Modeling and Control of Protein Crystal Shape and Size in Batch Crystallization

Joseph Sang-Il Kwon and Michael Nayhouse
Dept. of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095

Panagiotis D. Christofides
Dept. of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095

Dept. of Electrical Engineering, University of California, Los Angeles, CA 90095

Gerassimos Orkoulas
Dept. of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095

DOI 10.1002/aic.14039
Published online March 4, 2013 in Wiley Online Library (wileyonlinelibrary.com)

In this work, the modeling and control of a batch crystallization process used to produce tetragonal hen egg white lysozyme crystals are studied. Two processes are considered, crystal nucleation and growth. Crystal nucleation rates are obtained from previous experiments. The growth of each crystal progresses via kinetic Monte Carlo simulations comprising of adsorption, desorption, and migration on the (110) and (101) faces. The expressions of the rate equations are similar to Durbin and Feher. To control the nucleation and growth of the protein crystals and produce a crystal population with desired shape and size, a model predictive control (MPC) strategy is implemented. Specifically, the steady-state growth rates for the (110) and (101) faces are computed and their ratio is expressed in terms of the temperature and protein concentration via a nonlinear algebraic equation. The MPC method is shown to successfully regulate both the crystal size and shape distributions to different set-point values. © 2013 American Institute of Chemical Engineers *AIChE J*, 59: 2317–2327, 2013

Keywords: protein crystallization, kinetic Monte Carlo simulation, process control, model predictive control, crystal shape

Introduction

Protein crystallization is a central activity in the production of pharmaceuticals. Additionally, the production of highly ordered, high quality protein crystals through batch crystallization processes is vital in devising proteins for therapeutic purposes. Protein structure can be found via nuclear magnetic resonance and X-ray crystallography.[1–3] However, nuclear magnetic resonance can only be used for proteins of small molar mass (less than 30,000). For cases with large molar mass, X-ray crystallography[4] can be used as long as protein crystals are of desired shape and high quality. In this work, we investigate tetragonal hen egg white (HEW) lysozyme protein crystals which is a naturally occurring enzyme with antibacterial activity. It is a widely used model for the study of protein crystallization and is composed of 129 amino acids with a molecular weight of 14,388.

More specifically, in the present work, we focus on both the nucleation of lysozyme crystals in a batch crystallizer and the crystal growth that follows after each crystal is nucleated until the end of the batch simulation. Several attempts aimed at modeling protein nucleation[4,5] and growth[6–8] have been made. These efforts make it possible to manipulate the size distribution and morphology of the protein crystal, which is a very critical variable for pharmaceutical products. For the tetragonal form of lysozyme, experiments indicate that at low supersaturation growth depends on a lattice defect mechanism, whereas at high supersaturation growth proceeds via two-dimensional (2-D) nucleation on the crystal surfaces.[6,9,10] Kinetic Monte Carlo (kMC) simulation methods will be used to model the crystal growth since it is a nonequilibrium process. kMC algorithms are based on a dynamic interpretation of the Master equation.[11,12] These algorithms form the basis for applying the Monte Carlo method to simulate dynamic processes.[13–24] One more assumption that we will use in the present work is the solid-on-solid lattice model.[25] This approximation will cause the crystal to be very compact by avoiding voids and overhangs while depositing particles onto the crystal lattice.

To implement our kMC methodology in a way that will consider the entire crystal lattice, in addition to being computationally efficient, we extend the methodology of Ref. 26 to the rate equations originally developed by Durbin and Feher.[25] This methodology has been discussed in our recent work[27] for single crystal growth while allowing molecular attachment, detachment, and migration events on the (110) and (101) faces. In the present work, this method is expanded to account for many crystals in a batch crystallization process. Similar to our previous work,[27] all attachments

Correspondence concerning this article should be addressed to G. Orkoulas at makis@seas.ucla.edu.

will be done in monomer units. We also assume that the monomer is not attached to water, but is a pure protein solute molecule. In the future, we can consider both the aggregates of lysozyme and water bonded protein molecules.[28] Assuming all surface sites are available for attachment, the attachment rate is considered to be independent for each lattice site. Detachment and migration events, however, are dependent on the local surface micro-configuration. For the local micro-configuration, we classify each lattice site based on the number of nearest neighbors surrounding that site. We consider nearest neighbors to be on the north, south, east, and west directions which are of the same height or higher compared to the current lattice site. This will allow each lattice site to have from zero to four nearest neighbors, giving a total of five classes. These classes are used to lessen the computational cost when calculating the rates for our three microscopic events. Since the solid-on-solid model is assumed, every lattice site has a nearest neighboring lattice site below (i.e., no voids or overhangs), and this nearest neighbor will be taken into account in the pre-exponential factor of the rate equations, described in the following section. Owing to the dependence of detachment and migration rates on the surface configuration, kMC simulations are needed to compute the net crystal, steady-state, growth rate as a function of temperature and protein concentration in the continuous phase.

In the next section, we will summarize the methodology for our kMC simulation method, including open-loop simulations to fit the data of Durbin and Feher.[6] Next, we will present our model predictive control (MPC) methodology to regulate crystal size and shape. After that we will present and discuss our simulation results for batch crystal nucleation and growth under closed-loop conditions. We will finish with a short conclusion and some ideas for future research.

Modeling and Simulation

As previously stated, we assume the solid-on-solid lattice model to carry out the crystal growth process simulation. For this work, we will focus on square lattice models of length and width $N=50$ sites with periodic boundary conditions represented by a 2-D array. Previous work[29] reports that no finite size effects were found between systems of sizes $N=30$, $N=60$, and $N=120$ sites. Furthermore, the present work will use the value at each lattice site of our 2-D array to be the number of particles (i.e., height) at the current location. Each event of our kMC simulation is chosen randomly based on the rates of the three microscopic phenomena.

Surface kinetics

The following description of the surface kinetics for our model was described previously by Nayhouse et al.[27] As noted in the previous work, the following description of the surface kinetics follows that of Durbin and Feher,[25] which was extended by Ke et al.[29] for migration events. As emphasized earlier, since each surface site is available for attachment, the attachment rate is independent of each lattice site and defined as

$$r_a = K^+(\Delta\mu) = K_0^+ \exp\frac{\Delta\mu}{k_B T} \tag{1}$$

where K_0^+ is the attachment coefficient, k_B is the Boltzmann constant, T is the temperature in Kelvin, and $\Delta\mu = k_B T \ln(c/s)$, where c is the solute concentration and s is the protein solubility and this term is the crystallization driving force. The protein solubility is dependent on temperature (°C) and defined for pH = 4.5 and 4% (w/v) NaCl by Refs. 30 and 31 with the following third-order polynomial

$$s(T) = 2.88\times10^{-4}T^3 - 1.65\times10^{-3}T^2 + 4.619\times10^{-2}T + 6.008\times10^{-1} \tag{2}$$

It is reported that Eq. 2 gives an error of 6.8% for the selected pH and salt concentration.[31] Returning back to the adsorption rate, we define the total rate of adsorption as

$$W_a = N^2 r_a \tag{3}$$

since each lattice site is considered to be independent.

