

Group Contribution Method for Rapid Estimation of Crystal Growth Rates

Anish V. Dighe¹, Prem K.R. Podupu¹, Vamsi Vikram Gande¹, Urmila Diwekar^{2,3,*}, and Meenesh R. Singh^{1,*}

¹Department of Chemical Engineering, University of Illinois Chicago, Chicago, IL 60607, United States of America

²Vishwamitra Research Institute, Clarendon Hills, IL 60514, United States of America

³Richard and Loan Hill Department of Biomedical Engineering, University of Illinois Chicago, Chicago, IL 60607, United States of America

***Corresponding Author:**

Prof. Meenesh R. Singh
Associate Professor
Department of Chemical Engineering
929 W. Taylor St
University of Illinois Chicago
Chicago, IL 60607
Tel: (312) 996-3424
Email: mrsingh@uic.edu

Dr. Urmila Diwekar
President
Vishwamitra Research Institute
368 5th Street
Clarendon Hills, IL 60514
Email: urmila@vri-custom.org

Keywords: Growth Rates, Crystallization, Group Contribution Method, Stochastic Optimization, Organic Crystals

Abstract

The morphological evolution of organic crystals during crystallization depends on the face-specific growth rates. Classical growth rate models relate the face-specific growth rates to the crystal lattice, energy of stable facets, growth mechanism, and supersaturation. The complexities of these models have increased over time to account accurately for solution conditions, the structure of growth units, and their attachment rates. Such advanced growth rate models require several layers of computations to obtain attachment energies of facets, nucleation rates, kink density, and attachment rates. Among these, the most intensive and time-consuming computation is for attachment rates, which require molecular dynamic simulations. This substantially increases the overall computation time to predict the absolute growth rate for even one crystallization condition. Since it is nearly impossible to iterate such a growth rate model, optimization schemes cannot be implemented to identify solution conditions that favor specific crystal growth. To reduce the computational time for attachment rate calculations, we implement a group contribution method (GCM) that relates the properties of functional groups in a molecule to their attachment rates to the crystal lattice, thereby rapidly estimating the growth rates of organic crystals. The process of molecular attachment involves partial desolvation of a solvated molecule, referred to as a transition state, followed by total desolvation via spontaneous attachment to a crystal facet. The first step in GCM is to identify the equilibrium states of fully solvated and partially desolvated solute molecules. The degree of supersaturation dictates the extent of this equilibrium and, thereby, the activation barrier for the growth of crystals, according to transition state theory. Identifying this equilibrium phenomenon allows for capturing the functional-group-specific interactions that depend on molecular motion, which could be related to operating conditions such as temperature and pressure. The stochastic optimization technique with Monte-Carlo sampling allows an efficient optimization problem solution to obtain the group interaction parameters. The GCM approach is first validated for the estimation of growth rates of glutamic acid and L-histidine, and then extended to predict growth rates of alanine and glycine rapidly. The optimized parameters and GCM scheme can be used to estimate growth rates in other crystallization systems.

1. Introduction

Physical properties of pharmaceutical organic crystals, such as dissolution, mechanical strength, and bioavailability, and the post-processability during drying, filtration, and tableting, depend on the morphology of crystals.^{1,2} Morphology, in turn, depends upon the growth rates of various faces during the process of crystallization.³ The self-assembly of molecules affects the growth rates and is sensitive to the degree of supersaturation and the solvent environment.⁴⁻⁷ Currently, there is no generalized method to estimate growth rates rapidly. The limited understanding of molecular self-assembly and the large number of molecules involved during the process of crystallization contributes to the complexity of the problem. Most commonly, crystallization is understood with the help of operating conditions, such as temperature, pressure, and supersaturation. Although controlling supersaturation is the most common and effective method to control crystal growth rates, a large number of experiments are required to estimate the growth rates of a single system.

Estimating the absolute growth rates of small organic molecules using computational approaches is a common practice in the field of crystallization. In general, molecular simulations or empirical relations are used to determine the activation energy⁸ and this activation energy is used in the mechanistic growth rate model to determine the growth rate. In our previous work, a series of computational simulations were conducted to investigate the behavior of solid-liquid interfaces in supersaturated solutions.^{9,10} We employed a comprehensive computational approach that involved characterizing solvation dynamics, calculating binding energies, and determining activation barriers to model the crystal growth rates and shapes in supersaturated solutions. Tanaka et al.¹¹ analyzed the crystal growth of urea with all-atom molecular dynamics (MD) simulation for the (001) and (110) faces. The number of atoms was several times larger in the MD of crystal growth than in the free-energy calculation with the energy-representation method. Accordingly, the computational load was less by 2–3 orders of magnitude for the adsorption free energy than for the growth rate. Kontkanen et al.¹² used molecular-resolution cluster population simulations to determine growth rate of atmospheric particles involving sulfuric acid and organic vapors which result from the interplay of factors such as the concentration and evaporation rates, particle population dynamics, and stochastic fluctuations. Li et al.¹³ have created detailed mechanistic models for crystal growth and introduced a prototype software tool named ADDICT (Advanced Design and Development of Industrial Crystallization Technology). ADDICT needs electron density calculations from Gaussian¹⁴ as an input. Although ADDICT methodology is applicable to any choice of forcefield, Generalized Amber Force Field (GAFF) parameter database is employed to assess intermolecular interactions within the supercell. Further with these intermolecular interactions ADDICT compiles a list of low Miller index planes based on the crystal space group, identifying the most stable slice for each one. Subsequently, it establishes periodic bond chains (PBCs) for each slice and compute energies, which are then used in the mechanistic growth model to determine the growth rate. ADDICT focuses only on implementing a model for spiral growth. These growth rates are crucial for understanding the kinetics and thermodynamics of chemical reactions, as well as for designing and optimizing chemical processes and materials. Tilbury et al.¹⁵ established a connection between pre-existing mechanistic expressions for spiral growth and two-dimensional nucleation growth regimes. They accomplished this by applying

stationary nucleation rate theory to integrate insights from both spiral and two-dimensional nucleation growth regimes. This innovative approach serves to bridge various growth regimes, extending beyond the spiral growth model previously developed by Li et al.¹³ Han et al.¹⁶ performed experiments and constant chemical potential molecular dynamics simulations to explore the solvent-mediated crystal morphology selection of the anti-tuberculosis drug isoniazid, revealing insights into the influence of solvent polarity on crystal aspect ratio and providing a mechanistic understanding of surface growth control. Piana and Gale¹⁷ performed dynamical atomistic simulation to investigate the dissolution and growth of urea crystals, revealing a nucleation and growth mechanism on the [001] face, with the removal of surface defects identified as the rate-limiting step, and the evolution of crystal growth characterized by a rough surface topography rather than a clean layer-by-layer mechanism. Turner et al.¹⁸ employed mechanistically based digital workflow for solvent-mediated crystal morphologies of the α and β polymorphic forms predictions. It involved the calculation of crystal lattice energy, intermolecular synthons, interaction energies, and their pivotal role in comprehending and predicting crystal morphology.

