Available online at www.sciencedirect.com

Chemical Engineering Research and Design

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

Process modelling of protein crystallisation: A case study of lysozyme

Hamish M. Mitchell[a], *Derrick Jovannus*[a], *Ian Rosbottom*[a], *Frederik J. Link*[a], *Niall A. Mitchell*[b], *Jerry Y.Y. Heng*[a,c,*]

[a] *Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

[b] *Siemens Process Systems Engineering, London W6 7HA, United Kingdom*

[c] *Institutute for Molecular Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

ARTICLE INFO

Article history:
Received 4 October 2022
Received in revised form 11 February 2023
Accepted 13 February 2023
Available online 15 February 2023

Keywords:
Population balance modelling
Protein crystallisation
Lysozyme
Nucleation
Crystal growth
Parameter estimation

ABSTRACT

With the rise in interest of protein crystallisation as a purification step in downstream processing, there is significant interest in the process modelling of these crystallisation steps. Herein, we demonstrate and compare the applicability of "traditional" nucleation and growth models, commonly used to model small molecule crystallisation, for the successful population balance modelling of lysozyme crystallisation at the 100 mL and 1 L scales. Results show that both empirical power-law and first-principles models for nucleation and growth provide good fits to experimental data. Results from parameter estimation highlight a high degree of model sensitivity to initial guesses and stress the importance of providing particle size estimates in order to extract sensible data from the models. Estimates obtained for the 100 mL scale provided suitable initial guesses for the 1 L scale, despite significant differences in the final values obtained at each scale. For future work, further investigation into model validation upon scale-up is recommended. The work performed demonstrates the effectiveness of population balance modelling in the prediction of protein crystallisation behaviour, regardless of the underlying physical phenomena.

1. Introduction

The optimisation of upstream processing of biologically relevant protein materials has resulted in the economic bottleneck of biopharmaceutics being shifted to the downstream purification steps. Some studies estimate that up to 80% of the costs of biopharmaceutical product development and manufacture are associated with the purification steps, often related to chromatography (D'Souza et al., 2013).

Crystallisation remains the preferred purification step in the small molecule pharmaceutical industry, owing to both the physical stability and purity of the solids produced (Kirwan and Orella, 2002). However, the high molecular weight and degrees of freedom of protein molecules, as well as their tendency to have highly hydrated structures, (Chen et al., 2020) has resulted in the crystallisation of proteins

* Corresponding author at: Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom.
E-mail address: jerry.heng@imperial.ac.uk (J.Y.Y. Heng).

https://doi.org/10.1016/j.cherd.2023.02.016

predominantly for the purpose of structural determination (Chayen and Saridakis, 2008; Yang et al., 2019). However, recent advances in continuous processing (Chen et al., 2020; Yang et al., 2019; Kwon et al., 2014a; Neugebauer and Khinast, 2015; Pu and Hadinoto, 2020, 2021) and the utilisation of nucleation enhancing seeds, (Pu and Hadinoto, 2021; Kertis et al., 2012; Chen et al., 2021a) as well as soluble additives (Link and Heng, 2021) has pointed the way towards robust, reliable and scalable protein crystallisation. Furthermore, understanding about the role of pH in the crystallisation kinetics of proteins (Link and Heng, 2022) and the successful crystallisation of monoclonal antibodies (Chen et al., 2021b) has shown that protein crystallisation as a purification process has transitioned from a proof-of-concept to the beginning of industrial implementation (Roque et al., 2020). However, successful implementation is hindered by factors such as the complexity of the fermentation broths from which the protein is to be purified, as well as challenges in the successful scale-up of protein crystallisation from the nanolitre to the litre scale (Roque et al., 2020).

Process modelling utilises numerical models to predict material behaviour under unit processes associated with pharmaceutical manufacture, such as crystallisation, filtration, and granulation. Process models can effectively identify and be used to optimise the conditions for materials processing, allowing for efficient design of equipment, (Zobel-Roos et al., 2020; Lübbert and Simutis, 1994) as well as assisting scale-up of industrial processes, (Rosenbaum et al., 2019) thereby minimising material usage. The reduction of complex atomic and molecular level behaviour also allows the modelling of these phenomena over process relevant timescales, as compared to atomic and molecular scale modelling such as quantum mechanics (QM) or molecular dynamics (MD) (Paquet and Viktor, 2015). In turn, this can allow for model predictive control of pharmaceutical processes (Zobel-Roos et al., 2020). The computational expense of QM and MD of proteins is increased by the large size and complexity of the molecules, which leads to a degree of flexibility, restricting the feasible timescale for simulation (Condic-Jurkic et al., 2018). It is worth noting that the inverse is also true; process modelling cannot be said to accurately describe the molecular level phenomena occurring, and therefore does not provide a wholly accurate description of what is occurring within a crystallising system. However, the information provided by process models of crystallisation processes is invaluable in reducing material strain and allowing for more efficient design of experiments, which can then be used to further validate the models developed.

The process modelling of small molecule crystallisation has identified optimal experimental parameters, such as stirring speed, (Trampuž et al., 2021) reactor geometry, (Zhao et al., 2015) and temperature (Kwon et al., 2014b; Mitchell, 2012) to maximise the yield and particle properties of the crystallised material. First principles and empirical models of nucleation, growth, breakage, and seeding (amongst many others) have been utilised to create a more complete picture of the small molecule crystallisation process using population balance modelling, which can bring about the more accurate prediction of the particle properties of the material (Orehek et al., 2021; Gong et al., 2021; Szilágyi, 2021; Mozdzierz et al., 2021). However, there are comparatively fewer studies which examine the process modelling of protein crystallisation, with relevant literature only focusing on modelling the growth kinetics of proteins using process models (Liu et al., 2009, 2010a, 2010a, 2013; Zhou et al., 2022). Notably, while these studies are successfully implemented to predict the morphology of the lysozyme crystals used, they only implement semi-empirical growth rate expressions. There is little published research on implementing both nucleation and growth equations into the process modelling of protein crystallisation, (Sevilla et al., 2005) and relevant literature does not consider or discuss the choice of equation used.

Here, we examine and compare the feasibility of using both first principles and semi-empirical mechanistic models for both nucleation and crystal growth to model the isothermal crystallisation of lysozyme at the 100 mL and 1 L scales, using gPROMS FormulatedProducts. The concentration and particle size distribution are modelled, and the impact of different models and the initial guesses in the parameter estimation are discussed with respect to obtaining a model which can describe the system. This work will lead to improvements in the prediction of protein crystallisation at industrially relevant scales. The investigation into the transfer of kinetics across process scales is especially pertinent to industrial applications, as performing experiments at smaller scales provides superior material efficiency.

2. Background and theory

2.1. Population balance modelling

Rigorous process modelling of crystallisation processes is often performed using a governing population balance model, which describes the time evolution of the particle size distribution (PSD). The general one-dimensional population balance in the case of a batch system takes the form: (Myerson and Ginde, 2002)

$$\frac{\partial n(L,t)}{\partial t} + \frac{\partial (n(L,t)G)}{\partial L} = B - D \tag{1}$$

Here, $n(L,t)$ is the number density of crystals of a characteristic length L at time t. Crystal growth is accounted for by the linear growth rate, G, and the appearance and disappearance of crystals via processes such as primary and secondary nucleation, alongside agglomeration and dissolution, is accounted for by the birth and death terms, B and D, respectively. The population balance is accompanied by relevant heat and mass balances for both the liquid and crystalline phase, to bring about a complete picture of the crystallisation process. In the case of an isothermal system as studied herein, the heat balance can be disregarded. The accompanying mass balance assumes that the depletion of concentration is purely due to crystal growth, as the forming nuclei are too small to affect the concentration significantly.

$$\frac{dC}{dt} = -3k_v\rho_c \int_0^\infty Gn(L,t)L^2 dL \tag{2}$$

The concentration is dependent on k_v, the volumetric shape factor (assumed here to be 0.81, equivalent to a cubic crystal habit), and ρ_c, the crystal density, as well as the growth rate and number density of the crystals. To account for system hydrodynamics, these balances are often coupled with empirical equations to calculate relevant hydrodynamic parameters, or with Computational Fluid Dynamics (CFD), (Camacho Corzo et al., 2017) which can lead to a more accurate description of the hydrodynamics of the system, at the expense of increased computational complexity. In this case, the crystalliser is assumed to be well-mixed and

therefore treated as a lumped system, negating the need for CFD.

