Supporting information for:

Robust Crystallization Process Development for the Metastable δ-form of Pyrazinamide

Martin Wijaya Hermanto,[*,†] Alvin Yeoh,[†] Beatrice Soh,[†] Pui Shan Chow,[†] and Reginald B. H. Tan[*,†,‡]

Institute of Chemical and Engineering Sciences Limited, 1 Pesek Road, Jurong Island, Singapore 627833, and Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

E-mail: martin_hermanto@ices.a-star.edu.sg; reginald_tan@ices.a-star.edu.sg

Phone: +65-67963851. Fax: +65-63166183

Quantification of Polymorphic Composition from Powder-XRD Pattern

Suppose we have a sample mixture of m pure phase components where its XRD pattern comprises N data points, then a system of linear equations can be constructed as follows:

$$\mathbf{x_1}p_1 + \mathbf{x_2}p_2 + \mathbf{x_3}p_3 + \ldots + \mathbf{x_m}p_m = \mathbf{s} \tag{1}$$

[*]To whom correspondence should be addressed
[†]Institute of Chemical and Engineering Sciences Ltd.
[‡]National University of Singapore.

where $\mathbf{s}$ is a $N \times 1$ vector containing the measured XRD pattern, $\mathbf{x_j}$ is a $N \times 1$ vector containing the XRD pattern of pure component j, and p_j is a scalar corresponding to the fraction arising from the XRD scattering power of sample due to component j. In matrix form, the above can be rewritten as

$$
\begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} & \cdots & \mathbf{x_p} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_m \end{bmatrix} = \mathbf{s} \tag{2}
$$

or

$$
\mathbf{X}\,\mathbf{p} = \mathbf{s}\,. \tag{3}
$$

Then, the estimated fractions of pure components $\mathbf{p_{est}}$ are obtained through the following minimization of ordinary least squares problem:

$$
\mathbf{p_{est}} = \arg \left\{ \min_{\mathbf{p}} \left(\mathbf{X}\,\mathbf{p} - \mathbf{s}\right)^T \left(\mathbf{X}\,\mathbf{p} - \mathbf{s}\right) \right\} \tag{4}
$$

subject to

$$
p_j \geq 0\,. \tag{5}
$$

The weight fraction of component n in a mixture comprising m components can be obtained by the following expression:

$$
c_n = p_n \frac{\mu^*}{\mu_n^*} \tag{6}
$$

where

$$
\mu^* = \sum_{j=1}^{m} c_j \mu_j^*\,, \tag{7}
$$

$$
\mu_j^* = \frac{\mu_j}{\rho_j}\,, \tag{8}
$$

μ_j and ρ_j are the atomic X-ray absorption coefficient and the density of component j, respectively. For this study, weight fraction measurement of α- and δ-forms are considered. The μ_j and ρ_j are available from the literature, where $\mu_j = 0.111$ and $0.113\,\mathrm{mm}^{-1}$, and $\rho_j = 1.496$ and $1.521\,\mathrm{g\,cm}^{-3}$ for the α- and δ-forms respectively.

One important issue with XRD quantification is the effect of preferred orientation due to minor variability in sample preparation. This is highlighted through the XRD patterns of 5 repeat samples of pure α- and δ-forms as shown in Figures S1a and S1b, respectively. All samples were ground by mortar and pestle as consistently as possible by the same individual. However, it is evident that the relative intensities of different peaks are inconsistent among samples, and cannot be resolved by normalization. Furthermore, there also appear to be slight peak shifts between samples.

In this study, two approaches were used to take into account such variability. Firstly, 5 repeat samples for each pure polymorph were used as reference XRD patterns ($\mathbf{x_j}$), that is $j = 1$ to 5 correspond to the α-form, and $j = 6$ to 10 correspond to the δ-form. The estimated weight fraction of each form is simply the summation of the corresponding c_n. Secondly, a weighted least squares method was used instead of the ordinary least squares described above. By using weighted least squares, it is possible to emphasize regions that are more relevant and deemphasize regions where variability is prominent due to the preferred orientation effect. Hence, the minimization problem (4) was modified as follows

$$\mathbf{p_{est}} = \arg\left\{\min_{\mathbf{p}}(\mathbf{X\,p} - \mathbf{s})^T\,\mathbf{W}^T\mathbf{W}\,(\mathbf{X\,p} - \mathbf{s})\right\} \tag{9}$$

where $\mathbf{W}$ is a $N \times N$ diagonal matrix in which each diagonal element corresponds to the weight given to each data point of the sample pattern. In other words, larger weights are assigned to emphasize relevant regions, and smaller weights to deemphasize less important regions. In this study, specific α-form regions $2\theta = 13 - 14$ and $14.8 - 15.8$ were considered important for the quantification of α-form and a weight of 2.5 was assigned to emphasize these regions. On the other hand, non-specific peaks around $2\theta = 25 - 29$ were subjected to significant variability due

to preferred orientation, especially in the XRD pattern for the δ-form (Figure S2). Therefore, these regions were deemphasized by assigning a weight of 0.3. The remaining regions were assigned weights of 1.

