CrystEngComm

Dynamic Article Links

Cite this: *CrystEngComm*, 2011, **13**, 1197

www.rsc.org/crystengcomm

PAPER

Nucleation and growth kinetics estimation for L-phenylalanine hydrate and anhydrate crystallization†

Nicholas C. S. Kee,[abc] Paul D. Arendt,[a] Li May Goh,[a] Reginald B. H. Tan[bc] and Richard D. Braatz*[ad]

Received 30th August 2010, Accepted 30th November 2010
DOI: 10.1039/c0ce00585a

A process model for the crystallization of L-phenylalanine crystals from mixed propanol–water solution, an enantiotropic system, is developed with nucleation and growth kinetics estimated for the anhydrate and monohydrate forms using *in situ* ATR-FTIR spectroscopy and laser backscattering. A challenging aspect of estimating kinetics for this system is the formation of large numbers of small crystals under certain conditions, which result in biases in the data collected from *in situ* ATR-FTIR and FBRM probes. Batch experiments were designed to follow particular trajectories in the phase diagram so that some kinetic phenomena are suppressed in some runs, which enabled the estimation of sets of kinetic parameters in stages, reducing the number of parameters to be estimated simultaneously. The model was validated by comparison of model predictions and experiments for the product crystals and metastable limits obtained from independent characterization and experiments. This combination of experimental design and process modeling may be emulated to facilitate process modeling and development for crystallizations involving multiple crystal structures.

Introduction

Polymorphism is the ability of a compound to adopt different crystalline arrangements.[1] Although chemically identical, different polymorphs of the same compound display a variation in physical properties such as crystal morphology, density, solubility, and color. Quality control for polymorphic or pseudo-polymorphic systems requires the control of the crystal structure in order to provide the desired product performance. Multiple crystal structures are more frequently encountered in the pharmaceutical industry, with the increase in structural complexity of high value-added products. A reliable production process for the targeted crystal form is critical for feasible economic yield and regulatory compliance. Numerous works have developed methods for selective crystallization[2–10] including our own[11] which utilized feedback control based on the solute concentration measured *in situ* to operate the crystallizer in appropriate domains that suppress secondary nucleation. This approach[12–14] robustly produces crystals of the desired crystal structure without requiring identification of the kinetics for crystal nucleation or growth.

Crystallization kinetics are useful for further optimization of the operating conditions, specifically to elucidate the combined effect of operating variables such as the temperature profile, antisolvent addition rates, and seeding conditions, as well as crystallizer baffling, mixing blade design, and feed pipe location.[15] In addition, process models of the transformations between crystal structures facilitate the understanding of the relative importance of various competing crystallization mechanisms, such as different mechanisms for secondary nucleation.

The transformation rate of the anhydrate to the monohydrate form in water below the transition temperature has been monitored previously based on the offline powder X-ray diffraction (PXRD) analysis.[6] More recent studies on monitoring the transformation of organic compounds have utilized *in situ* instruments such as ATR-FTIR and Raman spectroscopy, which have been applied in monotropic and enantiotropic systems.[16–21] *In situ* process monitoring and modeling with kinetics estimation have been applied to L-glutamic acid,[22–24] a monotropic dimorph system. Similar transformation analysis has been applied to the enantiotropic pseudo-dimorph of citric acid,[25] but only pertaining to the dissolution of the metastable form and nucleation and growth of the stable form.

This paper describes the staged design of experiments and parameter estimation to model the nucleation and growth of L-phenylalanine (L-phe) crystals in propanol–water solution, which is an enantiotropic system.[6] The crystallization model includes kinetic parameters for both the anhydrate and monohydrate

[a]*Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL, 61801, USA*
[b]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576*
[c]*Institute of Chemical and Engineering Sciences, 1 Pesek Road, Jurong Island, Singapore 627833*
[d]*Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA. E-mail: braatz@mit.edu; Fax: +1 617-258-0546; Tel: +1 617-253-3112*

† Electronic supplementary information (ESI) available: Additional experimental data with comparisons to model predictions. See DOI: 10.1039/c0ce00585a

forms (previously denoted as β and α-forms,[5,26] respectively, see Fig. 1), estimated from *in situ* experimental data, specifically, the solute concentration measured using Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) spectroscopy and moments of the chord length distribution (CLD) obtained using Focused Beam Reflectance Measurement (FBRM). The concentration feedback control approach[12–14] is used to implement experimental designs to follow specified trajectories in the solid–liquid phase diagram. The staged design enables individual sets of kinetic parameters to be estimated, as opposed to the simultaneous estimation of all kinetic parameters. The predictive ability of the model was evaluated based on comparisons with offline characterization of the product crystals as well as with the experimentally determined metastable limits. To the best of the authors' knowledge, this is the first comprehensive crystallization study for an enantiotropic system, which includes *in situ* monitoring and parameter estimation of the nucleation and growth kinetics of both forms.

Experimental

Materials and methods

A Dipper-210 ATR immersion probe (Axiom Analytical) with ZnSe as the internal reflectance element attached to a Nicolet Protégé 460 FTIR spectrophotometer was used to obtain the spectra of the L-phe solution. The chord length distribution of the crystals in solution was measured every 30 s using Lasentec FBRM with version 6.0b12 of the FBRM Control Interface software. *In situ* images of the slurry were collected using Lasentec Particle Vision and Measurement (PVM). The solution temperature was controlled by ratioing hot and cold water to the jacket with a control valve,[12] and was measured every 2 s using a Teflon-coated thermocouple attached to a Data Translation 3004 data acquisition board *via* a Fluke 80TK thermocouple module. PXRD patterns of L-phe crystals were collected offline using the Bruker General Area Detector Diffraction System (GADDS, Bruker AXS, Inc.) with Cu $K_{\alpha 1}$ and Cu $K_{\alpha 2}$ (weighted sum) radiation and a step size of 0.02°. The anhydrate form is orthorhombic while the monohydrate form is monoclinic; the space group is $P222_1$ for both.[6,26] L-Phe crystals obtained commercially (>98.5%, Sigma Aldrich) were verified using PXRD to be pure anhydrate form in which the characteristic peaks of the monohydrate form were not observed. The monohydrate form was produced by rapid cooling,[6,11] and was similarly verified using PXRD. Recrystallized anhydrate crystals were used in the metastable limit and batch crystallization experiments as seeds because these had a well-defined rhombic shape as opposed to the commercially available crystals which were irregular platelets. The recrystallized crystals were produced by transformation of the monohydrate form.[11] The solvent consisted of 75 wt% DI water and 25 wt% 2-propanol (ACS grade, Sigma Aldrich). Scanning Electron Microscopy (SEM) samples were sputtered with 4–8 nm of Au/Pd before being recorded with a JEOL 7000F SEM.

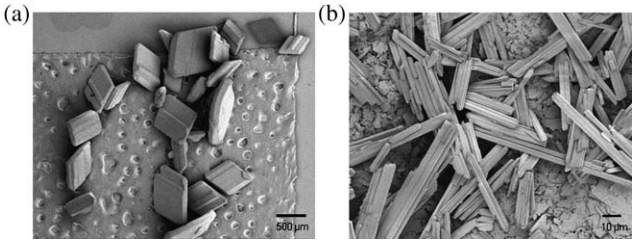

Fig. 1 Scanning electron micrographs of L-phe crystals: (a) anhydrate form and (b) monohydrate form.

Calibration for solute concentration

Different solute concentrations of L-phe and the mixed solvent (~400.00 g) were placed in a 500 ml jacketed round-bottom flask and heated until complete dissolution; an overhead mixer was used to agitate the solution with a stirring speed of 175 rpm. The solution was cooled at 0.5 °C min^{-1} while the IR spectra were collected. The IR spectra in the range 1100–1650 cm^{-1} were used to construct the calibration model, which was determined using various chemometrics methods.[27] The chemometrics forward selection PCR 2 (FPCR 2) method[28] was selected, which gave a prediction interval of ±0.00034 g per g solvent.

Solubility and metastable limit measurements

The IR spectra of L-phe slurries of each form were collected at different temperatures. At each evaluated temperature, the slurry was equilibrated for 30 min to an hour while recording IR spectra. The equilibrium solute concentration was then calculated using the calibration model. Care was taken to ensure that the monohydrate-form crystal slurry contained only a small amount of excess solids, to prevent significant solid-phase interference on the IR measurements. The solubility for the monohydrate form was also determined using a slow heating method to validate the result from the earlier measurements.[11] The metastable limit of L-phe solutions was determined based on the polythermal method,[29] using a constant cooling rate of 0.1 °C min^{-1} for various initial solute concentrations (Table 1). The same seeding conditions were used as that in the main crystallization experiments; anhydrate crystals (0.50 g, 220–520 μm) were added at 0.6 °C after the solution temperature, T[°C], crossed the saturation temperature of the anhydrate form. The CLD of the crystals was monitored using FBRM. A nucleation event was identified by the start of a sharp increase in the total counts per s profiles (ignoring the increase due to seeding).

Table 1 Initial solute concentrations in the metastable limit experiments

Run	Initial solute concentration/g per g solvent	Solvent mass/g
1m	0.01697	400.26 (deionised water: 300.18 and 2-propanol: 100.08)
2m	0.01900	400.26 (deionised water: 300.18 and 2-propanol: 100.08)
3m	0.02097	400.11 (deionised water: 300.09 and 2-propanol: 100.02)
4m	0.02299	400.11 (deionised water: 300.09 and 2-propanol: 100.02)
5m	0.02501	400.11 (deionised water: 300.09 and 2-propanol: 100.02)

PVM images were also recorded to check the appearance of the monohydrate needles. In both the solubility and metastable limit experiments, the solvent mass and composition as well as the stirring speed of the overhead mixer were similar to that for calibration.