The majority of research has primarily concentrated on straightforward organic compounds like urea and glycine, primarily due to the heightened computational demands associated with molecular dynamics (MD) simulations. To overcome some of the constraints of MD methods, researchers have turned to Kinetic Monte Carlo (kMC) techniques, allowing for simulations of crystal growth and dissolution processes at a closer approximation to the mesoscale.¹⁸ However, existing approaches do come with certain challenges and requirements like high computation power, knowledge of molecular topology, quantum mechanical calculations, appropriate force fields and appropriate computational methodology. Depending on the computational power available, computational time for these simulations vary accordingly. For example, MD simulations have typical computational times of around 0.3 hours per nanosecond for a system with 500 molecules on a laptop (quad-core 3.0 GHz processor).¹⁹ A summary of previous literature with simulation times and limitations is provided in **Table S1** of the supplementary information. Since it is nearly impossible to iterate such a growth rate model, optimization schemes cannot be implemented to identify solution conditions that favor specific crystal growth.

In this work, we extend the knowledge gained from such a computational approach to rapidly estimate crystal growth rates using the group contribution method (GCM) coupled with adsorbate solid-solution theory (ASST). GCM is used to estimate various properties or parameters of chemical compounds. In GCM molecular structure of a compound is broken down into groups (functional groups or fragments or building blocks or descriptors) and the property of a compound is estimated through the summation of the contributions of each group representing the molecule.²⁰ Property models operate under the assumption that the contribution of a specific group to a property remains consistent across all compounds containing that group. These models posit that a compound's property value is determined by the collective contributions of all the groups necessary to uniquely characterize its molecular structure. However, current GCM approaches used for crystallization hypothesize an equilibrium relationship between operating conditions (such as temperature, solvent concentration, and supersaturation) and the outcome of crystallization (such as crystalline yield or growth rates).^{21,22} This hypothesis is not valid since

crystallization is known to be kinetically driven and the outcome of crystallization is determined by kinetics and not the thermodynamics alone.²³⁻²⁵ In this work, we identify the equilibrium between the bulk and transition phases of the small organic molecules, to effectively use GCM for the estimation of crystal growth rates.

With the help of molecular dynamics (MD) simulations, the bulk and transition phases of organic molecules during the process of crystallization have been identified based on the number of solvent molecules around the solute molecules.⁹ In the bulk solvated phase, which is considered fully solvated, a solute molecule is surrounded by a large number of solvent molecules resulting in a thicker solvation shell compared to transition phase. At the transition phase, the number of solvent molecules around the solute molecules is lower, allowing higher solute-solute interactions and is considered partially desolvated. At or below supersaturation, the fully-solvated and partially-desolvated phases are in equilibrium. However, the extent of equilibrium shifts towards a partially desolvated phase in a supersaturated system. The formation of crystal from a partially desolvated state in a supersaturated system is instantaneous, and attaining the transition state is the rate-limiting step. Identification of such shift in equilibrium allows applying GCM to crystallization based on the molecular events.

The significant difference between the fully solvated and partially desolvated state is the number of solvent molecules around the solute molecule. It was shown that the removal of solvent molecules in the solvation shell contributes to the activation barrier for the attachment.⁹ Furthermore, the activation barrier and the change in the number of solvent molecules in the solvation shell follow a linear relationship.⁹ This knowledge allows the formulation of a central framework for the application of GCM to estimate crystal growth rates rapidly. The significant complexity of the crystallization process then boils down to estimating the number of solvent molecules removed from the solvation shell of the solute molecule. To estimate this number, we apply ASST, GCM, and modified UNIFAC model to relate the mole fraction of the solvent molecules to the group interaction parameters.^{26,27} We apply the GCM coupled with ASST to rapidly estimate growth rates of glutamic acid, histidine, glycine, and alanine. The molecular picture of the shift in equilibrium and the ASST framework is shown in **Figure 1**.

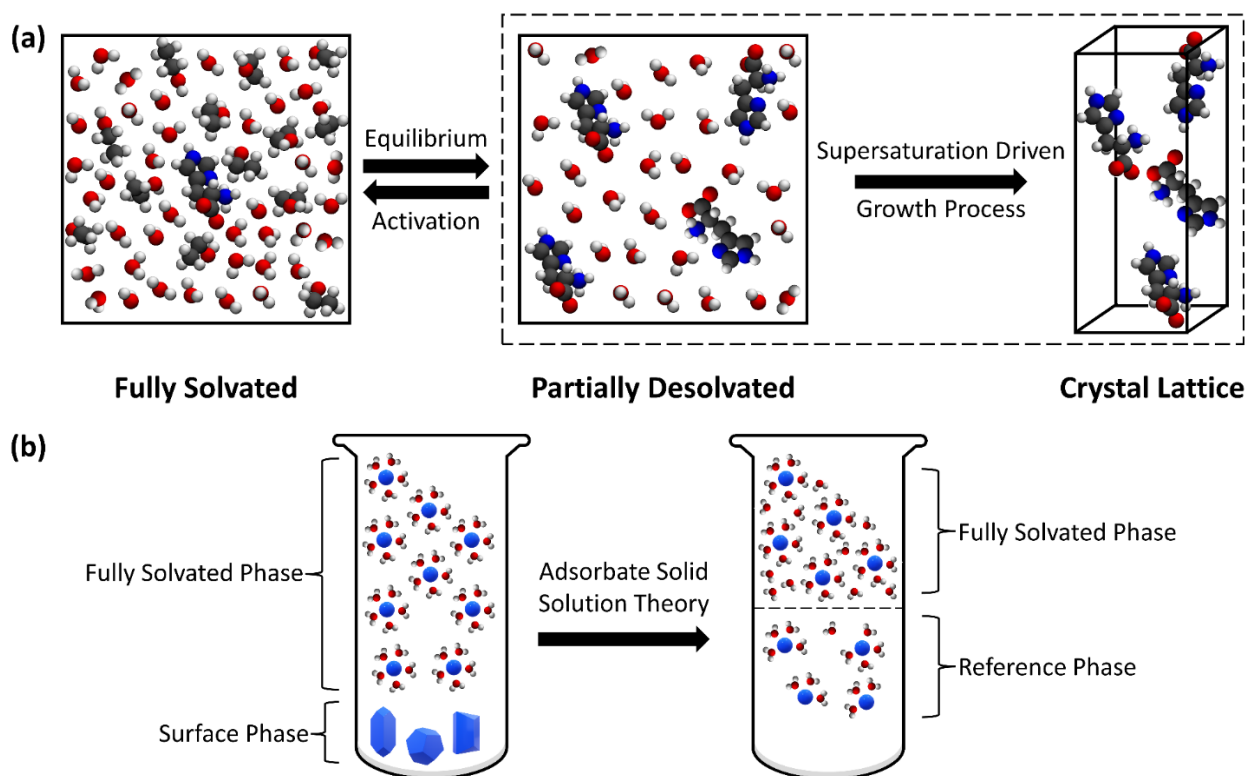


Figure 1: Central framework for the GCM approach for crystallization for rapid estimation of growth rates of organic molecules. (a) Molecular picture of the process of crystallization. In bulk, the solute molecule (histidine) is surrounded by the solvent (water) and antisolvent (ethanol) molecules and is fully solvated. Supersaturation causes higher solute-solute interactions resulting in the depletion of the solvation shell resulting in a partially desolvated solute molecule. The supersaturation-driven growth process is an instantaneous step in supersaturated systems. The extent of equilibrium between fully solvated and partially desolvated systems governs the activation barrier. (b) Adsorbate solid-solution theory: Instead of looking at the non-equilibrium behavior of crystallization, ASST looks at the crystal growth where the reference phase (partially desolvated molecules) undergoes crystallization. It allows for building a thermodynamic framework in terms of excess Gibbs free energies and relates mole fractions to group interaction parameters.

