Chemical Engineering Science 87 (2013) 216–223

Contents lists available at SciVerse ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Crystal shape modeling and control in protein crystal growth

Michael Nayhouse [a], Joseph Sang-Il Kwon [a], Panagiotis D. Christofides [a,b], Gerassimos Orkoulas [a,*]

[a] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095, USA
[b] Department of Electrical Engineering, University of California, Los Angeles, CA 90095, USA

HIGHLIGHTS

- Kinetic Monte Carlo simulation of protein crystal growth.
- Approximate model for evolution of protein crystal shape.
- Model predictive control of protein crystal shape.
- Disturbance handling using model predictive control.

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form
8 October 2012
Accepted 15 October 2012
Available online 26 October 2012

Keywords:
Crystallization
Molecular modeling
Process control
Optimization
Simulation
Lysozyme

ABSTRACT

Proteins play a key role as therapeutics in a number of diseases and protein crystallization is a central activity in the pharmaceutical industry. Protein crystals, usually produced through a batch crystallization process, are desired to be of high quality, of desired shape, and within a narrow size and shape distribution range. Motivated by the above considerations, the present work focuses on the modeling and control of protein crystal shape. The model protein used for this work is the tetragonal hen egg white lysozyme. The growth of an individual lysozyme crystal is modeled via kinetic Monte Carlo (kMC) simulations comprising adsorption, desorption, and migration events on the (110) and (101) faces, which are assumed to be independent. The expressions for the rate equations for each event type are similar to those of Durbin and Feher (1991). Extensive testing of the system parameters indicates crossover behavior between the growth rates of the two faces [i.e., (110) and (101)], a fact that has also been observed experimentally. A nonlinear algebraic equation that relates the steady-state growth rate ratios between the (110) and (101) faces, the temperature and concentration, is derived from the kMC simulation data. This nonlinear equation is then utilized by a model predictive controller which regulates the protein crystal to desired shapes subject to manipulated input constraints. The proposed method is shown to successfully regulate protein crystal shape, ranging from equidimensional to more elongated type of structures, in the presence of arbitrary variations of the protein concentration.

1. Introduction

Protein crystallization is a central activity in the production of pharmaceuticals. It is used to perform such processes as separating drug from solvent mixtures and to ensure that drug crystal products conform to size and morphology specifications. Protein structure can be found via nuclear magnetic resonance and X-ray crystallography (Rosenberger et al., 1996; Vekilov and Chernov, 2003; Wiencek, 1999). 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 (Vekilov and Chernov, 2003) can be used as long as protein crystals are of desired shape and high quality. In the present work, the model protein investigated is the tetragonal form of hen egg white lysozyme. This naturally occurring enzyme with antibacterial activity 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.

It is believed that protein crystallization proceeds in three stages: nucleation, crystal growth, and cessation of growth (Kierzek and Zielenkiewicz, 2001). The present work focuses on the second stage of protein crystallization, i.e., crystal growth, which will be modeled via kinetic Monte Carlo (kMC) simulations. Several attempts aimed at modeling protein nucleation (Galkin and Vekilov, 1999; Pusey and Nadarajah, 2002) and growth (Durbin and Feher, 1986; Forsythe et al., 1999; Kurihara et al., 1996) have also been made. These efforts make it possible to manipulate the size distribution and morphology of the protein

* Corresponding author. Tel.: +1 310 267 0169; fax: +1 310 206 4107.
E-mail address: makis@seas.ucla.edu (G. Orkoulas).

0009-2509/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ces.2012.10.020

crystal, which is a very critical variable for pharmaceutical products. 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 nucleation (Durbin and Carlson, 1992; Durbin and Feher, 1986; Vekilov et al., 1993). The present work focuses on crystals that already have been seeded via two-dimensional nucleation since a large supersaturation is normally required in a batch crystallization process (Asherie, 2004; Vekilov and Chernov, 2003) to achieve reasonable growth rates (Forsythe et al., 1999). It is noted that two-dimensional nucleation proceeds at supersaturation $\gtrsim 1.6$ (Vekilov et al., 1993). Since crystal growth is a non-equilibrium process, kMC simulation methods are used to model the growth. Kinetic Monte Carlo algorithms, which form the basis for applying the Monte Carlo method to simulate dynamic processes (Bortz et al., 1975; Dai et al., 2005, 2008; Gillespie, 1976, 1977, 1978, 1992, 2001, 2007; Rathinam et al., 2003; Reese et al., 2001; Snyder et al., 2005), are based on a dynamic interpretation of the Master equation (Fichthorn and Weinberg, 1991; Müller-Krumbhaar and Binder, 1973). As is a common practice in simulations of crystal growth, the solid-on-solid model (Durbin and Feher, 1991) is used in this work to interpret the growth of protein crystals from supersaturated solutions. In the solid-on-solid approximation, particles are deposited on the growing crystal lattice without voids or overhangs, resulting in a highly compacted crystal.

The kMC methodology (Christofides et al., 2008) proposed in the present work uses rate equations originally developed by Durbin and Feher (1991). Durbin and Feher (1991) found that different crystal faces produce different growth rates depending on the conditions of each independent face of the crystal. Their results show a crossover type of behavior between the growth rates of the (110) and (101) faces, a fact that is consistent with experimental findings (Durbin and Feher, 1986).

The simulations comprise three microscopic events, namely molecular attachment, detachment, and migration events on the (110) and (101) faces. All attachment events made in this work are implemented using monomer units. Assuming that all surface sites are available for attachment, the attachment rate is considered independent for each lattice site. Detachment and migration events, however, are dependent on their local environment. The local environment of a given lattice site is specified based on the number of nearest neighbors surrounding that site. The nearest neighbors of a lattice site are on the (N,S,E,W) directions which are of the same height or higher compared to the current lattice site. Another nearest neighbor is located directly below each surface particle, however no surface particle is without a nearest neighbor below itself due to the solid-on-solid model. Thus we will consider the nearest neighbor below each particle in the pre-exponential factor of both the desorption and the migration rate equations presented in the next section. That being said, the number of nearest neighbors we explicitly consider will only be on the (N,S,E,W) directions ranging from zero to four giving a total of five classes. These classes will be used to lessen the computational cost when calculating the rates for the three microscopic events, described in the following section. Owing to the dependence of detachment rate on the surface configuration, kMC simulation is 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, the methodology of the kMC simulation is provided followed by the Model Predictive Control methodology. The simulation results for crystal growth of the tetragonal form of hen egg white lysozyme under open-loop and closed-loop conditions are then presented and discussed followed by conclusions and directions for future research.