Seeded batch crystallization

An undersaturated solution with solute concentration ~0.02300 g per g solvent was cooled at 0.1 °C min^{-1} to 0.6 °C below the anhydrate form saturation temperature, upon which anhydrate seeds with similar crystal size distribution (CSD) as that in the metastable limit experiments were added. The seed masses were 0.51, 0.49, and 0.46 g for runs 1, 2, and 3, respectively. The stirring conditions were the same as that used previously. Preset supersaturation profiles were followed during crystallization using concentration feedback control[12,13] to adjust the cooling rate accordingly based on *in situ* solute concentration measurements. The control algorithm was started shortly after seeding. Supersaturation setpoint profiles were selected at different constant absolute supersaturations ($\Delta C_i = C - C_{\mathrm{sat},i}$) with respect to both the anhydrate and monohydrate forms, where C [g per g solvent] is the solute concentration, $C_{\mathrm{sat},i}$ [g per g solvent] is the solubility, and i indicates either the anhydrate (a) or monohydrate form (m). The solvent composition was identical to that used for calibration; the solvent masses were 400.16 g (300.11 g deionised water, 100.05 g 2-propanol), 400.20 g (300.17 g deionised water, 100.03 g 2-propanol), and 400.31 g (300.21 g deionised water, 100.10 g 2-propanol) for runs 1, 2, and 3, respectively.

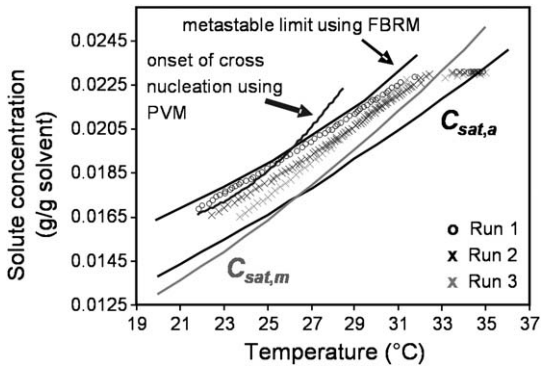

Fig. 2 Supersaturation profiles implemented in the seeded batch crystallization runs.

Experimental results

Solubility and metastable limit measurements

The solubility curves (each fitted to a second-order polynomial[11]) of both forms of L-phe and the metastable limits are shown in Fig. 2 (the data points used to construct the metastable limits are shown in Fig. S1, ESI†). An increase in the total number of particles counted per second measured by FBRM, also referred to as the total counts per s, indicates a nucleation event but cannot directly identify the nucleated form. The nucleated form was inferred depending on when monohydrate needles were detected *via* PVM images relative to the rise in total counts per s. At temperatures above the transition point, needles were not detected prior to or at the onset of an increase in total counts per s but were only observed at a substantially later time, indicating that the first nucleation event detected *via* FBRM was of the seeded anhydrate form. Below the transition temperature, needles were observed before an increase in FBRM counts, indicating cross-nucleation as the first nucleation event. Both of these metastable limits are shown in Fig. 2.

Concentration-controlled batch crystallization

Fig. 2 compares the supersaturation profiles of the three seeded concentration-controlled crystallization experiments (runs 1–3). Table 2 summarizes the operating conditions and results. The anhydrate seeds for all three runs were from the same batch, with crystal size characterized by the largest distance across a crystal measurable from microscopy images, L_a [μm], which occurs along the diagonal. The size distribution for the anhydrate product crystals was similarly constructed.

A crystallization model for L-phe was constructed to describe the nucleation and growth kinetics of both anhydrate and monohydrate forms. The model has a large number of kinetic parameters that are challenging to estimate simultaneously. In the above experiments, some kinetic phenomena were suppressed at certain segments of the experimental data, to reduce the number of kinetic parameters to facilitate efficient and accurate parameter estimation. The newly determined kinetic parameters were then utilized in the next set of data to facilitate the estimation of the remaining kinetic parameters. For example, the growth kinetics of the anhydrate form can be estimated from run 3 and the first 0–275 min of run 2 because nucleation of either form was not detected in these ranges (as cross-nucleation was detected in run 2 at 291 min, the data range was limited to 275 min to ensure no hydrate crystals were in the system). The

Table 2 Summary of operating conditions and results for the concentration-controlled crystallization experiments (the time the seeds were added is defined as $t = 0$ min)

Run	ΔC_i/g per g solvent	Anhydrate-form nucleation	Monohydrate-form nucleation	Monohydrate in the product crystals[b] (wt%)
1	$\Delta C_a = 0.00200$	Detected after $t \approx 36$ min	Detected after $t \approx 104$ min	9.45 ± 1.81
2	$\Delta C_a = 0.00150$	Not detected	Detected after $t \approx 291$ min	0.53 ± 1.81[a]
3	$\Delta C_m = 0.00110$	Not detected	Not detected	0.57 ± 1.81

[a] For run 2, the increase in the total counts per s and the detection of the monohydrate needles from PVM images towards the end of the run were substantial experimental evidence of cross-nucleation, although the amount of the monohydrate form in the product was below the detection limit using PXRD.[11] [b] Based on 95% confidence interval.[11]

Table 3 Segments of experimental data pertaining to different crystallization kinetics

Crystallization kinetics	Relevant experimental range
Anhydrate-form growth	Run 2 (0–275 min) and run 3 (entire duration)
Anhydrate-form nucleation	Run 1 (0–100 min)
Monohydrate-form nucleation and growth	Run 1 (100 min end) and run 2 (275 min end)

newly determined growth kinetics for the anhydrate form were then incorporated into the analysis of the next segment of the experimental data, for example, to account for the effects of anhydrate-form growth while estimating its nucleation kinetics from the first 100 min of run 1. Both sets of kinetic parameters for the anhydrate form were then utilized in the analysis of the remaining experimental data while estimating crystallization kinetics of the monohydrate form. The implementation of experiments and parameter estimation in stages reduces the number of parameters to be estimated simultaneously and thus decreases the chance of obtaining poorly conditioned parameter estimation optimizations.[24,30] Table 3 summarizes the segments of the experimental data relevant to the different kinetics. The development of the crystallization model along with the parameter estimation, corresponding confidence intervals, and model validation is described next.

Mathematical model of L-phe crystallization

This section presents the mathematical model for the crystallization of L-phe. This was based on the population balance equation (PBE), coupled with the kinetic expressions for nucleation and crystal growth for both the anhydrate and monohydrate forms. The kinetic parameters were estimated using weighted least-squares regressions applied to the data from the concentration-controlled experiments. The model was then validated based on the offline characterization of the product crystals and the metastable limit experiments.

Model equations

The model is based on the assumption that the crystallizer is well-mixed, that is, no spatial variations in T, C, or CSD. Such assumptions are common in the industrial crystallization literature for a lab-scale crystallizer.[30] The population balances that describe the time-dependent CSDs, $f_i(L_i,t)$ [# crystals/(μm g solvent)], are:

$$\frac{\partial f_i}{\partial t} + \frac{\partial\left(G_i\left(S_i, T, L_i; \theta_{g,i}\right) f_i\right)}{\partial L_i} = B_i(S_i, T, f_i;\ \theta_{b,i})\delta(L_i - L_{0,i}), \quad (1)$$

where i = anhydrate (a) or monohydrate (m) of L-phe, L_i [μm] is the characteristic crystal dimension and $L_{0,i} > 0$ is the characteristic size for nucleated crystals, T is temperature [°C], G_i [μm min^{-1}] is the crystal growth rate along the characteristic dimension, B_i [# crystals/(min g solvent)] is the nucleation rate, S_i [—] is the relative supersaturation $(C - C_{\mathrm{sat},i})/C_{\mathrm{sat},i}$, $\delta(\cdot)$ is the Dirac delta function, and $\theta_{g,i}$ and $\theta_{b,i}$ are sets of growth and nucleation parameters. The seed crystals consisted of pure anhydrate form. Thus the initial condition for the population balance is given by the CSD of the anhydrate seed crystals $f_a(L_a,0)$ with $f_m(L_m,0) = 0$. Assuming that there is no mechanism for producing either form of crystals of size $L_{0,i} = 0$ and that the initial distribution does not have anhydrate crystals of size $L_a = 0$, the boundary condition is

$$f_i(0,t) = 0,\ \forall\, t \in (0, \infty). \quad (2)$$

Applying the method of moments, the population balance equations were converted into the system of ordinary differential equations (ODEs) by multiplying both sides of eqn (1) by L^j, for $j = 0, 1, 2,\ldots$, and integrating each term of 0 to ∞:[31]

$$\frac{\mathrm{d}\mu_{0,i}}{\mathrm{d}t} = B_i \quad (3)$$

$$\frac{\mathrm{d}\mu_{j,i}}{\mathrm{d}t} = jG_i\mu_{j-1,i} + B_i L_{0,i}^j \quad j = 1,\ 2,\ \ldots \quad (4)$$

where the j^{th} CSD moment is defined as:

$$\mu_{j,i} = \int_0^\infty L_i^j f_i(L_i, t)\mathrm{d}L_i. \quad (5)$$

The CSD moments relate to the physical properties of the crystals in the distribution; $\mu_{0,i}$ [# crystals/g solvent], $\mu_{1,i}$ [(# crystals μm)/g solvent], $\mu_{2,i}$ [(# crystals μm^2)/g solvent], and $\mu_{3,i}$ [(# crystals μm^3)/g solvent] are proportional to the total number, length, surface area, and volume of the crystals, respectively. The solute mass balance is:[32]

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \left(3G_a\mu_{2,a} + B_a L_{0,a}^3\right)\rho_a k_{v,a} + \left(3G_m\mu_{2,m} + B_m L_{0,m}^3\right)\rho_m k_{v,m}, \quad (6)$$

where ρ_i is the crystal density ($\rho_a = 1.29 \times 10^{-12}$ g μm^{-3} and $\rho_m = 1.724 \times 10^{-12}$ g μm^{-3}) and $k_{v,i}$ is the volumetric shape factor defined by $v_i = k_{v,i}L_i^3$ where v_i is the volume of an i-form crystal. Values for v_i [μm^3] were estimated from microscopy images of crystals of both forms, and $k_{v,i}$ subsequently calculated using L_a as the largest diagonal length and L_m as the thickness of the needle ($k_{v,a} = 0.0394$ and $k_{v,m} = 5.77$). The nucleated crystals were treated as being of negligible size $L_{0,i} \to 0$; using realistic nonzero values for $L_{0,i}$ produces essentially the same $\mu_{j,i}$ and C predictions because the terms that contain $L_{0,i}$ in the moment equations are many orders of magnitude smaller than the rest of the terms.[30,31] For parameter estimation, $\mu_{0,i}$, $\mu_{1,i}$, and $\mu_{2,i}$ were simulated for both forms, along with C, which are described by seven ordinary differential equations (ODEs, eqn (3) and (4) for both forms and eqn (6)). Values of $\mu_{3,i}$ were also simulated for comparison with the FBRM data.