2. Theoretical Methods:

2.1. Adsorbate Solid-Solution Theory (ASST):

In the classical description of adsorption, molecules from the bulk (or fully solvated) phase get adsorbed onto the surface phase, dividing the system into two phases, as shown in Fig 1b (left). Similarly, in the case of crystallization, a solute molecule from the bulk is integrated into the nucleated crystal. The integration refers to the adsorption of the molecule onto the crystal surface and reorientation for the formation of periodic arrangement. The rate-limiting step of the integration is adsorption, which requires solute molecules to shed their solvation shell. This step of the crystallization process can be analyzed from the Adsorbate solid-solution theory perspective. ASST describes the behavior of fluid on a solid surface in a thermodynamic framework. In the case of crystallization, the fluid is the supersaturated solution, and the nucleated crystal is the solid.

ASST treats the surface phase as a mixture of adsorbent and adsorbing species, and this is called as reference phase, as shown in Fig 1b (right). In this study, the reference phase is the partially desolvated solute molecules. Using this theory, mole fractions and activity coefficients of various components in the reference and bulk phases are related to the excess energies. These relations are given by Berti et al.,²⁰ and the phase equilibrium between reference and bulk phases for each component is shown below:

$$x_i^b \gamma_i^b = x_i^* \gamma_i^* \exp\left(\frac{\varphi^* - \varphi_{o,i}^*}{RT\Gamma^s}\right) \quad (1)$$

Where x_i is the mole fraction of i^{th} species, γ_i is the activity coefficient, φ represents chemical potential, R is the gas constant, T is the temperature, Γ^s is the adsorption capacity of the adsorbent. Superscript b represents bulk phase, $*$ represents reference phase, and subscript o, i represents pure phase component. The difference in chemical potentials between adsorbent before and after adsorption is given by $\varphi^* - \varphi_{o,i}^*$, and is further elaborated in terms of excess Gibbs energies as:

$$\varphi^* - \varphi_{oi}^* = \frac{1}{m_0} (G^{E*} - G^{Es} - G_{oi}^{E*}) \quad (2)$$

Where, G^{E*} represents excess Gibbs free energy of the reference phase, G^{Es} is the excess Gibbs free energy of the surface phase, G_{oi}^{E*} refers the excess Gibbs free energy of component's adsorption, and m_0 is the molar mass of adsorbent. The excess Gibbs free energy is then related to activity coefficient as follows:

$$\frac{G^{E*}}{RT} = \sum_{i=0}^k n_i^* \ln \gamma_i^* \quad (3)$$

$$\frac{G^{Es}}{RT} = \sum_{i=1}^k n_i^s \ln \gamma_i(x_i^s) \quad (4)$$

Where, index zero is the adsorbent and n_i , molar quantity of component i , is calculated as:

$$n_i^s = x_i^s n_T^* (1 - x_o^*) \quad (5)$$

The mole fractions of component i in the surface phase are calculated as follows:

$$x_i^s = \frac{x_i^*}{1 - x_o^*} \quad (6)$$

$$\gamma_{o,i}^* = 1 + \frac{1}{\Gamma^s M_0} \quad (7)$$

2.2. Group Contribution Method (GCM):

While pure component activity coefficient is calculated as shown in Equation 7 (where M_0 is the molecular weight of the adsorbent), group contribution method can be used to obtain the activity coefficients of various components in the mixture.²⁸ In the GCM, components are divided

into groups of atoms and the parameters are calculated based on the frequency of that group times its contribution. These parameters can then be used to estimate the interaction parameters and thereby activity coefficients, even when there is no experimental data. GCM combined with modified UNIFAC as G^E model can be used to calculate the excess Gibbs energy of the reference phase.²⁰ For this the activity coefficient of the reference phase is divided to two parts: a pure component part and concentration dependent part as shown:

$$\ln\gamma_i^* = \ln\gamma_{o,i}^* + \ln\gamma_{GE,i}^* \quad (8)$$

Where $\gamma_{GE,i}^*$ is the concentration dependent activity coefficient given by the UNIFAC model. It is then further broken down into combinatorial (superscript C) and residual (superscript R) parts:

$$\ln\gamma_{GE,i}^* = \ln\gamma_i^{C*} + \ln\gamma_i^{R*} \quad (9)$$

Similarly, activity coefficient for the bulk phase is given by:

$$\ln\gamma_i^b = \ln\gamma_i^{Cb} + \ln\gamma_i^{Rb} \quad (10)$$

Residual part of the above equation for bulk phase (γ_i^{Rb}) could be calculated as:

$$\ln\gamma_i^{Rb} = Q_g \left[1 - \ln \left(\sum_1^m \theta_m \psi_{mg} \right) - \left(\sum_1^m \frac{\theta_m \psi_{gm}}{\sum_1^n \theta_n \psi_{nm}} \right) \right] \quad (11)$$

$$\psi_{nm} = \exp \left(\frac{-a_{nm}}{T} \right) \quad (12)$$

Where Q_g represents van der Waals surface area, θ_m is the surface area fraction, and ψ_{nm} is the temperature-dependent parameter, which further depends on the group interaction parameter a_{nm} between the groups of type n and m , respectively. Further elaborations for γ_i^{C*} , γ_i^{R*} , γ_i^{Cb} , and θ_m are given in **Section S3** of the supplementary information.

3. Computational Approach:

3.1. Sequence of theoretical calculations:

In brief, the method is divided into three stages: gathering data from literature, performing optimization to get missing group interaction parameter values, and estimating growth rates. The initial two stages are dedicated to estimating the missing group interaction parameter values, employing the methods and data already published in our previous works. The final stage involves the computation of growth rates, utilizing either the interaction parameters derived from the preceding stages or those extracted from existing literature.

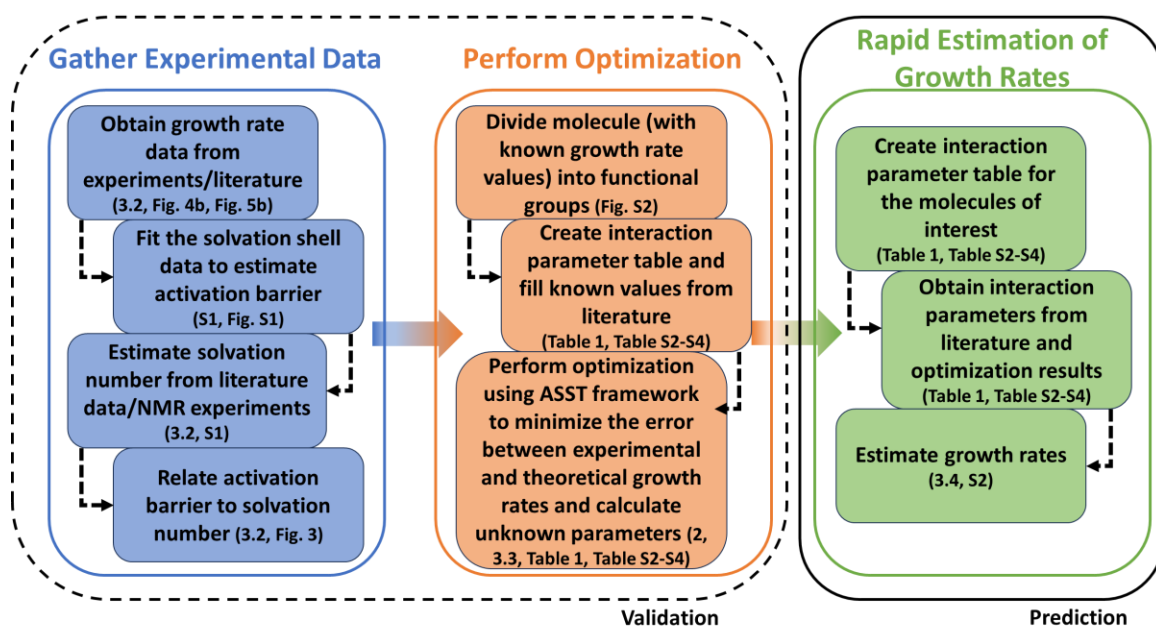


Figure 2: Flowchart of the overall approach, with references to the locations pertaining to the manuscript’s information (Content in Supplementary information is denoted with prefix “S”)