2. Methodology

As already emphasized, the solid-on-solid model, which is a simple way to look at the crystallization process, is used to model the growth of lysozyme. As noted by Ke et al. (1998), the system size does not largely affect crystal growth. They report that no finite-size effects were found on systems of sizes 30×30, 60×60, and 120×120 sites. In the present work, a periodic square lattice of length and width of $N=50$ sites is used. The height at a given location within the lattice is defined as the number of particles in the growth direction. Each simulation comprises 4 million events, or approximately 1600 events per lattice site on average. At the beginning of each simulation, the lattice is initialized to a flat surface. To ensure that the initial configuration does not have a noticeable impact on the results, the first 50,000 events are discarded in order to allow the surface to roughen. Each event of our kMC simulation is chosen randomly based on the rates of the three microscopic phenomena, described below.

2.1. Surface kinetics

The following description of the surface kinetics for the present model follows closely that of Ke et al. (1998) which is based on the work by Durbin and Feher (1991). As emphasized earlier, since each surface site is available for attachment, the attachment rate is independent of each lattice site and defined as

$$r_a = K^+(\Delta\mu) = K_0^+ \exp\left(\frac{\Delta\mu}{k_BT}\right), \tag{1}$$

where K_0^+ is the attachment coefficient, k_B is the Boltzmann constant, T is the temperature in Kelvin, and $\Delta\mu = k_BT\ln(c/s)$, where c is the concentration and s is the protein solubility and this term is the crystallization driving force. It is noted that $r_a \propto c$. Since the total number of lattice sites is N^2, the total rate of adsorption, W_a, is defined as

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

As already commented, the desorption rate of a surface particle depends on the local environment. Thus, the desorption rate of a lattice site with i nearest neighbors is given by

$$K^-(E_b) = K_0^- \exp\left(-\frac{E_b}{k_BT}\right) = K_0^- \exp\left(-i\frac{E_{pb}}{k_BT}\right), \tag{3}$$

where K_0^- is the desorption coefficient, i is the number of bonds, E_{pb} is the average binding energy per bond, and $E_b = iE_{pb}$ is the total binding energy. The bond-dependent desorption rate, $r_d(i)$, is thus defined as

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

It can be seen that with less nearest neighbors, the desorption rate becomes higher. The total rate of desorption is computed by

$$W_d = \sum_{i=0}^{4} W_{d_i} \quad \text{with } W_{d_i} = M_i r_d(i), \tag{5}$$

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. Similar to the desorption rate, Ke et al. (1998) defined the migration rate the following way:

$$r_m(i) = K_0^- \exp\left(-i\frac{E_{pb}}{k_BT} + \frac{E_{pb}}{2k_BT}\right). \tag{6}$$

The extra term added in Eq. (6) will cause migration events to be favored versus desorption events. The total migration rate is

computed as

$$W_m = \sum_{i=0}^{4} W_{m_i} \quad \text{with } W_{m_i} = M_i r_m(i). \tag{7}$$

The total rate, W_{tot}, is computed by summing over all rates of the three microscopic events, i.e., $W_{\text{tot}} = W_a + W_d + W_m$.

As was pointed out by Durbin and Feher (1991) and Ke et al. (1998), K_0^+ and K_0^- are not independent. At equilibrium, $\Delta\mu = 0$ and also the attachment and detachment rates are equal, i.e.,

$$K^+(\mu = 0) = K^-(\phi), \tag{8}$$

where ϕ is the binding energy per molecule of a fully occupied lattice. Thus, thermodynamic equilibrium provides the following connection between the adsorption, K_0^+, and desorption, K_0^-, coefficients:

$$K_0^+ = K_0^- \exp\left(-\frac{\phi}{k_B T}\right) \tag{9}$$

or by rearranging

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

Plugging this result back into the original equations yields the following expressions for the desorption and migration rates:

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

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

The binding energies cannot be accessed by experiments (Durbin and Feher, 1991; Feher and Kam, 1985). 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 (Durbin and Feher, 1991). The same approach is also followed in the present work. For this work we use $K_0^+ = 0.4704\ \text{s}^{-1}$ following from similar numbers of Durbin and Feher (1991) and Grimbergen et al. (1999). Additionally, we use $c = 7.8$ mg/ml, $E_{pb}/k_B = 218.99$ K and $\phi/k_B = 734.78$ K for the (110) face, and $E_{pb}/k_B = 259.34$ K and $\phi/k_B = 564.77$ K for the (101) face. It is noted that $E_{pb}/k_B T$ and $\phi/k_B T$ are required in Eqs. (11) and (12) and these quantities must be divided by the temperature in Kelvin. Finally, it is important to emphasize that the computation of the growth rate for each face requires the use of kMC simulations owing to the dependence of the detachment rate on the surface micro-configuration and it cannot be computed by simply subtracting the attachment and detachment rates.

2.2. Event execution

In order to execute an event, a uniform random number, $\zeta_1 \in [0,1)$, is generated. If $\zeta_1 \le W_a/W_{\text{tot}}$, then an adsorption event is executed. If $W_a/W_{\text{tot}} < \zeta_1 \le (W_a + W_d)/W_{\text{tot}}$, then a desorption event is executed. Lastly, if $\zeta_1 > (W_a + W_d)/W_{\text{tot}}$, then a migration event is executed. For the case of adsorption, a lattice site is chosen at random for the adsorption 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

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

Once the class is determined, a second random number, ζ_2, is generated to select a random lattice site within 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:

$$\frac{W_a + W_d + \sum_{i=0}^{k-1} W_{m_i}}{W_{\text{tot}}} < \zeta_1 \le \frac{W_a + W_d + \sum_{i=0}^{k} W_{m_i}}{W_{\text{tot}}}. \tag{14}$$

In Durbin and Feher's work (1991), it was 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 (Durbin and Feher, 1991). In the present work, this behavior is modeled 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. Similar to Gilmer and Bennema (1972), each available migration site is given equal probability to accept the displaced particle.

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

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

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

3. Linking growth rate ratio with concentration and temperature

The data generated from the kMC simulations was used to construct a nonlinear algebraic equation (visualized by a 3-D plot). This equation was utilized in the model predictive control (MPC) formulation to relate the crystal growth rate ratio to the temperature and protein concentration in the continuous phase. The shape of the crystal is determined by the two independent faces and their relative growth rates. Therefore, it is possible to control the evolution of crystal shapes by controlling the ratio between the growth rates of the two faces. The growth rate ratio curve (discussed in Section 4) and the measurement of current temperature as well as of the concentration are assumed to be available to the controller.