For the system of ODEs (eqn (3) and (4)), the initial conditions were calculated from the CSD of the anhydrate seed crystals using:

$$\mu_{j,a}^s = \sum_{k=1}^{N-1}\left(\frac{L_{k+1,a} + L_{k,a}}{2}\right)^j f_a^s(L_{k+1,a} - L_{k,a}, 0), \quad (7)$$

where $L_{1,a}$ and $L_{N,a}$ refer to the minimum and maximum sizes (200 and 520 μm, respectively), and $f_a^s(L_{k+1,a} - L_{k,a},0)$ [# crystals] is the number of anhydrate crystals in a sample with sizes between $L_{k+1,a}$ and $L_{k,a}$ at time $t = 0$. In the model, $\mu_{j,a}$ (and $\mu_{j,m}$)

were expressed as per unit mass of solvent, while that calculated from eqn (7) was for the number of crystals sampled. These quantities are related by a proportionality constant, λ [1/g solvent]:

$$\mu_{j,a} = \lambda \mu^{s}_{j,a}. \tag{8}$$

The initial value for $\mu_{3,a}$ can be calculated from mass balance:

$$m_{\mathrm{sd}} = \mu_{3,a}(t=0) k_{v,a} \rho_a m_{\mathrm{solv}}, \tag{9}$$

where m_{sd} [g] is the mass of seed crystals and m_{solv} [g] is the solvent mass. The value for λ was then calculated by comparing $\mu^{s}_{3,a}(t=0)$ from eqn (7) and $\mu_{3,a}(t=0)$ from eqn (9). With λ known, the $\mu_{j,a}(t=0)$ was calculated from $\mu^{s}_{j,a}(t=0)$ for $j = 0$, 1, and 2. The initial solute concentration, $C(t=0) = C_0$, was the value measured immediately after seeding. This study utilized the power-law expressions for the nucleation and growth kinetic equations, which have been widely used for a large number of systems in literature.[30,31] Specifically, the growth rate of the anhydrate form, G_a, was assumed to be size independent, and expressed as:

$$G_a = k^{'}_{g,a} S_a^{g_a} \tag{10}$$

where $k^{'}_{g,a}$ [μm min^{-1}] and g_a [—] are the parameters to be estimated. The Arrhenius relationship was used to account for the temperature dependence using

$$k^{'}_{g,a} = k_{g,a} \exp\left(\frac{E_{g,a}}{R(T+273.15)}\right), \tag{11}$$

where $E_{g,a}$ [J mol^{-1}] is the activation energy for the growth rate of the anhydrate form, R [J mol^{-1} K^{-1}] is the gas constant, and $k_{g,a}$ has the same dimensions as $k^{'}_{g,a}$. An expression for size-dependent growth (linear dependence on size) was also tested:

$$G_a = k^{'}_{g,a} S_a^{g_a} (1 + \gamma L_a) \tag{12}$$

where the constant γ [1/μm] may be positive or negative.[33] Using eqn (12) leads to a different form for the moment equations:

$$\frac{\mathrm{d}\mu_{j,a}}{\mathrm{d}t} = j G_a \left(\mu_{j-1,a} + \gamma \mu_{j,a}\right), \quad j = 1,\ 2, \ldots \tag{13}$$

The solute mass balance eqn (6) was modified in the same manner. For the monohydrate form, the growth rate G_m was assumed to be independent of both size and temperature:

$$G_m = k^{'}_{g,m} S_m^{g_m} \tag{14}$$

where, analogous to that of the anhydrate form, $k_{g,m}$ [μm min^{-1}] and g_m [–] are the pre-exponential and exponential factors to be determined. In general, nucleation is classified into two categories, primary or secondary, for nucleation in the absence and in the presence of crystals, respectively. In this study, secondary nucleation from crystal surfaces was considered, as this is the dominant mechanism in seeded crystallizers. For the anhydrate form, the nucleation rate was expressed as:

$$B_a = k^{'}_{b,a} S_a^{b_a} \mu_{2,a} \tag{15}$$

where $k^{'}_{b,a}$ [1/(min μm^2)] and b_a [—] are the parameters to be estimated. The Arrhenius temperature dependence was assumed:

$$k^{'}_{b,a} = k_{b,a} \exp\left(\frac{E_{b,a}}{R(T+273.15)}\right), \tag{16}$$

where $E_{b,a}$ [J mol^{-1}] is the activation energy for the nucleation rate for the anhydrate form and $k_{b,a}$ has same dimensions as $k^{'}_{b,a}$. The nucleation rate of the monohydrate form:

$$B_m = k_{b1,m} S_m^{b_1} \mu_{2,m} + k_{b2,m} S_m^{b_2} \mu_{2,a} \tag{17}$$

where $k_{b1,m}$ [1/(min μm^2)], $k_{b2,m}$ [1/(min μm^2)], $b_{1,m}$ [–], and $b_{1,m}$ [–] are the nucleation parameters to be estimated. This includes a self- and a cross-nucleation term (*i.e.*, dependence on $\mu_{2,a}$) as the secondary nucleation of the monohydrate form on the surfaces of anhydrate crystals has been reported.[6]

Relationship between CSD and CLD moments

The FBRM probe emits a laser beam which moves in a circular path at a high frequency using rotating optics. As the laser beam 'crosses' a particle, a chord length or a count is registered, *i.e.*, a chord length is the distance across the crystal surface as observed by the laser beam. The CLD measured *in situ* is related to but not the same as the CSD. The model eqn (3) and (4) utilize the CSD moments; these data are not available in real-time, but can be estimated from the measured CLD. Specifically the CSD is first recovered from the CLD before calculating its moments. Numerous previous works[34–36] have studied how the CSD can be recovered from the measured CLD by geometric modeling. The relationship between the CLD and CSD moments has also been modeled by:[37,38]

$$\mu_j = \xi_j \mu^{c}_{j-1}, \tag{18}$$

where μ^{c}_{j-1} is the $(j-1)^{\mathrm{th}}$ moment of the CLD and ξ_j is a shape-dependent scaling factor (a function of the single-particle CLD). Models that attempt to construct the CSD from the CLD require many assumptions,[39] including perfect backscattering of the laser at all angles and that all particles have a known shape. While these may hold true for some particulate systems, these assumptions are not applicable for other systems, and will not be true for most pharmaceutical systems, in which the refractive indices of crystal and solution are similar. Furthermore, unmodeled factors such as the optical properties of solids and solvents, as well as the positioning of the probe in relation to the surrounding flow, can significantly impact the CLD and subsequently the estimated CSD.[39] An alternative approach is to use the low-order CLD moments, without first obtaining a CSD estimate; this replaces the first-principles model for the CSD with a gray-box model for the CLD, in which the structure of the first-principles model for the low-order CSD moments is used to parametrize that of the CLD.[25,40] The low-order CLD moments should follow the same dynamic trends as the low CSD moments because the mapping between the CLD and CSD is static (most of the models mentioned earlier assume that the mapping is not only static but also linear). This study follows a modification of that approach to establish the relationship between the CSD and CLD moments. At any experimental time point, the solute mass balance can be written as:

$$m_{\mathrm{sd}} + C_0 m_{\mathrm{solv}} = (\mu_{3,a} \rho_a k_{v,a} + \mu_{3,m} \rho_m k_{v,m} + C) m_{\mathrm{solv}}. \tag{19}$$

For run 3, $\mu_{3,m}$ was set to 0 because cross-nucleation was not detected. Simplifying eqn (19) and using the values of the solute concentration accurately determined *in-situ* from ATR-FTIR spectroscopy, $\mu_{3,a}$ was calculated at various time points and compared to the second-, third-, and fourth-order CLD moments for the anhydrate form, $\mu^c_{2,a}$ [(# counts μm^2)/s], $\mu^c_{3,a}$ [(# counts μm^3)/s], and $\mu^c_{4,a}$ [(# counts μm^4)/s], respectively. The CLD moments were estimated based on the superposition principle in which both forms contribute additively to the overall CLD, *i.e.*, $\mu^c_j = \mu^c_{j,a} + \mu^c_{j,m}$. From the overall (measured) CLD,

$$\mu^c_j = \sum_{k=1}^{N-1} \left(\frac{L^c_{k+1} + L^c_k}{2} \right)^j f^c\left(L^c_{k+1} - L^c_k\right), \tag{20}$$

where L^c_1 and L^c_N refer to the minimum and maximum chord lengths (1 and 1000 μm, respectively) and $f^c(L^c_{k+1} - L^c_k)$ [# counts/s] is the number of chord lengths between L^c_{k+1} and L^c_k. In the absence of cross-nucleation in run 3, the CLD data were due solely to the anhydrate form; $\mu^c_{j,m}$ was set to zero and $\mu^c_{j,a}$ calculated directly from eqn (20). Each order of $\mu^c_{j,a}$ was scaled accordingly to give the best fit to $\mu_{3,a}$. In Fig. 3, $\mu^c_{4,a}$ shows the best fit to $\mu_{3,a}$, so the relationship between CLD and CSD moments for the anhydrate form was:

$$\mu_{j,a} = \Phi_{j,a}\mu^c_{j+1,a}, \tag{21}$$

where $\Phi_{j,a}$ [(# crystals s)/(# counts μm g solvent)] is a proportionality constant. The corresponding relationship for the monohydrate form is similar. The empirical correlation (eqn (21)) differs from some theoretical predictions (eqn (18)). Recall that theoretical predictions are based on many restrictive assumptions that do not apply to all particulate systems. It is interesting that eqn (21) is an expression commonly used in industrial practice (called "length-weighting"), whereas eqn (18) is not.

