a particular basin and away from the “true” $p_B = 0.5$ isosurface, the acceptance rate of trajectories reduces, and the only way for a trajectory to be accepted is if the shooting point is reverted in the opposite direction in the neighborhood of the transition state ensemble. However, this does not affect in any way the likelihood maximization procedure utilized to determine the approximate reaction coordinate.

3.4. Likelihood Maximization. The majority of previous studies have focused their attention on the role of the crystalline cluster size as the reaction coordinate to describe the nucleation mechanism. Using likelihood maximization, we attempt to compute the relative importance of this parameter versus the other molecular OPs discussed previously. The first step of likelihood maximization involves finding the scores for individual OPs. In Table 2, we report the scores for averaged distance order, averaged bond orientation, averaged relative orientation OPs, and cluster size as computed in previous sections. Contrary to previous studies, the cluster size serves as a fairly poor reaction coordinate as compared even to the space-averaged OPs utilized in our study. We calculated the likelihood scores for cluster size based on different definitions such as much stronger or weaker criteria for defining a molecule to be crystalline based on bond orientation, but found them to be equivalent to (or worse than) the ones listed in Table 2. Also, utilizing relative orientation to define a cluster size proved to be futile in improving the likelihood score for the cluster size.

A low value of likelihood score suggests that cluster size (or critical nucleus from CNT) might not be too important in describing the transition pathways of the nucleation mechanism. One obvious reason for the apparent unimportance of the cluster size lies in the arbitrary definition of cluster size itself. It is not *a priori* clear as to which OPs should be considered relevant in defining the cluster size, i.e., when is a molecule considered to be crystalline enough? The relative importance of distance, bond orientation, relative orientation (and possibly other OPs) in defining a crystalline cluster for a system is not very obvious. Hence, such an OP serves as a poor reaction coordinate.⁵⁹ Another issue with using the cluster size as a relevant OP for the reaction coordinate is the issue of the connectivity of molecules within the cluster. It is not apparent whether we should include the molecules to be a part of cluster based on their proximity to the main crystalline molecule and/or based on their actual value of bond or relative orientation OP. Even though cluster size serves as good metric to distinguish the solid and liquid basins (Figure 3 and Figure 6a), it does not serve as good reaction coordinate for the nucleation mechanism. To answer these questions and to alleviate the need to define a critical cluster size, one needs to look at the values of local

Table 3. 1-OP Screening of Local OPs^a

number of OPs	system size	OP	log likelihood ($\ln(L)$)
1	$2 \times 2 \times 2$	distance order	-785.3
1	$2 \times 2 \times 2$	bond orientation	-765.7
1	$2 \times 2 \times 2$	relative orientation	-748
1	$3 \times 3 \times 3$	distance order	-744.7
1	$3 \times 3 \times 3$	bond orientation	-722.3
1	$3 \times 3 \times 3$	relative orientation	-727.1
1	$4 \times 4 \times 4$	distance order	-749.6
1	$4 \times 4 \times 4$	bond orientation	-720.4
1	$4 \times 4 \times 4$	relative orientation	-723.5
1	$5 \times 5 \times 5$	distance order	-769.4
1	$5 \times 5 \times 5$	bond orientation	-734
1	$5 \times 5 \times 5$	relative orientation	-724.2

^a BIC = 3.39.

OPs and evaluate the relative importance of these OPs at different spatial locations within the system.