3.1. Modeling steady-state growth rate ratio dependence on temperature and solution concentration

The three variables (i.e., temperature, growth rate ratio, and protein concentration) used for the model construction are obtained from the open-loop simulations using the kMC model in Section 2. The protein solution concentration (temperature) is fixed during each open-loop simulation to observe the dependence of the crystal growth rate ratio on temperature (protein solution concentration).

A nonlinear algebraic equation relating the relative growth rate ratio versus protein concentration and temperature is derived to quantify the evolution of the crystal growth accounting for the effect of protein concentration variations and temperature changes in supersaturated protein solutions. The concentration variation results from the mixing problem of the batch process as well as the technical difficulty in the measurement of protein concentration. Therefore, a Gaussian noise with the following mean and covariance is added to the concentration measurements, $c(t)$, in simulations:

$$\langle c(t) \rangle = c_n, \quad \langle c(t)c(t') \rangle = \sigma^2 c_n^2, \tag{16}$$

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. In order to suppress the uncertainty in the protein concentration, the kMC simulations are repeated to compute expected values [i.e., averages denoted as $\langle \cdot \rangle$ see, e.g., Eq. (16)]. Also, it is important to note that the concentration variation results in a change of the attachment rate [cf. Eq. (1)].

Different operating conditions strongly affect the crystallization process and, consequently, the behavior of the growth rate ratio profile. Fig. 1 shows the ratio data collected from an open-loop simulation. In this work, the temperature, T, is chosen as the manipulated input and the pH and NaCl concentration are maintained fixed at 4.6 and 5.0%, respectively. The protein concentration as well as the temperature dependence of the growth rate ratio is obtained by generating a 3-D plot for a variety of protein concentrations ranging from 7.2 to 8.4 mg/ml, and temperatures ranging from 4 to 25 °C. It should be noted that 25 °C is the maximum temperature at which tetragonal crystals can be obtained (Berthou and Jolles, 1974). There is a maximum in the growth rate ratio located at ~15 °C. Thus the ratio tends to decrease as the temperature increases or decreases from 15 °C and the protein concentration decreases. Fig. 1 is the 3-D surface plot of $r=f(T,C)$, where f is a nonlinear function of the growth rate ratio, to better illustrate the effects of changing temperatures and protein concentrations. To this end, in Fig. 1, the data was interpolated to fit a regularly spaced grid.

It can be clearly seen that the growth rate ratio is a strong function of the temperature and protein concentration, and this dependence is the basis for using temperature, T, to control protein crystal shape.

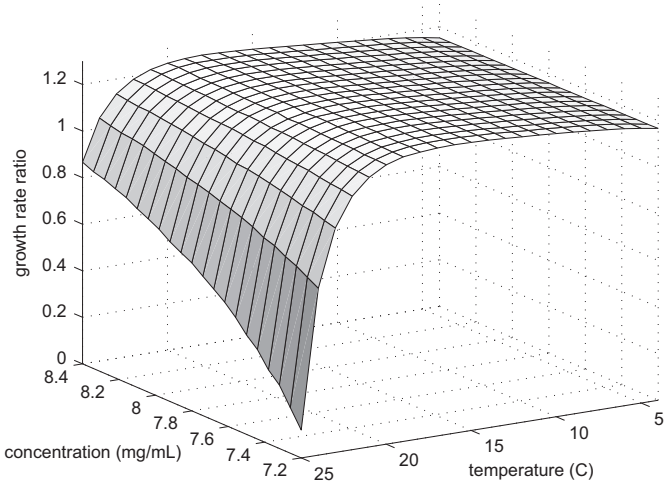

Fig. 1. Surface plot of the growth rate ratio data for tetragonal lysozyme at pH 4.6 and 5% NaCl. The data from the open-loop kMC simulation are plotted to demonstrate the effect of temperature and concentration variations on this ratio. Protein concentration and temperature range from 7.2 to 8.4 mg/ml and from 4 to 25 °C, respectively.

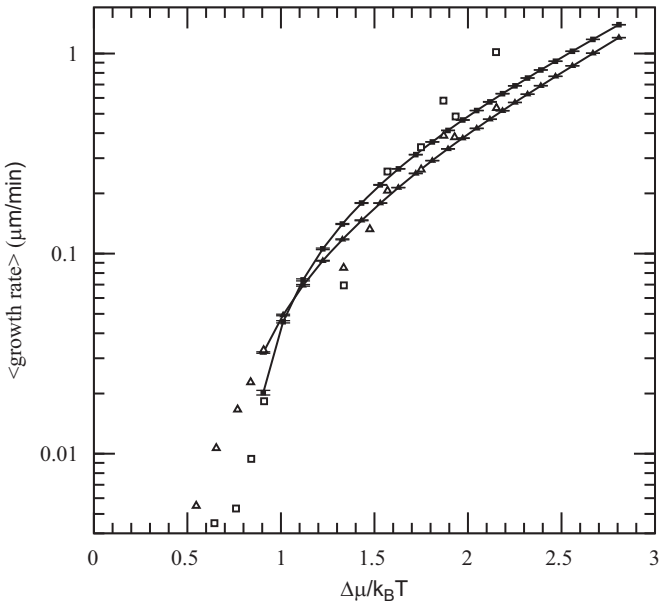

Fig. 2. The expected growth rate versus the degree of supersaturation at $c=7.8$ mg/ml. The (■) represents the (110) face with $E_{pb}/k_B=218.99$ K and $\phi/k_B=734.78$ K. The (▲) represents the (101) face with $E_{pb}/k_B=259.34$ K and $\phi/k_B=564.77$ K. The error bars represent two standard deviations of the growth rate. The (□) and (△) represent the (110) and (101) faces respectively, obtained from Durbin and Feher (1986).