3.2. Gather Experimental Data:

In this stage, growth rate data obtained from literature or experiments is used to estimate the activation energy. Additionally, solvation number is obtained from the literature or experiments, which is then related to the activation energy. Solvation number is the number of solvent molecules that remain associated with the solute during its movement through the solution. (It is called hydration number when the solvent is water.) The change in the number of water molecules (hydration number) during the cooling crystallization of glutamic acid and antisolvent crystallization of histidine has been reported in the literature.^{9,10} Hydration numbers for fully solvated and partially desolvated phases were obtained from the MD simulations, and the activation barrier was calculated from the Double-Well approach for both works. Activation energy here is the energy required for a fully solvated state to become a partially desolvated state or transition state. After the transition state, it is a downhill process where the molecules readily integrate into the kink sites on the crystal surface. Additional information regarding solvation number and Double-Well approach (**Figure S1**) are provided in **Section S1** of the supplementary information.

Furthermore, it was shown that the relationship between the change in hydration number and the activation barrier for crystal growth rates is linear and follows an Evans-Polanyi relationship, as shown in **Figure 3**. Apart from this, it should be noted that while the change in hydration number from fully solvated to partially desolvated phases affects the growth rates, functional group specific solvation shells result in polymorph selection.²⁹ In short,

$$\dot{g} = f_1(\Delta E_{act}) \quad (13)$$

$$\Delta E_{act} = f_2(N_{b,w} - N_w^*) \quad (14)$$

where, $\dot{g}$ is the growth rate and ΔE_{act} is the activation barrier. $N_{b,w}$ is the number of water (solvent) molecules in the solvation shell of the fully solvated molecule (bulk solvated phase), and N_w^* is the number of water molecules in the solvation shell of the partially desolvated molecule (reference phase). Further details about equation (13) are given in **Section S2** of the supplementary information. The relationship shown in equation (14) and the data obtained from computations is depicted in **Figure 3**.

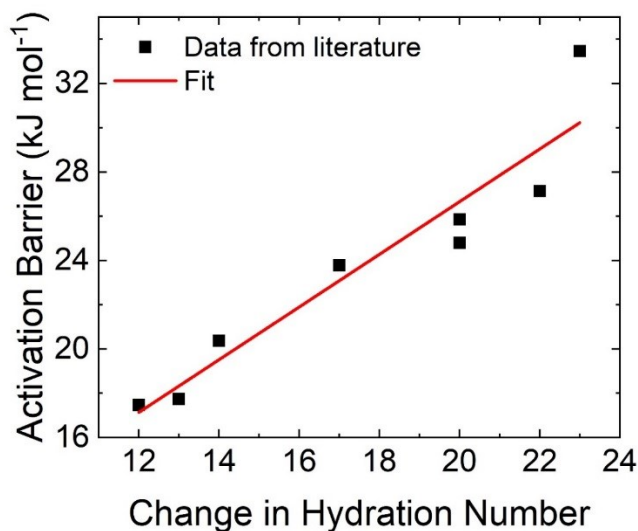


Figure 3: The relationship between activation barrier and change in the number of water molecules (hydration number) obtained from the literature.^{9,10} The scatter points are data points obtained from literature, and the solid red line is the linear fit.

The mole fraction of water in the solvation shell of the solute molecules in the bulk and reference phase is given by:

$$x_{b,w} = \frac{N_{b,w}}{N_{T,b,w}} \quad (15)$$

$$x_w^* = \frac{N_w}{N_{T,w}^*} \quad (16)$$

where, x refers to mole fraction, N refers to the number of molecules, subscript b refers to bulk solvated phase, subscript w refers to water, and subscript T refers to the total number.

3.3. Optimization:

In the second stage, we perform optimization to estimate the unknown group interaction parameters for glutamic acid and histidine (Functional groups for the solute molecules are shown in **Figure S2** of the supplementary information). ASST relates the mole fraction to the activity coefficient, as shown in equations 1-7. GCM and modified UNIFAC model relate the activity

coefficient to group interaction parameters (equations 8-12). The optimization problem to obtain the unknown interaction parameters is then formulated as:

$$\min Err = \left[(x_{b,w} - x_w^*)^{calc} - (x_{b,w} - x_w^*)^{literature} \right]^2 \quad (17)$$

where, superscript *calc* represents the values calculated with the help of ASST and GCM and superscript *literature* represents values obtained from literature. The efficient Ant Colony Optimization (EACO) technique with Monte-Carlo sampling was used to solve the optimization problem.^{30,31} The Monte-Carlo method operates by randomly sampling parameter values from the input probability distribution, computing the corresponding output. This process is repeated multiple times to generate multiple output values. A significant advantage of this method is that the precision of the output distribution can be estimated using conventional statistical techniques. However, it should be noted that the pseudorandom number generator may produce samples that are clustered in specific regions of the population, resulting in non-uniform samples.

3.4. Growth Rate Estimation:

In the final stage, the optimized interaction parameters obtained from analyzing glutamic acid and histidine growth rates were used to estimate glycine and alanine crystal growth rates.^{32,33} Once the face specific growth rates are known, they are normalized to obtain the relative growth rates. These relative growth rates and space group are provided as input to the WinXMorph package³⁴ which gives morphology as the output. Additional details regarding growth rate calculations are provided in **Section S2** of the supplementary information. The growth rates are then compared to the literature values. Furthermore, the group interactions are analyzed to understand which interaction contributes the most to the crystal growth rates.

4. Results and Discussion

The solution technique for the optimization problem begins with UNIFAC group interaction parameters. The interaction parameter table allows visualizing the interaction parameters that need to be optimized. Glutamic acid, histidine, glycine, and alanine are the molecules used in this study. Optimized interaction parameters for the cooling crystallization of glutamic acid are shown in **Table 1**.

Table 1: Optimized interaction parameters for the case of glutamic acid cooling crystallization.

Groups	COOH(b)	CH2(b)	CHNH2(b)	H2O(sol,b)	H2O(sol,s)	COOH(s)	CH2(s)	CHNH2(s)
COOH(b)	0	315.30	-330.48	-66.17	0	1.09	0.47	-0.14
CH2(b)	663.50	0	391.50	1318	0	1.35	-0.59	-0.75
CHNH2(b)	202.50	-30.48	0	-330.48	0	0.23	0.67	1.34
H2O(sol,b)	-14.09	300	48.89	0	0	0	0	0
H2O(sol,s)	0	0	0	0	0	0.49	0.60	-0.38
COOH(s)	-0.25	-0.18	1.86	0	-0.28	0	0.54	2.26
CH2(s)	-0.06	-0.07	-1.14	0	-0.10	-1.54	0	0.31
CHNH2(s)	0.67	1.71	-0.64	0	0.51	0.13	-0.85	0

In **Table 1**, the rows and columns represent the functional groups on the molecule. The letter in the brackets is to distinguish different phases. The letter b represents the bulk solvated phase, s represents the surface (reference) phase, and sol represents solvent. The interaction parameters for the bulk solvated phase are directly obtained from the literature.³⁵ The values highlighted are the optimized values obtained from ASST coupled with GCM in this study. Solvent molecules in the solvation shell do not interact with the solvent molecules in the solvation shell of the alternate phase. Hence all such interaction parameters are set to zero. All the interaction parameters are non-dimensional.