(a) (b)
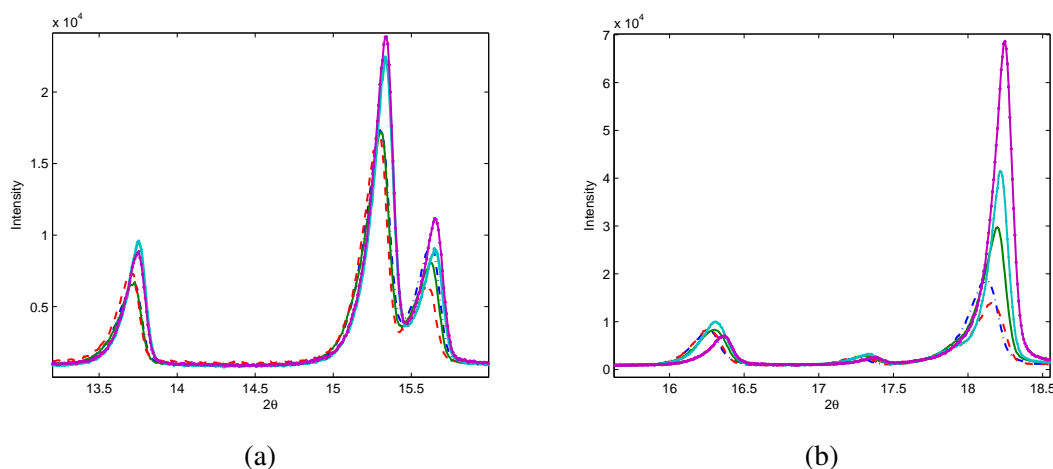
Figure S1: Selected XRD patterns of 5 different samples of (a) pure α-form and (b) pure δ-form
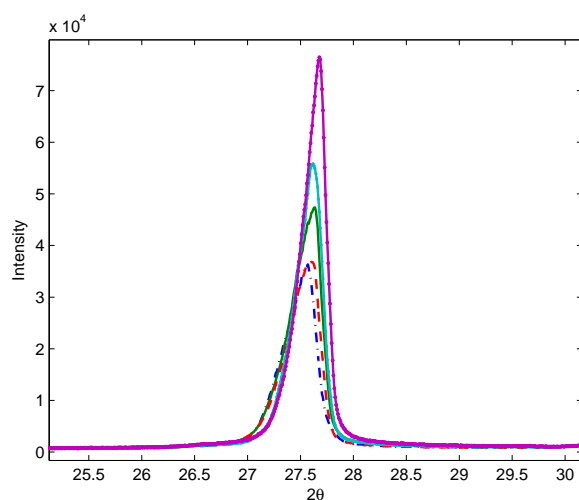

Figure S2: Variability in non-specific peaks of pure δ-form XRD pattern

Before being used for quantification, all XRD sample patterns were preprocessed to remove noise and correct baseline. To remove noise, Savitsky-Golay smoothing was utilized with polynomial order = 5 and frame size = 15. To correct baseline, local 5th order polynomial functions were fitted to the data, and then subtracted to produce a pattern with a flat baseline.