Conversely, the desorption rate is dependent on the surface micro-configuration surrounding that surface particle. A lattice site with i nearest neighbors has desorption rate

$$K^-(E_b) = K_0^- \exp\left(\frac{E_b}{k_B T}\right) = K_0^- \exp\left(-i\frac{E_{pb}}{k_B T}\right) \tag{4}$$

where K_0^- is the desorption coefficient, and E_{pb} is the average binding energy per bond, and $E_b = iE_{pb}$ is the total binding energy. From this point forward, we will refer to the desorption rate which is dependent on the number of nearest neighbors as

$$r_d(i) = K_0^- \exp\left(-i\frac{E_{pb}}{k_B T}\right) \tag{5}$$

As expected, the desorption rate is lower when i is higher. The total rate of desorption is computed by

$$W_d = \sum_{i=0}^{4} W_{d_i}; \qquad W_{d_i} = M_i r_d(i) \tag{6}$$

where W_{d_i} is the total rate of desorption for each class and M_i is the number of lattice sites with i nearest neighbors. To define the migration rate, Ke et al.[29] used a rate similar to the desorption rate which includes an additional term. This term causes migration to have a higher rate compared to the desorption rate and is defined the following way

$$r_m(i) = K_0^- \exp\left(-i\frac{E_{pb}}{k_B T} + \frac{E_{pb}}{2k_B T}\right) \tag{7}$$

The total migration rate is computed as

$$W_m = \sum_{i=0}^{4} W_{m_i}; \qquad W_{m_i} = M_i r_m(i) \tag{8}$$

Now that there is a rate expression for each of the microscopic events, the total rate, W_{tot}, can be computed by summing over all rates, i.e., $W_{tot} = W_a + W_d + W_m$

As pointed out by Durbin and Feher[25] and Ke et al.,[29] K_0^+ and K_0^- are not independent. At equilibrium, $\Delta\mu = 0$, and the attachment and detachment rates must be equal, or

$$K^+(\Delta\mu = 0) = K^-(\phi) \tag{9}$$

where ϕ is the binding energy per molecule of a fully occupied lattice. Plugging these two values into each equation will produce the following relationship

$$K_0^- = K_0^+ \exp\left(\frac{\phi}{k_B T}\right) \tag{10}$$

With one additional substitution, the desorption and migration rate equations take the following form

$$r_d(i) = K_0^+ \exp\left(\frac{\phi}{k_B T} - i\frac{E_{pb}}{k_B T}\right) \tag{11}$$

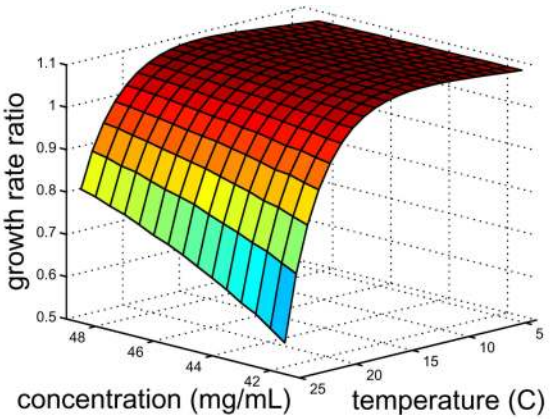
$$r_m(i) = K_0^+ \exp\left(\frac{\phi}{k_B T} - i\frac{E_{pb}}{k_B T} + \frac{E_{pb}}{2k_B T}\right) \tag{12}$$

We note that the binding energies cannot be accessed by experiments.[25,32] Previous simulation work assigned a range of values to E_{pb} and ϕ, until satisfactory agreement between the calculated and the experimental growth rates was achieved.[25] We follow the same approach in this work to obtain growth rate results consistent with the work of Durbin and Feher.[6] Finally, as mentioned previously, the computation of the growth rate for each face requires the use of kMC simulations owing to the dependence of the detachment and migration rates on the surface micro-configuration and it cannot be computed by simply subtracting the attachment and detachment rates.

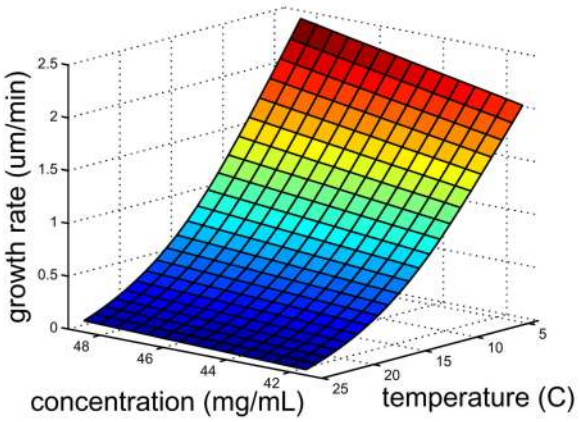
Event execution

To execute an event we generate a uniform random number, $\zeta_1 \in [0,1)$. If $\zeta_1 \leq W_a/W_{tot}$, then an adsorption event will be executed. If $W_a/W_{tot} < \zeta_1 \leq (W_a+W_d)/W_{tot}$, then a desorption event will be executed. Last, if $\zeta_1 > (W_a+W_d)/W_{tot}$, then a migration event will be executed. For adsorption, a lattice site is chosen at random for the event to take place. For desorption and migration, the specific class needs to be determined. In the case of a desorption event, the kth class is determined to be an integer from $[0,4]$ such that

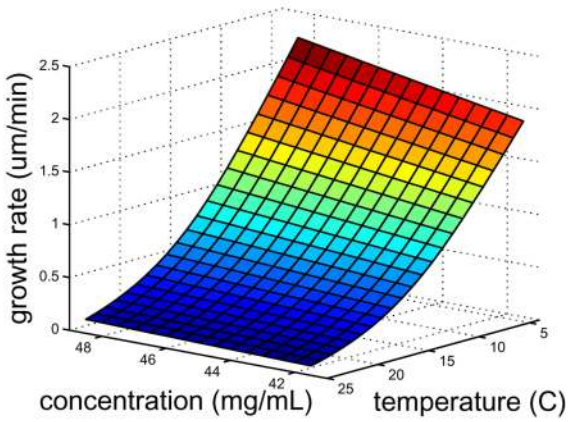
$$\frac{W_a + \sum_{i=0}^{k-1} W_{d_i}}{W_{tot}} < \zeta_1 \leq \frac{W_a + \sum_{i=0}^{k} W_{d_i}}{W_{tot}} \tag{13}$$



(a) Growth rate ratio



(b) Growth rate of the (110) face



(c) Growth rate of the (101) face

Figure 1. Plots of the growth rate data obtained for the (110) face, the (101) face, and the growth rate ratio between (110) and (101) faces for tetragonal lysozyme protein crystals at pH =4.5.

It is noted that the growth rate ratio, G, is equal to $\frac{G_{110}(t)}{G_{101}(t)}$ and is dimensionless. The data from the open-loop kMC simulation are plotted to demonstrate the effect of temperature and concentration variations on growth rates. Protein concentration and temperature range from 41.5 to 48.5 mg mL^{-1} and 4 to 25°C, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Once the class is determined, a second random number, ζ_2, is generated to select a random lattice site in class k to execute the desorption event. Migration events work in an analogous way to desorption events with a minor modification to Eq. 13 as follows

$$\frac{W_a+W_d+\sum_{i=0}^{k-1}W_{m_i}}{W_{tot}}<\zeta_1\leq\frac{W_a+W_d+\sum_{i=0}^{k}W_{m_i}}{W_{tot}} \quad (14)$$

In Durbin and Feher's original work,[25] they found that only half the lattice sites were available for adsorption on the (101) face, whereas every lattice site is available on the (110) face. This is due to the fact that only half the molecules on the (101) face have dangling bonds (i.e., the points of attachment for incoming molecules), whereas every molecule on the (110) face has dangling bonds.[25] In the present work, we model this behavior by accepting 50% of adsorption events on the (101) face, compared to 100% on the (110) face. In the case of desorption, the event is always accepted for both faces. Similar to desorption events, migration events are always accepted as long as there exists at least one available migration site. For this work, an available migration site is a nearest neighboring site which is lower in height than the current lattice site where the migration event is taking place. We assume each available migration site has equal probability to accept the displaced particle, which follows from.[33]

After each event is executed, a time increment, Δt, is computed based on the total rate of the microscopic events as follows

$$\Delta t=-\ln(\zeta)/W_{tot} \quad (15)$$

where ζ is a uniform random number, $\zeta\in[0,1)$. Events will continue to take place until the simulation has ended.