4. Model predictive control of crystal shape

As emphasized previously, in the kMC simulations molecular attachment, detachment, and migration events are considered. Depending on the relative attachment energy of the crystal faces and assuming that the independent crystal faces that appear during the crystal growth are the (110) and (101) faces, kMC simulations reproduce the experimentally observed crossover behavior in the crystal growth rates between the two faces [cf. Fig. 2]. In this section, a model predictive controller is designed based on the 3-D plot (i.e., the non-linear growth rate ratio equation), Fig. 1, to suppress the concentration variation and achieve the desired set-point values by manipulating the temperature. A desired set-point value of the growth rate ratio is included in the cost function in the MPC formulation. Regarding the choice of MPC for the controller design, it is noted that classical control schemes like proportional (P) control cannot be employed to explicitly account for input/state constraints, optimality considerations, and the nature of the attachment, detachment, and migration processes. Also, dynamic open-loop optimization may be used but it does not provide robustness against model inaccuracies and fluctuation in the protein concentration.

4.1. Model predictive control formulation

To this end, consider the control problem of the growth rate ratio by using an MPC design. The expected value of the growth rate ratio, $\langle G_{\text{ratio}} \rangle$, is chosen as the control objective, where G_{ratio} is defined as a ratio G_{110}/G_{101}. Here G_{110} and G_{101} signifies the growth rate on the (110) and (101) faces, respectively. The temperature is used as the manipulated input. The pH and NaCl concentration are fixed during all closed-loop simulations. The proposed modeling and control methods do not depend on the specific number of the manipulated variables and can be easily extended to the case of multiple inputs. 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 such that $4 \le T \le 25$ °C. This constraint ensures validity for the kMC model by imposing a range that will not damage the protein crystal. Another constraint is imposed on the rate of change of the temperature to account for actuator limitations. The control action at time t is obtained by solving a finite-horizon optimal control problem. The cost function in the optimal control problem includes a penalty on the deviation of $\langle G_{\text{ratio}} \rangle$ from its set-point value which is determined based on the desired crystal shape. Since the protein crystallization process is a batch process, a desired minimum thickness

(i.e., minimum amount of growth on each face of the crystal) may be required to ensure that the crystal has the properties necessary for its desired application at the end of the crystallization process. The thickness may be obtained by adjusting the time that growth takes place since the growth rates can be estimated. However, in this work to simplify the development and focus on crystal shape control, imposing a minimal crystal size of the two surfaces is disregarded in the MPC formulation. The optimal temperature profile is calculated by solving a finite-dimensional optimization problem in a receding horizon fashion. The MPC problem is formulated as follows:

$$\begin{aligned} \underset{T_1,\ldots,T_i,\ldots,T_p}{\text{minimize}} \quad & \sum_{i=1}^{p} F_i, \\ \text{subject to} \quad & F_i = (\langle G_{\text{ratio}} \rangle - G_{set})^2, \\ & G_{\text{ratio}} = f(T,C), \\ & T_{\min} \le T_i \le T_{\max}, \\ & \left|\frac{T_{i+1}-T_i}{\Delta}\right| \le R_T, \\ & i = 1,2,\ldots,p, \end{aligned} \tag{17}$$

where t is the current time, F_i is the cost function expressing the deviation of $\langle G_{\text{ratio}} \rangle$ from its set-point ratio, G_{set}, Δ is the sampling time, p is the number of prediction steps, $p\Delta$ is the specified prediction horizon, t_i, $i=1,2,\ldots,p$, is the time of the ith prediction step, $t_i = t+i\Delta$, respectively, $T_i, i=1,2,\ldots,p$, 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, and R_T is the limit on the rate of change of the temperature. The optimal set of control actions $(T_1,T_2,\ldots,T_p)$, is obtained from the solution of the multi-variable optimization problem of Eq. (17), 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, when a new measurement of protein concentration in the continuous phase is received from the kMC simulation and the MPC problem of Eq. (17) is re-solved 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 (Cacioppo et al., 1991). In a previous work, empirical expressions were obtained for the growth rates by fitting algebraic expressions to the available experimental data (Shi et al., 2005). In the present work, the growth rates are computed following the kMC methodology from Section 2. Furthermore, the stochastic nature of the system and the model uncertainty will be accounted for in the protein concentration variations; see, e.g., Shi et al. (2006) and Chiu and Christofides (2000), for results on model predictive control and robust-control of crystallization systems, respectively.

5. Results

5.1. Open-loop simulations

For a given set of the simulation conditions composed of temperature, pH, salt and protein solution concentrations, the method described in the Section 2 results in averaged lysozyme face growth rates at various values of supersaturations.

In Fig. 2, crystal growth is modeled at supersaturation $0.9 \lesssim \ln(c/s) \lesssim 2.8$, where c (mg/ml) is the solution protein concentration and s (mg/ml) is the solubility. The growth rates produced are the average growth rate for each set of conditions over 10 independent kMC runs and are compared against data from Durbin and Feher (1986). For Fig. 2, $c=7.8$ mg/ml. The solubility is determined by Eq. (18) using a third order polynomial of solubility in terms of temperature T (°C) at pH = 4.6 and 5%(w/v) NaCl (Cacioppo et al., 1991; Cacioppo and Pusey, 1991).

$$s(T) = 3.506 \times 10^{-4} T^3 - 9.046 \times 10^{-3} T^2 + 1.303 \times 10^{-1} T + 7.209 \times 10^{-2}. \tag{18}$$

The above equation allows for accurate modeling of the solubility in terms of temperature at the selected pH and salt concentration with an error of 5.4% (Cacioppo and Pusey, 1991). As is evident in Fig. 2, crossover behavior between the (110) and (101) faces does, indeed, occur. Specifically, the crossover value of the growth rate from Fig. 2 is ~ 0.06 μm/min and Durbin and Feher (1986) show ~ 0.1 μm/min. The same growth rates are shown in terms of temperature in Fig. 3. This figure directly shows the relationship between the growth rates and the temperature. As shown in Fig. 3, as temperature increases, the growth rates on both faces decrease. The growth rate for the (110) face decreases at a higher rate than the growth rate for the (101) face. These results follow previous experimental work for pH and salt concentration near our given values (Nadarajah et al., 1995, 1997).

Furthermore, the kMC simulations were tested over a varying range of protein concentrations in the liquid solution. Concentrations were taken at finite values of 6.8, 7.8, 8.8, and 9.8 mg/ml. As anticipated, higher protein concentrations and lower temperatures yield higher growth rates. However, as temperature rises, the difference between the growth rate of the constant concentration curves becomes smaller. Each of the points in Figs. 4 and 5 is taken from averaging over 10 kMC simulations to compute accurate expected values, where the error bars represent two standard deviations. Since the rates for desorption and migration change after the execution of each microscopic event, the steady-state growth rate must be computed by averaging over the individual growth rates obtained from several kMC processes. Fig. 5 displays the four finite values of concentration at $T = 14$ °C located in the rectangular box from Fig. 4. There are minimal fluctuations in Fig. 5, that allow one to deduce reliable estimates for the growth rate.