The obvious choices to consider for local OPs are the three OPs defined for every molecule. But that might not be wise and effective on two accounts: first, there would be too many OPs to screen, and second, we might be looking at the relative importance of individual molecular fluctuations that might not have significant physical meaning to describe the actual nucleation pathway. Our approach to obtain likelihood scores of local OPs consists of these following steps: divide the entire box into $N_k \times N_k \times N_k$ cells (where $N_k = 1, 2, 3, \dots$) and have a single OP for distance (θ_C^d), bond orientation (θ_C^{bo}), and relative orientation (θ_C^{ro}) within each cell, C. This reduces the number of OPs to screen with the likelihood maximization algorithm and lets us evaluate the effect of size of the cell (denoted in terms of N_k) and the nature of OP (distance vs bond orientation vs relative orientation) important for describing the nucleation mechanism. The obvious case of $N_k = 1$ corresponds to the space-averaged OPs, and we have already accounted for them in Table 1. In Table 3, we show the highest likelihood scores for each of the distance order, bond orientation, and relative orientation for different values of N_k . We can clearly see the marked improvement in the likelihood scores for the local OPs in comparison with the space-averaged OPs and cluster size (Table 1). This again reinforces the fact that local orientational order and local distance OPs are far more relevant in describing the nucleation mechanism rather than cluster size and space-averaged OPs. We observe the best single OP (1-OP) likelihood score for local bond orientation OP for the case of $N_k = 4$. In the case of $4 \times 4 \times 4$ division of cells, the best likelihood scores are obtained for all three types of local OPs. There is no rigorous explanation for this observation. But we can rationalize this by considering that for the case of bigger cells (say the case of a $2 \times 2 \times 2$ system), we are still looking at a coarser picture of nucleation activity, whereas for a very tiny cell (say the case of $5 \times 5 \times 5$), where each cell consists of just five or six molecules, we are trying to evaluate the relative importance of molecular level fluctuations that might be too small or irrelevant to account for any nucleation activity. For this system, it might be more relevant to look at individual subcells that have few tens of molecules. Our further results for likelihood scores for increasing the combination of local OPs will be confined to the case of 64 subcells. We will attempt to find the combination of relevant OPs from the list of 192 (64×3) OPs to

Table 4. Likelihood Scores for Combination of OPs for the 4 × 4 × 4 Case^a

number of OPs	OP	log likelihood ($\ln(L)$)
1	θ_{39}^{bo}	-720.4
2	$\theta_{39}^{\text{bo}} + \theta_{43}^{\text{ro}}$	-704
3	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{43}^{\text{ro}}$	-697.2
4	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{27}^{\text{ro}} + \theta_{43}^{\text{ro}}$	-692.4
5	$\theta_{43}^{\text{d}} + \theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{27}^{\text{ro}} + \theta_{43}^{\text{ro}}$	-690.8

^a BIC = 3.39.**Table 5. Combination of Local OPs with Equivalent Likelihood Scores^a**

number of OPs	OP	log likelihood ($\ln(L)$)
4	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{27}^{\text{ro}} + \theta_{43}^{\text{ro}}$	-692.4
4	$\theta_{43}^{\text{d}} + \theta_{39}^{\text{bo}} + \theta_{24}^{\text{ro}} + \theta_{27}^{\text{ro}}$	-695.7
4	$\theta_{43}^{\text{d}} + \theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{43}^{\text{ro}}$	-694.3
4	$\theta_{24}^{\text{bo}} + \theta_{27}^{\text{ro}} + \theta_{39}^{\text{bo}} + \theta_{43}^{\text{ro}}$	-694.9
4	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{43}^{\text{bo}} + \theta_{27}^{\text{ro}}$	-694.5
4	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{43}^{\text{ro}} + \theta_{27}^{\text{ro}}$	-695.8
4	$\theta_{24}^{\text{bo}} + \theta_{39}^{\text{bo}} + \theta_{23}^{\text{ro}} + \theta_{43}^{\text{ro}}$	-695.3
4	$\theta_{39}^{\text{bo}} + \theta_{24}^{\text{ro}} + \theta_{27}^{\text{ro}} + \theta_{43}^{\text{ro}}$	-693.8

^a BIC = 3.39.

find a reaction coordinate model that best fits the shooting point data.

