Supplementary Material for: Controlling the crystal polymorph by exploiting the time dependence of nucleation rates

Laurie J. Little, 1 Alice A. K. King, 1, 2 Richard P. Sear, 1, a) and Joseph L. Keddie 1

1) Department of Physics, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom
2) Department of Physics and Astronomy, University of Sussex, Brighton, BN1 9RH, United Kingdom

(Dated: 25 September 2017)

This supplementary material has six sections:

(I) We discuss literature values of glycine crystal growth rates in the main text. In this section we plot the rates as a function of supersaturation and we give details of the experimental conditions in these studies.

(II) This contains plots of the solubility of glycine in water, as a function of temperature.

(III) This section contains Raman spectra of the α and γ polymorphs of glycine. We also show the X-ray diffraction (XRD) results that validate Raman spectroscopy’s ability to distinguish between the two polymorphs of glycine.

(IV) Background to the statistics needed to understand the nucleation of competing polymorphs, and details of the models we used.

(V) Additional experimental results for nucleation.

(VI) Additional experimental results for growth.

---

a) Electronic mail: r.sear@surrey.ac.uk
Figure S1. Glycine crystal growth rates, are plotted as a function of supersaturation. This is data from five earlier studies: [Li] (Li et al.) \(^1\), [Han-A] (Han et al.) \(^2\), [Han-B] (Han et al.) \(^3\), [Dow] (Dowling et al.) \(^4\) and [Sul] (Sultana et al.) \(^5\). These growth rates have been obtained from graphs, so are approximate. Colour coding is as follows, \(\alpha\) b-axis, \(\alpha\) c-axis, \(\gamma\) b-axis and \(\gamma\) c-axis are pink, green, gold and dark blue respectively. [Sul] used an anti-solvent, other studies are purely aqueous.

I. LITERATURE STUDIES OF THE GROWTH OF GLYCINE CRYSTALS

In the main text we discussed literature values for the growth rates of glycine crystals. These rates are plotted in Figure S1. There was one growth rate \(\approx 80,000 \mu\text{m/min}\), which was too large to fit on our graph. The experimental details for these experiments are summarised in Table S1.

Crystal growth rates appear to be very sensitive to crystal and solution properties, such as crystal size and solvent. The growth rates of \(\alpha\) glycine measured by Sultana et al. are lower than those measured by Han et al. and Dowling et al. In the work of Sultana et al., the glycine solution is supersaturated by the addition of the anti-solvent methanol, while in the experiments of Han et al. and Dowling et al. only water and glycine are present. Also, both Dowling and Han work with crystals with sizes of the order of mm, while Sultana et al. work with crystals of sizes of the order tens of \(\mu\text{m}\). It may be that the presence of
Table S1. The experimental details of a number of studies on glycine growth. The growth rates from these studies are shown in Figure S1.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Supersaturation range</th>
<th>Experimental set-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toldy et al.</td>
<td>3.5 – 6.5</td>
<td>Crystals nucleate in supersaturated droplets within an emulsion at 84°C. Crystals sizes are of the order of tens of µm.</td>
</tr>
<tr>
<td>Dowling et al.</td>
<td>1.3 – 1.45</td>
<td>Individual crystal seeds are placed in supersaturated solution, and observed under a microscope at 20°C. Crystal sizes are of the order of mm. Solubility $c_s$ is given as 212 g/l.</td>
</tr>
<tr>
<td>Han et al.</td>
<td>1.15 – 1.6</td>
<td>Individual crystal seeds are placed in supersaturated solution and observed under a microscope at 23°C. Crystal sizes are of the order of mm. Solubility $c_s$ is given as 226 g/l.</td>
</tr>
<tr>
<td>Sultana et al.</td>
<td>1.08 – 1.75</td>
<td>Supersaturated solutions containing crystal seeds flow through a PDMS (poly-dimethylsiloxane) microfluidic device. Solutions are supersaturated using the anti-solvent methanol. Crystal sizes are of the order tens of µm</td>
</tr>
<tr>
<td>Li et al</td>
<td>1.01 – 1.08</td>
<td>Crystal seeds are placed in a glass cell at 23°C while supersaturated solution flows through the cell, seed sizes are of the order of hundreds of µm</td>
</tr>
<tr>
<td>Han et al</td>
<td>1.5</td>
<td>Set-up is the same as in Dowling et al.</td>
</tr>
</tbody>
</table>

methanol is affecting the growth of crystals in the experiments of Sultana et al, and the crystal growth rate may also be changing with crystal size.