Fitting experimental data via open-loop simulations

For a given set of the simulation conditions comprised of temperature, pH, salt, and protein solute concentrations, the method described earlier in this section results in averaged lysozyme face growth rates at various values of supersaturation.

In Figure 2, crystals have been grown at supersaturation $2.1\lesssim\ln(c/s)\lesssim3.95$, where c (mg mL^{-1}) is the protein solute concentration and s (mg mL^{-1}) is the solubility, and the estimated growth rate at 4.0% NaCl is compared with the experimentally measured data at 3.5% and 5.0% NaCl from Ref. 10, respectively. The parameters for the kMC simulation in Figures 2–9 are listed in Table 1. The original parameters were taken from Ref. 25 and modifications were performed by making small changes to these parameters to ensure both validity in the crossover behavior of the growth rates of the (110) and (101) faces, and that the data were in the correct regime for the 4% NaCl to make use of available nucleation rate data.[34] The growth rates produced are the average growth rate for each set of conditions over 10 independent kMC runs. For Figure 2, c=45.0 mg mL^{-1}. The solubility is determined by Eq. 2 in terms of temperature (°C) at pH = 4.5 and 4% (w/v) NaCl. As is evident in Figures 2 and 3, crossover behavior between the (110) and (101) faces does, indeed, occur. It is also apparent in Figure 2 that the kMC simulation models the experimental result by Ref. 25. We also note that with a different set of experimental data, the parameters can be fit again to achieve a satisfactory agreement between simulation and experiment.

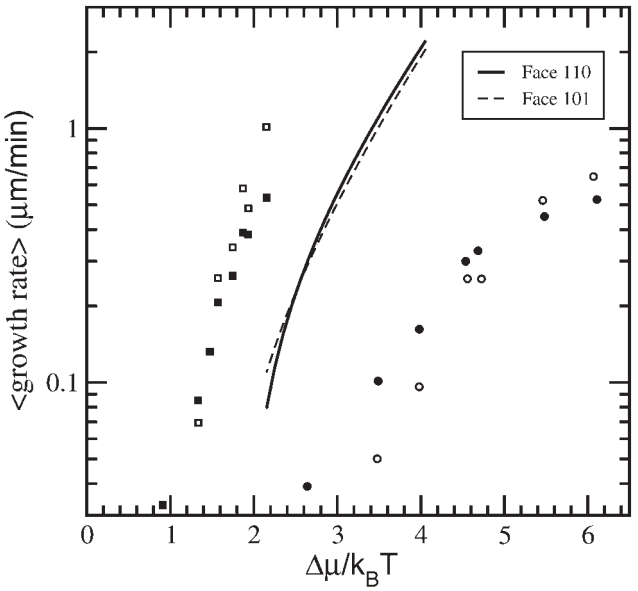


Figure 2. The expected growth rate vs. the degree of supersaturation at $c=45$ mg mL^{-1} and 4% NaCl are shown as the solid (110 face) and the dashed (101 face) lines.

The (■) and (□) represent the measured experimental data for 101 and 110 faces with 5% NaCl; (•)/(○) represent the measured experimental data with 3.5% NaCl; extracted from Ref. 7 at pH =4.6.

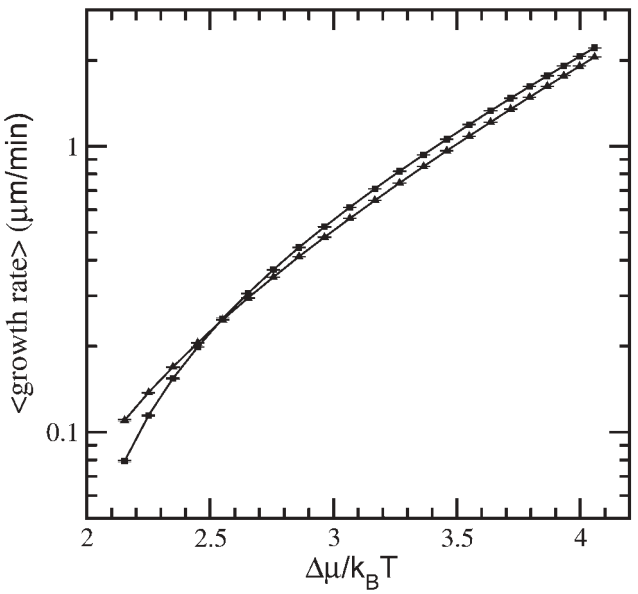
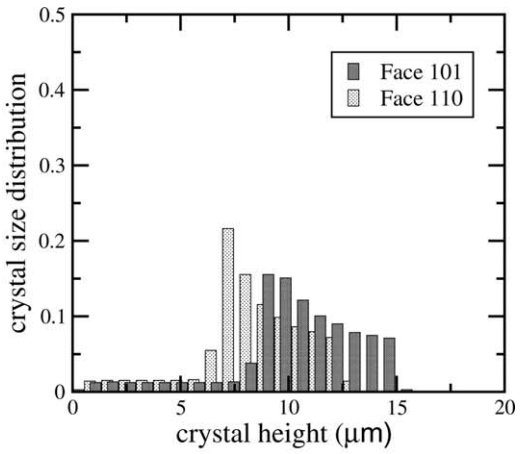
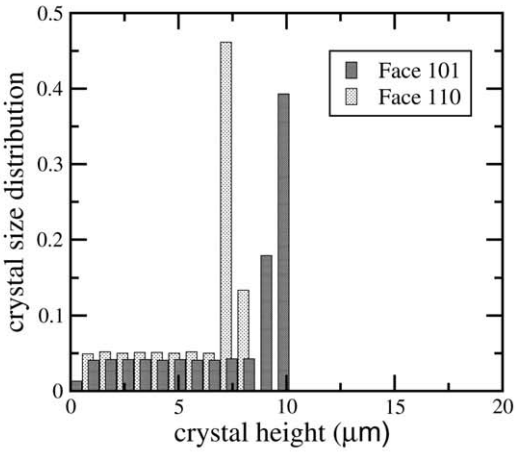


Figure 3. The expected growth rate vs. the degree of the supersaturation at $c=45$ mg mL^{-1} with $E_{pb}/k_BT=(0.78/0.83)$ and $\phi/k_BT=(3.7/2.75)$ for 110 (■) and 101 faces (▲), respectively.

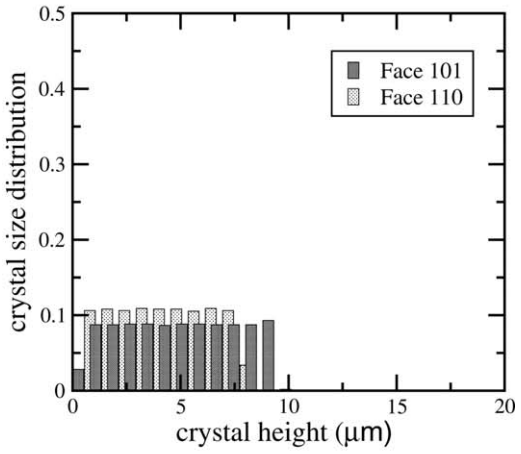
Error bars represent two standard deviations of 10 simulations for each point.



(a) $\langle G \rangle = 0.82 \quad T_o = 5.0^o\mathrm{C}$



(b) $\langle G \rangle = 0.82 \quad T_o = 15.0^o\mathrm{C}$



(c) $\langle G \rangle = 0.82 \quad T_o = 23.0^o\mathrm{C}$

Figure 4. The final crystal size distribution at the end of the batch simulation for each face under closed-loop operation starting from different initial temperature and aiming at growth rate ratio set-point value, $\langle G \rangle$=0.82.