2.2. Primary nucleation kinetics

Multiple kinetic models are available for modelling primary nucleation, which consider not only the spontaneous formation of crystals from a supersaturated system, but also heterogeneous nucleation (occurring on foreign surfaces, such as insoluble impurities or the walls of the crystalliser). To model primary nucleation, the kinetics considered are the Classical Nucleation Theory and a power-law model.

2.2.1. Classical nucleation theory

The Classical Nucleation Theory (CNT) is one of the most widely used kinetic models for homogeneous nucleation and began with the work of Volmer and Weber (Volmer, 1926). The basic principle behind CNT is that the nucleation rate (the rate of formation of spherical crystal nuclei per unit volume) follows a relationship analogous to the Arrhenius equation:

$$J = A_J \exp\left(\frac{-16\pi\gamma^3 v^2}{3k^3T^3\ln^2(S)}\right) \tag{3}$$

The two kinetic parameters which govern the classical nucleation theory are the pre-exponential factor A, and the solid-liquid interfacial energy, γ, and these can be estimated using methodologies such as induction time analysis, (Kulkarni et al., 2013) as well as parameter estimation. The interfacial energy can also be estimated through other methods such as contact angle measurements (Good and Girifalco, 2002). S refers to the relative saturation of the system ($S = \frac{C}{C^*}$, where C and C^* denote the system concentration and the solubility, respectively).

2.2.2. Power law nucleation

While the CNT employs a solid thermodynamic basis for its derivation, the underlying assumptions of CNT do not hold true for many real-world systems. As a result, the CNT tends to mis-predict the true nucleation rate, especially in industrial processes. In the context of proteins, it has been shown that the CNT overestimates the nucleation rate of protein crystals by as many as ten orders of magnitude (Vekilov, 2010; Galkin and Vekilov, 1999a). As such, the CNT is often replaced by empirical power-law nucleation equations: (Myerson and Ginde, 2002)

$$J = k_J\left(\frac{C - C^*}{C^*}\right)^j = A_J \exp\left(-\frac{E_{A,J}}{RT}\right)(S-1)^j \tag{4}$$

The values of A_J, $E_{A,J}$, and j are empirical constants and are tuned to experimental data via methods such as parameter estimation. S refers to the relative supersaturation ($S = C/C^*$). This nucleation equation acts as a 'catch-all' and encompasses both homogeneous and heterogeneous primary nucleation (common on industrial scales) in a single term.

2.3. Growth kinetics

Four crystal growth models are available in gPROMS FormulatedProducts (Process Systems Enterprise Ltd, 2022). As an initial approach, the two models taken into further consideration were a two-step Mersmann kinetics, as well as a power law model.

2.3.1. Two-step growth

The Mersmann two-step growth model (Garside et al., 2002) is a common kinetic model for describing crystal growth. The first step described in the model is the mass transfer of solute from the bulk solution ($C_{bulk} = C$) to the crystal surface (C_{int}), the kinetics of which are given by Eq. 5:

$$G(L) = k_d(L)\left[\frac{C_{bulk} - C_{int}(L)}{\rho_c}\right] \tag{5}$$

The size-dependent mass transfer coefficient (k_d) of this mass transfer step is difficult to determine analytically, and so is calculated using a Sherwood correlation: (Garside et al., 2002)

$$k_d(L) = \frac{D_{AB}}{L}\left[2 + 0.8\left(\frac{\bar{\varepsilon}L^4}{\upsilon_L^3}\right)^{\frac{1}{5}}\left(\frac{\upsilon_L}{D_{AB}}\right)^{\frac{1}{3}}\right] \tag{6}$$

This correlation for the mass transfer coefficient is dependent on the kinematic viscosity (υ_L), the energy dissipation rate ($\bar{\varepsilon}$), and the diffusion coefficient (D_{AB}), which can be estimated as: (Einstein, 1905)

$$D_{AB} = \alpha\frac{kT}{6\pi\eta\frac{d_m}{2}} \tag{7}$$

The diffusion coefficient is therefore a function of the temperature, T, the dynamic viscosity of the liquid, η, and the molecular diameter, d_m. As this equation is an approximation for the diffusion coefficient, is it multiplied by a correction factor, α.

The second step corresponds to surface integration of the solute on the crystal surface into the crystal lattice. This step accounts for mechanisms such as surface diffusion and orientation of the solute molecule, as well as the integration step, and the kinetics of this are given by Eq. 8.

$$G(L) = A_G \exp\left(\frac{-E_{A,G}}{RT}\right)\left[\frac{C_{int}(L) - C^*}{\rho_c}\right] \tag{8}$$

For undersaturated solutions, surface integration does not occur, and hence the rate determining step shifts to mass transfer. To account for this, both equations are modified by the simplification $C_{int}(L) = C_{sat}$, which effectively sets the growth rate of the integration step to zero, and allows for a negative growth rate (i.e., dissolution) in Eq. 5.

2.3.2. Power law growth

As with nucleation, a power law growth model is also proposed for crystal growth, which is akin to the surface integration step and analogous to the power-law model for primary nucleation (Myerson and Ginde, 2002):

$$G = k_g(S-1)^g = A_G \exp\left(\frac{-E_{A,G}}{RT}\right)(S-1)^g \tag{9}$$

As this equation is empirical in nature, the constants A_G, $E_{A,G}$ and g do not convey much physical significance but are instead fitted to experimental data. However, Mullin hypothesized that the value of $E_{A,g}$ can provide information on the controlling mechanism of crystal growth – a value of 10–20 kJ mol^{-1} indicates surface diffusion-controlled growth, whereas a value closer to 40–60 kJ mol^{-1} indicates surface integration as the rate limiting step (Mullin, 2001).

Table 1 – Sources for relevant thermophysical data for the liquid and crystal phases.

Species	Density (kg m^{-3})	Specific Heat Capacity (J kg^{-1} K^{-1})
Crystallisation	1072 (Leung et al., 1999)	1398 (Gómez et al., 1995)
Lysozyme crystal	1240 (Leung et al., 1999)	1750 (Imaizumi et al., 1979)
Sodium Acetate Buffer	1026 (Kharat, 2008)	3000 (Araki et al., 1995)
Sodium Chloride	1012 (Green and Perry, 2008)	864 (National Instutite of Standards and Technology, 1998)
Water	998 (National Instutite of Standards and Technology, 2022)	4184 (National Instutite of Standards and Technology, 2022)

3. Methodology

3.1. *Thermophysical data*

The thermophysical data required for the successful modelling of lysozyme crystallisation in gPROMS FormulatedProducts includes molecular weight, density, and specific heat capacity, for both the liquid and crystal phase, and the solubility, which is the thermodynamic factor underpinning crystallisation. Sources for most of the relevant thermophysical data are given in Table 1. The enthalpy of crystallisation was found to be – 13.2 kcal mol^{-1} under similar process conditions (Schall et al., 1996). As the experimental data collected under isothermal conditions, the specific heat capacities and enthalpy of crystallisation are not wholly necessary but included for completeness.

As the liquid phase is composed of four different species, ideal combining rules were used to calculate relevant parameters within gPROMS. The dynamic viscosity of the mixture was assumed constant and identical to the dynamic viscosity of water at 20°C (~10^{-3} Pa s), as the liquid phase is primarily water.

For proteins, the solubility is a function of various system properties such as temperature, pH, and precipitant concentration. As such, solubility data for lysozyme at process relevant conditions was sourced from literature (Cacioppo and Pusey, 1991). Details of the solubility values used are given in Section 3.3.