II. GLYCINE SOLUBILITY

For glycine in water the solubility varies as shown in Figure S2. We can see that glycine’s solubility is very sensitive to temperature. This is beneficial in that it allows us to easily create highly supersaturated solutions (by cooling), but problematic in that to perform an experiment at constant supersaturation, temperature must be controlled very precisely. The curve in Figure S2(a), is used for our supersaturation calculations. We include solubility data from several sources in Figure S2(b) to show that values for glycine solubility in the literature can vary significantly between studies.
Figure S2. Solubility in water of the α and γ polymorphs of glycine as a function of temperature. (a) Data from Yang et al. We have fit a second order polynomial (black curve), so that the solubility can be approximated between data points. This fit is solubility (g/kg) = 0.0301T^2 + 2.96T + 109, with T in °C. (b) Comparison of γ glycine solubility from several sources: Yang et al., Dowling et al., Han et al. and Yi et al.

III. VALIDATION OF RAMAN SPECTRA FOR IDENTIFYING THE α AND γ POLYMORPHS OF GLYCINE

The α and γ crystals have distinct Raman spectra, which can be identified by a set of distinctive peaks. Examples of spectra from each polymorph are given in Figure S3(b). Both polymorphs have a very intense peak at 886 cm\(^{-1}\). There are clear differences between the two spectra in the region 100-200 cm\(^{-1}\) where the peaks can be attributed to intermolecular vibrations. The α polymorph has several small peaks (relative to the 886 cm\(^{-1}\) peak) at 118 cm\(^{-1}\), 164 cm\(^{-1}\), 171 cm\(^{-1}\) and 203 cm\(^{-1}\), which the γ polymorph does not have. The 203 cm\(^{-1}\) peak is often of low intensity and the peaks at 164 cm\(^{-1}\) and 171 cm\(^{-1}\) are often merged. Spectra where these peaks are more clearly visible can be seen in Figure S3(c) and (d). The γ polymorph has a very intense peak (comparable in intensity to the 886 cm\(^{-1}\) peak) at 157 cm\(^{-1}\), which the α polymorph does not have. There are also differences between the polymorphs at \(\approx 500\) cm\(^{-1}\). The α polymorph has two low intensity peaks at 492 cm\(^{-1}\) (NH\(_3\) torsional mode) and 502 cm\(^{-1}\) (CO\(_2\) rocking mode) while the γ polymorph has just one high intensity peak at 504 cm\(^{-1}\).

X-ray diffraction (XRD) was used as a test to confirm the results of our Raman spectroscopy analysis. A sample of crystals was analysed with Raman spectroscopy (the two example spectra shown in Figure S3(b)). Immediately after the Raman spectroscopy had been carried out, the crystals were removed from the microplate and separated into two
Figure S3. (a) XRD patterns of two powdered samples, one of α glycine and one of γ glycine as identified with Raman spectroscopy. The glycine identified as α via Raman spectroscopy is shown in blue and glycine identified as γ is shown in green. The circles represent known α XRD peaks while the triangles represent known γ XRD peaks. (b)(i) Raman spectra for a typical α glycine crystal and a typical γ glycine crystal. The spectra are normalised to the intensity of their highest peak at 886 cm⁻¹ (b)(ii) The spectra of (b)(i) in the region 100 - 250 cm⁻¹. (b)(iii) The spectra of (b)(i) in the region 450 - 550 cm⁻¹. (c) and (d) Two additional α glycine spectra where the characteristic peaks at 164 cm⁻¹, 171 cm⁻¹ and 203 cm⁻¹ (which cannot be easily seen in (b)(ii)) can be more clearly seen.