We adopt the approach of Pan et al.⁴⁰ to choose the relevant set of OPs for higher order likelihood maximization screening. In order to reduce the combinatorial problem, only the relevant OPs obtained after single OP screening are used for higher order analysis. The benefit of adding an extra OP to the reaction coordinate model represented in eq 7 is evaluated until the improvement in likelihood score is less than the BIC.¹⁵ For our system, the value of the BIC is 3.39 ($N_R = 874$). For the case of 4 × 4 × 4 division of subcells, we present the best n-OP results in Table 4. We can clearly see that the best score for a combination of five OPs is not statistically any different from that of the best score for a combination of four OPs, since the improvement in the scores upon addition of an extra OP to the reaction coordinate model is less than the BIC. So for our system of 720 molecules of benzene, we obtain the combination of these four OPs to give us the best likelihood score. The subscripts denote the cell number (C) for which the respective local OP is evaluated. Another thing to note is that all combinations of four OPs are statistically equivalent if they lie within the BIC of the best likelihood score (-692.4). In Table 5, we display all possible combinations of four local OPs that are statistically indistinguishable from each other. We observe a much higher prevalence for local bond orientation and local relative orientation OPs in comparison to local distance order in describing a reaction coordinate model. Also we observe OPs from only 5 cells out of a total of 64. These cells (23, 24, 27, 39, 43) along with the local OPs calculated within them, seem to serve as a good model for the reaction coordinate for the set of data obtained from the aimless shooting simulations. In Figure 7, we color the system according to whether the molecules belong to these five cells or not. The relevant cells are colored in blue. It is very clear that all these five cells lie adjacent to one another in space. This is a good sign that local OPs calculated within a cluster of neighboring cells contribute to

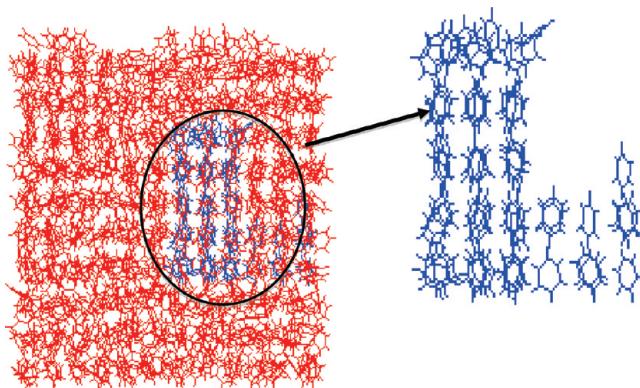


Figure 7. Benzene molecules colored according to the important cells obtained from OP screening by likelihood maximization. Cells 23, 24, 27, 39, and 43 (colored in blue) and their corresponding OPs are relevant for this system of melt crystallization.

the reaction coordinate model rather than individual cells spread across the system. When higher order likelihood screening was carried out for different divisions of the box (e.g., 8, 27, 125, etc.), the relevant local OPs belonged to similar spatially located clusters. This gives us confidence in our likelihood maximization analysis that the key OPs were correctly identified. It should be noted that the exact location of these cells within the box does not matter, and their location is an artifact of the nature of shooting points obtained from the aimless shooting simulations. Such a collection of cells and their corresponding orientational OPs could in principle be located anywhere in the box. The key implication is the role of local orientational OPs in describing the reaction coordinate for homogeneous nucleation.

One way of testing our reaction coordinate model is to obtain p_B histograms for these reaction coordinate models obtained from the likelihood maximization screening. We locate shooting points that satisfy the reaction coordinate models obtained by 1-OP, 2-OP, 3-OP, and 4-OP likelihood screening. Newer shooting points were obtained by performing BOLAS simulations⁵⁶ so that we have a collection of configurations describing the $p_B = 0.5$ isosurface for our reaction coordinate models. From each of these points, 20 half trajectories (of 400 ps) were initiated, and their end points were evaluated to see whether they reached the product (crystal) basin. In Figure 8, we plot the p_B histograms obtained for our four reaction coordinate models. The comparison with a binomial distribution is also shown in these plots (red line). Each histogram represents a collection of 600 MD trajectories. The mean and variance for each histogram and the corresponding errors in their values are calculated by performing the analysis given in a previous paper⁵⁷ and are presented in Table 6. We can observe that for every addition of a new OP to our reaction coordinate model, the mean of the p_B histogram improves. In Figure 8c,d, the means of histogram for our reaction coordinate models for combination of 3-OP and 4-OP are 0.538 and 0.508, respectively. Obtaining a p_B histogram peaked quite close to 0.5 is a good indication that our reaction coordinate model obtained from likelihood maximization screening serves to represent the actual transition state well. Our limited sampling of the transition state ensemble using aimless shooting and likelihood maximization is represented in the errors obtained for means and variances of these histograms (refer Table 6). If in principle, we could obtain an infinite number of reactive trajectories from aimless shooting simulations, the likelihood maximization