It is noted that the crystal size distribution is a dimensionless variable and is normalized over the entire crystal population so that summing over all histogram bars for each face will add up to 1, for each different set of growth rate ratio and starting temperature.

Furthermore, kMC simulations were carried out to evaluate the effect of nucleation (i.e., crystals nucleated at different time as the batch crystallization process proceeds). The nucleation rates were extracted from Ref. 34, which will be discussed in MPC Formulation section. As anticipated, for a given temperature, higher solute concentrations in the continuous phase result in higher supersaturation, and yield both higher growth rates and higher nucleation rates. However, for a given concentration, as temperature rises, the supersaturation decreases resulting in a lower nucleation rate and a lower growth rate. The supersaturation dependence of the nucleation rate can be seen in Eq. 17, which is presented in the next section and used to calculate the number of newly formed crystals. The final crystal shape and size distribution can be driven to a desired range by the controller design described in the following section.

MPC of Crystal Size and Shape

In the kMC simulations, crystal nucleation is considered alongside crystal growth via molecular attachment, detachment, and migration events. Since the nucleation rate depends on the supersaturation, it has been manipulated by changing the temperature for a given concentration. In Table 1, parameters of crystal growth have also been appropriately chosen for kMC simulations to capture the experimentally observed crossover behavior in the crystal growth rates between the (110) and (101) faces.[6,25] In this section, a model predictive controller is designed based on the nonlinear growth rate equations (i.e., 3-D plots) for both faces and the growth rate ratio which has been shown in Figure 1a–c to suppress the uncertainty in the solute concentration, and achieve the desired set-point values by manipulating the temperature. A desired set-point value of the growth rate ratio and minimum crystal sizes for the (110) and (101) faces are included in the cost function in the MPC formulation. MPC resolves the drawbacks of the classical control schemes like proportional integral control, which cannot explicitly take into account input/state constraints, optimality considerations, and the nature of the nucleation, attachment, detachment, and migration processes. Furthermore, dynamic open-loop optimization may be used. However, open-loop optimization does not provide robustness against model inaccuracies and stochastic variation in the protein concentration and the operating environment.

The population balance equation for protein crystallization

The evolution of the particle size and shape distribution in a protein crystallization process is typically described by a population balance equation (PBE). More specifically, for the batch crystallizer considered in this work, a population balance model can be used to describe the evolution of the crystal size distribution (CSD), $n(h_{110}, h_{101}, t)$, which represents the number of crystals at time t with height h_{110} and h_{101} on each face. To describe the behavior of the CSD in a crystallization process, it is necessary to know the nucleation rate as well as the growth rate. In Liu's work,[35] they only considered crystal growth from seeds. However, in the present work, we consider the nucleation process as well. We assume that the height of nuclei for the (110) and (101) faces, h_{110} and h_{101}, is negligible, and the number of nuclei newly formed at time t is denoted as $n(0, 0, t)$. We have two reasons supporting this assumption. First, a nucleus consists

of 3 (to 4) HEW lysozyme molecules, whose size is negligible relative to the final crystal size.[36] In addition to this, nuclei cannot be observed until they reach the resolution limit of the optical microscope, ~0.5 μm.[37] The resulting PBE has the following form:

$$\begin{aligned}
&\frac{\partial n(h_{110},h_{101},t)}{\partial t}+\partial\left(\frac{(G_{110}(h_{110},t)n(h_{110},h_{101},t))}{\partial h_{110}}\right)\\
&+\partial\left(\frac{(G_{101}(h_{101},t)n(h_{110},h_{101},t))}{\partial h_{101}}\right)=0\\
&G=f_G(T,C)\\
&G_{110}(h_{110},t)=f_{110}(T,C,t)\\
&G_{101}(h_{101},t)=f_{101}(T,C,t)\\
&n(h_{110},h_{101},t)=\sum_{k=0}^{t-1}\frac{\Delta}{J(0,0,k)\cdot V}\\
&J(0,0,t)=f_{\text{nucleation}}(\sigma(t))
\end{aligned} \tag{16}$$

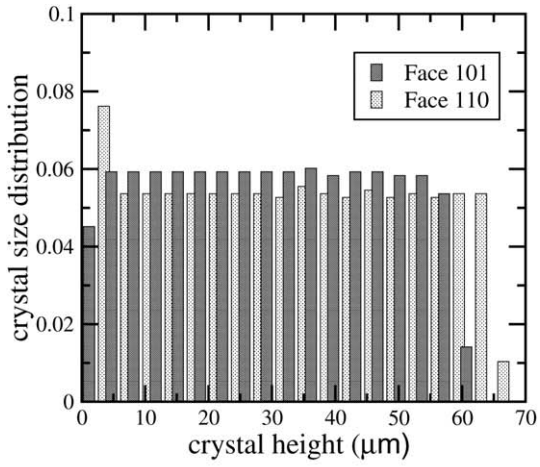
where $G_{110}(h_{110},t)$ and $G_{101}(h_{101},t)$ are the growth rates for (110) and (101) faces, Δ is the sampling time, V is the volume of the crystallizer, $G=\frac{G_{110}(t)}{G_{101}(t)}$, is the growth rate ratio. The nonlinear equations, f_G, f_{110}, and f_{101}, show their dependencies on temperature, solute concentration, and time, respectively. The number of crystals nucleated at time, t, is obtained from Ref. 34, where the homogeneous nucleation rate is determined by the supersaturation and NaCl concentration, as well as pH level. In this work, we assume that heterogeneous nucleation, i.e., nucleation on a surface, is negligible. The following equation for the nucleation rate, $J(0,0,t)(\sigma)$, at time t (with units [$\text{cm}^{-3}\cdot\text{s}^{-1}$]), was extracted from Ref. 34 at pH = 4.5 and 4% (w/v) NaCl

$$J(0,0,t)(\sigma)=\begin{cases}0.041\sigma+0.063 & \text{if } \sigma\geq 3.11\\ 8.0\times10^{-8}\exp(4.725\sigma) & \text{if } \sigma<3.11\end{cases} \tag{17}$$

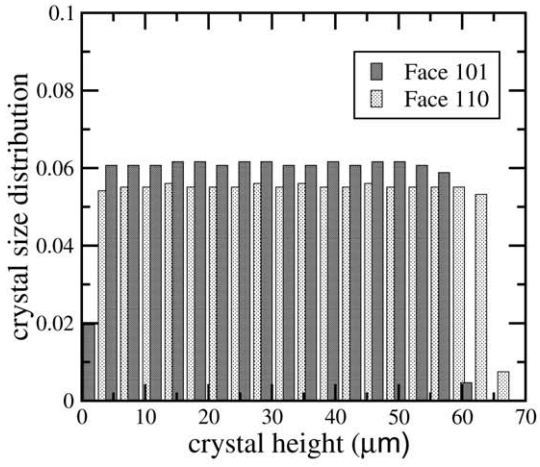
Instead of solving Eq. 16 directly, we execute several kMC simulations with nuclei formed at different times. This is considered to be equivalent to solving Eq. 16. Also, for a given supersaturation, the nucleation rate is obtained from the experimental results by Galkin et al.[34] Then, a controller design method, which was developed by Nayhouse et al.,[27] is further generalized to the case with nucleation of crystals, which is described in the following section.