groups: Those that had been identified as α glycine and those that had been identified as γ glycine. The two samples were then prepared for XRD using the same procedure as described in earlier work[10]. The crystals were ground up into a fine powder before analysis. Powder XRD was carried out using a PANalytical XPer Pro diffractometer across a 2θ
Figure S4. Experimental XRD patterns of two powdered samples, one of α glycine (blue dotted curve) and one of γ glycine (green dotted curve), as identified with Raman spectroscopy. The reference peak positions for all three polymorphs are shown as solid vertical lines. Our PANalytical XPert Pro diffractometer uses JCPDS reference files, numbers 00-032-1702 (α), 00-002-0171 (β) and 00-006-0230 (γ). We show a 2θ range where there are three strong β peaks that are all well separated from any α or γ peak. None of the experimentally observed peaks match the β reference file. All five patterns are normalised such that the highest peak has a height of one.

As expected, the set of crystals which we identified as α with Raman spectroscopy are found to show only α glycine peaks with powder XRD. The Raman spectra and XRD were also in agreement for the set of γ crystals. We never observed the third polymorph of glycine, β glycine, in our XRD patterns. The β polymorph can form needle-like crystals\textsuperscript{[11]}.
resembling those we do see, but the XRD patterns rule out significant amounts of the β polymorph, at the point at the end of the experiment when XRD patterns are taken. In Figure S4 we have plotted our XRD patterns together with reference α, γ and β patterns. Note that this range of 2θ includes the strongest peak of the β polymorph (at 2θ = 18.1°) which is absent in both our sets of diffractograms.

During the course of our experiments, we do not observe habit changes of our crystals. This suggests that once formed, our crystals do not transform into another polymorph. However, both Raman and XRD analyses are made at the end of our runs, which is days after many of the crystals formed. Thus, although it seems unlikely, we cannot rule out a polymorph-to-polymorph transformation during our experiment that does not affect habit.

A. Raman spectra for needle-like and non-needle-like crystals

We were surprised by our observation that both polymorphs formed two very different crystal habits, in each case a needle-like habit and a much more compact habit. So we looked carefully at the Raman spectra we obtained from these crystals. Raman spectra for needle-like and non-needle crystals are shown in Figure S5 for the α polymorph and in Figure S6 for the γ polymorph. In Figure S5(a) we see Raman spectra from 50 needle and non-needle-like crystals, of the α polymorph. Figure S5(c) is a scatter plot of estimates of the height of the peak nominally located at 118 cm$^{-1}$, versus its position, and the height of the combined 164 and 171 cm$^{-1}$ peaks, versus their location. The peak heights and positions are obtained by fitting the sum of two Gaussians plus a constant background term to the spectra over the range 100 to 200 cm$^{-1}$. Note that the Raman spectrum of the β polymorph is distinct from that of the α and γ polymorphs\textsuperscript{12,13}. The Raman spectrum of the β is flat between approximately 120 and 200 cm$^{-1}$, whereas we see peaks in that range, in both Figures S5 and S6.

Note that there is clearly significant variability in the peak height, but that almost all estimates of the peak position are within a narrow range. The distributions of the peak heights and positions of the needle and non-needle crystals are indistinguishable. As far as the Raman spectra are concerned, the needle-like and non-needle-like both have the characteristic α polymorph peaks at the same position, and in both cases with variable intensity. Whatever the cause of the fast growth along one axis of the needles, it is not
Figure S5.  (a) Raman spectra for 24 non-needle like (solid curves) and 26 needle-like (dashed curves) crystals, all of which we identified as the $\alpha$ polymorph from its Raman spectrum. All crystals were obtained from run number one at a salt concentration of 250 mg/ml. In two cases (green and orange dashed curves) there are anomalous spikes at single values in the data, which may be due to energetic events (cosmic rays) in the detector.  (b) Four each of the non-needle (solid curves) and needle spectra (dashed curves) of (a), plotted in the range we use for fitting. (c) Plot of peak height versus peak position, obtained by a fit of two Gaussians plus a constant background to the part of the Raman spectra in (a) between wavenumbers 100 and 200 cm$^{-1}$. We use two Gaussians as for $\alpha$ we expect one peak at around 118 cm$^{-1}$, plus two peaks at 164 and 171 cm$^{-1}$, where these two peaks are merged into one in the spectra of most crystals. Note that in a few instances, a peak is very weak, fitting fails, and an anomalous point, e.g., slightly negative height, is produced.