Parameter estimation

Eqn (3), (4), and (6) describing the nucleation and growth of both forms of L-phe contain 12 parameters that were estimated using the *in situ* experimental solute concentration recorded by ATR-FTIR spectroscopy and moments μ^c_j (for j = 1, 2, and 3) measured by FBRM for runs 1, 2, and 3. The kinetic parameters were estimated by minimizing the weighted difference between the experimental measurements y^{ex} and model predictions y^{md}:

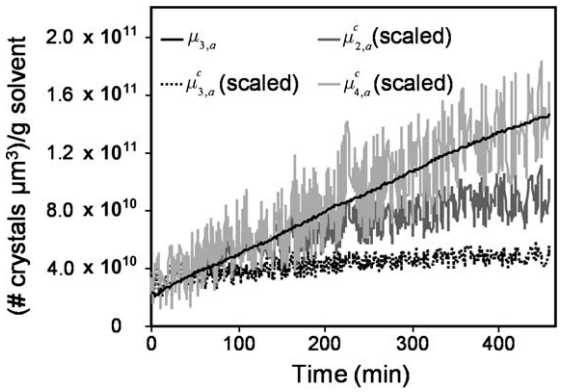

Fig. 3 Comparison of the CSD moment $\mu_{3,a}$ to various CLD moments.

$$\Psi = \min_{\theta} \sum_{q=1}^{N_v} \sum_{n=1}^{N_s} \frac{w_q}{N_s} \left(y^{ex}_{q,n} - y^{md}_{q,n}(\theta) \right)^2, \tag{22}$$

where θ is the vector of the estimated parameters, N_v (= 4) is the number of measured variables, N_s is the number of sampling instances, and w_q is a weighting factor that is the inverse of an estimate of the error variance σ^2_q. Initially, σ_q was set to the standard deviation of the actual measurement. A set of parameters was estimated by solving the system of ODEs (eqn (3), (4), and (6) coupled with the optimization in eqn (22), using the built-in Matlab codes 'ode45'[41] and 'fmincon'.[42] The standard deviations for the errors in the q^{th} measured variable were then estimated and used for σ_q in a re-solving of the parameter estimation optimization. The iterative procedure was repeated until no significant difference between the given and estimated standard deviations was detected.

The parameter estimation for the different sets of kinetic parameters was carried out in stages. In Stage 1, the parameters for the anhydrate form growth were estimated first using the experimental data from run 2 (up to 275 min) and run 3 (for the entire experiment, see Tables 2 and 3). In these durations, no nucleation was detected, and the time evolution of the solute concentration and CLD moments $\mu^c_{j,a}$ were solely due to the growth of anhydrate crystals. The system of ODEs was simplified by setting B_a, B_m, and $G_m = 0$. The proportionality constants $\Phi_{j,a}$ in eqn (21) were calculated from:

$$\Phi_{j,a} = \frac{1}{N_s} \sum_{n=1}^{N_s} \frac{\mu^{md}_{j,a,n}}{\mu^{c,ex}_{j+1,a,n}}. \tag{23}$$

With these $\Phi_{j,a}$, the model CLD moments $\mu^{c,md}_{j+1,a}$ calculated from eqn (21) were used for weighted least-squares regression in eqn (22). Fig. 4 and S2† show very good agreement between the Stage 1 model and experimental data (lack of fit from 300 min onwards for run 2 in Fig. S2† was due to cross-nucleation not accounted for in Stage 1).

Although not utilized for regression, μ^{md}_3 was also simulated, scaled, and compared to $\mu^{c,ex}_4$, as shown in Fig. 4d and S2d†. The agreement between the Stage 1 model and data is very good. Variants of the power-law model for the growth rate for the anhydrate form were also investigated (see Table 4), along with the nominal model (eqn (10) and (11)). The *F*-test was used to compare the first and second model based on the weighted sum-of-squared residuals, Ψ_1 and Ψ_2, respectively. The *F*-test statistic is given by:[43]

$$F = \frac{(\Psi_1 - \Psi_2)/(N_{\theta,2} - N_{\theta,1})}{\Psi_2/(N_t - N_{\theta,2})}, \tag{24}$$

where $N_{\theta,2}$ and $N_{\theta,1}$ are the number of parameters in the corresponding models and N_t = 1424 is the total number of independent measurements from ATR-FTIR spectroscopy and FBRM. The values for *F*-critical listed in Table 4 are based on the rejection probability of 0.05, with ($N_{\theta,2} - N_{\theta,1}$, $N_t - N_{\theta,2}$) degrees of freedom. As $F >$ *F*-critical, the reduction in Ψ due to the additional parameter in model 2 is statistically significant. Similar analysis comparing models 2 to 3 indicated that the reduction in Ψ with the addition of the parameter, γ, for size-dependent growth was not statistically significant. Comparison of the CSD of the anhydrate seed and product crystals of run 3

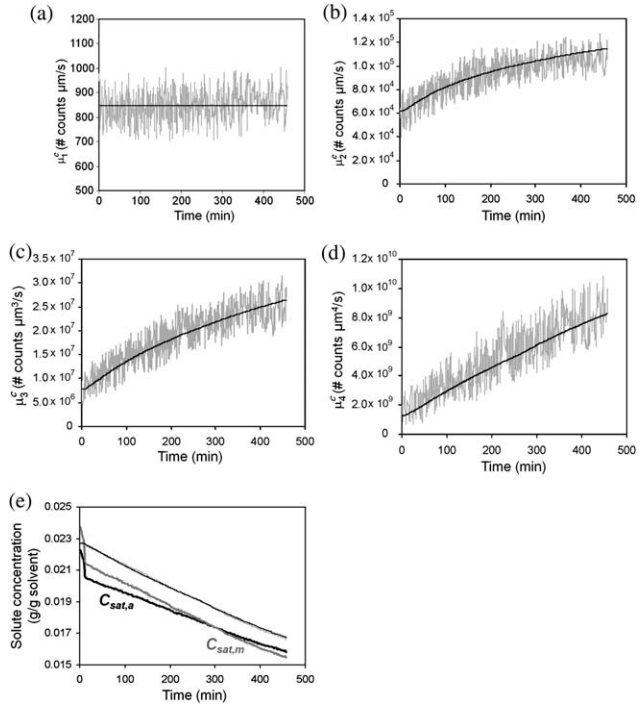

Fig. 4 Run 3 moments and solute concentration, experimental (—); Stage 1 model (−): (a) μ_1^c, (b) μ_2^c, (c) μ_3^c, (d)μ_4^c, and (e) C. The solubility of each crystal form computed from the measured temperature is shown as thick grey and black lines in (e).

suggested some size-dependent growth because the latter has a wider spread. Nonetheless, the available *in situ* experimental data did not statistically justify the estimation of γ, possibly due to the noise in the CLD measurements. The noise level could be reduced by increasing the number of seed crystals (from the current seed loading at ~15%); but this would also reduce the amount of growth that occurs during the batch run, such that the overall signal-to-noise ratio is not significantly improved.

In Stage 2 of the parameter estimation, the parameters for the nucleation of the anhydrate form were determined from the experimental data from run 1 (up to 100 min, where cross-nucleation was negligible, see Tables 2 and 3). The same set of ODEs was used with $B_m = G_m = 0$ with the kinetics for G_a inserted. The proportionality constants $\Phi_{j,a}$ for the CLD moments were estimated in the same manner using eqn (23). Their values were nearly the same as in Stage 1, for example, $\Phi_{0,a}$ was 0.487 and 0.519 in Stages 1 and 2, respectively. The minor difference may have been due to a slight variation in the crystal habit of the anhydrate seed crystals used in the different experiments. The Stage 2 model and experimental data show good agreement for the modeled time range (see Fig. S3†, with the lack of fit starting from 100 min onwards due to cross-nucleation which was not modeled in Stage 2).

A simpler variant of the power-law model for the nucleation of the anhydrate form was also studied. Similar analysis using the F-test statistics (with $N_t = 251$ and F-critical based on the same rejection probability as before) indicated that the improvement in the fits between experimental data and the model was statistically significant with the addition of $E_{b,a}$ to describe the temperature dependence (Table 5).