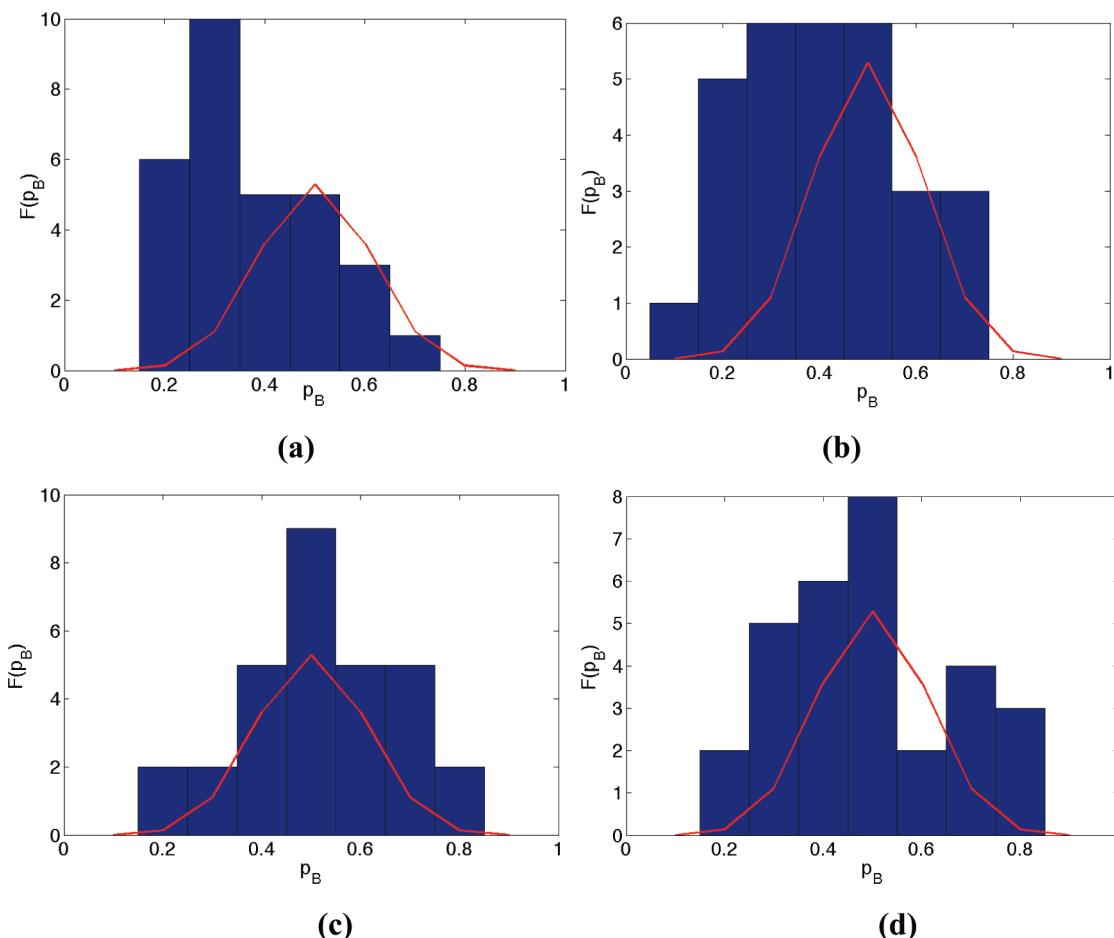


Figure 8. p_B histograms for best reaction coordinate models obtained from (a) 1-OP, (b) 2-OP, (c) 3-OP and (d) 4-OP. For plots c and d, the means of histograms are fairly close to the *ideal* value of 0.5. The red line represents the binomial distribution.⁵⁷