MPC Formulation

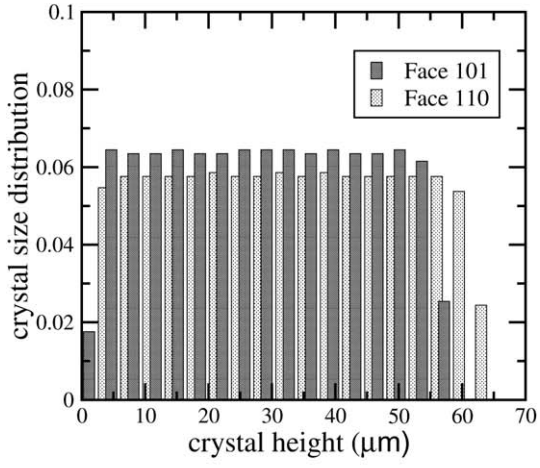
We consider the control of the crystal size and shape for a crystal population nucleated at different times as the batch crystallization process proceedes by using a MPC design. The expected value of the growth rate ratio, $\langle G\rangle$, is chosen as the control objective, where G is defined as the ratio G_{110}/G_{101}. A desired minimum crystal size is also included in the cost function of the MPC formulation and the temperature is used as the manipulated input. Although there are various factors that affect the protein solubility and crystal morphology during



(a) $\langle G\rangle = 1.10 \quad T_o = 5.0^o\text{C}$



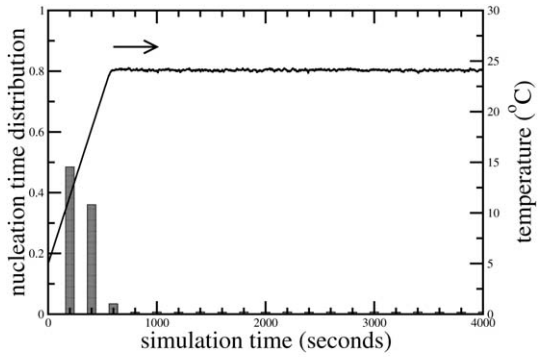
(b) $\langle G\rangle = 1.10 \quad T_o = 15.0^o\text{C}$



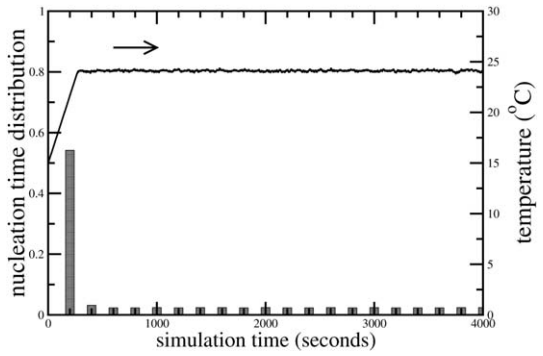
(c) $\langle G\rangle = 1.10 \quad T_o = 23.0^o\text{C}$

Figure 5. The final crystal size distribution at the end of the batch simulation for each face under closed-loop operation starting from different initial temperature and aiming at growth rate ratio set-point value, $\langle G\rangle$=1.10.

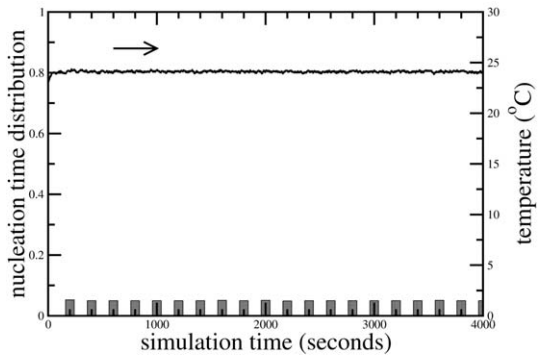
It is noted that the crystal size distribution is a dimensionless variable and is normalized over the entire crystal population so that summing over all histogram bars for each face will add up to 1, for each different set of growth rate ratio and starting temperature.



(a) $\langle G \rangle = 0.82 \quad T_o = 5.0^o\text{C}$



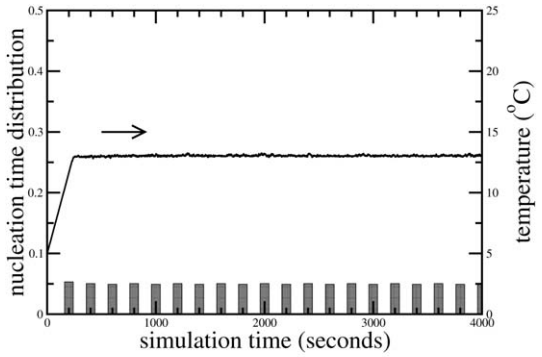
(b) $\langle G \rangle = 0.82 \quad T_o = 15.0^o\text{C}$



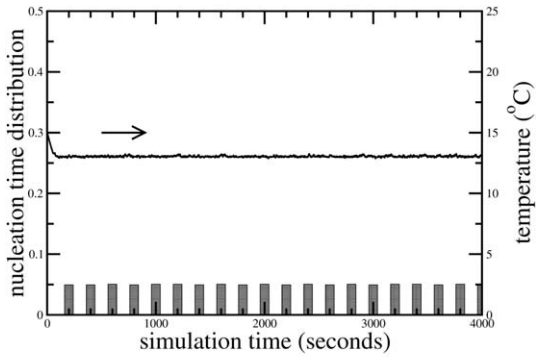
(c) $\langle G \rangle = 0.82 \quad T_o = 23.0^o\text{C}$

Figure 6. Profiles of nucleated crystals and temperature with time during the batch run under closed-loop operation for varying initial temperature and for the growth rate ratio set-point value, $\langle G \rangle = 0.82$.

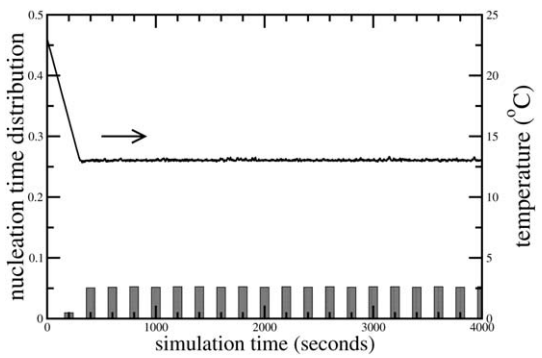
It is noted that the nucleation time distribution is a dimensionless variable and is normalized over the entire crystal population so that summing over all histogram bars, for each different set of growth rate ratio and starting temperature, will add up to 1.



(a) $\langle G \rangle = 1.10 \quad T_o = 5.0^o\text{C}$



(b) $\langle G \rangle = 1.10 \quad T_o = 15.0^o\text{C}$



(c) $\langle G \rangle = 1.10 \quad T_o = 23.0^o\text{C}$

Figure 7. Profiles of nucleated crystals and temperature with time during the batch run under closed-loop operation for varying initial temperature and for the growth rate ratio set-point value, $\langle G \rangle = 1.10$.

It is noted that the nucleation time distribution is a dimensionless variable and is normalized over the entire crystal population so that summing over all histogram bars, for each different set of growth rate ratio and starting temperature, will add up to 1.

the crystallization process,[38–42] we assume only temperature and solute concentration vary, while all other parameters remain constant for the closed-loop simulations (e.g., pH, NaCl concentration, buffer concentration, etc.). The proposed modeling and control methods can be easily extended to the case of multiple inputs since they do not depend on the specific number of the manipulated variables. To account for a number of practical considerations, several constraints are added to the control problem. First, there is a constraint on the range of variation of the temperature to ensure validity for the kMC model by imposing a temperature range that will not damage the protein crystal. Specifically, $4^\circ\text{C} \leq T \leq 25^\circ\text{C}$ is the working range for temperature values. Second, another constraint is imposed on the rate of change of the temperature to account for actuator limitations. The controller can change temperature at a maximum of 2°C min^{-1}. The other constraint restricts the number of crystals nucleated during, for example, the second half of the batch run, which limits the nucleation

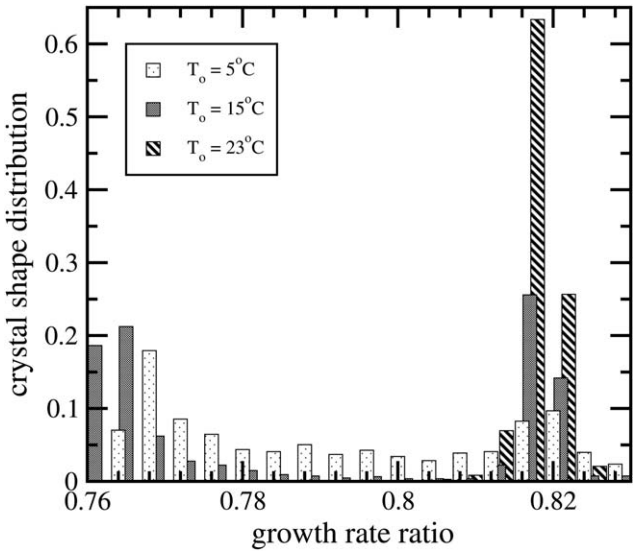


Figure 8. The final crystal shape distribution at the end of the batch simulation under closed-loop operation for varying initial temperature and for the growth rate ratio set-point value, $\langle G \rangle = 0.82$.