In Stage 3 of the parameter estimation, the parameters for both the nucleation and growth of the monohydrate form were determined using the relevant range of experimental data in run 1 (100 to 200 min) and run 2 (from 275 min onwards, see Tables 2 and 3). Although cross-nucleation occurred in run 1 from 100 min onwards, the modeled time range was restricted up to 200 min because preliminary modeling gave extremely poor fits in the CLD moments from about 200 min onwards. The FBRM probe likely became partially clogged at increasingly high solids density of the monohydrate needles. In addition, despite the appearance of distinct individual needles in the PVM images, there was possible formation of aggregates or 'bundles' which would produce erroneous high counts of the chord length. The model for parameter estimation was the ODEs with the newly estimated kinetics for B_a and G_a to account for the nucleation and growth of anhydrate form crystals. The measured CLD moments, $\mu_{j+1}^{c,\mathrm{ex}}$, were due to both anhydrate and monohydrate forms. The rates B_a and G_a were used to simulate the CLD moments due solely to the anhydrate form, $\mu_{j+1,a}^{c,\mathrm{md}}$ from eqn (21); the excess moments after subtracting from $\mu_{j+1}^{c,\mathrm{ex}}$ were attributed only to the monohydrate form and were then used to estimate its kinetics. The proportionality constants were calculated using:

$$\Phi_{j,m} = \frac{1}{N_s}\sum_{n=1}^{N_s}\frac{\mu_{j,m,n}^{\mathrm{md}}}{|\mu_{j+1,n}^{c,\mathrm{ex}} - \mu_{j+1,a,n}^{c,\mathrm{md}}|}. \quad (25)$$

The model CLD moments $\mu_{j+1,m}^{c,\mathrm{md}}$ calculated from these $\Phi_{j,m}$ were used for regression in eqn (22). Although not used in regression, $\mu_{3,m}^{\mathrm{md}}$ was simulated and scaled to give $\mu_{4,m}^{c,\mathrm{md}}$. Fig. S2 and S3† show good agreement between the model and experimental data for modeled range (the overall model CLD moments

Table 4 Variants of power-law growth model for anhydrate L-phe

Growth model	Corresponding expressions for G_a	Parameters to be estimated	Ψ	F	F-critical
(1) Size and temperature independent	$k_{g,a}S_a^{g_a}$	$k_{g,a}$; g_a	3.981	—	—
(2) Size independent, temperature dependent	$k_{g,a}\exp\left(\frac{E_{g,a}}{R(T+273.15)}\right)S_a^{g_a}$	$k_{g,a}$; g_a; $E_{g,a}$	3.953	9.983	3.848
(3) Size and temperature dependent	$k_{g,a}\exp\left(\frac{E_{g,a}}{R(T+273.15)}\right)S_a^{g_a}(1+\gamma L_a)$	$k_{g,a}$; g_a; $E_{g,a}$; γ	3.946	2.840	3.848

Table 5 Variants of power law nucleation model for anhydrate L-phe

Nucleation model	Corresponding expressions for B_a	Parameters to be estimated	Ψ	F	F-Critical
(1) Temperature independent	$k_{b,a}S_a^{b_a}\mu_{2,a}$	$k_{b,a}$; b_a	4.094	—	—
(2) Temperature dependent	$k_{b,a}\exp\left(\frac{E_{b,a}}{R(T+273.15)}\right)S_a^{b_a}\mu_{2,a}$	$k_{b,a}$; b_a; $E_{b,a}$	3.967	7.722	3.879

$\mu_{j+1}^{c,\mathrm{md}} = \mu_{j+1,a}^{c,\mathrm{md}} + \mu_{j+1,m}^{c,\mathrm{md}}$). Although not used in the regression, $\mu_4^{c,\mathrm{md}}$ agreed well with $\mu_4^{c,\mathrm{ex}}$ in the modeled range (Fig. S2d and S3d†). As mentioned earlier, the lack of fit for the moment profiles in run 1 starting from 200 min onwards was likely due to some sticking of crystals on the FBRM probe tip or formation of aggregates. The solution concentration profile in Fig. S3e† showed a poorer fit to the experimental profile from 200 min onwards. This could also be associated with the progressively higher solid density of the monohydrate needles. In most crystal–solvent systems, the penetration depth of the evanescent field (~0.5 to 2.5 μm, for the experimental setup in this study) into the solution is smaller than the liquid phase barrier between the probe and the particles. Then when the ATR probe is inserted into the crystal slurry, the substance in immediate contact with the probe is nearly all liquid, with negligible interference from the crystals. In contrast, in this specific case the monohydrate needles were of dimensions comparable to the penetration depth; although initially negligible at low solid density, the interference from this solid phase became increasingly significant as run 1 progressed with the rising monohydrate form solid density. The kinetic expressions for the crystallization of the monohydrate form (eqn (14) and (17)) contain a total of 6 parameters to be estimated simultaneously, fairly larger than that for the anhydrate form. Unlike the anhydrate-form growth kinetic parameters that were determined from anhydrate-form seeded experiments, it is difficult to reduce the number of estimated parameters for the monohydrate form by first evaluating its growth kinetics solely in a separate experiment using monohydrate seed crystals. This is because it is very difficult to estimate the CSD (and its moments) of the monohydrate seeds reliably, given its smaller dimensions (thickness typically 2–10 μm, see Fig. 1b) and the tendency of the needles to aggregate when in the dried state and ease of breaking during handling. Inaccurate estimation of CSD moments as the initial conditions in the system of ODEs will lead to unreliable models and kinetic parameters. The variants of the power-law model for both the nucleation and growth of the monohydrate form were also not investigated. Incorporating the temperature dependence using the Arrhenius equation will lead to a total of 9 parameters, or more if the parameter describing size-dependence growth is included as well. Simultaneous estimation of a large number of crystallization parameters results in a wide range of parameters that produce similar quality fit to the data with highly correlated uncertainty descriptions.[32]

Table 6 lists all the parameter estimates from Stages 1, 2, and 3. The growth exponent ranges from 1 to 2 for diffusion-limited and surface integration-limited growth, respectively; the estimates indicate that the growth of anhydrate crystals is likely be controlled more by the latter, while the growth of the monohydrate form is affected by both. In the secondary nucleation of the monohydrate form, the larger estimate for $b_{1,m}$ relative to $b_{2,m}$ suggests that nucleation on its own surfaces may have a stronger dependence on S_m compared to nucleation on the crystal surfaces of the anhydrate form.

Confidence intervals for the parameter estimates

Parameter uncertainties were estimated by linearization of the model[43,44] around the vector of the parameter estimates:

$$Y^{\mathrm{md}}(\theta) \approx Y^{\mathrm{md}}(\theta^*) + F_J(\theta - \theta^*), \tag{26}$$

where $\theta^T = [\ln(k_{g,a}), g_a, \ln(E_{g,a}),\ldots, g_m]$, $p = 1, 2,\ldots, 12$, the Jacobian:

$$F_J = \left.\frac{\partial Y^{\mathrm{md}}}{\partial\theta}\right|_{\theta^*}, \tag{27}$$

Y^{md} is a vector of model outputs C^{md} and $\mu_j^{c,\mathrm{md}}$ for $j = 1$, 2, and 3. The Jacobian was computed using central differences; this matrix contains the first derivative of the model outputs with respect to θ, arranged so that derivatives with respect to a particular parameter, θ_p, are in the same column. The parameter covariance matrix for the linearized problem is given by:[43]

$$V_\theta^{-1} = F_J^T V_Y^{-1} F_J, \tag{28}$$

where V_Y is the covariance matrix for Y^{md}. The $100(1-\alpha)\%$ confidence interval[44] for each parameter is described by:

$$\theta_p^* - t_{1-\alpha/2}(N_\mathrm{t} - N_\theta)\sqrt{V_{\theta,\mathrm{pp}}} \le \theta_p \le \theta_p^* + t_{1-\alpha/2}(N_\mathrm{t} - N_\theta)\sqrt{V_{\theta,\mathrm{pp}}}, \tag{29}$$

which are listed in Table 6 for all parameter estimates, where t is the Student t-distribution. The F_J matrices from all three parameter estimation stages satisfied the identifiability condition,[43]

$$|F_J^T F_J| \neq 0, \tag{30}$$

indicating that all of the parameters could be identified from the experimental data.

Further analysis was performed by evaluating the linear correlation matrix R_c which measures the degree of linear dependence between a pair of parameters, where a magnitude of 1 indicates perfect correlation and a value of 0 indicates no correlation. Whenever all the off-diagonal elements have magnitude exceeding 0.8, the estimates are highly correlated and tend to be inaccurate. These correlation coefficients in R_c are given by:[43]

Table 6 Parameter estimates with 95% confidence intervals

Kinetic parameter	Parameter estimates, θ^*	95% Confidence interval	Number of independent measurements, N_t
Parameter estimation Stage 1: anhydrate form growth			
$\ln k_{g,a}$	42.330	[41.6, 43.0]	1424
g_a	1.760	[1.61, 1.91]	
$\ln E_{g,a}$	11.472	[11.45, 11.49]	
Parameter estimation Stage 2: anhydrate form nucleation			
$\ln k_{b,a}$	73.502	[71.0, 76.0]	251
b_a	1.952	[1.83, 2.08]	
$\ln E_{b,a}$	12.301	[12.23, 12.37]	
Parameter estimation Stage 3: monohydrate form nucleation and growth			
$\ln k_{b1,m}$	−2.295	[−2.58, −2.01]	584
$b_{1,m}$	2.502	[2.19, 2.82]	
$\ln k_{b2,m}$	−10.200	[−10.8, −9.62]	
$b_{2,m}$	1.998	[1.74, 2.25]	
$\ln k_{g,m}$	1.135	[0.96, 1.30]	
g_m	1.497	[1.44, 1.55]	

$$R_{c,ij} = V_{\theta,ij}(V_{\theta,ii}V_{\theta,jj})^{-1/2}. \quad (31)$$

The parameter covariance and correlation matrices for Stage 1 of the parameter estimation were:

$$V_\theta = \begin{bmatrix} \ln k_{g,a} & g_a & \ln E_{g,a} \\ 1.341 \times 10^{-1} & 8.189 \times 10^{-3} & 2.025 \times 10^{-3} \\ 8.189 \times 10^{-3} & 6.095 \times 10^{-3} & -1.939 \times 10^{-4} \\ 2.025 \times 10^{-3} & -1.939 \times 10^{-4} & 8.699 \times 10^{-5} \end{bmatrix}$$

$$R_c = \begin{bmatrix} 1.000 & 0.286 & 0.593 \\ 0.286 & 1.000 & -0.266 \\ 0.593 & -0.266 & 1.000 \end{bmatrix}$$