The accuracy of the improved method was compared to that of the original least squares method using XRD patterns of polymorphic mixtures with different known compositions. The predicted versus actual weight fractions of the α-form for both methods are shown in Figures S3a and S3b. It is evident that the accuracy of the original method becomes poorer as the actual weight fraction increases. On the other hand, the proposed method remains consistently accurate. The root mean square prediction error of the proposed method is about one fifth smaller (1.15%) than the original method (5.48%). Hence, the proposed method was used to quantify polymorphic composition throughout this study.
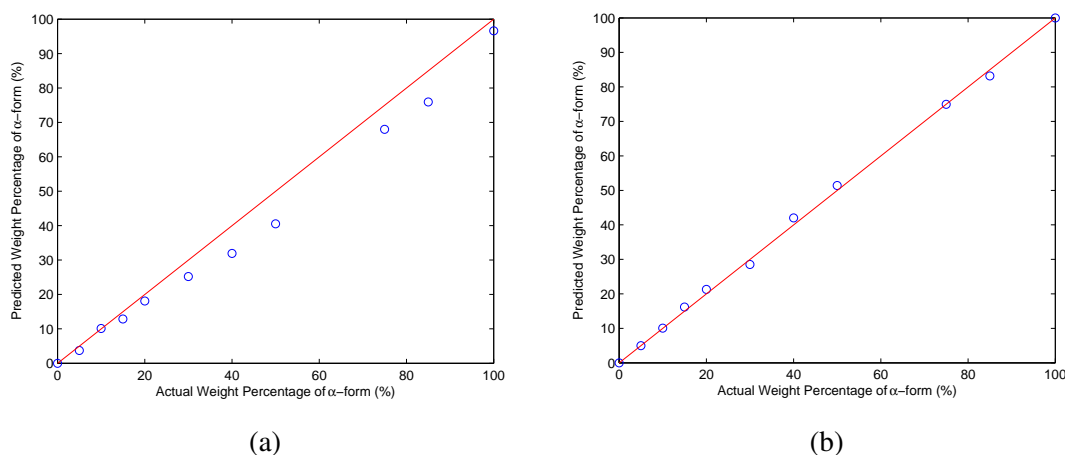

(a) (b)

Figure S3: Actual versus predicted weight percentage of α-form using the (a) original least squares method and (b) the proposed weighted least squares method

Solubility Measurements of the Metastable Forms

The solubility of the metastable δ and γ-forms were measured using a different technique due to the risk of conversion to the stable α-form taking place before complete dissolution. Multiple vials of solution saturated with a known quantity of the α-form were first prepared and held at different temperatures at intervals of 0.5 °C. 35 mg of the corresponding metastable form was then added to each of these vials. By monitoring the transmissivity profile of each vial in the Crystalline, we were able to determine temperature at which the additional mass of material dissolved. As the quantity

of material added to each vial was small, the dissolution time was shorter than the time required for polymorphic transformation to take place. This ensured that we were measuring the true solubility of the metastable form rather than that of the α-form or a mixture of both. Figure S4 shows a schematic of transmissivity plotted against the cells held at different temperatures, showing the change in transmissivity as the temperature increases across each cell.
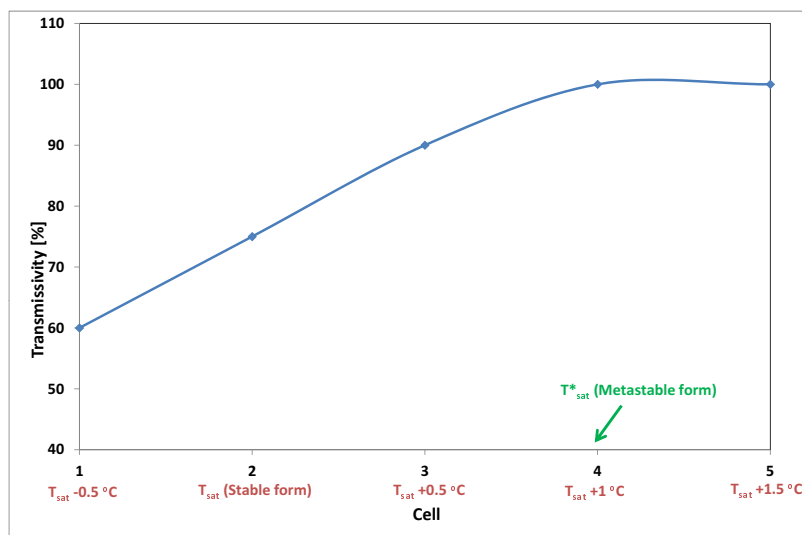

Figure S4: The transmissivity in cells 1 to 5, each held at ta different temperature. (T_{sat} is the saturation temperature of the stable α-form)

Response Surface Curves Development

Response surface curves for polymorphic impurity level were developed from the experimental data. In order to alleviate the non-constant variance problem and to improve model fit to the response variable (α-form impurity), Box-Cox transformation ($\hat{y}$) was performed on the response variable (y)

$$\hat{y} = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y}\ln(y) & \lambda = 0 \end{cases} \tag{10}$$

where $\dot{y}$ is the geometric mean of all the observations and λ is the Box-Cox transformation parameter to be determined (λ between -1 and 1 at an interval of 0.1 were used in this study).