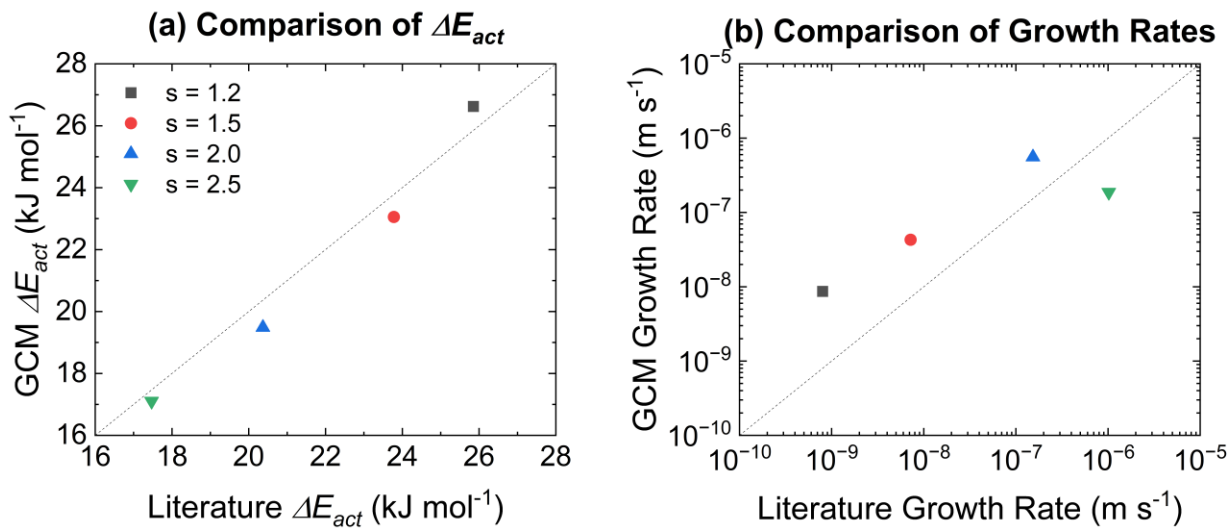


Figure 4: Activation barrier and growth rates of glutamic acid cooling crystallization calculated using GCM and ASST are compared with growth rates obtained from the literature.⁹ (a) Comparison of activation barrier and (b) comparison of growth rates. The dashed black line in both panels represents the reference where values match exactly.

The interaction parameters for the glutamic acid cooling crystallization case were used to solve the reverse problem where mole fractions are obtained from the optimized interaction parameters. For this case, the activation barrier values calculated using the relationship shown in **Figure 3** are compared with the literature values in **Figure 4a**. Variable “s” in **Figure 4a** refers to supersaturation (ratio of solute concentration to solubility at equilibrium). The growth rates calculated using the activation barrier values obtained from the GCM method are compared with the literature values of growth rates in **Figure 4b**. The concentration, hydration number, and volume values are obtained from the literature⁹ that allows converting mole fractions from the GCM approach to the change in hydration number. The activation barrier values match closely with the literature.⁹ However, since the relationship between activation barrier and growth rate is exponential, the minor difference in the prediction is amplified exponentially during the calculation of growth rates. The morphology obtained from these growth rate values is validated in the literature and hence the average growth rate values are shown.⁹

The interaction parameter table for the case of antisolvent crystallization is more complex as the antisolvent (ethanol) adds more functional groups. The interaction parameters for this case are shown in **Table S2Error! Reference source not found.** of the supplementary information.

The role of antisolvent in crystallization is to reduce solubility by reducing the interaction of solute molecules with the solvent. The energy required to replace the antisolvent molecules is significantly lower; hence, the reference phase is based on the number of water molecules in the solvation shell. Hence, all the interaction parameters between antisolvent and surface phase are also set to zero.¹⁰ Glutamic acid and histidine have some of the common functional groups. All of the common interaction parameter values in **Table S3** are obtained from **Table 1**. The values that are not common with glutamic acid and optimized using GCM and ASST approach are highlighted.

With the help of optimized interaction parameters, the activation barrier for the growth of histidine crystal was predicted and compared with literature values.¹⁰ Since the activation barrier values for various mole fractions of ethanol are reported in the literature, the growth rates at various mole fractions were calculated. The comparison is shown in **Figure 4**.

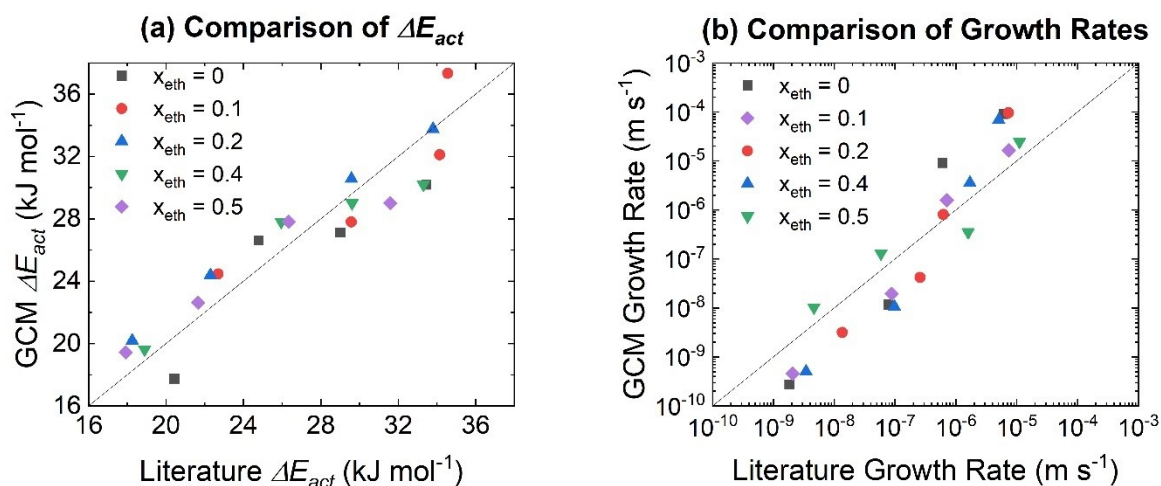


Figure 5: Comparison of activation barrier and growth rate values calculated from the GCM approach with the values reported in the literature¹⁰ for the case of histidine antisolvent crystallization. (a) Comparison of activation barrier and (b) comparison of growth rates. The solid line represents the reference where the values match exactly.

For the case of antisolvent crystallization, the activation barrier values show a slightly higher mismatch than in the case of glutamic acid. The mismatch can be attributed to the added complexity due to the addition of antisolvent functional groups. The mismatch is similarly amplified in the growth rate calculations. However, with only a few interaction parameters, GCM predicted growth rates from 24 antisolvent crystallization experiments. Performing and controlling such experiments is not trivial, as it requires equipment like DLS (Dynamic Light Scattering) or FBRM (Focused Beam Reflectance Measurement) probe³⁶ where the particle chord length can be measured every 2 seconds. But depending on orientation and aspect ratio the predictions might differ. Further, in our previous work,^{10,37} we calculated growth rate experimentally by capturing time-lapse images (every 3 mins) of the crystallization compartment using a built-in color camera (LC 30, Olympus America Inc.) of the optical microscope under the reflected light mode which were processed using MATLAB.