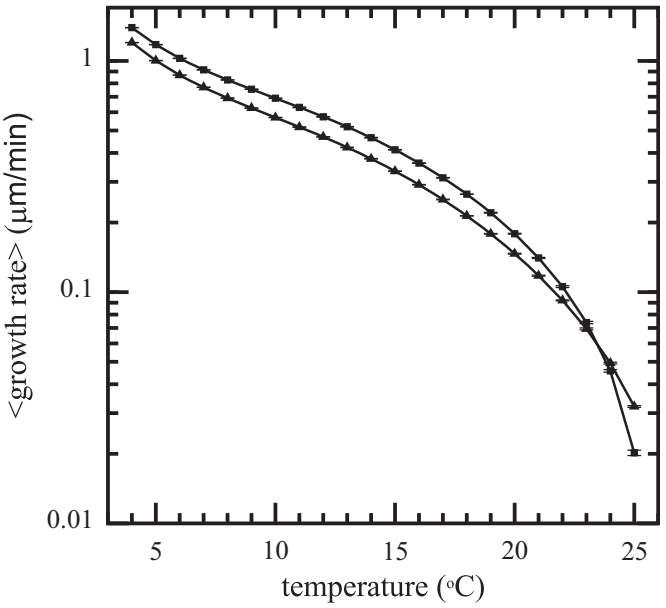

Fig. 3. The expected growth rate versus temperature of $c=7.8$ mg/ml. The (■) represents the (110) face with $E_{pb}/k_B = 218.99$ K and $\phi/k_B = 734.78$ K. The (▲) represents the (101) face with $E_{pb}/k_B = 259.34$ K and $\phi/k_B = 564.77$ K. The error bars represent 2 standard deviations of the growth rate.

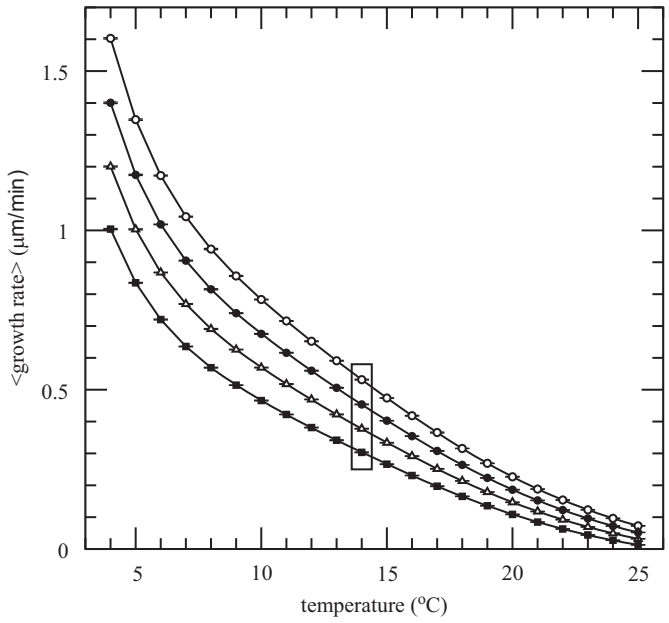

Fig. 4. The expected growth rate on the (101) face of pH 4.6, 5% NaCl, and various values of lysozyme concentration. The labeling is as follows: (○), c=9.8 mg/ml; (●), c=8.8 mg/ml; (△), c=7.8 mg/ml; and (■), c=6.8 mg/ml. For these simulations, $E_{pb}/k_B = 259.34$ K and $\phi/k_B = 564.77$ K. The error bars represent two standard deviations of the growth rate.

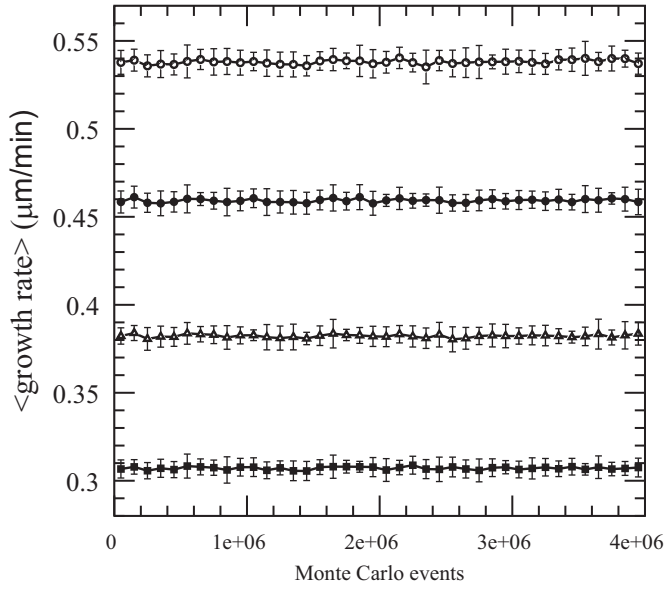

Fig. 5. The expected growth rate on the (101) face at pH 4.6, 5% NaCl, $T = 14$ °C, and various values of lysozyme concentration. The labeling is as follows: (○), c=9.8 mg/ml; (●), c=8.8 mg/ml; (△), c=7.8 mg/ml; and (■), c=6.8 mg/ml. For these simulations, $E_{pb}/k_BT = 0.903$ and $\phi/k_BT = 1.967$. The error bars represent two standard deviations of the growth rate.

5.2. Closed-loop simulations

In this section, the proposed model predictive controller of Eq. (17) is applied to the kMC model described in Section 2. These closed-loop simulations will be performed to test the ability of the model predictive controller to drive the growth rate ratio to desired set-point values in the presence of significant variation/disturbance in the operating conditions. At each sampling time (1 s), the optimal temperature, obtained by solving the optimization problem of Eq. (17), 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 model described previously to predict the dependence of the crystal growth rate ratio on temperature and protein concentration.