For each value of λ, multi linear regressions were performed between the transformed response variable and the decision variables (i.e., cooling rate, seeding temperature, seed loading, solvent composition) to obtain various models (up to second order model). The combination of λ and model that minimized the sum of squared error (SSE) was selected. Figures S5a and S5b show that the optimum λ for both ultrasonic and ballmilled seed preparation techniques are 0.4 and 0.5, respectively.
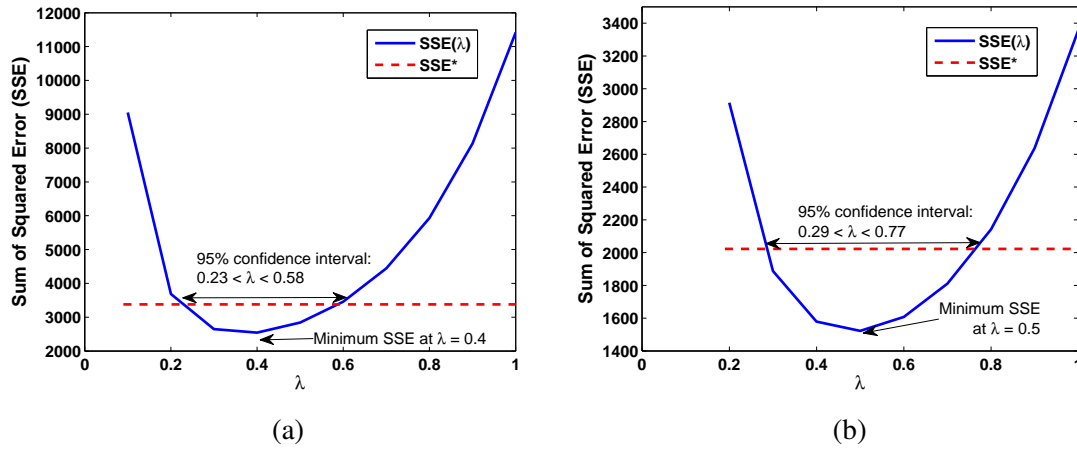

Figure S5: Plot of sum of squared error of prediction against Box-Cox transformation parameter λ for (a) ultrasonic seeds and (b) ballmilled seeds

To ensure the validity of Box-Cox transformation, it is important to estimate the confidence interval of λ to determine whether $\lambda = 1$ is included. If so, Box-Cox transformation is not required. Hence, the confidence interval for λ was estimated by computing

$$SSE^* = SSE_{\min}\left(1 + \frac{t_{\theta/2,v}^2}{v}\right) \tag{11}$$

where $t_{\theta/2,v}$ is the Student's t-distribution value with θ significance level and v degree of freedom. The confidence limits were obtained by locating the points on the λ axis where SSE^* cuts the curve $SSE(\lambda)$ (Figures S5a and S5b). Results show that both 95% confidence intervals do not include $\lambda = 1$, which justify the use of Box-Cox transformation. The final models that explain

the relations between the decision variables and the transformed response variables for both seed preparation techniques are tabulated in Table S1. The (non-transformed) response variable was then calculated by the inversion of equation (10).

Table S1: Model components and their coefficients for the transformed response variables for both seed preparation techniques (where A: Cooling Rate, B: Seeding temperature, C: Seed loading, D: Solvent composition, E: 1/Cooling Rate)

No.	Model Components	Coefficients (Ultrasonic)	Coefficients (Ballmilled)
1	Intercept	3.7069×10^3	3.4015×10^3
2	A	-3.4390×10^1	-5.9709×10^1
3	B	-1.7247×10^2	-1.5185×10^2
4	C	-2.6076×10^1	-5.0069×10^1
5	D	-3.5909×10^{-1}	2.0739×10^0
6	E	6.2032×10^0	-2.2924×10^1
7	AB	3.0296×10^0	1.1515×10^1
8	AD	1.0803×10^{-1}	-7.1000×10^{-1}
9	BC	5.0235×10^{-1}	7.8915×10^{-1}
10	BE	-1.6872×10^{-1}	4.8525×10^{-1}
11	CE	4.2874×10^{-1}	9.4371×10^{-1}
12	B^2	2.0289×10^0	1.7279×10^0
13	C^2	-1.4417×10^{-2}	3.8325×10^{-1}
14	D^2	-3.4237×10^{-2}	-1.1436×10^{-1}