Table 6. Means and Standard Deviations of p_B Histograms Shown in Figure 6

reaction coordinate model	μ_h	σ_h
best 1-OP coordinate	0.39 ± 0.027	0.13 ± 0.28
best 2-OP coordinate	0.43 ± 0.024	0.14 ± 0.028
best 3-OP coordinate	0.538 ± 0.028	0.15 ± 0.029
best 4-OP coordinate	0.508 ± 0.03	0.16 ± 0.026

algorithm would provide us with a reaction coordinate model that would fit the binomial distribution well.

4. DISCUSSION

This study presents an important application of aimless shooting and likelihood maximization techniques to study rare events. These techniques allow us to obtain unbiased reaction pathways connecting two basins separated by a free energy barrier or large time scales. The likelihood maximization technique allows us to systematically analyze the relative importance of different collective variables in describing the transition state. This study improves upon previous simulation works on homogeneous nucleation^{19,30} by taking into account various aspects of orientational order, and also by accounting for their relative importance in understanding the nucleation mechanism. The works of Frenkel and co-workers considered the nucleation

mechanism along a fixed reaction coordinate comprising a cluster size calculated on the basis of Steinhardt bond orientational OP. Previous utility of path sampling calculations involving melt crystallization of sodium halides³⁰ were used to obtain relatively few reactive trajectories, without quantifying the role of different collective variables in describing the actual mechanism. This study clearly elucidates the role of orientational order, both bond and relative orientational OPs, in describing the nucleation pathway from liquid to solid benzene. It is again pertinent to note that the spatial location of the important local OPs itself is not that important. The results of likelihood maximization presented in Table 4 and Table 5 should be interpreted to understand the role of local orientational OPs in describing the nucleation mechanism. It is very surprising to note from our likelihood maximization results that even for a relatively flat and planar molecule such as benzene, relative orientation acts as an important collective variable in describing the reaction coordinate (Table 5). It can be speculated that for more complex molecules and systems involving solvent degrees of freedom, relative orientation might even play a much more important role in elucidating the nucleation mechanism. This study also justifies the utility and application of recently developed OPs for molecular crystals in understanding nucleation and crystallization.⁴⁶ Specifically with methods such as these, the role of different aspects of crystalline order, viz. distance versus bond orientation versus relative orientation, can be rigorously quantified in

description of nucleation mechanism in molecular crystals. It would be interesting to extend these methodologies and the OP framework to the complex problem of solution crystallization of small organic molecules and study whether distance order indeed precedes orientational ordering in the nucleation pathways, as speculated by the two-step nucleation theory.²

A notable finding of this study is the quantification of the role of cluster size as a possible reaction coordinate for nucleation in realistic systems, as assumed in CNT and other studies. Our results suggest that it does not play an important role in describing the transition state. This is in direct contradiction of assumptions inherent in CNT and other studies,¹⁹ that presumed the reaction coordinate to be the cluster size. Instead our results focus on importance of the local orientational OPs rather than the precise number of orientationally ordered molecules constituting a given cluster. It is very difficult to calculate the size of the crystalline cluster without knowing the relative importance of different OPs necessary to describe the crystallinity of a molecule. Also, it is not obvious whether we should exclude molecules from the cluster that have very low values of OPs but are extremely close to a strongly ordered molecule. Instead our results indicate that it might be useful to not consider one single definition of cluster size but instead focus on the individual local OPs. Previous simulation study on the Lennard-Jones liquid showed that in addition to cluster size, the shape and structure are also relevant variables in the reaction coordinate.⁷ We notice that there is not one specific shape for a critical cluster of molecules at the different shooting points and that the shape is definitely not spherical (Figure 7). Another observation from the collection of shooting points is that we never observe a fully formed crystalline cluster surrounded by liquid molecules. Instead we obtain molecules with partial (and varying) degrees of crystalline order. This highlights the fact that there is not a single configuration for the transition state, but an ensemble that encompasses a range of sizes and range of values for OPs. Thus another major assumption of CNT, which assumes a sharp interface between the two phases, becomes suspect. Similar insights were also obtained in previous theoretical studies employing density functional theory.^{9,60}