The desired values (set-points) in the closed-loop simulations are $\langle G_{\text{ratio}} \rangle = 1.19$ and $\langle G_{\text{ratio}} \rangle = 0.67$. The protein concentration randomly varies following the Gaussian distribution of Eq. (16) and the pH and NaCl concentration are fixed at 4.6 and 5.0%, respectively, and the initial temperature is 15 °C. The nominal concentration is 7.8 mg/ml with $\sigma = 2.5\%$ ($\sigma \times c_n = 0.195$ mg/ml). We would like to note that no material balance was included for the solute concentration since the focus of this work is on modeling and control of the crystal shape. The maximum rate of change of the temperature is 2 °C/min, and the minimum and maximum temperature allowed by the controller is 4 °C and 25 °C, respectively. Since the MPC formulation uses the steady-state growth rates, the number of prediction steps is set to be p=1. The time interval between two sampling times is 1 s. The prediction horizon of each step is fixed at $p\Delta = 1$ s. The concentration varies every 0.333 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 is 3000 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, $\langle G_{\text{ratio}} \rangle = 1.19$ and $\langle G_{\text{ratio}} \rangle = 0.67$ respectively. Thus, the cost function of this problem contains penalty on the deviation of the expected growth rate ratio from the set-point value.

The results of the closed-loop simulations are shown in Figs. 6–9. From Fig. 6, for the low set-point ratio $\langle G_{\text{ratio}} \rangle = 0.67$, it can be seen that soon after the initial rise, the growth rate ratio decreases constantly, then fluctuates for the rest of time towards the end of the simulation. Although disturbance results in the fluctuation of the concentration, the MPC can successfully drive the growth rate ratio to the desired set-point. Also, as is shown in Figs. 8 and 9, the use of expected values (ie., averages obtained form 100 independent simulations for the same set of conditions) in the control formulation, suppresses fluctuations. Again, the

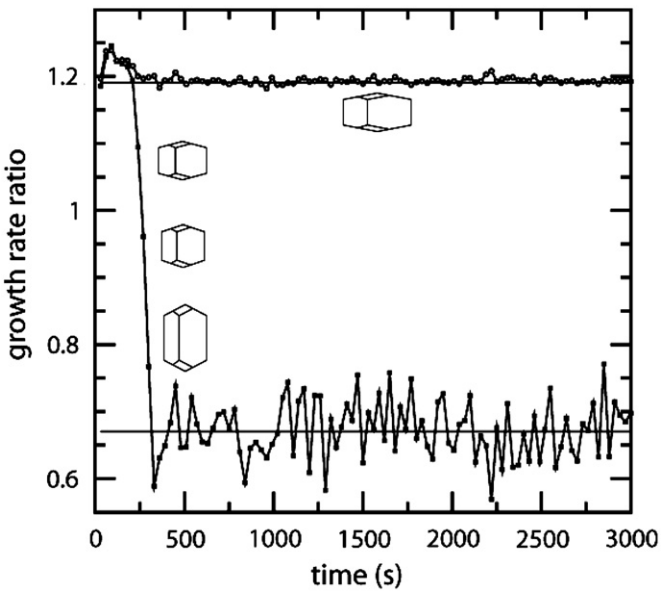

Fig. 6. Profiles of growth rate ratio under closed-loop operation. The growth rate ratio set-point values are (○), $\langle G_{\text{ratio}} \rangle = 1.19$ and (■), $\langle G_{\text{ratio}} \rangle = 0.67$, respectively.

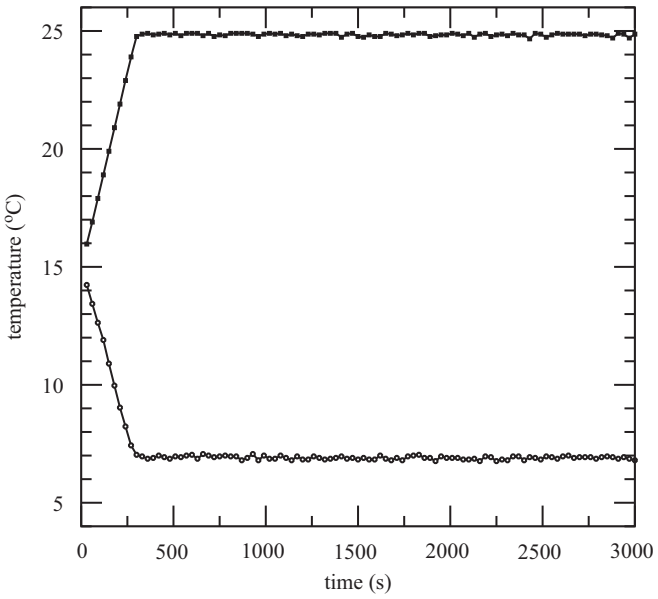

Fig. 7. Profiles of temperature under closed-loop operation. The growth rate ratio set-point values are (○), $\langle G_{ratio} \rangle = 1.19$ and (■), $\langle G_{ratio} \rangle = 0.67$, respectively.

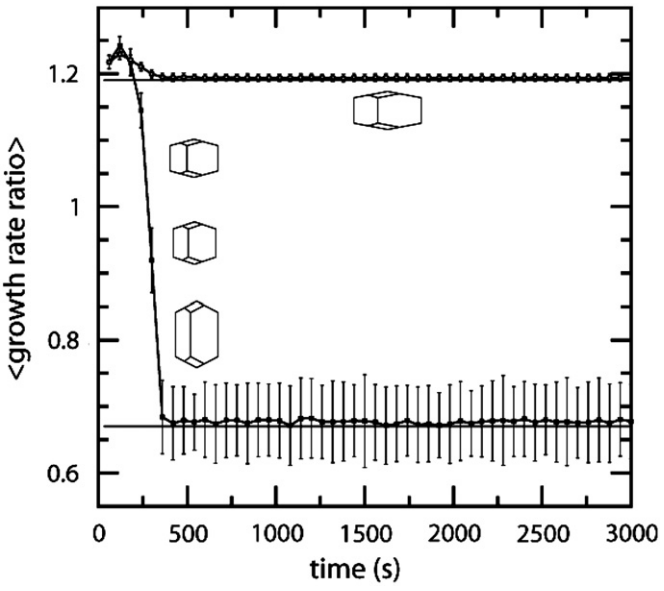

Fig. 8. Profiles of the expected values of growth rate ratio under closed-loop operation. The growth rate ratio set-point values are (○), $\langle G_{ratio} \rangle = 1.19$ and (■), $\langle G_{ratio} \rangle = 0.67$, respectively. The error bars represent two standard deviations of the growth rate ratio.