The off-diagonal terms of R_c indicate that, while there is some correlation between parameter estimates, the correlations are not excessive. The correlations between parameter estimates in Stage 2 are comparable in magnitude:

$$V_\theta = \begin{bmatrix} \ln k_{b,a} & b_a & \ln E_{b,a} \\ 1.577 & 3.620 \times 10^{-2} & 2.725 \times 10^{-2} \\ 3.620 \times 10^{-2} & 4.047 \times 10^{-3} & -9.662 \times 10^{-4} \\ 2.725 \times 10^{-2} & -9.662 \times 10^{-4} & 1.259 \times 10^{-3} \end{bmatrix}$$

$$R_c = \begin{bmatrix} 1.000 & 0.453 & 0.612 \\ 0.453 & 1.000 & -0.428 \\ 0.612 & -0.428 & 1.000 \end{bmatrix}$$

The corresponding matrices for Stage 3 are below.

$$R_c = \begin{bmatrix} 1.000 & 0.727 & -0.083 & 0.267 & 0.287 & -0.336 \\ 0.727 & 1.000 & 0.015 & 0.657 & 0.771 & -0.301 \\ -0.083 & 0.015 & 1.000 & 0.675 & -0.232 & -0.104 \\ 0.267 & 0.657 & 0.675 & 1.000 & 0.355 & -0.461 \\ 0.287 & 0.771 & -0.232 & 0.355 & 1.000 & 0.225 \\ -0.336 & -0.301 & -0.104 & -0.461 & 0.225 & 1.000 \end{bmatrix}$$

While the off-diagonal elements of R_c do not exceed 0.8, some of the elements are very near to that value. This was expected, as the number of parameters estimated in Stage 3 is significantly larger than Stages 1 and 2. Correlation coefficients larger than 0.8 would likely occur if more parameters were estimated, which would occur by including the temperature dependence in the nucleation and growth expressions for the monohydrate form.

Model validation

The crystallization model for L-phe was validated in two parts. The first validation compares the model prediction of the average size of the anhydrate crystals and the composition of the product crystals to offline measured data for runs 1, 2, and 3. The second validation compares the metastable limit predicted from the model to that determined experimentally.

A plot of the average size of the i-form crystals, $L_i = \mu_{1,i}/\mu_{0,i}$, from the model indicates that the three runs covered a range of growth rates (Fig. 5a and b), which was required to estimate independent values for the growth exponents and pre-exponential factors. The associated crystal composition is shown

$$V_\theta = \begin{bmatrix} \ln k_{b1,m} & b_{1,m} & \ln k_{b2,m} & b_{2,m} & \ln k_{g,m} & g_m \\ 2.115 \times 10^{-2} & 1.703 \times 10^{-2} & 3.585 \times 10^{-3} & 5.024 \times 10^{-3} & 3.710 \times 10^{-3} & -1.331 \times 10^{-3} \\ 1.703 \times 10^{-2} & 2.590 \times 10^{-2} & 7.046 \times 10^{-4} & 1.369 \times 10^{-2} & 1.101 \times 10^{-2} & -1.319 \times 10^{-3} \\ 3.585 \times 10^{-3} & 7.046 \times 10^{4} & 8.751 \times 10^{-2} & 2.584 \times 10^{-2} & -6.108 \times 10^{-3} & -8.402 \times 10^{-4} \\ 5.024 \times 10^{-3} & 1.369 \times 10^{-2} & 2.584 \times 10^{-2} & 1.675 \times 10^{-2} & 4.080 \times 10^{-3} & -1.623 \times 10^{-3} \\ 3.710 \times 10^{-3} & 1.101 \times 10^{-2} & -6.108 \times 10^{-3} & 4.080 \times 10^{-3} & 7.888 \times 10^{-3} & 5.434 \times 10^{-4} \\ -1.331 \times 10^{-3} & -1.319 \times 10^{-3} & -8.402 \times 10^{-4} & -1.623 \times 10^{-3} & 5.434 \times 10^{-4} & 7.416 \times 10^{-4} \end{bmatrix}$$

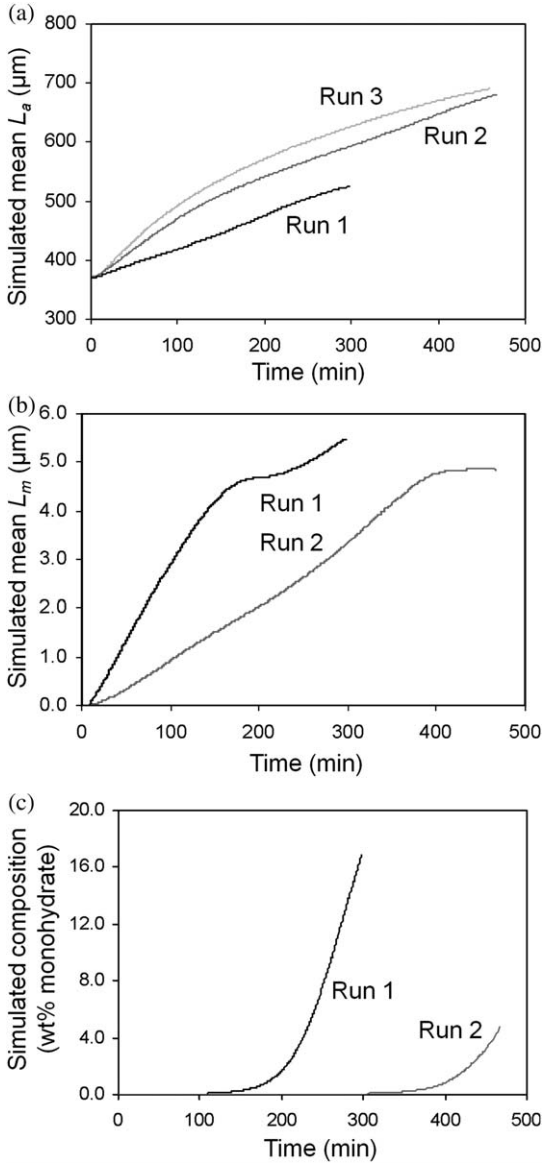

Fig. 5 Simulated mean lengths (a) L_a and (b) L_m, and (c) overall crystal composition (wt% monohydrate). The predicted mass of the solid phase for the *i*-form, $\mu_{3,i}k_{v,i}\rho_i m_{\text{solv}}$, was used to calculate the polymorph composition from $\mu_{3,m}k_{v,m}\rho_m m_{\text{solv}}/(\mu_{3,m}k_{v,m}\rho_m m_{\text{solv}} + \mu_{3,a}k_{v,a}\rho_a m_{\text{solv}})$.

in Fig. 5c. The average % difference between the model and experimental anhydrate product size is ~11% (Table 7), while crystal compositions had somewhat larger deviations. The product thickness L_m of the monohydrate form predicted by the model was ~5 μm (Fig. 5b), which is consistent with that observed from the distinct individual monohydrate needles in a PVM image obtained towards the end of run 1.[11] No comparison is shown for the CSD for the monohydrate needles in the final product due to their tendency to aggregate when in the dried state and as ease of breaking during handling. The CSD for the monohydrate form was difficult to determine using offline microscopy; any experimental data estimated in this manner would be highly inaccurate and meaningless for comparison.

The second validation compares the model predictions to the results from seeded metastable limit experiments (runs 1m to 5m; Fig. 6 with operating conditions listed in Table 1). The anhydrate seeds were from the same group of crystals as used in runs 1, 2, and 3. The initial conditions for the moments in the model were calculated using eqn (7)–(9), and the same operating conditions were applied (seeding temperatures and temperature profiles) as in the experiments. The model and experimental solute concentration and μ_1^s (due to both anhydrate and monohydrate forms) have good agreement, providing some support for the predictive ability of the model.

Table 7 Comparison of simulated and experimental results: mean L_a and product composition

Run	Simulation	Experimental
	Mean size of anhydrate product crystals/μm	
1	524	597, bin range 540–630
2	679	614, bin range 540–630
3	690	621, bin range 540–630
	Monohydrate of product crystals (wt%)	
1	16.75	9.45 ± 1.81
2	4.50	0.53 ± 1.81

[a] The experimental results were determined from independent offline characterization of sample product crystals, with the mean size computed from $\mu_{1,a}^s/\mu_{0,a}^s$ with $\mu_{j,a}^s$ calculated from eqn (7) applied to the anhydrate product size distribution determined from optical microscopy.