apparent in the Raman spectra. This would be consistent with the difference in growth rates being due to different defects, as these defects would be unlikely to show up in the Raman spectra.
Figure S6. (a) Raman spectra for 37 non-needle like (solid curves) and 7 needle-like (dashed curves) crystals, all of which we identified as the $\gamma$ polymorph from their Raman spectra. All crystals were obtained from run number one at a salt concentration of 250 mg/ml. In three cases (blue dashed curve, magenta and red solid curves) there are anomalous spikes at single values in the data, which may be due to energetic events (possibly cosmic rays) in the detector. (b) Four each of the non-needle (solid curves) and needle (dashed curves) spectra of (a), plotted in the range we use for fitting. (c) Plot of peak height versus peak position, obtained by a fit of a Gaussian plus a constant background to the part of the Raman spectra in (a) between wavenumbers 100 and 200 cm$^{-1}$. We use one Gaussian as for $\gamma$ we expect one peak at around 157 cm$^{-1}$.

The corresponding data for the $\gamma$ polymorph needle-like crystals and non-needle-like crystals are in Figure S6. As with the $\alpha$ polymorph, there is no apparent difference between the set of needle spectra and the set of non-needle Raman spectra.
IV. STATISTICS AND MODELS FOR THE NUCLEATION OF COMPETING POLYMORPHS

As we only observe one polymorph or the other in a single well, not both, nucleation of the two polymorphs are mutually exclusive events in our system. Thus our data on the kinetics of nucleation of single crystals of competing polymorphs consist of the pair of observations, \( (t_{nuc}, i = \alpha, \gamma) \), for each well where crystallisation occurred. We want to analyse this quantitative data to build the most robust model with the greatest predictive power.

The statistics of the nucleation of competing polymorphs is subtle, but fortunately, analogous data sets occur in a number of other fields, in particular in medical statistics. Typically in mortality studies, there are competing illnesses or causes of death. An example might be a study of, say, 100 patients at risk of dying of cancer or of heart disease, where date and cause of death are recorded. As the two causes of death, like our competing polymorphs, are mutually exclusive, the data is also of the form of a pair of observations: a time, and one of a number of competing outcomes.

A. Models that include only observables

Our observables are the \( I_i \), together with their derivatives \( h_i \), and \( P = 1 - I_\alpha - I_\gamma \). We can construct models using only these functions. We can create a model by specifying the two cause specific hazard (CSH) functions.

1. Fitting procedure for models that include only observables

The model we fit is defined by the definitions of \( I_i \) and of \( P \):

\[
\frac{dI_i(t)}{dt} = P(t)h_i(t) \quad i = \alpha, \gamma \quad (1)
\]

\[
P(t) = 1 - I_\alpha(t) - I_\gamma(t) \quad (2)
\]

with two boundary conditions \( I_i(t = 0) = 0 \). We can obtain an equation for \( \frac{dP}{dt} \) and integrate it, to obtain

\[
P(t) = \exp \left[ -\int_0^t \left( h_\alpha(t') + h_\gamma(t') \right) dt' \right] \quad (3)
\]
which when the \( h_i \) are Weibull CSHs, becomes

\[
P(t) = \exp \left[ -\frac{t}{\tau_\alpha}^{\alpha} - \frac{t}{\tau_\gamma}^{\beta} \right]
\]

(4)