3.2. *Process flowsheet*

To model the batch crystallisation of lysozyme in a stirred vessel, a relatively simple flowsheet was assembled, as shown in Fig. 1, utilising gPROMS FormulatedProducts 2.3.1.

The flowsheet used consists of a global specifications block (Specifications), which allows for input of the major components of the system and their molecular weights, as well as relevant thermophysical data.

The main block in the flowsheet is the mixed-solids mixed-product removal crystalliser (MSMPR), which is where relevant hydrodynamics of the system are specified, such as equipment volume, impeller diameter, and impeller speed. The MSMPR crystalliser also allows for the input of crystallisation mechanisms, including both primary and secondary nucleation, alongside crystal growth. There is also the option to include advanced crystallisation mechanisms, such as impurity inclusion and crystal agglomeration. As a first approximation and for the purposes of this work, secondary nucleation and agglomeration kinetics were set to inactive, as well as the effects of impurities.

Finally, the MSMPR requires input of the initial conditions within the crystalliser (e.g. initial amount of liquid/solids in the vessel, temperature, liquid composition). To model specifically batch crystallisation, the liquid and solid inflow and outflow rates were set to zero. In order to monitor and control crystallisation, there are three sensors attached to the MSMPR. Firstly, there is a temperature sensor (TC01), which allow for temperature control within the crystalliser. There is then a liquid composition sensor (CT01) and a particle size sensor (PT01), which allows for liquid phase concentration measurements and PSD measurements within the crystalliser, respectively.

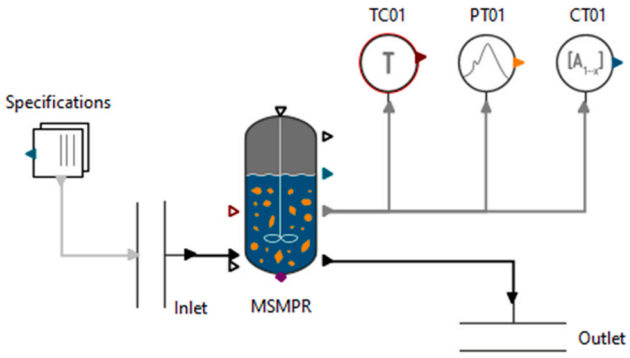

Fig. 1 – Flowsheet of the batch crystallisation process as constructed in gPROMS FormulatedProducts 2.3.1.

3.3. *Experimental data*

To both assess model accuracy and perform parameter estimation, it is necessary to obtain a trajectory of the liquid phase concentration of lysozyme as a function of time. As well as this, it is desirable to have an estimate for the mean particle size of the crystals formed, an important quality attribute for crystallisation to assess the relative rates of nucleation and growth. The PSD is also usually a critical quality attribute for the final product from crystallisation steps, at it determines mechanical behaviour of the particles and therefore the flowability and tabletability of the powder, (Li et al., 2004) as well as the dissolution rate of API crystals (Sun et al., 2012).

For this, the work of Tang et al (Tang et al., 2018). was used, where they assessed the influence of parameters such as stirrer speed and cooling rate on the crystallisation of lysozyme. They reported the liquid phase concentration profile for the crystallisation of lysozyme at the 100 mL scale, alongside the PSD at various points in time throughout. This data was reported for various cooling rates between 0 and ~0.1 °C min^{-1}, cooling from 20 °C and 0 °C. For this work, the experimental data used was for the cooling rate of 0°C min^{-1} (i.e., thermostatic crystallisation). The relevant concentration profile, alongside initial concentrations of lysozyme, sodium

chloride, and sodium acetate, as well as pH and temperature, were sourced from the methodologies provided.

To ensure the efficacy of the model in predicting scale-up, concentration data were also taken from the work of Smejkal et al., (Smejkal et al., 2013) who assessed the efficacy of scaling up of protein crystallisation processes (specifically, *Gallus gallus* lysozyme and Canakinumab Fab-fragment) using the maximum local energy dissipation as a scaleup rule. They tracked the concentration of lysozyme with time to assess the impact of agitation at the 1 L scale. The relevant concentration profile, alongside initial conditions, were again sourced from their methodologies.

It is worth noting that both studies used discuss the extent of agglomeration or aggregation of lysozyme crystals during crystallisation. For the experiments performed at the 100 mL scale, micrograph images show that higher cooling rates of the crystallisation solution result in larger extents of aggregation of lysozyme crystals. This is reinforced by PSDs at different cooling rates, which show that at high cooling rates, the average particle size is significantly increased, as well as the distribution being much broader (Tang et al., 2018). As only the thermostatic data was used for parameter estimation, the degree of agglomeration is not high, as evidenced by micrograph images. Similarly, experimental data at the 1 L scale shows that the degree of agglomeration increases with decreasing stirrer speed (Smejkal et al., 2013). As such, only data at the highest stirrer speed was used, for which micrograph images again show that the degree of agglomeration is low.

As well as concentration data, the MSMPR module also requires estimates for key equipment specifications, namely the volume, impeller diameter, and stirrer speed. For the work performed at the 100 mL scale, the stirrer speed was given as 210 rpm, and the impeller diameter was estimated to be around 10 mm. For the litre-scale crystallisation, the impeller diameter and stirring speed was given as 24.8 mm and 200 rpm, respectively. Both the impeller power number and pumping number (required by gPROMS) were kept at their default value of 0.5 and 0.8, respectively. Experiments at the 100 mL scale were conducted in a sodium acetate buffer at pH 4.5 and 20°C, with 5% w/v NaCl as the precipitant. At these conditions, the solubility was evaluated as 2.16 mg mL^{-1} (Cacioppo and Pusey, 1991). For data at the 1 L scale, experiments were performed in a sodium acetate buffer with a pH of 4 and 20°C, with 4% w/v NaCl as the precipitant, giving a solubility value of 2.47 mg mL^{-1} (Cacioppo and Pusey, 1991).

3.4. *Parameter estimation*

Parameter estimation was performed using the built-in parameter estimation capabilities of FormulatedProducts. The experimental data shown earlier was used for parameter estimation. For experimental measurements of both particle size and liquid concentration, a constant variance model was assumed for the sensors (0.5 µm and 0.1 g L^{-1}, respectively). A list of the parameters required to be estimated is given in Table 2. Parameter estimation is a necessary step as the nucleation and growth kinetics of the system are highly dependent on process conditions. As well as this, complete sets nucleation and growth parameters are scarcely reported in literature for lysozyme and show limited transferability between systems.

Table 2 – The list of parameters to be estimated as part of this study.

Parameter	Symbol	Units
Nucleation pre-exponential factor	A_J	(no. of nuclei) m^{-3} s^{-1}
Nucleation activation energy	$E_{A,J}$	J mol^{-1}
Nucleation exponent	j	-
Surface energy	γ	mJ m^{-2}
Growth pre-exponential factor	A_G	µm min^{-1}
Growth activation energy	$E_{A,G}$	J mol^{-1}
Growth exponent	g	-
Effective diffusivity correction factor	α	-

The DAE solver and MINLP solver (required for solving systems of differential and algebraic equations, and nonlinear programs, respectively) were kept at their default choices of DAEBDF and NLPSQP. It is worth noting that NLPSQP is not necessarily guaranteed to find a globally optimal solution for parameter estimation; a solver such as NLPMSO, which implements multi-start optimisation to attempt to find a global solution, may be a more robust choice. However, multi-start optimisation is also more likely to result in solver errors depending on start-point, as well as increasing the computational time significantly. As this paper is focused on the proof of concept of modelling protein crystallisation, NLPMSO was not explored in further detail.