Another insight obtained from these simulations is that nucleation in organic crystals is a highly diffusive process.^{59,61} It was frequently observed that shooting trajectories spent a lot of time sampling the region near the transition state ensemble, before committing to either liquid or crystal basin. This led to a relatively low rate of conclusive trajectories (53%) in our aimless shooting simulations. One way of alleviating this issue would have been to study the system with very long MD trajectories. But again, this would severely limit the number of trajectories that can be sampled in a reasonable amount of time. For the case of shooting trajectories associated with the 4-OP reaction coordinate model (refer to Figure 8d), we calculated the number of recrossing events (every 4 ps) observed within the $p_B \sim 0.5$ isosurface, before they reached $p_B \sim 0$ or $p_B \sim 1$ isosurfaces. It was observed that the average number of recrossing events into the $p_B \sim 0.5$ isosurface was more than 5 for 400 ps shooting trajectories (detailed plot provided in the Supporting Information, Figure S1). This analysis provides suitable quantification for the diffusive nature of the nucleation mechanism. Regular TPS calculations^{16,62} were also tested out for this system, where only the Langevin thermostat was used to obtain a new trajectory from an old one, without perturbing the momenta of atoms.⁶³ Due to the highly diffusive nature of this problem, it was immensely

difficult to get new reactive trajectories from old ones. In contrast, aimless shooting methodology serves as a much better alternative among existing path sampling algorithms to obtain higher acceptance rates for highly diffusive transitions.⁵⁹

This study was carried out for melt and was conducted for a very small system (720 molecules). The computational requirements for a bigger and more complex system (involving solvents and/or other surfaces) might get prohibitively expensive to obtain a reasonable number of accepted trajectories while performing aimless shooting (or other path sampling) techniques. Recently methods such as string method in collective variables (SMCV)⁶⁴ and Voronoi dynamics⁶⁵ have been developed to obtain minimum free energy pathways in rare events. Such methods need to be tested and validated to study the nucleation mechanism in systems of pharmaceutical importance. However these methods (like umbrella sampling) require *a priori* knowledge of the important collective variables. A new method developed by Bolhuis and co-workers,⁵⁵ that is capable of combining the string method with the likelihood maximization technique, seems to be a promising way to study such systems. Equilibrium path sampling³⁹ or BOLAS⁵⁶ also serve as attractive options in studying such systems in order to gain insights into transitions states and obtaining free energy of activation. Even with the issue of computational efficacy in terms of time, aimless shooting and likelihood maximization are quite powerful methods to extract insights into transitions states and quantify the relative importance of several collective variables in modeling the reaction mechanism. We are in the process of extending its utility in understanding the more complicated problems of solution crystallization and heterogeneous nucleation.

5. SUMMARY AND CONCLUSIONS

In this research, we present the utility of techniques such as aimless shooting and likelihood maximization to understand the nucleation mechanism in melt crystallization in benzene. The key OPs responsible for nucleation were found to be local OPs, specifically bond and relative orientational OPs, localized in a specific portion of the overall system. However, the size of biggest crystalline cluster was itself found to be fairly unimportant in describing the reaction coordinate for nucleation mechanism. Other features of CNT were also found to be in contradiction with our results. We did not observe a fully crystalline critical nucleus surrounded by liquid at the shooting points. The shooting points consist of clusters of molecules with partial degrees of order. This implies that the use of a single interfacial tension becomes a flawed assumption in the CNT. In concurrence with previous simulation results, our study shows the importance of local orientational order in understanding nucleation. The role of relative orientation between neighboring molecules to describe the nucleation transition has been shown for the first time. Our results also provide justification for the utility of recently developed OPs for molecular crystals in practical applications of gaining molecular level understanding of nucleation and crystallization.

