ABSTRACT

Crystallization is widely used in many industries including specialty chemicals and pharmaceuticals for separation and purification purposes. Many solids exhibit polymorphism, the ability of a material to exist in multiple crystal structures, and crystal structure affects key material properties like stability, solubility, and dissolution rate. Thus, controlling the crystal structure in a crystallization process is essential for the production of materials with desired properties. Crystal nucleation, with which the formation of the solid phase begins, is the key event that governs the kinetics of crystallization and plays an important role in determining the resulting crystal structure. Hence, knowledge of nucleation kinetics is necessary for the design and optimization of crystallization processes.

Despite its importance, crystal nucleation is still not fully understood. This mainly results from the difficulty in observing nucleation with experimental techniques, both due to the small time and length scales involved in the formation of critical clusters and the stochastic nature of nucleation as an activated process. Thus, there is a lot of trial-and-error that goes into crystallization process design, taking significant time and resources. Theoretical predictions of nucleation kinetics, on the other hand, can accelerate process design but require a good understanding of the nucleation mechanism. Classical nucleation theory (CNT) provides a kinetic expression that can be used to predict nucleation rates using the thermodynamic properties of bulk crystal slabs, assuming that these properties remain the same for clusters at the nanoscale. However, there is growing evidence in the literature showing that the cluster structure at the time of nucleation can be wildly different than that when the crystal has grown to its bulk form. There is still much to be learned about predicting nucleation rates and incorporating nucleation kinetics into the macroscopic models of industrial crystallizers when nucleation occurs through these non-classical mechanisms.

Molecular simulations can help with both understanding the nucleation mechanism at the atomic scale and predicting nucleation rates. This thesis demonstrates the use of molecular simulations and specifically multidimensional free energy calculations as a powerful tool to understand non-classical nucleation mechanisms. With a motivation to develop a compu-

tational methodology that can predict nucleation rates and identify the kinetically favored polymorph without relying on CNT, two studies have been conducted:

First, to develop the multidimensional approach for calculating nucleation rate, NaCl nucleation from concentrated aqueous solutions was studied, based on literature evidence for 2-step nucleation. By calculating the free energy of nucleation as a function of structure specific nucleus size coordinates, it was shown that at high concentrations, NaCl nucleation occurs through the formation of composite clusters that have a crystalline core surrounded by a layer of amorphous particles. The effect of supersaturation on the mechanism was investigated by rerunning the calculations at a higher concentration and a shift was observed in the nucleation mechanism from single step to 2-step nucleation. The shift was shown to be resulting from a change in the relative stability of the intermediate amorphous solid phase. This study demonstrated the use of multidimensional free energy surfaces for understanding the nucleation mechanism but did not include multiple polymorphs.

Next, to explore the free energy surface of nucleation in the presence of competing crystal structures, crystallization of the Lennard-Jones (LJ) fluid from the melt phase was investigated. The choice of LJ fluid as the crystallization system allowed for quick sampling of the free energy while still exhibiting non-classical nucleation behavior. The free energy surface of nucleation, plotted against polymorph-specific nucleus size variables, revealed that polymorph selection in the LJ fluid does not happen during nucleation, but when the emerging clusters are much larger than the critical cluster size, in contrast with the classical nucleation theory assumption. It was observed that the post-nucleation events like crystal growth and polymorphic transformations played a significant role in determining the final structure of the grown crystals. Since a CNT based crystallization model is not able to track the structural changes within the nucleated clusters, a novel population balance modeling framework was proposed. The proposed model connects the structure-specific nucleation rates obtained from atomic-scale simulations with the post-nucleation events and was able to predict the correct polymorphic distribution obtained from 500 nucleation trajectories.

This thesis represents a step forward to a complete *in silico* design of crystallization processes for producing materials with the desired crystal structure. Next steps include application of the methods described here to more complex molecules that are of interest in the pharmaceutical industry, to serve the ultimate goal of accelerating the drug production pipeline. To this end, a deep learned interaction energy model for predicting inter-molecular interactions, which can potentially make multidimensional free energy calculations of complex molecules attainable, is currently under development for glycine, a small yet industrially relevant molecule. From this dissertation, the multidimensional approach to nucleation combined with the trained interaction energy model appears to be promising towards resolving the long standing problem of polymorph prediction in the pharmaceutical industry.