The optimized interaction parameters can be used for other systems as well. Glycine and alanine are the two molecules that have similar functional groups, have growth rate values reported in the literature, and have important applications.^{32,33} For glycine, the face-specific growth rates are reported allowing the prediction of morphology based on the growth rates obtained from GCM and ASST. Since the hydration number for glycine and alanine is not reported in the literature, the reverse problem needs to be solved twice: for the bulk solvated phase and then for the reference phase. In the case of glutamic acid and histidine, solving for the bulk solvated phase mole fraction was not necessary as the values were already reported in the literature.^{9,10} The interaction parameters for glycine and alanine are shown in **Table S3** and **Table S4** of the supplementary information, respectively.

The interaction parameters for alanine are exactly the same as glutamic acid. The only difference between glutamic acid and alanine is the CH₃ functional group. According to the UNIFAC group interaction parameter database, the simple alkane functional groups are similar. Hence the interaction parameter values do not change. However, the group-specific van der Waals radius and volume fraction are different for alanine, resulting in variation in the prediction of mole fraction from the same interaction parameters.

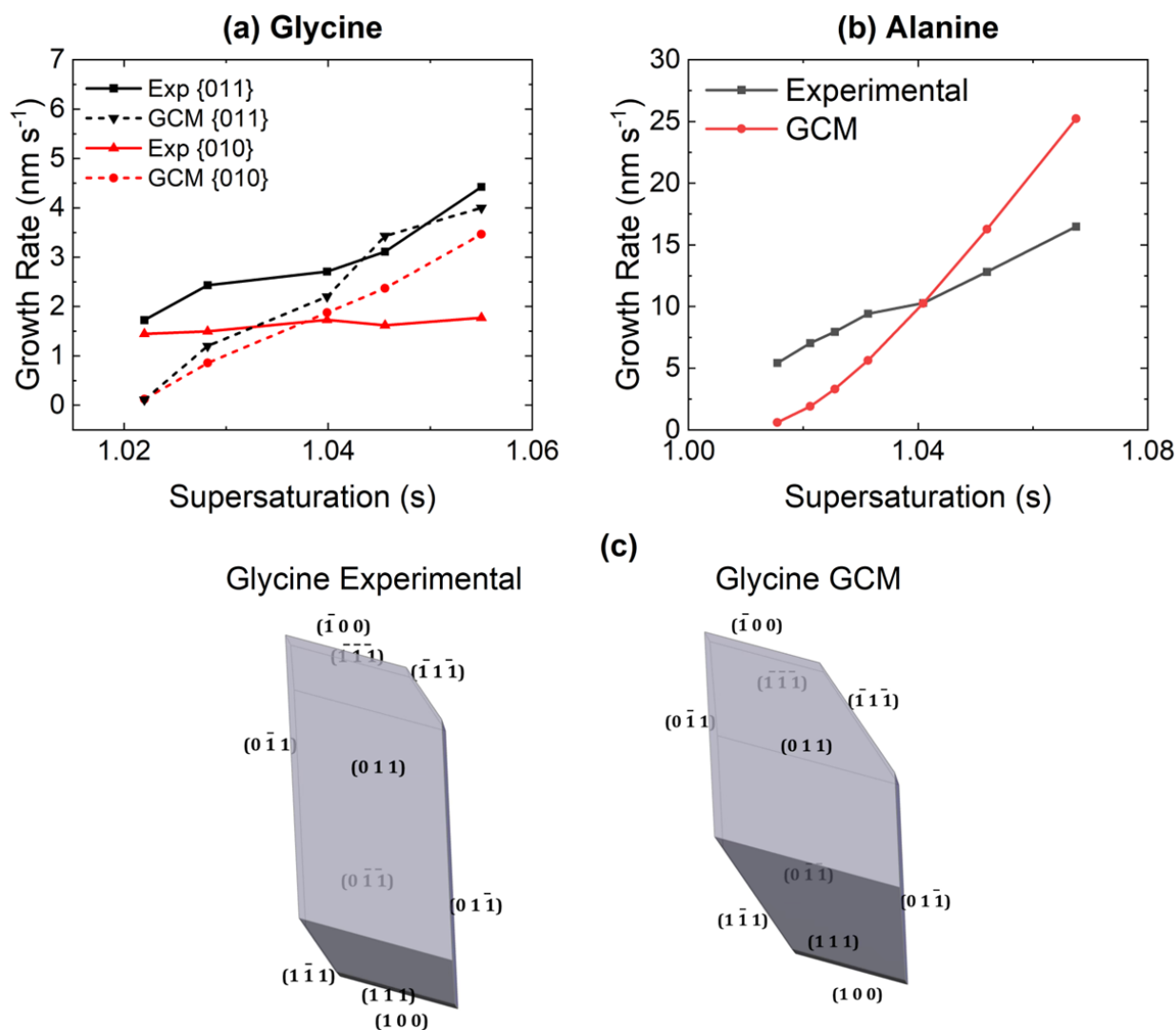


Figure 6: Growth rates of glycine and alanine crystals calculated using GCM and ASST and compared with literature.^{32,33} Face-specific growth rates of glycine allow predicting and comparing the morphology as well. (a) Glycine face-specific growth rates, (b) alanine growth rates (dominant face), (c) Glycine morphology obtained from growth rates calculated from GCM and ASST and compared with experimental morphology.

The growth rates obtained from the GCM and ASST approach for glycine and alanine crystals with the help of optimized interaction parameters and the glycine crystal morphology are shown in **Figure 6**. GCM and ASST reproduce the experimental growth rates reasonably. The critical length of h-vectors of glycine and alanine required to obtain face-specific parameters for the calculation of growth rates are given in **Figure S3** of the supplementary information. Although the plots show greater mismatch at high supersaturations, the morphology predicted from these growth rates is not significantly different from the experimental morphology. This phenomenon is shown in glycine crystals.³³ The mismatch is only due to the relative area of the $\{011\}$ face. The $\{111\}$ family of faces were assumed to have similar growth rates to the $\{011\}$ family to complete

the crystal hull of glycine crystal. Growth rates of $\{111\}$ family of faces are not explicitly reported in the literature.