Even though aimless shooting along with reaction coordinate validation can prove to be extremely computationally intensive for fairly complex systems, it is still quite a powerful technique to study rare events in chemical and biological systems. Together with likelihood maximization, it offers a systematic and unbiased way of screening collective variables to obtain a reaction coordinate.

■ ASSOCIATED CONTENT

S Supporting Information. The Supporting Information document contains results and analysis pertaining to the diffusive nature of the nucleation mechanism and the ensemble of states visited during the path sampling simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Mail: Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E19-502B, Cambridge MA, 02139. Phone: 617-258-5021. Fax: 617-253-2272. E-mail: trout@mit.edu.

■ ACKNOWLEDGMENT

We would like to thank the Novartis-MIT Center for Continuous Manufacturing for their generous financial support for this study. We would like to acknowledge Gregg Beckham for helpful discussions. We would also like to acknowledge the help from TeraGrid (Project # CTS10033) and Texas Advanced Computing Center (Ranger) for their computational support. We would also like to thank our reviewers for their suggestions and insightful comments.

■ REFERENCES

- (1) Mullin, J. W. *Crystallization*, 4th ed.; Butterworth-Heinemann: Oxford, 2001.
- (2) Erdemir, D.; Lee, A.; Myerson, A. *Acc. Chem. Res.* **2009**, *42*, 621–629.
- (3) Vekilov, P. *Cryst. Growth Des.* **2010**, *10*, 5007–5019.
- (4) Debenedetti, P. G. *Metastable Liquids*; Princeton University Press: Princeton, NJ, 1996.
- (5) Bernstein, J. *Polymorphism in Molecular Crystals*, 1st ed.; Oxford University Press: New York, 2002.
- (6) Anwar, J.; Zahn, D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1996–2013.
- (7) Moroni, D.; ten Wolde, P.; Bolhuis, P. *Phys. Rev. Lett.* **2005**, *94*, 235703.
- (8) Gasser, U.; Weeks, E.; Schofield, A.; Pusey, P.; Weitz, D. *Science* **2001**, *292*, 258–262.
- (9) Oxtoby, D. *J. Phys.: Condens. Matter* **1992**, *4*, 7627–7650.
- (10) Oxtoby, D. *Philos. Trans. R. Soc., A* **2003**, *361*, 419–427.
- (11) Duff, N.; Peters, B. *J. Chem. Phys.* **2009**, *131*, 184101.
- (12) Vatamanu, J.; Kusalik, P. G. *Phys. Chem. Chem. Phys.* **2010**, *12*, 15065.
- (13) Yau, S.; Vekilov, P. *Nature* **2000**, *406*, 494–497.
- (14) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: New York, 2001.
- (15) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125*, 054108.
- (16) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (17) Best, R.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.
- (18) Weinan, E.; Ren, W.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2005**, *413*, 242–247.
- (19) Van Duijneveldt, J.; Frenkel, D. *J. Chem. Phys.* **1992**, *96*, 4655–4668.
- (20) ten Wolde, P.; Ruiz-Montero, M.; Frenkel, D. *Phys. Rev. Lett.* **1995**, *75*, 2714–2717.
- (21) ten Wolde, P.; Ruiz-Montero, M.; Frenkel, D. *J. Chem. Phys.* **1996**, *104*, 9932–9947.
- (22) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
- (23) Steinhardt, P.; Nelson, D.; Ronchetti, M. *Phys. Rev. B* **1983**, *28*, 784–805.
- (24) Dellago, C.; Bolhuis, P.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 9236–9245.
- (25) Bolhuis, P.; Dellago, C.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5877–5882.