error bars displayed are 2 standard deviations in Figs. 8 and 9. The shape of the protein crystal can be estimated from Fig. 6 and it is slightly elongated along the (101) direction for the lower set-point ratio while it is more equidimensional to slightly elongated along the (110) direction for the higher set-point ratio. In Figs. 7 and 9, temperature profiles are displayed to show how the optimal input changes over time. The temperature reaches steady-state with minor fluctuation once the growth rate ratio settles onto the desired ratio. The expected growth over 3000 s for the set-point ratio $\langle G_{ratio} \rangle = 1.19$ on the (110) and (101) faces were approximately 45.69 and 38.22 μm/min, respectively giving a ratio of 1.195. In contrast, the expected growth for the set-point ratio $\langle G_{ratio} \rangle = 0.67$ on the (110) and (101) faces were approximately 2.07 and 2.42 μm/min, respectively giving a ratio of 0.855. The reason this ratio is much further away from the set-point

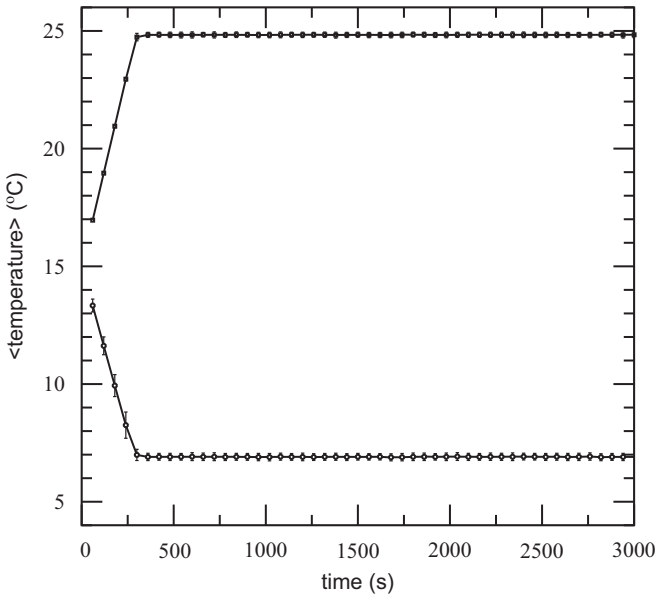

Fig. 9. Profiles of the expected values of temperature under closed-loop operation. The growth rate ratio set-point values are (○), $\langle G_{ratio} \rangle = 1.19$ and (■): $\langle G_{ratio} \rangle = 0.67$, respectively. The error bars represent two standard deviations of the temperature.

value is due to the starting temperature at 15 °C compared to the steady state value of approximately 25 °C causing the growth rate at the beginning of the run to be much higher than at the end. If we remove the first 300 s from the growth, the expected growth for the set-point ratio $\langle G_{ratio} \rangle = 0.67$ on the (110) and (101) faces were approximately 1.07 and 1.58 μm/min, respectively giving a ratio of 0.677, which is in good accord with the set-point. This shows the importance of the starting temperature in the batch process to control the crystal shape. The larger error bars are due to the higher sensitivity of the crystal growth rate ratio in the presence of solute concentration disturbance at higher temperatures. (We note that the magnitude of the solute concentration disturbance is independent of the nominal solute concentration value and thus the disturbance has a more pronounced effect at low solute concentrations resulting in larger error bars given the gaussian nature of the disturbance.)

6. Conclusions and future work

The present work is associated with application of modeling, simulation, and control of a batch protein crystallization process in order to produce a single lysozyme protein crystal. Based on the assumption that the two independent crystal faces are the (110) and (101) faces, dependence of the growth rates of the two faces on temperature and protein solution concentration, obtained from kMC open-loop simulations, was observed. The shape of the resulting lysozyme crystals was controlled through temperature manipulation. This was achieved via a nonlinear steady-state model generated from open-loop kMC simulations. The non-linear model captures the protein solution concentration and temperature dependence of the growth rate ratio, thereby describing the key elements of the protein crystallization process. An MPC strategy, which uses the steady-state model, was then designed to drive the growth rate ratio to the desired value while satisfying constraints on the magnitude and rate of change of temperature, which is chosen as the manipulated input in accordance with standard batch crystallization practice. Simulation results showed that the proposed controller was able to

control the crystal shapes by appropriately manipulating the desired ratio ranging from 0.67 to 1.19. The present methodology shows that both equidimensional and elongated crystals can be produced, as shown in Figs. 6 and 7. Furthermore, the proposed kMC methodology offers a number of advantages compared to previous work. In previous simulations, such as Ke et al. (1998), a lattice site was chosen at random, and then the rates were calculated for that lattice site only. In contrast, in the present methodology, the entire lattice is considered. To avoid the recalculation of the total rates for the entire lattice in a site by site manner after each event, we update the five classes and then use them to easily calculate the total rates. This allows our method to remain computationally efficient.

In addition to monomer growth, aggregates of two or more molecules can be considered during adsorption events onto a crystal surface. Ke et al. (1998) did this, but only for one face of the crystal lattice, and were able to observe growth rates closer to experimental results (e.g., growth rates increase more at moderate supersaturation levels and decrease at very high supersaturation levels). In order to select either a monomer or an aggregate in every attachment event, another random number is generated and this consideration can produce more realistic growth rates.

In the present work, lysozyme growth was studied on flat crystal face surfaces for which 50 thousand Mote Carlo events were executed to allow for the surface to roughen. The influence of initial seed shape on the crystal evolution has been considered by Zhang and Doherty (2004). In the future, asymmetrical seeds for initiation of the simulation will be employed in order to better achieve a variety of crystal shapes.

Finally, it is noted that the present work considered the growth of a single lysozyme crystal only. A future work will look to model an entire crystal population while controlling the crystal shape in a batch crystallizer. Also, a population balance model (PBM) can be adopted to better describe the variation of the crystal shape and size distribution for multiple crystals under the consideration of concentration disturbances and this problem will be considered in a future work.

Acknowledgments

Financial support from the National Science Foundation (NSF) CBET-0967291 is gratefully acknowledged. This material is also based on the work supported by the NSF Graduate Research Felloship 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.