The modeled metastable limit was determined by assessing the evolution of the number and mean size of the nucleated crystals. Experimentally, the onset of cross-nucleation was evaluated using the PVM, which has a detection limit ≈ 3 μm.[35,45] Considering the possible spread or distribution in L_m, the modeled metastable limit was determined when its mean size, $\mu_{1,m}/\mu_{0,m}$ reached 2.5 μm (Fig. S4a†), slightly smaller than the experimental detection limit. In each of the five runs, the corresponding solute concentration C and temperature T were determined at the time point at which $\mu_{1,m}/\mu_{0,m}$ reached this value. Fig. S5a† shows good agreement between the model and experimental metastable limits, supporting the predictive ability of the model for the monohydrate form. For the anhydrate form, the system of ODEs was modified slightly before carrying out similar analysis as above. The nucleation rate B_a was set to zero in the original set of ODEs pertaining to the anhydrate form, so that this set of ODEs only described the growth of the anhydrate seeds. Another set of ODEs (eqn (3) and (4)) was incorporated to describe the nucleated anhydrate crystals. The model metastable zone for the anhydrate form determined in similar fashion (Fig. S4b†) as above was very narrow and inconsistent with the experimental results which were determined using FBRM (detection limit ≥ 1 μm). Further analysis revealed that the corresponding increase in the model $\mu_{0,a}$ was <0.5% from an initial value of 450 (Fig. S4d†). This indicated an exceedingly small number of nucleated anhydrate crystals, unlike the monohydrate form which had corresponding simulated values of $\mu_{0,m}$ ranging from 2000 to 3000 (Fig. S4c†), when the simulated mean $L_m = 2.5$ μm. In this case, evaluating the predicted metastable limit based solely on the simulated mean size of the nucleated crystals will not yield realistic results comparable to experimental findings given the noise in the measured CLD moments and possible masking effect in delaying the detection of nucleation. A more suitable

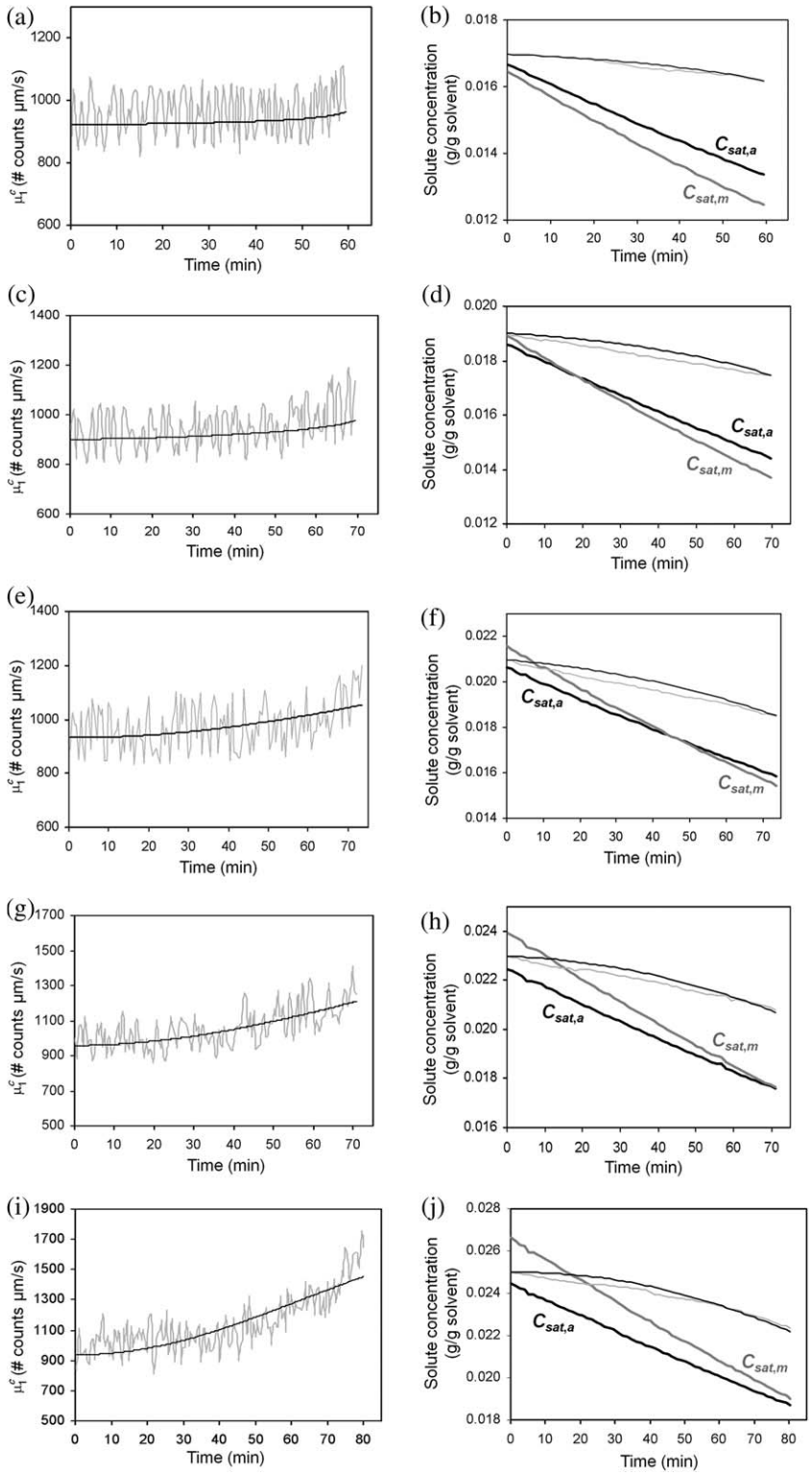

Fig. 6 First-order CLD moment μ_1^c and solution concentration for the metastable limit experiments, experimental (—); model (—): (a and b), run 1m; (c and d), run 2m; (e and f), run 3m; (g and h) run 4m; (i and j) run 5m. The solubility of each crystal form computed from the measured temperature is shown as thick grey and black lines in (b,d,f,h,j).

criterion is to apply a certain threshold for the increase in the model $\mu_{0,a}$ (assumed to be 10% in this study) in which enough nuclei are produced to be detected by FBRM. The corresponding values of the solute concentration C and temperature T were used to plot the model metastable limit for the anhydrate form in Fig. S5b†, which shows good agreement to the experimental metastable limit. Given the arbitrary selection for the threshold increase in $\mu_{0,a}$, this analysis does not fully

validate the crystallization model for the anhydrate form, but does demonstrate predictive ability to some extent.

Staged *vs.* simultaneous parameter estimation

Any person who has constructed a process model for a sufficiently complex system has used some sort of staged approach, in which simplified experiments are designed to enable the estimation of individual sets of parameters. An advantage of the staged approach is that numerical optimizations for eqn (22) are more likely to converge to the global solution, due to the smaller number of parameters. The optimization in eqn (22) is rarely convex for nonlinear dynamic processes with more than 10 parameters, such as those that occur in crystallizations involving transformation between different crystal structures, in which case numerical optimization algorithms can converge to a local solution that may produce both a poor fit to experimental data and a model with poor predictive ability. Numerical optimization over three parameters such as in Stages 1 and 2 is more likely to converge to the global solution by numerical optimization, and could even be solved graphically by gridding over the kinetic parameters.

Another advantage of the staged approach is that it reduces the effect of model bias on the initial sets of estimated parameters. To understand this argument, consider that all process models have limitations.[46–49] For example, most crystallization models assume that a lab-scale crystallizer is well-mixed and that all crystals of a particular structure have exactly the same crystal shape, which are only approximations. Due to these limitations, parameters estimated by fitting data to the outputs of these models are biased.[50] A goal of process modeling is for this bias to be sufficiently small that the process model is accurate enough for its intended purposes (such as design). An advantage of the staged approach to experimental design is that the bias in the kinetic parameters associated with parameter estimation in an early stage is only affected by model assumptions for the crystallization phenomena occurring at that and earlier stages—not on model assumptions associated with crystallization phenomena that only occur in later stages. Alternatively, the simultaneous estimation of parameters can produce very large biases. Actually, before applying the staged approach we attempted to simultaneously fit all parameters to all experimental data, which resulted in poor fits to the Stage 1 and Stage 2 data, as the simultaneous parameter estimation shifted those kinetic parameters to achieve a better fit to the remaining data. The staged approach also made it easier to identify the faulty data produced by the *in situ* ATR-FTIR and FBRM probes (discussed earlier).

A weakness of the staged approach is that any biases in the parameter estimates at any particular stage will bias the parameter estimates in subsequent stages. Also, confidence intervals computed in subsequent stages (such as in Table 6) will underestimate the amount of uncertainty in the parameter estimates. These effects should be small in this study due to the small size of the confidence intervals in Stages 1 and 2 (see Table 6). While the effects of uncertainties on parameter estimates in one stage on the uncertainty in parameter estimates on subsequent stages can be quantified using similar techniques,[51–53] these methods rely on the underlying model structure being known and fully characterized mathematically, which is not true in the presence of model bias. For these reasons, it is important to evaluate model predictions by comparing with experiment data collected from a separate set of experiments (as done here).

The staged experimental designs implemented here are very different from the D-optimal designs that have been applied extensively to crystallizations involving one crystal structure.[32,40,53,54] D-optimal design can result in experimental designs in which all crystallization phenomena occur simultaneously, forcing the simultaneous estimation of all of the parameters. It should be possible, however, to add constraints in the crystallization phase diagram to the D-optimal approach so that staged experimental designs can be implemented that minimize the uncertainties in the parameters while retaining the advantages of stage design. This would combine the best features of the two approaches.

Conclusions

A process model for the crystallization of L-phenylalanine, an enantiotropic pseudo-dimorph system, was developed. The kinetic parameters for nucleation and growth for both the anhydrate and monohydrate forms were estimated using *in situ* probes (solute concentration obtained from ATR-FTIR spectroscopy and CLD moments measured using FBRM) from experiments in which feedback controlled motion in the crystallization phase diagram suppressed certain kinetic events (for example, cross-nucleation) in particular segments of the experimental data. This simplified the crystallization model and reduced the number of parameters to be estimated simultaneously, by focusing on the relevant segments of the experimental data pertaining only to certain kinetics. This facilitated the estimation of the parameters in a stage-wise manner for each set of kinetics. Such an approach is advantageous for crystallization models with a large number of parameters such as for pseudo-polymorphic and polymorphic systems. The predictive ability of the crystallization model was evaluated based on the model properties of the product crystals and metastable limit which were in good agreement with that from independent characterization and experiments. This combination of experimental design and process modeling may facilitate process modeling and development for other pharmaceutical compounds involving transformation between different crystal structures.

Acknowledgements

Mitsuko Fujiwara is acknowledged for providing technical advice on this project. The authors thank Professor Emeritus Ted White from the University of Queensland, Australia for providing helpful references. The authors thank Scott Wilson and Maya Ramesh from the 3M Materials Laboratory at the University of Illinois for the PXRD data collection and James Mabon for the SEM imaging carried out in the Frederick Seitz Materials Research Laboratory Central Facilities at the University of Illinois, which are partially supported by the US Department of Energy under grants DE-FG02-07ER46453 and DE-FG02-07ER46471. Financial support is also acknowledged from the Singapore Agency for Science, Technology and Research.