Goodness-of-fit was measured using the objective function used for parameter estimation in gPROMS, which is the log likelihood function for a Gaussian distribution, and is given as:

$$\Phi = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\min_{\theta}\left\{\sum_{i=1}^{N_E}\sum_{j=1}^{N_{V_i}}\sum_{k=1}^{N_{M_{ij}}}\ln\left(\sigma_{ijk}^2\right) + \frac{\left(\bar{z}_{ijk} - z_{ijk}\right)^2}{\sigma_{ijk}^2}\right\} \quad (10)$$

Here, N denotes the total number of measurements taken across all experiments, θ is the set of parameters to be estimated, and subscripts i, j, and k denote the experiment, variable, and measurement number, with $(1,1,1) \le (i, j, k) \le (N_E, N_{V_i}, N_{M_{ij}})$. $\bar{z}_{ijk}$ and σ_{ijk}^2 denote the measured value and variance, respectively, while z_{ijk} denotes the model prediction.

As there are two nucleation and two growth models, this gave a total of four different kinetic models for the system. Each of these cases underwent parameter estimation within gPROMS to obtain estimates for the nucleation and growth parameters. For future reference, the possible combinations are referred to as Case 1 (CNT + two-step growth), Case 2 (Power-law nucleation and growth), Case 3 (CNT + Power-law growth), and Case 4 (Power law nucleation + two-step growth).

4. Results and discussion

4.1. *Initial approaches*

As a first approach, classical nucleation and growth models were tested (Case 1). However, it was found that simulations failed to converge to any solution, reasoned to be due to the diffusion parameter within the two-step growth kinetics. As lysozyme is a large molecule (with a molecular weight of 14.3 kDa), the equation used to calculate the diffusion coefficient D_{AB} (Eq. 9) was likely not applicable, resulting in the

Table 3 – Units and estimated values/bounds for nucleation and growth rate parameters for initial trials. [a]These values were sourced from the default bounds in gPROMS.

Parameter	Units	Value/Bounds	Ref.
$\ln(A_J)$	$\ln(\#\ m^{-3}\ s^{-1})$	0–100[a]	-
$E_{A,J}$	J mol^{-1}	1000–1500	(Su et al., 2017)
j	-	1–10[a]	-
γ_{SL}	mJ m^{-2}	0.64	(Dimitrov et al., 2015)
A_G	nm min^{-1}	600–1500	(Dimitrov et al., 2015;Darcy and Wiencek, 1998;Liu et al., 2010b)
$E_{A,G}$	J mol^{-1}	0	(Su et al., 2017)
α	-	0–10[a]	-
g	-	1–3	(Darcy and Wiencek, 1998)

failure of simulations. As a result, the cases using first principles growth mechanisms (Cases 1 and 4) were disregarded during initial trials.

Where possible, initial guesses for the relevant nucleation and growth parameters (such as pre-exponential factor, activation energy, and supersaturation order) were sourced from literature, and suitable bounds were set around these initial guesses for parameter estimation. In the case that either no single value was given for a parameter, or no such data was available from literature, a value within either the ranges given or the default limits in gPROMS was used as an initial guess. Due to the novelty of this work, some of the parameters required were not available for lysozyme, and as such, were sourced from other systems, such as glycine crystallisation (Su et al., 2017). While the kinetics for glycine crystallisation are likely to be quite different to those of lysozyme, the estimates obtained were nevertheless useful in providing an order-of-magnitude initial guess. A list of these initial guesses is given in Table 3.

As an alternative approach to first principles models, a power-law equation was initially employed to model both the nucleation and growth rates (Case 2). The temperature dependence term ($E_{A,G}$) in the growth rate was set at zero, as there is no temperature change in the system and therefore a single temperature independent variable (A_G) would suffice. However, when analysing the results of this parameter estimation, it was found that that there was minimal convergence to experimental data, as shown in Fig. 2. Through adjustments to the kinetic parameters of this model, it was found that the growth rate constant (A_G) and nucleation activation energy ($E_{A,J}$) had the most significant impact on the model. To assess their impact, the initial guesses for each of these variables were manually changed, to quantify their effect on the model solution. The results of this can be seen in Table 4. All other variables were kept at their default initial guesses and bounds.

As shown in Fig. 2 and Table 4, it was found that iteration 3 gave the poorest fit to experimental data, as the heightened value of the nucleation rate constant gave rise to instant nucleation, resulting in an almost instantaneous drop in lysozyme concentration. The other three iterations (1, 2, and 4) resulted in better fits to experimental data, with iteration 4 giving a marginally better fit to the experimental data when compared to iterations 1 and 2, as judged by the values of the objective function. It was noticed, however, that the predicted initial concentration (at t = 0 min) was lower than expected for all trials, which was reasoned to be due to the presence of the salts and buffer, increasing the density and lowering the concentration of lysozyme in the initial solution. For subsequent parameter estimations, the initial

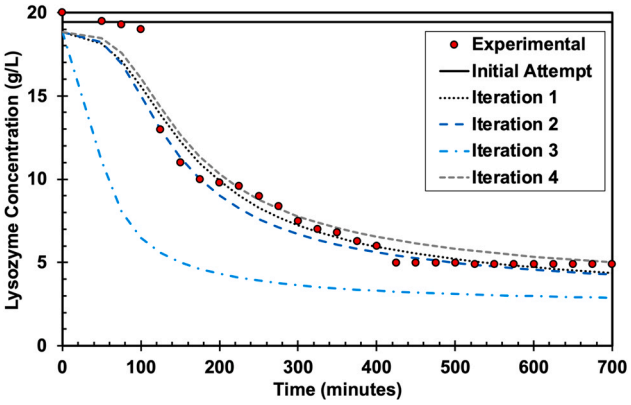

Fig. 2 – Concentration trajectories predicted by each iteration produced by manual tweaking of initial guesses and bounds.

Table 4 – Effects of initial guesses on final values of nucleation activation energy and growth rate constant, as well as the log-likelihood function of parameter estimates and the estimate obtained for the final volume-weighted mean particle size, $D_{4,3}$.

Iteration Number		1	2	3	4
Initial Guess	A_G (nm min^{-1})	600	2000	2000	100,000
	$E_{A,J}$ (J mol^{-1})	1000	1000	5000	1000
Final Value	A_G (nm min^{-1})	1000	2670	3220	26,800
	$E_{A,J}$ (J mol^{-1})	2130	3.49	6740	689
Φ / 10^6		8.28	8.28	43.0	5.76
$D_{4,3}$ (nm)		5.16	11.4	3.63	21.2

lysozyme concentration was adjusted in gPROMS until it matched that of experimental data.

Despite providing a good fit to concentration data, the crystal size in all iterations was still very small (in the range of 3.63–21.22 nm), indicating that the results of parameter estimation overpredicted the nucleation rate when compared to the growth rate. As a comparison, the hydrodynamic diameter for a single lysozyme molecule is around 3.5 nm, (Weichsel et al., 2017) indicating that the crystals formed would contain fewer than ten lysozyme molecules at the end of crystallisation experiments, which is highly unlikely to be feasible.

4.2. *Hybrid modelling approach*

The next kinetic model attempted was a hybrid combination of classical nucleation kinetics and a power law growth model (Case 3).

Table 5 – Initial guesses, final values, and parameter estimation bounds for initial classical nucleation parameter estimation.

Parameter	Units	Initial	Final	Bounds
$\ln(A_J)$	$\ln(\#\ m^{-3}\ s^{-1})$	40	48	0–100
γ	$mJ\ m^{-2}$	0.1	0.2	0.001–1000
A_G	$nm\ min^{-1}$	600	3	0–1000
$E_{A,G}$	$J\ mol^{-1}$	0	0	0
g	-	1.5	1	1–3

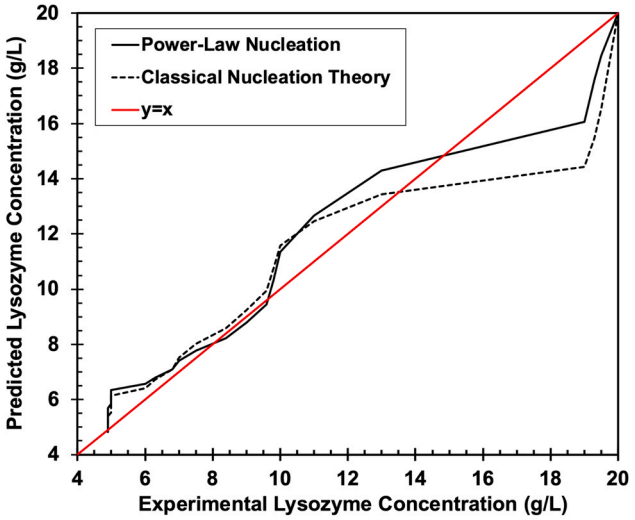

Fig. 3 – Parity plot comparing choice of nucleation equation (Classical Nucleation Theory vs power-law) with a power-law growth model.