Equation (1) then becomes

\[
\frac{dI_i(t)}{dt} = \exp \left[ -\frac{t}{\tau_\alpha}^{\alpha} - \frac{t}{\tau_\gamma}^{\beta} \right] \beta_i \left( t^{\beta_i - 1} / \tau_i^{\beta_i} \right) = \alpha, \gamma
\]

(5)

For given values of the four (two \( \tau_i \) and two \( \beta_i \)) parameters, we integrate the coupled ordinary differential equations for the \( I_i \), Eq. (5), to get the two \( I_i(t) \) functions. To fit to data we simply vary the four parameters, and minimize the sum of the squares of the difference between the modelled \( I_i(t) \) and observed \( I_i(t) \). This is done using a Python program\(^{21}\).

**B. Models with latent nucleation times**

Models can also be constructed that rely on latent nucleation times, i.e., on hypothetical nucleation times \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \) for each well. As we only observe one polymorph for a well, for those wells where we observe nucleation we only measure the shorter one of these two times, the other one is not observable. Thus in all cases one of these times is hidden, hence the name latent time. See Beyersmann et al.\(^{17}\) and Geskus\(^{20}\) for discussion of the advantages and disadvantages of models that rely on latent times. Tsiatis\(^{14}\), Peterson\(^{15}\), and Slud and Byar\(^{16}\) both discuss the limits of what can be inferred about \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \), from data of our type. In general, the two nucleation times for a single well will be correlated, for example there may be a tendency that if one nucleation time is small in a well, that the other polymorph’s nucleation time in the same well, is also small.

We can write the probability that no nucleation has occurred in a droplet at time \( t \) as

\[
P(t) = \Pr(t_{\text{nuc,}\alpha} > t, t_{\text{nuc,}\gamma} > t)
\]

(6)

i.e., the probability that both \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \) are greater than \( t \).
1. Model with independent latent nucleation times

If the two nucleation times are independent then Eq. (6) simplifies to

\[ P(t) = \Pr(t_{\text{nuc},\alpha} > t)\Pr(t_{\text{nuc},\gamma} > t) \]  (7)

If the two independent latent nucleation times are both modelled by Weibulls, then for the probability density function for \( t_{\text{nuc},i} \), we have

\[ p(t_{\text{nuc},i}) = \beta_i \left( t^{\beta_i-1} / \tau_i^{\beta_i} \right) \exp \left[ - (t/\tau_i)^{\beta_i} \right] \quad i = \alpha, \gamma \]  (8)

and the cumulative probabilities are

\[ \Pr(t_{\text{nuc},i}) = \exp \left[ - (t/\tau_i)^{\beta_i} \right] \quad i = \alpha, \gamma \]  (9)

When the latent times are independent, the rate at which \( \alpha \) is observed to nucleate is just \( p(t_{\text{nuc},\alpha}) \) times the probability that nucleation of the \( \gamma \) polymorph has not yet occurred, so we have for \( \alpha \) nucleation

\[
\frac{dI_\alpha(t)}{dt} = \beta_i \left( t^{\beta_i-1} / \tau_i^{\beta_i} \right) \exp \left[ - (t/\tau_\alpha)^{\beta_\alpha} \right] \times \exp \left[ - (t/\tau_\gamma)^{\beta_\gamma} \right] \\
= \exp \left[ - (t/\tau_\alpha)^{\beta_\alpha} - (t/\tau_\gamma)^{\beta_\gamma} \right] \beta_\alpha \left( t^{\beta_\alpha-1} / \tau_\alpha^{\beta_\alpha} \right) 
\]  (10)

plus an analogous equation for \( dI_\gamma/dt \).