- (26) Dellago, C.; Bolhuis, P. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*; Reiher, M., Ed.; Topics in Current Chemistry; Springer: New York, 2007; Vol. 268, pp 291–317.
- (27) Pan, A.; Chandler, D. *J. Phys. Chem. B* **2004**, *108*, 19681–19686.
- (28) Radhakrishnan, R.; Trout, B. *J. Chem. Phys.* **2002**, *117*, 1786–1796.
- (29) Radhakrishnan, R.; Trout, B. *J. Am. Chem. Soc.* **2003**, *125*, 7743–7747.
- (30) Zahn, D. *J. Phys. Chem. B* **2007**, *111*, 5249–5253.
- (31) Zahn, D.; Leoni, S. *Phys. Rev. Lett.* **2004**, *92*, 250201.
- (32) Zahn, D. *Phys. Rev. Lett.* **2004**, *93*, 227801.
- (33) Peters, B.; Beckham, G. T.; Trout, B. L. *J. Chem. Phys.* **2007**, *127*, 034109.
- (34) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769–6779.
- (35) Beckham, G.; Peters, B.; Starbuck, C.; Variakaval, N.; Trout, B. *J. Am. Chem. Soc.* **2007**, *129*, 4714–4723.
- (36) Beckham, G.; Peters, B.; Trout, B. *J. Phys. Chem. B* **2008**, *112*, 7460–7466.
- (37) Juraszek, J.; Bolhuis, P. *Biophys. J.* **2008**, *95*, 4246–4257.
- (38) Vreede, J.; Juraszek, J.; Bolhuis, P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2397–2402.
- (39) Peters, B.; Zimmermann, N.; Beckham, G.; Tester, J.; Trout, B. *J. Am. Chem. Soc.* **2008**, *130*, 17342–17350.
- (40) Pan, B.; Ricci, M.; Trout, B. *J. Phys. Chem. B* **2010**, *114*, 4389–4399.
- (41) Raiteri, P.; Martonak, R.; Parrinello, M. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 3769–3773.
- (42) Bacon, G.; Curry, N.; Wilson, S. *Proc. R. Soc. London A* **1964**, *279*, 98–110.
- (43) Akella, J.; Kennedy, G. *J. Chem. Phys.* **1971**, *55*, 793–796.
- (44) Cansell, F.; Fabre, D.; Petitet, J. *J. Chem. Phys.* **1993**, *99*, 7300–7304.
- (45) Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. *J. Am. Chem. Soc.* **2004**, *126*, 14278–14286.
- (46) Santiso, E.; Trout, B. *J. Chem. Phys.* **2011**, *134*, 064109.
- (47) Allen, F. H. *Acta Crystallogr. B: Struct. Sci.* **2002**, *58*, 380–388.
- (48) MacKerell, A.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
- (49) MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (50) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (51) Xu, S.; Bartell, L. S. *J. Phys. Chem.* **1993**, *97*, 13544–13549.
- (52) Kinney, K. E.; Xu, S.; Bartell, L. S. *J. Phys. Chem.* **1996**, *100*, 6935–6941.
- (53) Mardia, K. V.; Jupp, P. E. *Directional Statistics*, 2nd ed.; Wiley: New York, 1999.
- (54) Radhakrishnan, R.; Schlick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5970–5975.
- (55) Lechner, W.; Rogal, J.; Juraszek, J.; Ensing, B.; Bolhuis, P. *J. Chem. Phys.* **2010**, *133*, 174110.
- (56) Nocedal, J.; Wright, S. *Numerical Optimization*, 2nd ed.; Springer: New York, 2006.
- (57) Radhakrishnan, R.; Schlick, T. *J. Chem. Phys.* **2004**, *121*, 2436–2444.

- (58) Peters, B. *J. Chem. Phys.* **2006**, *125*, 241101.
(59) Peters, B. *Mol. Simul.* **2010**, *36*, 1265.
(60) Oxtoby, D. *Acc. Chem. Res.* **1998**, *31*, 91–97.
(61) Zahn, D.; Hochrein, O.; Kawska, A.; Seifert, G.; Grin, Y.; Kniep, R.; Leoni, S. *Sci. Technol. Adv. Mater.* **2007**, *8*, 434–441.
(62) Geissler, P.; Dellago, C.; Chandler, D. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1317–1322.
(63) Bolhuis, P. *J. Phys.: Condens. Matter* **2003**, *15*, S113–S120.
(64) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106.
(65) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, 194101.