Initial guesses, bounds, and the final values from parameter estimation of this nucleation and growth model are given in Table 5. From this, it was found that the combination of classical nucleation kinetics and power-law growth modelling provided a similarly reasonable fit to experimental data compared to the power-law nucleation model used previously. However, while the overall fit was comparable, this combination of kinetics led to a less than satisfactory prediction of the onset of nucleation, which is a crucial aspect of crystallisation processes. This is shown visually in F. At higher experimental concentrations (i.e., closer to the onset of nucleation) the CNT provides a poorer fit to experimental data. As a result, the power law model was further analysed (Case 2). As with the previous power-law models, the particle size was still predicted to be very small (in the order of nanometres) despite the good fit to concentration data. (Fig. 3)

4.3. *Improving particle size estimation*

As previously noted, the results of parameter estimation were sensitive to the initial guesses of the kinetic parameters. The average particle size was overly underpredicted by the models used, indicating that the nucleation rate predicted by the models is far too high, and as a result, the crystals do not grow to substantial sizes due to the number of nuclei formed. Without an estimate for the end-point particle size, it is difficult for parameter estimation to give an accurate representation of the relative effects of nucleation and growth. To take this into account, an estimate for the final particle size was included in the experimental data from which parameter estimation was performed, to ensure adequate growth rates within the model. This estimate was taken again from the work of Tang et al., (Tang et al., 2018) who obtained average particle sizes of around 10 µm at a timepoint of 200 min. Therefore, a representative value of 10 µm was taken as an estimate for the median particle size (D_{50}) at 200 min.

It was also observed that for the full ranges of data used, the onset of nucleation was the most critical part of the experimental data to be modelled, as it exhibits the largest drop in concentration and corresponds to the point of formation of nuclei. As a result, parameter estimation was repeated using only the first 200 min of experimental data, to improve the prediction of the onset of nucleation. Assuming a representative value of 10 µm for the mean particle size at the onset of nucleation, the predicted particle side was around 2 µm, which is a much closer prediction than previous models, and lies around the same order of magnitude. This result stresses the importance of including measurements for the PSD, to accurately capture the relative magnitudes of nucleation and growth rates using population balance modelling.

4.4. *Final models*

With the model now providing improved estimates for the final particle size, the power-law model was revisited with the refined initial guesses for parameter estimates, and further tuned with all available D_{50} values to account for further growth of the crystals beyond the first 200 min. The D_{50} values were used instead of the full PSDs as the PSDs were shown to be unimodal, and the shape of the PSD was nearly identical at all time points (Tang et al., 2018). The main difficulty at this stage was that the parameter estimates were exceptionally sensitive to initial guesses, likely due to the presence of local optima in the system due to the complexity of the model. At this stage, it was noticed that the temperature-dependent activation energy parameters ($E_{A,J}$ and $E_{A,G}$) were highly correlated with their corresponding pre-exponential factors. This was reasoned to be due to the experimental data being at isothermal conditions, meaning that parameter estimation would be unable to extract a temperature dependence on nucleation or growth. To simplify the systems, both of these parameters were therefore fixed at zero.

As can be seen in Fig. 4, all models provided an almost identical fit to experimental data. To ascertain the best fitting model, the χ^2 values for each combination of kinetics were evaluated and compared (shown in Table 6). As can be seen from the χ^2 values for both concentration and particle size, the order of best fit is Case 2 > Case 4 > Case 1 > Case 3. However, the differences in χ^2 value and predicted final particle size are minimal. As such, it would be reasonable to conclude that all models tested provide a good fit to experimental data, as well as providing a reasonable estimate for the final particle size. The final parameter estimation results for each of the cases analysed are presented in Table 7.

When comparing the parameter estimates for the nucleation and growth models, there are a few conclusions that can be drawn. Firstly, the power-law nucleation model has significantly larger confidence intervals than the Classical Nucleation Theory, and the use of power-law nucleation induces larger confidence intervals for the growth model used.

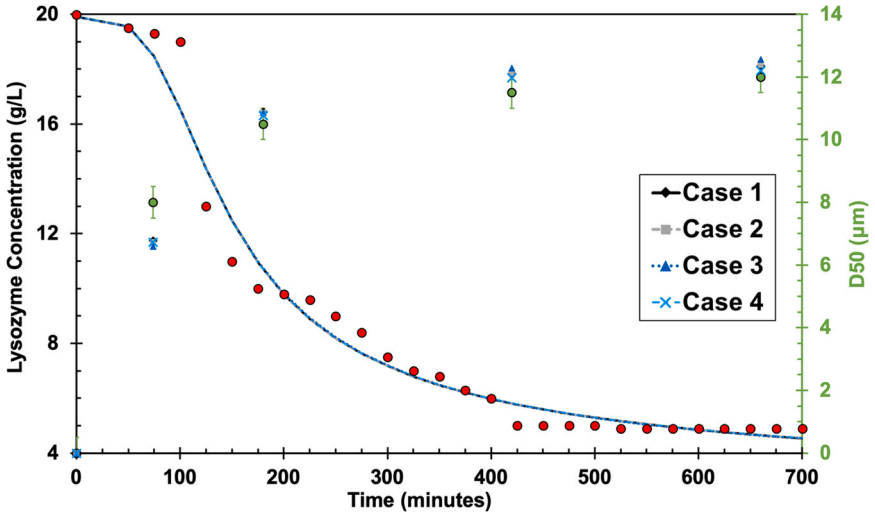

Fig. 4 – Concentration trajectories for the four final models. Experimental data for concentration (red dots) and D_{50} (green dots) provided as a reference.

Table 6 – Comparison of goodness-of-fit (as measured by chi-squared testing) and predicted end-point particle sizes for all four model combinations tested.

Case	1	2	3	4
Concentration χ^2	1520.8	1520.2	1532.3	1520.8
D_{50} χ^2	9.08	8.58	12.1	7.92
Predicted $D_{50,\ final}$ / µm	12.4	12.4	12.6	12.2
Actual $D_{50,\ final}$ / µm	12 ± 0.5			

Analysis of the correlation matrices between parameters within gPROMS revealed that power-law nucleation models suffered from a high degree of correlation (> 99%) not only between the two nucleation parameters, but also between the supersaturation order and both growth parameters, regardless of growth model used. Both nucleation parameters, $\ln(A_j)$ and j, had individual 95% t-values higher than the reference value, indicating that there was not sufficient experimental data to estimate these values accurately. Classical Nucleation Theory did not have the same issue – despite high correlation (> 99%) between nucleation parameters (which explains the large confidence intervals for nucleation parameters), there was lower correlation between nucleation and growth parameters. which also explains the smaller confidence intervals for growth parameters. As such, while all four combinations provide near-identical goodness-of-fit, the models utilising Classical Nucleation Theory may be more reliable at this stage.

This result is interesting as the Classical Nucleation Theory has been shown to be a poor predictor of protein nucleation experimentally, and it was expected that power-law nucleation models would not only provide a better prediction but act also as a 'catch-all' nucleation equation, accounting for other nucleation phenomena such as heterogeneous nucleation and secondary nucleation. As the systems studied are both stirred systems, it would also be interesting to consider the effects of other nucleation mechanisms, such as secondary nucleation via attrition due to crystal-impeller and crystal-crystal collision. The consideration of stirring is especially important for proteins, as they are often sensitive to shear and can easily denature, which would perhaps also need to be included. However, despite these phenomena not being considered, the population balance model implanted in gPROMS was able to give a reasonable estimate for both solution concentration and particle size.