These two equations are identical to Eq. (5). So our model in the previous section that modelled the observable CSHs, via Weibulls (equation (3) of the main text), can be obtained starting with latent nucleation times, and then assuming that they are both Weibull distributed, and are independent. Thus the fits of our model using only observables, are also what we would obtain from a model of independent latent times. However, it should be noted that, although for every set of observables \( I_i \) there is a corresponding model with independent latent times, for the same \( I_i \) there are an infinite number of models with differently correlated latent times that yields this same \( I_i \).
2. Computational generation of models with independent latent nucleation times

As we just discussed, a model with independent nucleation times is the same as our model based on observables. Nonetheless, we now briefly describe how to generate a model based on independent latent nucleation times computationally here. Then in the next sub-section we describe how we introduce correlations. To generate the model with independent times, we simulate the behaviour of a large number samples, and determine the fraction of those samples at time \( t \), where nucleation has not occurred, \( N(t) \). When the number of samples is sufficiently large \( N(t) \approx P(t) \). For each well we need both a \( t_{nuc,\alpha} \) and \( t_{nuc,\gamma} \), we then select the shorter time to be the nucleation time of the well.

For this model we need to generate two sets of variables, \( t_{nuc,\alpha} \) and \( t_{nuc,\gamma} \), such that both sets of variable are Weibull distributed. We can do this by generating a set of uniform random variables for each polymorph, and putting each random number through the inverse cumulative distribution function (CDF) of a Weibull distribution.

\[
y = \tau_i \left( -\ln(1-u) \right)^{\frac{1}{\beta_i}} \quad i = \alpha, \gamma
\]  

(11)

Where \( u \) is a uniform random variable such that \( 0 \leq u < 1 \) and \( y \) is the output variable which is Weibull distributed. In this way we can easily create two independent sets of Weibull distributed nucleation times. We can then vary the \( \beta \) and \( \tau \) values to fit our data.

3. Model with correlated latent nucleation times

We do not know if nucleation of the two polymorphs is correlated. So to understand the role of correlations, it is useful to have a model with variable correlations between the nucleation times of the two polymorphs. To generate the model, we start, as in the previous section, with two sets of Weibull distributed numbers for \( t_{nuc,\alpha} \) and \( t_{nuc,\gamma} \). In this case however we want those two sets of numbers to be correlated such that for a well with a short \( \alpha \) nucleation time there is a high probability of a short \( \gamma \) nucleation time.

We set about generating the correlated nucleation times as follows. We generate two correlated sets of numbers, both Gaussian distributed. We then put each number into a Gaussian CDF which transforms them to uniformly distributed numbers between zero
and one. After that, we put the uniformly distributed numbers into an inverse Weibull CDF, Eq. (11). At this point, we have two Weibull distributed sets of numbers that are correlated. By generating a large number of nucleation time pairs, we obtain \( I_\alpha \) and \( I_\gamma \) pairs where the \( \alpha \) and \( \gamma \) nucleation processes have the desired correlation. We used these generated distributions to fit our data, in Figure 2 of the main text.

4. Measuring correlation

We use Spearman’s rank coefficient\(^{22}\) to measure correlation in this model. We briefly explain how Spearman’s rank coefficient works. For each sample we have two times: \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \). We rank all the \( t_{\text{nuc,}\alpha} \), for all samples in order of increasing length. We do the same with the set of \( t_{\text{nuc,}\gamma} \). We then measure for each pair of times the difference between the \( t_{\text{nuc,}\alpha} \) rank, \( r_\alpha \) and the \( t_{\text{nuc,}\gamma} \) rank, \( r_\gamma \). For example, if for one sample the nucleation times are \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \), and if \( t_{\text{nuc,}\alpha} \) is the 6th longest nucleation time \( (r_\alpha = 6) \) and \( t_{\text{nuc,}\gamma} \) is the 11th longest nucleation time \( (r_\gamma = 11) \), then the difference in ranks, \( r_\gamma - r_\alpha \) is 5. Note that if \( t_{\text{nuc,}\alpha} \) and \( t_{\text{nuc,}\gamma} \) are perfectly correlated all their ranks will be the same.