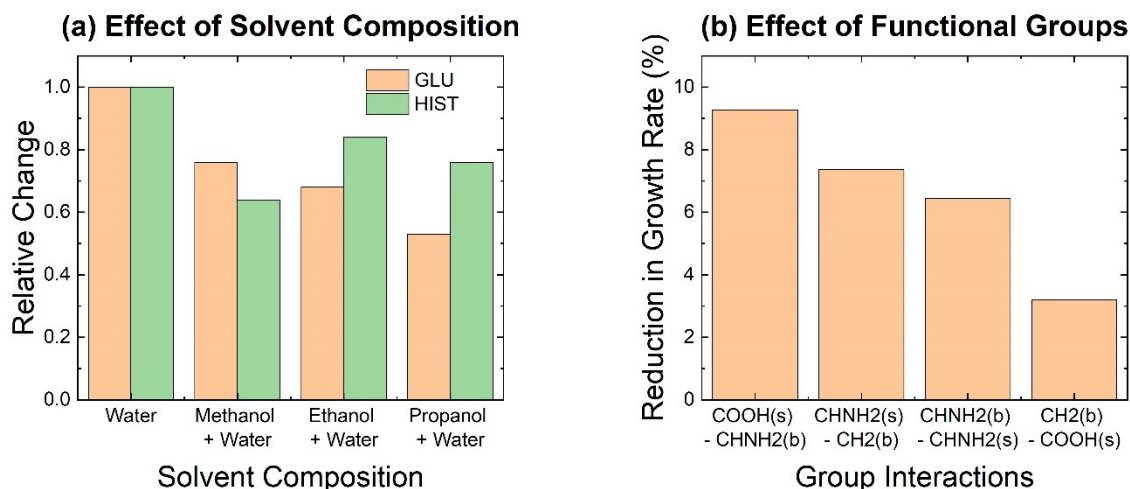


Figure 7: Effect of solvent composition and group interactions on the crystal growth rates.

The approach described in this article can also be used to understand the effect of solvent composition and group interactions on the crystal growth rates, as shown in **Figure 7**. The effect of 0.5 mole fraction of methanol, ethanol, and propanol on the growth rates of glutamic acid and histidine molecules was analyzed using GCM and ASST approaches, as shown in **Figure 7a**. For both molecules, GCM predicts the highest growth rate in water. For glutamic acid, the growth rate goes down as higher molecular weight antisolvent is used. However, the growth rate for histidine is higher for ethanol as an antisolvent than methanol and propanol. Such behavior can be directly attributed to solubility. Glutamic acid has the highest solubility in water, and the solubility reduces as antisolvent with higher molecular weight is used. For histidine, the solubility is higher when ethanol is used as antisolvent than methanol or propanol as antisolvents. This result shows that GCM and ASST effectively capture the equilibrium properties even when crystallization is inherently a non-equilibrium process. The effect of group interaction parameters is shown in **Figure 7b**. Error in estimating GCM parameters would propagate to the absolute growth rates prediction, as shown in **Figure S4** of supplementary information. Since the errors are propagated proportionally, morphology predictions will not be affected. Certain pairs of interactions were removed from the interaction parameter table to analyze which interactions contribute most to the growth rates. As expected, the charged functional groups such as carboxylic acid and quaternary amine have the most effect on the growth rates. It further corroborates the GCM and ASST approach.

5. Conclusion

The article presents a newer approach to understanding crystallization and studying crystal growth kinetics.³⁸ Crystallization is inherently stochastic, involves a large number of molecules, and is difficult to control. However, understanding the dynamics of fully solvated and partially desolvated molecules reduces the degree of freedom required to apply known thermodynamic

models. More importantly, understanding the change in the number of solvent molecules in the solvation shell and its effect on the activation barrier allows for forming an optimization problem to predict the change in the number of solvent molecules. This optimization problem can be solved in different ways, and the solution to this optimization problem using GCM and ASST is shown in this manuscript.

The rapid estimation of growth rates using this procedure exhibits certain limitations. It may not yield accurate results in cases of poor mixing, while the spiral growth model offers reasonable estimations under low supersaturation conditions. Additionally, the Group Contribution Method (GCM) approximation is worth noting, as interactions of a specific group within one molecule may not necessarily mirror those in another molecule. Apart from these limitations, the growth rates obtained from GCM and ASST match reasonably with the growth rates reported in the literature based on the predicted morphology. The computational approach described in this manuscript is also significantly resource inexpensive compared to estimating growth rates using molecular simulation techniques. The efficient Ant Colony optimization coupled with an effective sampling method allows finding minimum error in a large solution space.

The prediction of morphology is particularly important in pharmaceutical crystallization to reduce the costs of post-processing the crystals after the purification steps. The functional groups studied in this manuscript are most commonly seen on active pharmaceutical ingredient (API), as well as in the solvents used for crystallization of such APIs. Applying GCM and ASST to understand the effect of the solvent environment and predict the morphology can significantly reduce the costs involved in performing large-scale high throughput crystallization experiments. Furthermore, the GCM and ASST can be used to design the solvents for crystallization using Computer Aided Molecular Design (CAMD) approach.

Nomenclature:

x_i	–	Mole fraction of i^{th} species (obtained from previous literature MD simulations)
γ_i	–	Activity coefficient of i^{th} species (estimated using ASST and GCM)
φ	–	Chemical Potential (estimated using ASST)
R	–	Gas Constant
T	–	Temperature (operating condition)
Γ^s	–	Adsorption capacity of the adsorbent (parameter)
G^E	–	Excess Gibb's free energy (estimated using ASST)
m_o	–	Molar mass of adsorbent (parameter)
n_i	–	Molar quantity of component i (estimated using MD simulations)
M_o	–	Molecular weight of the adsorbent (parameter)
Q_g	–	van der Waals surface area (parameter)

θ_m	–	Surface area fraction (estimated as a part of GCM)
Ψ_{nm}	–	Temperature-dependent parameter (estimated as a part of GCM)
a_{nm}	–	Group interaction parameter between the groups of type n and m (estimated with optimization)

Subscript:

i	–	component i
o, i	–	pure phase component

Superscript:

b	–	bulk phase
s	–	surface phase
$*$	–	reference phase

Supporting Information

Details of the double well approach, group contribution methods, overall calculation schemes, and interaction parameters are provided in the supporting information.

Acknowledgments

This material is based on the work performed at the Materials and Systems Engineering Laboratory at the University of Illinois Chicago (UIC) in collaboration with the group of Dr. Urmila Diwekar. M.R.S. acknowledges funding support from UIC and the US National Science Foundation (NSF, EFRI 2132022).

Competing Interests:

The authors declare that there are no competing interests.

Author Contributions:

M.R.S. and U.D. conceptualized research; M.R.S., U.D., A.V.D., and P.K.R.P. designed methodology, A.V.D., and P.K.R.P., performed theoretical investigation, A.V.D., P.K.R.P., V.V.D., U.D. and M.R.S. performed validation; M.R.S., A.V.D., P.K.R.P., and U.D. wrote the original draft., M.R.S., V.V.D, A.V.D., P.K.R.P performed review and editing of the original draft.

References:

- 1 Singh, M. R. & Ramkrishna, D. in *Cryst Growth Des* Vol. 13 1397-1411 (2013).
- 2 Singh, M. R., Verma, P., Tung, H.-H., Bordawekar, S. & Ramkrishna, D. in *Cryst Growth Des* Vol. 13 1390-1396 (2013).
- 3 Myerson, A. S. *Handbook of industrial crystallization*. (Butterworth-Heinemann, 2002).