In terms of choice of growth equation, both power-law and classical growth equations gave similar estimates for both the pre-exponential factor and supersaturation exponent. The confidence intervals are much more reasonable than those obtained for the nucleation parameters, indicating that both models are suitable for parameter estimation. However, when comparing correlation matrices for Cases 1 and 3 (given in Table 8), the use of a classical growth equation results in higher correlation between all parameters. As well as this, the confidence intervals for all parameters, especially the interfacial energy, are significantly improved when using a power-law growth equation, as can be seen in Table 7.

Finally, reported literature values for the effective interfacial energy of lysozyme crystals range from 0.11 mJ m^{-2} to 0.64 mJ m^{-2} (Dimitrov et al., 2015; Galkin and Vekilov, 1999b, 1999c; Lin et al., 2017). These values lie within the bounds of parameter estimation, and indicates that while not identical, the values obtained as part of this work are in good agreement with other experimental data, despite the large confidence intervals around these parameter estimates.

4.5. *Process scaleup*

To assess model robustness, the final values for the model of Case 3 at the 100 mL scale were used as the starting values for the data taken for the 1 L scale. As explained previously, while all models provided similar goodness-of-fit to experimental data, the associated confidence intervals and correlation coefficients for power-law nucleation models was higher than that of the Classical Nucleation Theory, and the use of a power-law growth equation resulted in more statistically significant estimates for kinetic parameters. The kinetics obtained from previous parameter estimation gave a much poorer fit to experimental data, as can be seen Fig. 5.

Despite a poor fit to experimental data, the predicted particle size was around 6 µm, which is a much more representative estimate than those obtained during initial trials at the 100 mL scale. By using these values as a starting point for parameter estimation, it was possible to obtain

Table 7 – Final values for parameter estimates obtained for all four cases studied. [a]Denotes that this value reached the stipulated maximum in gPROMS. Uncertainty (±) in values is characterised by the 95% confidence intervals. Values highlighted in bold indicate that the 95% t-value was lower than the individual reference 95% t-value.

Case	Nucleation			Growth		
	$\ln(J_0)$	γ	I	k_g	j	α
	ln (# m^{-3} s^{-1})	mJ m^{-2}	-	nm min^{-1}	-	-
1	26.3 ± 11.7	**0.528 ± 0.640**	-	0.373 ± 0.057	2.43 ± 0.11	10[a]
2	**18.9 ± 19.9**	-	**2.14 ± 10.3**	0.367 ± 0.069	2.45 ± 0.14	-
3	32.8 ± 17.9	**0.766 ± 0.469**	-	0.394 ± 0.042	2.39 ± 0.08	-
4	**19.3 ± 19.8**	-	**1.96 ± 10.3**	0.360 ± 0.071	2.45 ± 0.15	10[a]

Table 8 – Correlation matrices for both models utilising classical nucleation theory. Bottom left: CNT + two-step growth. Top right: CNT + power-law growth. Correlation coefficients are calculated as $\rho_{xy} = \sigma_{xy}^2 / \sqrt{\sigma_{xx}^2 \sigma_{yy}^2}$.

Parameter	k_g	j	$\ln(J_0)$	γ
k_g		-0.893	0.815	0.819
j	-0.949		-0.946	-0.947
$\ln(J_0)$	0.910	-0.974		0.999
γ	0.914	-0.974	-0.999	

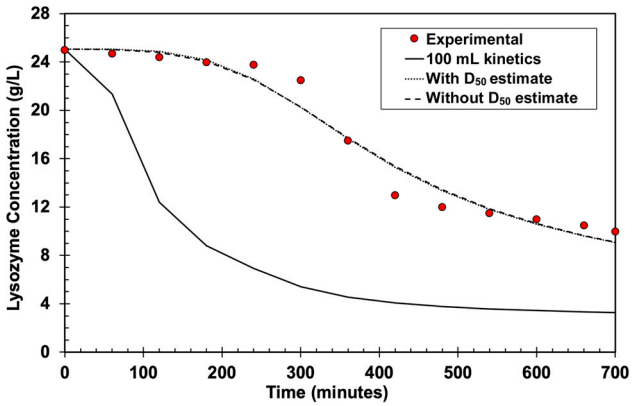

Fig. 5 – Comparison of the predictions for concentration profile at the 1 L scale.

revised parameter estimates and model fits for the experimental data at the 1 L scale, with and without an estimate of the D_{50}. It should be noted that no values for particle size were given for the concentration data used, and as such the particle size was estimated to be 10 µm from micrographs given by the authors. The results of this analysis are presented in Table 9 and Fig. 5.

It is evident that the values obtained for the kinetics at the 1 L scale differ significantly from those at the 100 mL scale. This suggests that the parameters may be overfitted to the experimental data at the 100 mL scale. The lack of estimates for particle size result in much larger confidence intervals in all kinetic parameters, likely due to the lack of particle size data, and as such the results obtained cannot be said to be statistically significant. Despite this, it was observed that this combination of kinetics resulted in a much better fit to concentration data (as shown in Fig. 5) and produced a reasonable estimate for the final D_{50}. Without an estimate for the D_{50} as an input to parameter estimation, the model appeared to underpredict the growth rate.

It is worth noting that while the crystallisation liquor is similar in both sources used, there are likely to be large discrepancies between experimental setups in terms of solution preparation and crystalliser hydrodynamics, amongst other factors, as the experiments were carried out by different research groups. It has also been shown that the solution pH is an important factor in the kinetics of protein crystallisation (Link and Heng, 2022; Zhang et al., 2013). This may partially explain the discrepancies in terms of parameter estimates and the lack of transferability between scales and highlights the challenges in reproducing protein crystallisation data. As well as this, the concentration profile for the 1 L crystallisation appears to exhibit very slow desupersaturation – at the 700-minute mark, the relative supersaturation is around 4, which should still drive both nucleation and crystal growth. The relatively slow desupersaturation may be due to the presence of impurities, retarding the growth of lysozyme crystals, or simply due to the slow nature of protein crystal growth.

As mentioned previously, it is likely that other phenomena (such as attrition and agglomeration) are present to some degree in both the 100 mL and the 1 L systems, and as such it may be differences in the extent of these phenomena which creates such discrepancies. It would also be interesting to consider alternative nucleation theories such as two-step nucleation or further semi-empirical adaptations to the conventional nucleation models used, as this may provide a better estimate for the onset of nucleation. To improve model robustness, it would be useful to include further sets of experimental data at multiple different temperatures, concentrations, and scales, amongst other factors. This will hopefully lead to the production of a more robust set of kinetics which can be applied more generally to lysozyme crystallisation. The derivation of robust kinetic parameters for lysozyme crystallisation allows for the digital design of more complex crystallisers, such as plug flow crystallisers or MSMPR cascades, which could then potentially be optimised to minimise solvent use and batch time, reducing both environmental and economic impacts.

It is hoped that estimates for kinetic parameters obtained as part of this work will provide useful starting points for the population balance modelling of protein crystallisation, not just for lysozyme but also other crystallisable proteins. Ultimately, this work contributes towards the successful adoption of crystallisation as a purification step for biopharmaceuticals. It would be interesting to assess the effects of further scale-up on the kinetic estimates obtained.

Table 9 – Comparison of kinetic parameters obtained from the 100 mL scale with the results of parameter estimation at the 1 L scale, with and without an estimate for the D_{50}. Values highlighted in bold indicate that the 95% t-value was lower than the individual reference 95% t-value.