References

1 J. M. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
2 N. Doki, M. Yokota, K. Kido, S. Sasaki and N. Kubota, *Cryst. Growth Des.*, 2004, **4**, 103–107.
3 N. Doki, H. Seki, K. Takano, H. Asatani, M. Yokota and N. Kubota, *Cryst. Growth Des.*, 2004, **4**, 949–953.
4 R. J. Davey, N. Blagden, G. D. Potts and R. Docherty, *J. Am. Chem. Soc.*, 1997, **119**, 1767–1772.
5 J. Yano, H. F. Milhofer, E. Wachtel and N. Garti, *Langmuir*, 2000, **16**, 10005–10014.
6 R. Mohan, K. K. Koo, C. Strege and A. S. Myerson, *Ind. Eng. Chem. Res.*, 2001, **40**, 6111–6117.
7 S. Gracin, M. Uusi-Penttila and A. C. Rasmuson, *Cryst. Growth Des.*, 2005, **5**, 1787–1794.
8 V. M. Profir and A. C. Rasmuson, *Cryst. Growth Des.*, 2004, **4**, 315–323.
9 M. Mirmehrabi and S. Rohani, *J. Pharm. Sci.*, 2005, **94**, 1560–1576.
10 M. Trifkovic and S. Rohani, *Org. Process Res. Dev.*, 2007, **11**, 138–143.
11 N. C. S. Kee, R. B. H. Tan and R. D. Braatz, *Cryst. Growth Des.*, 2009, **9**, 3044–3051; N. C. S. Kee, P. D. Arendt, R. B. H. Tan and R. D. Braatz, *Cryst. Growth Des.*, 2009, **9**, 3052–3061.
12 M. Fujiwara, P. S. Chow, D. L. Ma and R. D. Braatz, *Cryst. Growth Des.*, 2002, **2**, 363–370.
13 G. X. Zhou, M. Fujiwara, X. Y. Woo, E. Rusli, H.-H. Tung, C. Starbuck, O. Davidson, Z. Ge and R. D. Braatz, *Cryst. Growth Des.*, 2006, **6**, 892–898; Z. K. Nagy, J. W. Chew, M. Fujiwara and R. D. Braatz, *J. Process Control*, 2008, **18**, 399–407.
14 V. Liotta and V. Sabesan, *Org. Process Res. Dev.*, 2004, **8**, 488–494.
15 X. Y. Woo, R. B. H. Tan, P. S. Chow and R. D. Braatz, *Cryst. Growth Des.*, 2006, **6**, 1291–1303; G. P. Lian, S. Moore and L. Heeney, *Chem. Eng. Sci.*, 2006, **61**, 7819–7826; X. Y. Woo, R. B. H. Tan and R. D. Braatz, *Cryst. Growth Des.*, 2009, **9**, 156–164; G. di Veroli and S. Rigopoulos, *AIChE J.*, 2010, **56**, 878–892.
16 F. Wang, J. A. Wachter, F. J. Antosz and K. A. Berglund, *Org. Process Res. Dev.*, 2000, **4**, 391–395.
17 H. Groen and K. J. Roberts, *J. Phys. Chem. B*, 2001, **105**, 10723–10730.
18 C. Starbuck, A. Spartalis, L. Wai, W. Jian, P. Fernandez, C. M. Lindemann, G. X. Zhou and Z. Ge, *Cryst. Growth Des.*, 2002, **2**, 515–522.
19 Y. Hu, J. K. Liang, A. S. Myerson and L. S. Taylor, *Ind. Eng. Chem. Res.*, 2005, **44**, 1233–1240.
20 H. Qu, M. Louhi-Kultanen, J. Rantanen and J. Kallas, *Cryst. Growth Des.*, 2006, **6**, 2053–2060.
21 A. Caillet, F. Puel and G. Fevotte, *Int. J. Pharm.*, 2006, **307**, 201–208.
22 T. Ono, H. J. M. Kramer, J. H. Horst and P. J. Jansens, *Cryst. Growth Des.*, 2004, **4**, 1161–1167.
23 J. Schöll, D. Bonalumi, L. Vicum and M. Mazzotti, *Cryst. Growth Des.*, 2006, **6**, 881–891.
24 M. W. Hermanto, N. C. S. Kee, R. B. H. Tan, M. S. Chiu and R. D. Braatz, *AIChE J.*, 2008, **54**, 3248–3259.
25 A. Caillet, N. Sheibat-Othman and G. Fevotte, *Cryst. Growth Des.*, 2007, **7**, 2088–2095.
26 B. Khawas, *Indian J. Chem., Sect. A: Inorg., Phys., Theor. Anal.*, 1985, **59**, 219–226.
27 T. Togkalidou, M. Fujiwara, S. Patel and R. D. Braatz, *J. Cryst. Growth*, 2001, **231**, 534–543; T. Togkalidou, R. D. Braatz, B. K. Johnson, O. Davidson and A. Andrews, *AIChE J.*, 2001, **47**, 160–168; T. Togkalidou, H.-H. Tun, Y. Sun, A. Andrews and R. D. Braatz, *Org. Process Res. Dev.*, 2002, **6**, 317–322.
28 Y. Xie and J. Kalivas, *Anal. Chim. Acta*, 1997, **348**, 19–27.
29 J. Nyvlt, O. Sohnel, M. Matuchova and M. Broul, *The Kinetics of Industrial Crystallization*, Elsevier, New York, 1985.
30 J. B. Rawlings, S. M. Miller and W. R. Witkowski, *Ind. Eng. Chem. Res.*, 1993, **32**, 1275–1296; R. Gunawan, D. L. Ma, M. Fujiwara and R. D. Braatz, *Int. J. Mod. Phys. B*, 2002, **16**, 367–374; Z. K. Nagy, M. Fujiwara, X. Y. Woo and R. D. Braatz, *Ind. Eng. Chem. Res.*, 2008, **47**, 1245–1252.
31 A. D. Randolph and M. A. Larson, *Theory of Particulate Processes*, Academic Press, San Diego, 2nd edn, 1988.
32 S. M. Miller and J. B. Rawlings, *AIChE J.*, 1994, **40**, 1312–1327.
33 H. L. MacGillivray, *J. Appl. Probab.*, 1980, **17**, 956–967.
34 A. Tadayyon and S. Rohani, *Part. Part. Syst. Charact.*, 1998, **15**, 127–135.
35 E. J. Hukkanen and R. D. Braatz, *Sens. Actuators, B*, 2003, **96**, 451–459.
36 M. Li, D. Wilkinson and K. Patchigolla, *Part. Part. Syst. Charact.*, 2006, **23**, 170–174.
37 E. J. W. Wynn, *Powder Technol.*, 2003, **133**, 125–133.
38 A. Vaccaro, J. Šefcík and M. Morbidelli, *Part. Part. Syst. Charact.*, 2007, **23**, 360–373.
39 J. Worlitschek and M. Mazzotti, *Cryst. Growth Des.*, 2004, **4**, 891–903; M. N. Pons, K. Milferstedt and E. Morgenroth, *Chem. Eng. Sci.*, 2006, **61**, 3962–3973; L. Ehrl, M. Soos and M. Morbidelli, *Part. Part. Syst. Charact.*, 2007, **23**, 438–447; N. Kail, H. Briesen and W. Marquardt, *Part. Part. Syst. Charact.*, 2007, **24**, 184–192; N. Kail, H. Briesen and W. Marquardt, *Powder Technol.*, 2008, **185**, 211–222; N. Kail, H. Briesen and W. Marquardt, *Chem. Eng. Sci.*, 2009, **64**, 984–1000.
40 T. Togkalidou, H.-H. Tung, Y. Sun, A. T. Andrews and R. D. Braatz, *Ind. Eng. Chem. Res.*, 2004, **43**, 6168–6181.
41 www.mathworks.com/help/techdoc/ref/ode23.html.
42 www.mathworks.com/help/toolbox/optim/ug/fmincon.html.
43 J. V. Beck and K. J. Arnold, *Parameter Estimation in Engineering and Science*, Wiley, New York, 1977.
44 E. Rusli, F. Xue, T. O. Drews, P. Vereecken, P. Andracacos, H. Deligianni, R. D. Braatz and R. C. Alkire, *J. Electrochem. Soc.*, 2007, **154**, D584–D597.
45 E. J. Hukkanen and B. Smith, *Proceedings of the AIChE Spring National Meeting, Fifth World Congress on Particle Technology*, AIChE Press, New York, 2006, Abstract 156e.
46 G. E. P. Box, in *Robustness in Statistics*, ed. R. L. Launer and G. N. Wilkinson, Academic Press, New York, 1979.
47 K. Denbigh, *The Thermodynamics of the Steady State*, Mathuen & Co., New York, 1951.
48 R. Aris, *Mathematical Modeling Techniques*, Pitman Publishing, London, 1978.
49 M. M. Denn, *Process Modeling*, Longman, New York, 1986.
50 L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ, 1987.
51 D. L. Ma, S. H. Chung and R. D. Braatz, *AIChE J.*, 1999, **45**, 1469–1476; D. L. Ma and R. D. Braatz, *IEEE Trans. Control Syst. Technol.*, 2001, **9**, 766–774.
52 Z. K. Nagy and R. D. Braatz, *IEEE Trans. Control Syst. Technol.*, 2003, **11**, 694–704; Z. K. Nagy and R. D. Braatz, *J. Process Control*, 2004, **14**, 411–422.
53 H. B. Matthews, PhD thesis, University of Wisconsin-Madison, April 1997; H. B. Matthews and J. B. Rawlings, *AIChE J.*, 1998, **44**, 1119–1127.
54 S. H. Chung, D. L. Ma and R. D. Braatz, *Chemom. Intell. Lab. Syst.*, 2000, **50**, 83–90.