We measure the difference in rank for each pair of variables. The sum of the square rank differences is the covariance of \( r_\alpha \) and \( r_\gamma \). The formula we use to calculate Spearman’s rank correlation coefficient, \( R_s \), is

\[
R_s = \frac{\text{cov}(r_\alpha, r_\gamma)}{\sigma_{r_\alpha} \sigma_{r_\gamma}}
\]

(12)

where \( \sigma_{r_\alpha} \) and \( \sigma_{r_\gamma} \) are the standard deviations of all \( r_\alpha \) and \( r_\gamma \) respectively and \( \text{cov}(r_\alpha, r_\gamma) \) is the covariance of \( r_\alpha \) and \( r_\gamma \). When data are strongly positively correlated \( R_s \) approaches 1, when data are uncorrelated \( R_s \) is close to zero and when data are strongly negatively correlated \( R_s \) approaches -1. The Spearman’s rank correlation coefficient between the two functions in the model used in the fit in Figure 2 of the main text, is 0.95, i.e, \( \alpha \) and \( \gamma \) nucleation are strongly correlated.

V. ADDITIONAL RESULTS ON NUCLEATION

A. CIF plots for the five salt concentrations between 60 and 250 mg/ml

In Figure S7, we have plotted the CIFs: \( I_\alpha \), \( I_\gamma \) and the sum \( I_\alpha + I_\gamma \), for five different salt concentrations. We do not plot results for the experiments at 300 mg/ml NaCl because
there is little nucleation at this salt concentration. Note the change of time scale between Figure S7(b) and (c). The experimental timescale is longer at higher NaCl concentrations because as salt is added, nucleation slows. The plots for 250 mg/ml are also shown in Figure 2 of the main text. The fit parameters for the Weibull fits are in Table S2.

Figure S7. Plots of CIFs for five salt concentrations: (a) 60 mg/ml, (b) 90 mg/ml, (c) 150 mg/ml, (d) 200 mg/ml, and (e) 250 mg/ml. The legend in (c) applies to all plots. The points are our data: $I_\alpha$ (blue), $I_\gamma$ (green) and $I_\alpha + I_\gamma$ (red), respectively. Purple curves are fits of models with Weibull CSHs to the data.
Table S2. Best-fit values for fits of models with Weibull CSHs to the CIFS for α and γ nucleation.

<table>
<thead>
<tr>
<th>NaCl (mg/ml)</th>
<th>$\tau_\alpha$ (h)</th>
<th>$\beta_\alpha$</th>
<th>$\tau_\gamma$ (h)</th>
<th>$\beta_\gamma$</th>
<th>$R^2_\alpha$</th>
<th>$R^2_\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>7.49</td>
<td>0.26</td>
<td>4.74 x 10^5</td>
<td>0.28</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>90</td>
<td>57.2</td>
<td>0.49</td>
<td>5.84 x 10^5</td>
<td>0.26</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>150</td>
<td>245</td>
<td>0.49</td>
<td>3.35 x 10^4</td>
<td>0.40</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>200</td>
<td>1650</td>
<td>0.89</td>
<td>1250</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>250</td>
<td>244</td>
<td>1.41</td>
<td>610</td>
<td>0.48</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

$R^2_i$, $i = \alpha, \gamma$, is the $R^2$ value for comparison of the fit $I_i$ to the data.

Figure S8. The final fraction of wells that contain the γ polymorph, $f_\gamma$, plotted as a function of NaCl concentration. The dark green circles are the overall fraction averaged over all runs at a single NaCl concentration. The smaller green diamonds are the fractions in individual runs. The solid curve is a fit of a logistic function to $f_\gamma = n_\gamma / [n_\alpha + n_\gamma]$, where $n_\alpha$ and $n_\gamma$ are the number of wells containing α crystals and γ crystals, respectively. The error bars shown have a total height of $2[f_\gamma(1 - f_\gamma)/ (n_\alpha + n_\gamma)]^{1/2}$, where $n_\alpha$ is the number of wells containing α crystals and $n_\gamma$ is the number of wells containing γ crystals.

B. Increasing salt concentration favours the γ polymorph

We plot the final fraction of crystals that are in the γ polymorph, $f_\gamma$, in Figure S8. This is for experiments were carried out at NaCl concentrations $c_{NaCl}$ from 60 mg/ml to 300 mg/ml. At least two runs were carried out at each salt concentration. We define a run as the set of nucleation times recorded from one 96-well microplate. On completing each run, the fraction of wells containing each polymorph was determined.