- 4 Al-Zoubi, N. & Malamataris, S. in *International Journal of Pharmaceutics* Vol. 260 123-135 (2003).
- 5 Chakraborty, J., Singh, M. R., Ramkrishna, D., Borchert, C. & Sundmacher, K. Modeling of crystal morphology distributions. Towards crystals with preferred asymmetry. *Chemical Engineering Science* **65**, 5676-5686 (2010). <https://doi.org/10.1016/j.ces.2010.03.026>
- 6 Coliaie, P. *et al.* Advanced continuous-flow microfluidic device for parallel screening of crystal polymorphs, morphology, and kinetics at controlled supersaturation. *Lab on a Chip* **21**, 2333-2342 (2021). <https://doi.org/10.1039/d1lc00218j>
- 7 Coliaie, P., Kelkar, M. S., Nere, N. K. & Singh, M. R. Continuous-flow, well-mixed, microfluidic crystallization device for screening of polymorphs, morphology, and crystallization kinetics at controlled supersaturation. *Lab Chip* **19**, 2373-2382 (2019). <https://doi.org/10.1039/C9LC00343F>
- 8 Tilbury, C. J., Green, D. A., Marshall, W. J. & Doherty, M. F. Predicting the effect of solvent on the crystal habit of small organic molecules. *Cryst Growth Des* **16**, 2590-2604 (2016).
- 9 Dighe, A. V. & Singh, M. R. Solvent fluctuations in the solvation shell determine the activation barrier for crystal growth rates. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23954 (2019).
- 10 Dighe, A. V., Podupu, P. K. R., Coliaie, P. & Singh, M. R. Three-Step Mechanism of Antisolvent Crystallization. *Cryst Growth Des* **22**, 3119-3127 (2022). <https://doi.org/10.1021/acs.cgd.2c00014>
- 11 Tanaka, S., Yamamoto, N., Kasahara, K., Ishii, Y. & Matubayasi, N. Crystal Growth of Urea and Its Modulation by Additives as Analyzed by All-Atom MD Simulation and Solution Theory. *The Journal of Physical Chemistry B* **126**, 5274-5290 (2022).
- 12 Kontkanen, J. *et al.* What controls the observed size-dependency of the growth rates of sub-10 nm atmospheric particles? *Environmental Science: Atmospheres* **2**, 449-468 (2022).
- 13 Li, J., Tilbury, C. J., Kim, S. H. & Doherty, M. F. A design aid for crystal growth engineering. *Progress in Materials Science* **82**, 1-38 (2016).
- 14 Frisch, M. e. *et al.* (Gaussian, Inc. Wallingford, CT, 2016).
- 15 Tilbury, C. J. & Doherty, M. F. Modeling layered crystal growth at increasing supersaturation by connecting growth regimes. *AIChE Journal* **63**, 1338-1352 (2017).
- 16 Han, D., Karmakar, T., Bjelobrk, Z., Gong, J. & Parrinello, M. Solvent-mediated morphology selection of the active pharmaceutical ingredient isoniazid: Experimental and simulation studies. *Chemical Engineering Science* **204**, 320-328 (2019).
- 17 Piana, S. & Gale, J. D. Understanding the barriers to crystal growth: Dynamical simulation of the dissolution and growth of urea from aqueous solution. *Journal of the American Chemical Society* **127**, 1975-1982 (2005).
- 18 Turner, T. D. *et al.* A digital mechanistic workflow for predicting solvent-mediated crystal morphology: the α and β forms of L-glutamic acid. *Cryst Growth Des* **22**, 3042-3059 (2022).
- 19 Garrido, N. M., Jorge, M., Queimada, A. J., Economou, I. G. & Macedo, E. A. Molecular simulation of the hydration Gibbs energy of barbiturates. *Fluid Phase Equilibria* **289**, 148-155 (2010).
- 20 Berti, C., Ulbig, P., Burdorf, A., Seippel, J. & Schulz, S. Correlation and Prediction of Liquid-Phase Adsorption on Zeolites Using Group Contributions Based on Adsorbate–Solid Solution Theory. *Langmuir* **15**, 6035-6042 (1999).

- 21 Karunanithi, A. T., Achenie, L. E. & Gani, R. A computer-aided molecular design framework for crystallization solvent design. *Chemical Engineering Science* **61**, 1247-1260 (2006).
- 22 Watson, O. L. *et al.* Computer aided design of solvent blends for hybrid cooling and antisolvent crystallization of active pharmaceutical ingredients. *Organic Process Research & Development* **25**, 1123-1142 (2021).
- 23 Dombrowski, R. D., Litster, J. D., Wagner, N. J. & He, Y. Modeling the crystallization of proteins and small organic molecules in nanoliter drops. *AIChE journal* **56**, 79-91 (2010).
- 24 Goh, L. *et al.* A stochastic model for nucleation kinetics determination in droplet-based microfluidic systems. *Cryst Growth Des* **10**, 2515-2521 (2010).
- 25 Singh, M. R. & Ramkrishna, D. in *Chemical Engineering Science* Vol. 107 102-113 (2014).
- 26 Doshi, R. K., Mukherjee, R. & Diwekar, U. M. Application of adsorbate solid solution theory to design novel adsorbents for arsenic removal using CAMD. *ACS Sustainable Chemistry & Engineering* **6**, 2603-2611 (2018).
- 27 Shahmohammadi, M., Mukherjee, R., Takoudis, C. G. & Diwekar, U. M. Optimal design of novel precursor materials for the atomic layer deposition using computer-aided molecular design. *Chemical Engineering Science* **234**, 116416 (2021).
- 28 Fredenslund, A., Gmehling, J., Michelsen, M. L., Rasmussen, P. & Prausnitz, J. M. Computerized Design of Multicomponent Distillation Columns Using the UNIFAC Group Contribution Method for Calculation of Activity Coefficients. *Ind. Eng. Chem. Proc. Des. Dev.* **16**, 450-462 (1977). <https://doi.org/10.1021/i260064a004>
- 29 Dighe, A. V., Coliaie, P., Podupu, P. K. R. & Singh, M. R. Selective desolvation in two-step nucleation mechanism steers crystal structure formation. *Nanoscale* **14**, 1723-1732 (2022). <https://doi.org/10.1039/D1NR06346D>
- 30 Gebreslassie, B. H. & Diwekar, U. M. Efficient ant colony optimization for computer aided molecular design: Case study solvent selection problem. *Computers & Chemical Engineering* **78**, 1-9 (2015).
- 31 Dige, N. & Diwekar, U. Efficient sampling algorithm for large-scale optimization under uncertainty problems. *Computers & Chemical Engineering* **115**, 431-454 (2018).
- 32 Lechugaballesteros, D. & Rodriguezhorno, N. Growth and Morphology of L-Alanine Crystals - Influence of Additive Adsorption. *Pharmaceut Res* **10**, 1008-1014 (1993). <https://doi.org/10.1023/A:1018962722691>
- 33 Li, L. & Rodriguezhorno, N. Growth-Kinetics and Mechanism of Glycine Crystals. *Journal of Crystal Growth* **121**, 33-38 (1992). [https://doi.org/10.1016/0022-0248\(92\)90172-F](https://doi.org/10.1016/0022-0248(92)90172-F)
- 34 Kaminsky, W. WinXMorph: a computer program to draw crystal morphology, growth sectors and cross sections with export files in VRML V2. 0 utf8-virtual reality format. *J Appl Crystallogr* **38**, 566-567 (2005).
- 35 Gmehling, J., Li, J. & Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Industrial & Engineering Chemistry Research* **32**, 178-193 (1993).
- 36 Nowee, S. M., Abbas, A. & Romagnoli, J. A. Antisolvent crystallization: Model identification, experimental validation and dynamic simulation. *Chem. Eng. Sci.* **63**, 5457 (2008).

- 37 Coliaie, P., Kelkar, M. S., Nere, N. K. & Singh, M. R. Continuous-flow, well-mixed, microfluidic crystallization device for screening of polymorphs, morphology, and crystallization kinetics at controlled supersaturation. *Lab Chip* **19**, 2373 (2019).
- 38 Tang, S. K. *et al.* On Comparing Crystal Growth Rates: Para Substituted Carboxylic Acids. *Cryst Growth Des* **23**, 1786-1797 (2023).