It is noted that the crystal shape distribution is a dimensionless variable and is normalized over the entire population so that summing over all histograms will add up to 1 for each starting temperature.

of very small crystal fines. The control action (optimal temperature) at time t is obtained by solving a finite-dimensional optimization problem in a receding horizon fashion. The cost function in the optimal control problem includes a penalty on

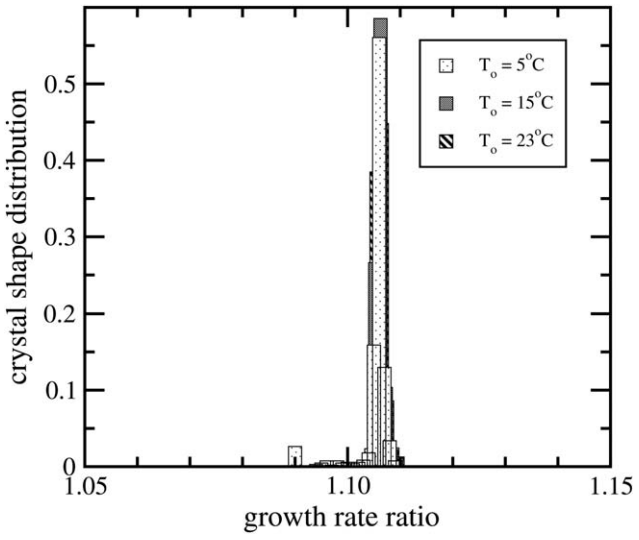


Figure 9. The final crystal shape distribution at the end of the batch simulation under closed-loop operation for varying initial temperature and for the growth rate ratio set-point value, $\langle G \rangle = 1.10$.

It is noted that the crystal shape distribution is a dimensionless variable and is normalized over the entire crystal population so that summing over all histogram bars for each different starting temperature will add up to 1.

Table 1. Parameters for Face (110) and (101) at 45 mg mL^{-1} NaCl and pH = 4.5 at T=18°C; additionally, $K_o^+ = 0.211$ s^{-1}

Face	ϕ/k_B	E_{pb}/k_B
(110)	1077.26	227.10
(101)	800.66	241.65

the deviation of $\langle G \rangle$ from its desired crystal shape. It also includes penalty costs on the negative deviation of the crystal size when the crystal size is less than the minimum. This helps to prevent crystal fines at the end of the crystallization process. For a given solute concentration and temperature, crystal growth during the time interval is estimated by using Figure 1, which was previously obtained from the open-loop simulations. The MPC problem is formulated as follows

$$
\begin{aligned}
&\underset{T_1,\ldots,T_i,\ldots,T_p}{\text{minimize}} \sum_{i=1}^{p} F_{\langle G\rangle,i} + F_{h_{110},i} + F_{h_{101},i} \\
&\text{subject to } \mathrm{F}_{\langle \mathrm{G}\rangle,\mathrm{i}} = (\langle \mathrm{G}\rangle - \mathrm{G}_{\mathrm{set}})^2 \\
&F_{h_j,i} = \begin{cases} \dfrac{h_{j,\min} - \langle h_j(t_i)\rangle}{h_{j,\min}} & \text{for } \langle h_j(t_i)\rangle < h_{j,\min} \\ 0 & \text{for } \langle h_j(t_i)\rangle \geq h_{j,\min} \end{cases} \\
&G_i = f_r(T, C, t_i) \\
&G_j(t_i) = f_j(T, C, t_i) \\
&T_{\min} \leq T_i \leq T_{\max} \\
&\left| \frac{T_{i+1} - T_i}{\Delta} \right| \leq R_T \\
&n(0,0,t) \leq n_{\text{limit}} \; \forall t \geq t_f/2 \\
&\langle h_j(t_i)\rangle = \frac{\langle h_j(t_{i-1})\rangle n(h_{110}, h_{101}, t_{i-1})}{n(h_{110}, h_{101}, t_i)} + R_j(t_{i-1})\Delta \\
&n(h_{110}, h_{101}, t_i) = n(h_{110}, h_{101}, t_{i-1}) + n(0,0,t_{i-1}) \forall i \\
&i = 1, 2, \ldots, p \\
&j \in \{110, 101\}
\end{aligned}
\tag{18}
$$

where t is the current time, t_i, $i=1,2,\ldots,p$, is the time of the ith prediction step, $t_i = t + i\Delta$, respectively, t_f is the length of the batch simulation, $F_{\langle G\rangle,i}$ is the cost function expressing the deviation of $\langle G \rangle$ from its set-point ratio, G_{set}, $F_{h_{110},i}$ and $F_{h_{101},i}$ are the cost functions expressing the penalty on the negative deviation of $\langle h_{110} \rangle$ and $\langle h_{101} \rangle$ from its minimum crystal size, $h_{110,\min}$ and $h_{101,\min}$, at time t_i, p is the number of prediction steps, $p\Delta$ is the specified prediction horizon, $T_i, i=1,2,\ldots,p$, is the temperature at the ith step, $T_i = T(t+i\Delta)$, respectively, $T_{\min}$ and $T_{\max}$ are the lower and upper bounds on the temperature, respectively, R_T is the limit on the rate of change of the temperature, n_{limit} limits the number of crystals nucleated during the latter half of the simulation time. The number of crystals, $n(h_{110}, h_{101}, t_i)$, and average height of the crystal face j, $\langle h_j(t_i) \rangle$, at time t_i are updated at every sampling time through the recursive equations (cf. Eq. 18), respectively. The optimal set of control actions $(T_1, T_2, \ldots, T_p)$ is obtained from the solution of the multivariable optimization problem of Eq. 18, and only the first value of the manipulated input trajectory, T_1, is applied to the protein crystallization process from time t until the next sampling time. At this point, a new measurement of protein concentration in the continuous phase is received from the kMC simulation and the MPC problem

of Eq. 18 is resolved for the computation of the next optimal input trajectory. The physical properties of the system (i.e., protein solubility and so on) were obtained from experimental data for the lysozyme protein solution.[30] In a previous work,[43] empirical expressions were obtained for the growth rates by fitting algebraic expressions to the available experimental data. In the present work, the growth rates are computed following the kMC methodology described previously. Furthermore, the stochastic nature of the system and the model uncertainty will be accounted for in the protein concentration variations. For results on robust-control of crystallization systems and MPC, see Refs. 44 and 45, respectively.

Batch Crystallization under Closed-Loop Operation

In this section, the proposed model predictive controller of Eq. 18 is applied to the kMC model described previously. At each sampling time (1 s), the optimal temperature, obtained by solving the optimization problem of Eq. 18, is applied to the closed-loop system until the next sampling time. The optimization problem is solved via a local constrained minimization algorithm using the nonlinear algebraic models described previously (cf. Figure 1) to predict the dependence of the crystal growth rate and growth rate ratio on temperature and solute concentration, respectively.