References

Asherie, N., 2004. Protein crystallization and phase diagrams. Methods 34, 266–272.
Berthou, J., Jolles, P., 1974. A phase transition in a protein crystal: the example of hen lysozyme. Biochem. Biophys. Acta 336, 222–227.
Bortz, A.B., Kalos, M.H., Lebowitz, J.L., 1975. New algorithm for Monte Carlo simulation of ising spin systems. J. Comput. Phys. 17, 10–18.
Cacioppo, E., Munson, S., Pusey, M.L., 1991. Protein solubilities determined by a rapid technique and modification of that technique to a micro-method. J. Cryst. Growth 110, 66–71.
Cacioppo, E., Pusey, M.L., 1991. The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4. J. Cryst. Growth 114, 286–292.
Chiu, T., Christofides, P.D., 2000. Robust control of particulate processes using uncertain population balances. AIChE J. 46, 266–280.
Christofides, P.D., Armaou, A., Lou, Y., Varshney, A., 2008. Control and Optimization of Multiscale Process Systems. Birkhäuser, Boston.
Dai, J., Kanter, J.M., Kapur, S.S., Seider, W.D., Sinno, T., 2005. On-lattice kinetic Monte Carlo simulations of point defect aggregation in entropically influenced crystalline systems. Phys. Rev. B 72, 134102.
Dai, J., Seider, W.D., Sinno, T., 2008. Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles. J. Chem. Phys. 128, 194705.
Durbin, S.D., Carlson, W.E., 1992. Lysozyme crystal growth studied by atomic force microscopy. J. Cryst. Growth 122, 71–79.
Durbin, S.D., Feher, G., 1986. Crystal growth studies of lysozyme as a model for protein crystallization. J. Cryst. Growth 76, 583–592.
Durbin, S.D., Feher, G., 1991. Simulation of lysozyme crystal growth by the Monte Carlo method. J. Cryst. Growth 110, 41–51.
Feher, G., Kam, Z., 1985. Nucleation and growth of protein crystals: general principles and assays. Methods Enzymol. 114, 77–112.
Fichthorn, K.A., Weinberg, W.H., 1991. Theoretical foundations of dynamical Monte Carlo simulations. J. Chem. Phys. 95, 1090–1096.
Forsythe, E.L., Nadarajah, A., Pusey, M.L., 1999. Growth of (101) faces of tetragonal lysozyme crystals: measured growth-rate trends. Acta Crystallogr. D 55, 1005–1011.
Galkin, O., Vekilov, P.G., 1999. Direct determination of the nucleation rates of protein crystals. J. Phys. Chem. B 103, 10965–10971.
Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22, 403–434.
Gillespie, D.T., 1977. Exact stochastic simulation of chemical reactions. J. Phys. Chem. 81, 2340–2361.
Gillespie, D.T., 1978. Monte Carlo simulation of random walks with residence time dependent transition probability rates. J. Comput. Phys. 28, 395–407.
Gillespie, D.T., 1992. A rigorous derivation of the chemical master equation. Physica A 188, 404–425.
Gillespie, D.T., 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115, 1716–1733.
Gillespie, D.T., 2007. stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. 58, 35–55.
Gilmer, G.H., Bennema, P., 1972. Simulation of crystal growth with surface diffusion. J. Appl. Phys. 43, 1347–1360.
Grimbergen, R.F.P., Boek, E.S., Meekes, H., Bennema, P., 1999. Explanation for the supersaturation dependence of the morphology of lysozyme crystals. J. Cryst. Growth 207, 112–121.
Ke, S.C., DeLucas, L.J., Harrison, J.G., 1998. Computer simulation of protein crystal growth using aggregates as the growth unit. J. Phys. D: Appl. Phys. 31, 1064–1070.
Kierzek, A.M., Zielenkiewicz, P., 2001. Models of protein crystal growth. Biophys. Chem. 91, 1–20.
Kurihara, K., Miyashita, S., Sazaki, G., Nakada, T., Suzuki, Y., Komatsu, H., 1996. Interferometric study on the crystal growth of tetragonal lysozyme crystal. J. Cryst. Growth 166, 904–908.
Müller-Krumbhaar, H., Binder, K., 1973. Dynamic properties of the Monte Carlo method in statistical mechanics. J. Statist. Phys. 8, 1–24.
Nadarajah, A., Forsythe, E.L., Pusey, M.L., 1995. The averaged face growth rates of lysozyme crystals: the effect of temperature. J. Cryst. Growth 151, 163–172.
Nadarajah, A., Li, M., Pusey, M.L., 1997. Growth mechanism of the (110) face on tetragonal lysozyme crystals. Acta Crystallogr. D53, 524–534.
Pusey, M.L., Nadarajah, A., 2002. A model for tetragonal lysozyme crystal nucleation and growth. Cryst. Growth Des. 2, 475–483.
Rathinam, M., Petzold, L.R., Cao, Y., Gillespie, D.T., 2003. Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. J. Chem. Phys. 119, 12784–12794.
Reese, J.S., Raimondeau, S., Vlachos, D.G., 2001. Monte carlo algorithms for complex surface reaction mechanisms: efficiency and accuracy. J. Comput. Phys. 173, 302–321.
Rosenberger, F., Vekilov, P.G., Muschol, M., Thomas, B.R., 1996. Nucleation and crystallization of globular proteins – what we know and what is missing. J. Cryst. Growth 168, 1–27.
Shi, D., El-Farra, N.H., Li, M., Mhaskar, P., Christofides, P.D., 2006. Predictive control of particle size distribution in particulate processes. Chem. Eng. Sci. 61, 268–281.
Shi, D., Mhaskar, P., El-Farra, N.H., Christofides, P.D., 2005. Predictive control of crystal size distribution in protein crystallization. Nanotechnology 16, S562–S574.
Snyder, M.A., Chatterjee, A., Vlachos, D.G., 2005. Net-event kinetic Monte Carlo for overcoming stiffness in spatially homogeneous and distributed systems. Comput. Chem. Eng. 29, 701–712.
Vekilov, P.G., 1993. Elementary processes of protein crystal growth. Prog. Cryst. Growth Charact. 26, 25–49.
Vekilov, P.G., Ataka, M., Katsura, T., 1993. Laser Michelson interferometry investigation of protein crystal growth. J. Cryst. Growth 130, 317–320.
Vekilov, P.G., Chernov, A.A., 2003. The physics of protein crystallization. Solid State Phys. 57, 1–147.
Wiencek, J.M., 1999. New strategies for protein crystal growth. Annu. Rev. Biomed. Eng. 1, 505–534.
Zhang, Y., Doherty, M.F., 2004. Simultaneous prediction of crystal shape and size for solution crystallization. AIChE J. 50, 2101–2112.