Parameter	Units	Initial Value	Without D_{50}	With D_{50}
$\ln(A_J)$	$\ln(\#\ m^{-3}\ s^{-1})$	32.8 ± 17.9	**26.9 ± 57.8**	**26.2 ± 49.7**
γ	$mJ\ m^{-2}$	**0.766 ± 0.469**	**0.506 ± 1.33**	**0.636 ± 2.33**
A_G	$nm\ min^{-1}$	0.394 ± 0.042	**0.104 ± 6.14**	**0.264 ± 0.205**
g	-	2.39 ± 0.08	**2.12 ± 1.88**	2.15 ± 0.425
Experimental D_{50}	μm	10 ± 1		
Predicted D_{50}	μm	5.75	3.83	10.2
Concentration χ^2	-	1.49×10^5	1650	1600

5. Conclusions

Herein, it has been demonstrated that population balance modelling, alongside "conventional" nucleation and growth kinetics, is able to provide an accurate depiction of protein crystallisation, using lysozyme as a model protein. This work acts as a proof-of-concept for the application of process modelling to protein crystallisation, and in particular the usage and comparison of conventional nucleation and growth models to describe protein crystallisation. The results obtained herein not only demonstrate the applicability of PBMs to proteins, but also provide insight into the rational choice of nucleation and growth equations for modelling protein crystallisation, reinforced by statistical analysis. All combinations of models tested (both first principles and empirical power-law) provided reasonable fits to experimental data, and results collected at 100 mL scales were able to be applied as initial guesses to separate experiments from different authors at the litre scale. The usage of the Classical Nucleation theory resulted in smaller confidence intervals and reduced correlation between all kinetic parameters, and as such may be more suitable for use for the modelling of lysozyme crystallisation. The values obtained for the thermodynamic parameter γ are consistent with those obtained elsewhere in literature. This work also stresses the importance of obtaining at least a representative value for the final PSD, as without this, parameter estimation has the tendency to underpredict the growth rate.

The parameter estimation solvers used within gPROMS was noted to be very sensitive to initial guesses. As such, it is important to perform multiple sets of parameter estimation with different initial guesses, to be surer that the set of kinetics obtained are likely to be the result of a global optimum, instead of local. It would also be possible to consider a different solver selection – a global solver or multi-start optimisation solver may provide a more rigorous output. Finally, the authors recommend further usage of experimental data over a range of supersaturations and temperatures, to reduce correlation between kinetic parameters. It is believed that the findings presented here are valuable for the implementation of protein crystallisation as a purification strategy, and that process modelling will aid in the design and prediction of industrial protein crystallisation processes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge the UKRI's Engineering and Physical Science Research Council for funding this work (EP/N015916/1). H.M. gratefully acknowledges financial support from Eli Lilly and Company and the Engineering and Physical Sciences Research Council of the UK via Prosperity Partnership (grant numbers EP/T005556/1 and EP/T518207/1). F.L. acknowledges the Department of Chemical Engineering at Imperial College London for a PhD Departmental Scholarship.

References

Araki, N., Futamura, M., Makino, A., Shibata, H., 1995. Measurements of thermophysical properties of sodium acetate hydrate. Int. J. Thermophys. 1995 16 6 (16), 1455–1466.

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.

Camacho Corzo, D.M., et al., 2017. Crystallisation Route Map. Engineering Crystallography: From Molecule to Crystal to Functional Form. NATO Science for Peace and Security Series A: Chemistry and Biology. Springer, Dordrecht, pp. 179–213. https://doi.org/10.1007/978-94-024-1117-1_11

Chayen, N.E., Saridakis, E., 2008. Protein crystallization: from purified protein to diffraction-quality crystal. Nat. Methods 5, 147–153.

Chen, W., Yang, H., Heng, J.Y.Y., 2020. Chapter 10: Continuous Protein Crystallization. The Handbook of Continuous Crystallization. Royal Society of Chemistry, pp. 372–392. https://doi.org/10.1039/9781788013581-00372

Chen, W., et al., 2021a. Protein purification with nanoparticle-enhanced crystallisation. Sep. Purif. Technol. 255, 117384.

Chen, W., et al., 2021b. Biopurification of monoclonal antibody (mAb) through crystallisation. Sep. Purif. Technol. 263, 118358.

Condic-Jurkic, K., Subramanian, N., Mark, A.E., O'Mara, M.L., 2018. The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. PLOS ONE 13, e0191882.

Darcy, P.A., Wiencek, J.M., 1998. Estimating lysozyme crystallization growth rates and solubility from isothermal microcalorimetry. Acta Crystallogr. Sect. D: Biol. Crystallogr. 54, 1387–1394.

Dimitrov, I.L., Hodzhaoglu, F. v, Koleva, D.P., 2015. Probabilistic approach to lysozyme crystal nucleation kinetics. J. Biol. Phys. 41, 327.

D'Souza, R.N., et al., 2013. Emerging technologies for the integration and intensification of downstream bioprocesses. Pharm. Bioprocess. 1, 423–440.

Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann. der Phys. 322, 549–560.