The solute concentration randomly varies following the Gaussian distribution of Eq. 19 to simulate the effect of disturbances at pH 4.5 and 4.0% NaCl, i.e.

$$\langle C(t)\rangle = C_n, \langle C(t)C(t')\rangle = \sigma^2 C_n^2, \tag{19}$$

where C_n is the nominal concentration of the system and σ^2 shows how far a set of measured concentrations deviates from its nominal value. Also, it is important to note that the concentration variation results in a change of the attachment rate (cf. Eq. 1). For all closed-loop simulations, the nominal concentration is 45 mg mL^{-1} with $\sigma = 2.5\%$ of C_n. The maximum rate of change of the temperature is 2°C min^{-1}. The volume of the crystallization batch is 1.0 L. The limit on the number of crystals nucleated during the latter half of the simulation time $n_{limit} = 500$. The following heuristic is also taken into account: It starts from a high enough supersaturation region in the beginning that many crystals are nucleated. Then, it moves on to a low enough supersaturation region to keep the rate of nucleation low to avoid the nucleation of the small crystal fines in the end of the batch run. This optimal strategy is taken into consideration in this work as one of the constraints restricting the number of crystals nucleated during the latter half of the simulation time. See, e.g., Figures 4a and 6a. Since the MPC formulation uses steady-state growth rates (cf. Figure 1b and c), the number of prediction steps is set to be $p = 1$. The time interval between the two sampling times is 1 s. The prediction horizon of each step is fixed at $p\Delta = 1$ s. The solute concentration varies every 1 s. The computational time that is used to solve the optimization problem with the current available computing power is negligible with respect to the sampling time interval. The closed-loop simulation duration $t_f = 4000$ s.

In the closed-loop simulations associated with controlling the growth rate ratio to the desired set-point values, the control objective is to separately regulate the expected ratio to the desired values, $\langle G\rangle = 0.82$ and $\langle G\rangle = 1.10$, respectively. Thus, the cost function of this problem contains a penalty on the deviation of the expected growth rate ratio from the set-point value. Also, the minimum thickness constraints described previously were imposed on each face to avoid undesirable crystal sizes.

The results of the closed-loop simulations are shown in Figures 4–9. Specifically, Figures 4 and 5 show results for six simulated batch runs under the different set-point growth rate ratios ($\langle G\rangle = 0.82$ and 1.10) and different starting temperatures ($T_o = 5$°C, $T_o = 15$°C, and $T_o = 23$°C). Note that we chose three starting temperatures, one at both extremes and one in the middle, in addition to two growth rate ratios at opposite ends to show a wide range of testing that can help guide practitioners while looking to achieve a desired shape and size distribution. In Figure 8, the lower desired crystal growth rate ratio, $<G> = 0.82$, has shown that final crystal shape distribution is very narrow when the initial temperature is close to the optimal temperature. Depending on the growth rate ratio set-point value and the initial temperature of the crystallizer, it takes a different amount of time until the temperature reaches the desired set-point value as it is seen in Figures 4 and 6. In this case, the optimal temperature is ~23°C. Figures 4a and 6a display that starting from a low initial temperature leads to nucleation of most of the crystals within the first 600 s of the batch simulation and thus a narrow size distribution but, in Figure 8, a broad shape distribution is achieved. In Figures 4b and 6b, since the initial temperature is closer to the optimal temperature, the system reaches its optimal temperature faster resulting in a less narrow size distribution and, in Figure 8, a less broad shape distribution. Last, Figures 4c and 6c display that the system reaches its optimal temperature very early and hence the size distribution is broader but, Figure 8, the shape distribution becomes very narrow.

In contrast to $<G> = 0.82$, the case of the higher desired crystal growth rate ratio, $<G> = 1.10$, has shown that the sensitivity of the system to the initial temperature is much less and the final crystal shape distribution is extremely narrow when the growth rate ratio set-point is 1.10 as it is seen in Figure 9. The reason is that the system reaches its optimal temperature faster, and thus the crystals uniformly nucleated along the batch go through the optimal temperature from the beginning. As stated earlier, a higher growth rate ratio as well as a higher nucleation rate is obtained at a low-temperature region.

In particular for Figures 5 and 7, the optimal temperature is ~13°C. The system reaches its optimal temperature faster, and crystals are uniformly nucleated along the batch and thus a broad size distribution but, in Figure 9, a narrow shape distribution is achieved. We note that if the temperature reaches its optimal state earlier in the batch simulation, undesirable effects of the nucleation and growth rates are minimized. In addition to this, as it is shown in Figure 1, variation in the solute concentration has less effect on the growth rates as well as the growth rate ratio for $\langle G\rangle = 1.10$, and thus the system reaches its steady-state much faster, which is shown in Figures 6 and 7.

Therefore, qualitatively speaking, MPC successfully drives the final crystal shape distribution to the desired set-point value, and a narrow size distribution can be achieved depending on the desired crystal morphology. For instance, for $<G> = 1.10$, it is much more likely to obtain a broad CSD, because the system reaches its optimal temperature faster regardless of the starting temperature. Although

disturbance results in the fluctuation of the solute concentration, the model predictive controller can successfully drive the growth rate ratio to the desired set-point. The desired distribution of the protein crystal shape can be estimated from Figure 8 and it is slightly elongated along the (101) direction for the lower set-point ratio while it is more equidimensional for the higher set-point ratio, shown in Figure 9.

Conclusions and Future Work

The present work focuses on the application of modeling, simulation, and control to a batch protein crystallization process to model multiple lysozyme protein crystals with the consideration of crystals nucleated at different times in the batch simulation. Based on the assumption that the two independent crystal faces are the (110) and (101) faces, dependence of the crystal growth of the two faces on temperature and protein solute concentration, obtained from kMC open-loop simulations, was observed. The crystal shape of the resulting lysozyme crystals as well as the nucleation rate was controlled through temperature variations. This was achieved via a nonlinear steady-state model generated from open-loop kMC simulations. Moreover, the nucleation rate data was obtained from experiments and is a function of supersaturation. The nonlinear model captures the protein solute concentration and temperature dependence of the growth rate ratio, thereby describing the key elements of the protein crystallization process. The other key achievement of the present work is that it takes nucleation into account resulting in different crystals nucleated at different times throughout the batch simulation.

An MPC strategy, which uses the steady-state model, was then designed to drive the final CSD to the desired set-point value while satisfying constraints on the magnitude and rate of change of temperature. As mentioned previously, the temperature is chosen as the manipulated input and this is in accordance with standard batch crystallization practice. The minimum thickness for each face is also imposed to avoid crystal fines. Simulation results demonstrated that the proposed controller was able to control the final crystal shape and size distribution by appropriately manipulating the temperature. The present methodology shows that more than 90% of the final crystals can be produced at the desired shape, either cubic or elongated crystals, as shown in Figures 8 and 9. Moreover, to implement this strategy in practice only protein solute concentration and temperature measurements are necessary. Once these measurements are available, the controller uses the algebraic model connecting the protein solute concentration and the temperature to the crystal growth rate ratio to compute the control action (i.e., temperature change); no crystal size/shape measurements are needed in the controller. These temperature changes are feasible due to the constraints put onto the controller.

Furthermore, the proposed kMC simulation offers a number of advantages compared to previous work. In previous simulations, such as Ke et al.,[29] they only considered the crystal seeds to initiate the crystallization. In contrast, in the present simulation, the nucleation rate was extracted from experiments and an appropriate region was picked and tested to model the experimental data. In the present work, a controller is designed to drive existing crystals as well as newly formed nuclei to desired crystal shape and size.

For the future work, in addition to monomer growth, aggregates of two or more molecules can be considered during adsorption events onto a crystal surface. To select either a monomer or an aggregate in every attachment event, another random number should be generated.[29] This consideration can potentially produce more realistic growth rates depending on the desired simulation range.

Acknowledgments

Financial support from the National Science Foundation (NSF) CBET-0967291 is gratefully acknowledged. This material is based upon work supported by the NSF Graduate Research Fellowship DGE-0707424 to Michael Nayhouse. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number TG-CCR120003.

Literature Cited

1. Rosenberger F, Vekilov PG, Muschol M, Thomas BR. Nucleation and crystallization of globular proteins—what we know and what is missing. *J Cryst Growth*. 1996;168:1–27.
2. Vekilov PG, Chernov AA. The physics of protein crystallization. *Solid State Phys*. 2003;57:1–147.
3. Wiencek JM. New strategies for protein crystal growth. *Annu Rev Biomed Eng*. 1999;1:505–534.
4. Galkin O, Vekilov PG. Direct determination of the nucleation rates of protein crystals. *J Phys Chem B*. 1999;103:10965–10971.
5. Pusey ML, Nadarajah A. A model for tetragonal lysozyme crystal nucleation and growth. *Cryst Growth Des*. 2002;2:475–483.
6. Durbin SD, Feher G. Crystal growth studies of lysozyme as a model for protein crystallization. *J Cryst Growth*. 1986;76:583–592.
7. Forsythe EL, Nadarajah A, Pusey ML. Growth of (101) faces of tetragonal lysozyme crystals: measured growth-rate trends. *Acta Cryst D*. 1999;55:1005–1011.
8. Kurihara K, Miyashita S, Sazaki G, Nakada T, Suzuki Y, Komatsu H. Interferometric study on the crystal growth of tetragonal lysozyme crystal. *J Cryst Growth*. 1996;166:904–908.
9. Durbin SD, Carlson WE. Lysozyme crystal growth studied by atomic force microscopy. *J Cryst Growth*. 1992;122:71–79.
10. Vekilov PG, Ataka M, Katsura T. Laser Michelson interferometry investigation of protein crystal growth. *J Cryst Growth*. 1993;130:317–320.
11. Fichthorn KA, Weinberg WH. Theoretical foundations of dynamical Monte Carlo simulations. *J Chem Phys*. 1991;95:1090–1096.
12. Müller-Krumbhaar H, Binder K. Dynamic properties of the Monte Carlo method in statistical mechanics. *J Stat Phys*. 1973;8:1–24.
13. Bortz AB, Kalos MH, Lebowitz JL. New algorithm for Monte Carlo simulation of Ising spin systems. *J Comput Phys*. 1975;17:10–18.
14. Dai J, Kanter JM, Kapur SS, Seider WD, Sinno T. On-lattice kinetic Monte Carlo simulations of point defect aggregation in entropically influenced crystalline systems. *Phys Rev B*. 2005;72:134102-1–134102-10.
15. Dai J, Seider WD, Sinno T. Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles. *J Chem Phys*. 2008;128:194705-1–194705-17.
16. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*. 1976;22:403–434.
17. Gillespie DT. Exact stochastic simulation of chemical reactions. *J Phys Chem*. 1977;81:2340–2361.
18. Gillespie DT. Monte Carlo simulation of random walks with residence time dependent transition probability rates. *J Comput Phys*. 1978;28:395–407.
19. Gillespie DT. A rigorous derivation of the chemical master equation. *Phys A*. 1992;188:404–425.
20. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys*. 2001;115:1716–1733.
21. Gillespie DT. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*. 2007;58:35–55.
22. Rathinam M, Petzold LR, Cao Y, Gillespie DT. Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J Chem Phys*. 2003;119:12784–12794.

23. Reese JS, Raimondeau S, Vlachos DG. Monte Carlo algorithms for complex surface reaction mechanisms: efficiency and accuracy. *J Comput Phys*. 2001;173:302–321.
24. Snyder MA, Chatterjee A, Vlachos DG. Net-event kinetic Monte Carlo for overcoming stiffness in spatially homogeneous and distributed systems. *Comput Chem Eng*. 2005;29:701–712.
25. Durbin SD, Feher G. Simulation of lysozyme crystal growth by the Monte Carlo method. *J Cryst Growth*. 1991;110:41–51.
26. Christofides PD, Armaou A, Lou Y, Varshney A. Control and Optimization of Multiscale Process Systems. Boston, MA: Birkhäuser, 2008.
27. Nayhouse M, Kwon JS, Christofides PD, Orkoulas G. Crystal shape modeling and control in protein crystal growth. *Chem Eng Sci*. 2013;87:216–223.
28. Müller C. How to describe protein crystals correctly? Case study of lysozyme crystals. Dissertation, Martin-Luther-Universitt Halle-Wittenberg, 2012.
29. Ke SC, DeLucas LJ, Harrison JG. Computer simulation of protein crystal growth using aggregates as the growth unit. *J Phys D: Appl Phys*. 1998;31:1064–1070.
30. Cacioppo E, Munson S, Pusey ML. Protein solubilities determined by a rapid technique and modification of that technique to a micromethod. *J Cryst Growth*. 1991;110:66–71.
31. Cacioppo E, Pusey ML. The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4. *J Cryst Growth*. 1991;114:286–292.
32. Feher G, Kam Z. Nucleation and growth of protein crystals: general principles and assays. *Methods Enzymol*. 1985;114:77–1112.
33. Gilmer GH, Bennema P. Simulation of crystal growth with surface diffusion. *J Appl Phys*. 1972;43:1347–1360.
34. Galkin O, Vekilov PG. Nulceation of protein crystals: critical nuclei, phase behavior, and control pathways. *J Cryst Growth*. 2001;232:63–76.
35. Liu JJ, Ma CY, Hu YD, Wang XZ. Modelling protein crystallisation using morphological population balance models. *Chem Eng Res Des*. 2010;88:437–446.
36. Nanev CN, Tsekova D. Heterogeneous nucleation of Hen-Egg-White Lysozyme-molecular approach. *Cryst Res Technol*. 2000;35:189–195.
37. Suzuki Y, Miyashita S, Komatsu H, Sato K, Yagi T. Crystal growth of Hen Egg White Lysozyme under high pressure. *Jpn J Appl Phys*. 1994;33:1568–1570.
38. Aldabaibeh N, Jones MJ, Myerson AS, Ulrich J. The solubility of orthorhombic lysozyme chloride crystals obtained at high pH. *Cryst Growth Des*. 2009;9:3313–3317.
39. Aldabaibeh N. Towards an improved understanding of the variables controlling crystal modification and phase transition in lysozyme crystals. Dissertation, Illinois Institute of Technology, 2010.
40. Weber M, Jones M, Ulrich J. Crystallization as a purification method for jack bean urease: on the suitability of poly(ethylene), Li_2SO_4 and NaCl as precipitants. *Cryst Growth Des*. 2008;8:711–716.
41. Müller C, Liu Y, Migge A, Pietzsch M, Ulrich J. Recombinant L-Asparaginase B and its crystallization—what is the nature of protein crystals?. *Chem Eng Technol*. 2011;34:571–577.
42. Müller C, Ulrich J. A more clear insight of the lysozyme crystal composition. *Cryst Res Technol*. 2011;46:646–650.
43. Shi D, Mhaskar P, El-Farra NH, Christofides PD. Predictive control of crystal size distribution in protein crystallization. *Nanotechnology*. 2005;16:S562–S574.
44. Shi D, El-Farra NH, Li M, Mhaskar P, Christofides PD. Predictive control of particle size distribution in particulate processes. *Chem Eng Sci*. 2006;61:268–281.
45. Chiu T, Christofides PD. Robust control of particulate processes using uncertain population balances. *AIChE J*. 2000;46:266–280.

Manuscript received Sept. 30, 2012, and revision received Dec. 29, 2012.