Galkin, O., Vekilov, P.G., 1999a. Are Nucleation Kinetics of Protein Crystals Similar to Those of Liquid Droplets? J. Am. Chem. Soc. 122, 156–163.
Galkin, O., Vekilov, P.G., 1999b. Direct Determination of the Nucleation Rates of Protein Crystals. J. Phys. Chem. B 103, 10965–10971.
Galkin, O., Vekilov, P.G., 1999c. Are Nucleation Kinetics of Protein Crystals Similar to Those of Liquid Droplets? J. Am. Chem. Soc. 122, 156–163.
Garside, J., Mersmann, Alfons, Nývlt, Jaroslav, 2002. Measurement of crystal growth and nucleation rates. Institution of Chemical Engineers.
Gómez, J., Hilser, V.J., Xie, D., Freire, E., 1995. The heat capacity of proteins. Protein.: Struct., Funct., Bioinforma. 22, 404–412.
Gong, W., Wu, Y., Lin, M., Rohani, S., 2021. Polymorphism control of l-Glutamic acid in a single-stage and a two-stage MSMPR crystallizer by different seeding strategies. Chem. Eng. Res. Des. 170, 23–33.
Good, R.J., Girifalco, L.A., 2002. A theory for estimation of surface and interfacial energies. III. Estim. Surf. Energ. Solids Contact angle data. *J. Phys. Chem.* 64, 561–565.
Green, D.W., Perry, R.H., 2008. Perry's Chemical Engineers' Handbook, eighth ed. McGraw-Hill Education.
Imaizumi, S., Matsuda, T., Hatta, I., 1979. Measurement of dynamic specific heat capacity of lysozyme crystals. J. Phys. Soc. Jpn. 47, 1643–1646.
Kertis, F., et al., 2012. Heterogeneous nucleation of protein crystals using nanoporous gold nucleants. J. Mater. Chem. 22, 21928–21934.
Kharat, S.J., 2008. Density, viscosity and ultrasonic velocity studies of aqueous solutions of sodium acetate at different temperatures. J. Mol. Liq. 140, 10–14.
Kirwan, D.J., Orella, C.J., 2002. Crystallization in the pharmaceutical and bioprocessing industries. In: Myerson, A.S. (Ed.), Handbook of Industrial Crystallization. Elsevier, pp. 249–266. https://doi.org/10.1016/B978-075067012-8/50013-6
Kulkarni, S.A., Kadam, S.S., Meekes, H., Stankiewicz, A.I., ter Horst, J.H., 2013. Crystal nucleation kinetics from induction times and metastable zone widths. Cryst. Growth Des. 13, 2435–2440.
Kwon, J.S.-I., Nayhouse, M., Orkoulas, G., Christofides, P.D., 2014a. Enhancing the Crystal production rate and reducing polydispersity in continuous protein crystallization. Ind. Eng. Chem. Res. 53, 15538–15548.
Kwon, J.S.-I., Nayhouse, M., Orkoulas, G., Christofides, P.D., 2014b. Crystal shape and size control using a plug flow crystallization configuration. Chem. Eng. Sci. 119, 30–39.
Leung, A.K.W., Park, M.M.V., Borhani, D.W., 1999. An improved method for protein crystal density measurements. J. Appl. Crystallogr. 32, 1006–1009.
Link, F.J., Heng, J.Y.Y., 2021. Enhancing the crystallisation of insulin using amino acids as soft-templates to control nucleation. CrystEngComm 23, 3951–3960.
Link, F.J., Heng, J.Y.Y., 2022. Unraveling the impact of ph on the crystallization of pharmaceutical proteins: a case study of human insulin. Cryst. Growth Des. 22, 3024–3033.
Lin, C., Zhang, Y., Liu, J.J., Wang, X.Z., 2017. Study on nucleation kinetics of lysozyme crystallization. J. Cryst. Growth 469, 59–64.
Liu, J.J., Hu, Y.D., Wang, X.Z., 2013. Optimization and control of crystal shape and size in protein crystallization process. Comput. Chem. Eng. 57, 133–140.
Liu, J.J., Ma, C.Y., Hu, Y.D., Wang, X.Z., 2010a. Effect of seed loading and cooling rate on crystal size and shape distributions in protein crystallization—a study using morphological population balance simulation. Comput. Chem. Eng. 34, 1945–1952.
Liu, J.J., Ma, C.Y., Hu, Y.D., Wang, X.Z., 2010a. Modelling protein crystallisation using morphological population balance models. Chem. Eng. Res. Des. 88, 437–446.
Liu, J.J., Ma, C.Y., Hu, Y.D., Wang, X.Z., 2009. Morphological population balance models for the dynamic evolution of particle shape and size distribution in protein crystallization. Comput. Aided Chem. Eng. 27, 1911–1916.
Liu, Y., Wang, X., Ching, C.B., 2010b. Toward further understanding of lysozyme crystallization: Phase diagram, protein-protein interaction, nucleation kinetics, and growth kinetics. Cryst. Growth Des. 10, 548–558.
Li, Q., Rudolph, V., Weigl, B., Earl, A., 2004. Interparticle van der Waals force in powder flowability and compactibility. Int. J. Pharm. 280, 77–93.
Lübbert, A., Simutis, R., 1994. Using measurement data in bioprocess modelling and control. Trends Biotechnol. 12, 304–311.
Mitchell, N.A., 2012. Numerical Modelling Of Cooling Crystallisation: Process Kinetics To Optimisation. University of Limerick.
Mozdzierz, N.J., et al., 2021. Mathematical modeling and experimental validation of continuous slug-flow tubular crystallization with ultrasonication-induced nucleation and spatially varying temperature. Chem. Eng. Res. Des. 169, 275–287.
Mullin, J.W., 2001. Crystal Growth. Crystallization. Elsevier, pp. 216–288. https://doi.org/10.1016/B978-075064833-2/50008-5
Myerson, A.S., Ginde, R., 2002. Crystals, crystal growth, and nucleation. In: Myerson, A.S. (Ed.), Handbook of Industrial Crystallization. Elsevier, pp. 33–65. https://doi.org/10.1016/B978-075067012-8/50004-5
National Instutite of Standards and Technology. NIST-JANAF Thermochemical Tables, 4th Edition. ⟨https://janaf.nist.gov/⟩ (1998) doi:10.18434/T42S31.
National Instutite of Standards and Technology. Thermophysical Properties of Fluid Systems. ⟨https://webbook.nist.gov/chemistry/fluid/⟩ (2022).
Neugebauer, P., Khinast, J.G., 2015. Continuous crystallization of proteins in a tubular plug-flow crystallizer. Cryst. Growth Des. 15, 1089–1095.
Orehek, J., Češnovar, M., Teslić, D., Likozar, B., 2021. Mechanistic crystal size distribution (CSD)-based modelling of continuous antisolvent crystallization of benzoic acid. Chem. Eng. Res. Des. 170, 256–269.
Paquet, E., Viktor, H.L., 2015. Molecular dynamics, Monte carlo simulations, and langevin dynamics: a computational review. BioMed. Res. Int. 2015.
Process Systems Enterprise Ltd. gPROMS FormulatedProducts. Preprint at ⟨https://www.psenterprise.com/products/gproms/formulatedproducts⟩ (2022).
Pu, S., Hadinoto, K., 2020. Continuous crystallization as a downstream processing step of pharmaceutical proteins: a review. Chem. Eng. Res. Des. 160, 89–104.
Pu, S., Hadinoto, K., 2021. Comparative evaluations of bulk seeded protein crystallization in batch versus continuous slug flow crystallizers. Chem. Eng. Res. Des. 171, 139–149.
Roque, A.C.A., et al., 2020. Anything but conventional chromatography approaches in bioseparation. Biotechnol. J. 15, 1900274.
Rosenbaum, T., Tan, L., Dummeldinger, M., Mitchell, N., Engstrom, J., 2019. Population balance modeling to predict particle size distribution upon scale-up of a combined antisolvent and cooling crystallization of an active pharmaceutical ingredient. Org. Process Res. Dev. 23, 2666–2677.
Schall, C.A., Arnold, E., Wiencek, J.M., 1996. Enthalpy of crystallization of hen egg-white lysozyme. J. Cryst. Growth 165, 293–298.
Sevilla, D.J.R., et al., 2005. Predictive control of crystal size distribution in protein crystallization. Nanotechnology 16, S562.
Smejkal, B., et al., 2013. Protein crystallization in stirred systems–scale-up via the maximum local energy dissipation. Biotechnol. Bioeng. 110, 1956–1963.
Sun, J., et al., 2012. Effect of particle size on solubility, dissolution rate, and oral bioavailability: evaluation using coenzyme Q10 as naked nanocrystals. Int. J. Nanomed. 7, 5733.
Su, Q., Rielly, C.D., Powell, K.A., Nagy, Z.K., 2017. Mathematical modelling and experimental validation of a novel periodic flow crystallization using MSMPR crystallizers. AIChE J. 63, 1313–1327.

Szilágyi, B., 2021. Modeling and analysis of MSMPR cascades involving nucleation, growth and agglomeration mechanisms with slurry recycling. Chem. Eng. Res. Des. 174, 42–56.
Tang, X.H., Liu, J.J., Zhang, Y., Wang, X.Z., 2018. Study on the influence of lysozyme crystallization conditions on crystal properties in crystallizers of varied sizes when temperature is the manipulated variable. J. Cryst. Growth 498, 186–196.
Trampuž, M., Teslić, D., Likozar, B., 2021. Crystal-size distribution-based dynamic process modelling, optimization, and scaling for seeded batch cooling crystallization of Active Pharmaceutical Ingredients (API). Chem. Eng. Res. Des. 165, 254–269.
Vekilov, P.G., 2010. The two-step mechanism of nucleation of crystals in solution. Nanoscale 2, 2346–2357.
Volmer, M., Weber, A., 1926. Keimbildung in übersättigten Gebilden. Z. für Phys. Chem. 119U, 277–301.
Weichsel, U., Segets, D., Thajudeen, T., Maier, E.M., Peukert, W., 2017. Enhanced crystallization of lysozyme mediated by the aggregation of inorganic seed particles. Cryst. Growth Des. 17, 967–981.
Yang, H., Chen, W., Peczulis, P., Heng, J.Y.Y., 2019. Development and workflow of a continuous protein crystallization process: a case of lysozyme. Cryst. Growth Des. 19, 983–991.
Zhang, C.Y., et al., 2013. A strategy for selecting the pH of protein solutions to enhance crystallization. Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun. 69, 821–826.
Zhao, Y., et al., 2015. Kinetic identification and experimental validation of continuous plug flow crystallisation. Chem. Eng. Sci. 133, 106–115.
Zhou, H.Y., Zhou, G.Z., Wang, X.Z., 2022. Multi-objective optimization of protein cooling crystallization with morphological population balance models. J. Cryst. Growth 588, 126664.
Zobel-Roos, S., et al., 2020. Digital twins in biomanufacturing. In: Herwig, C., Pörtner, R., Möller, J. (Eds.), Advances in Biochemical Engineering/Biotechnology, vol. 176. Springer, Cham, pp. 181–262.