In Figure S8, the variation of the final fraction of crystals in the γ polymorph, $f_\gamma$, is modelled using a logistic function. The fit is shown as a black curve in Figure S8. The logistic function is

$$f_\gamma (c_{NaCl}) = \frac{1}{1 + \exp \left[ -(c_{NaCl} - c_{1/2}) / c_{SW} \right]}$$  \hspace{1cm} (13)
and it has two parameters: $c_{1/2}$ and $c_{SW}$. As we can see in Figure S8, this functional form fits our data well. The best-fit parameters are $c_{1/2} = 215$ mg/ml, and $c_{SW} = 100$ mg/ml. The parameter, $c_{1/2}$, is an estimate for the salt concentration at which half the samples are $\alpha$ and half $\gamma$. There is broad region, a few hundred mg/ml, over which we go from a region with very small, but non-zero, amounts of the $\gamma$, to a large majority of the crystals in the $\gamma$ polymorph. The $c_{SW}$ parameter can be used as an estimator of the width of this region. The data is in Table S3.

Table S3. Values of $f_\gamma$ at six NaCl concentrations.

<table>
<thead>
<tr>
<th>NaCl (mg/ml)</th>
<th>$f_\gamma$</th>
<th>$n_\alpha$</th>
<th>$n_\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.04</td>
<td>143</td>
<td>6</td>
</tr>
<tr>
<td>90</td>
<td>0.10</td>
<td>109</td>
<td>12</td>
</tr>
<tr>
<td>150</td>
<td>0.14</td>
<td>102</td>
<td>17</td>
</tr>
<tr>
<td>200</td>
<td>0.54</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>250</td>
<td>0.56</td>
<td>80</td>
<td>103</td>
</tr>
<tr>
<td>300</td>
<td>0.89</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

$f_\gamma$ is the fraction of wells at the end of the experiment, where the the crystal(s) are of the $\gamma$ polymorph. Also shown are the total numbers of wells with $\alpha$ and with $\gamma$ crystals, $n_\alpha$ and $n_\gamma$, respectively.

C. Rates of nucleation, and total amount of nucleation are not well reproducible, relative rates and fractions are reproducible

The nucleation time cumulative incidence functions (CIFs) are plotted for each individual run in Figure S9. The CIFs in Figure S7 were obtained by combining the data in these individual runs. The number of wells in which nucleation occurs, varies from run to run, i.e., reproducibility of the amount of nucleation is poor. What is reproducible, is the fraction of nucleation events of each polymorph. In Figure S9, $\gamma$ nucleation corresponds to the difference between the all-nucleation CIFs and the $\alpha$ nucleation CIFs for each run. We see at low NaCl concentrations nucleation is dominated by the $\alpha$ polymorph and this phenomenon is reproducible between runs. Even when two runs have a significantly different amount of nucleation occurring, the relative amount of the nucleation of each polymorph is similar. For example at 150 mg/ml NaCl, see Figure S9(c), one run (dark red) has about three times more nucleation events than the other (light red). However, for both runs, about 15% of the total nucleation events give the $\gamma$ polymorph.
Figure S9. CIFs for individual runs at NaCl concentrations: (a) 60 mg/ml, (b) 90 mg/ml, (c) 150 mg/ml, (d) 200 mg/ml and (e) 250 mg/ml. The key in (e) applies to all 5 plots. There are 2 runs in (a) to (d), and 3 runs in (e). For each run we plot $I_{\alpha}$ as closed blue symbols, and $I_{\alpha} + I_{\gamma}$ as open red/brown symbols. For example, in (a) for the first run no $\gamma$ crystals form, so the solid and open symbols are on top of each other as $I_{\gamma} = 0$, while for the second run a small number of $\gamma$ crystals start to form after a few hours, and so the pale blue closed and open red symbols move apart.

It should also be noted that the distribution of nucleation times is reproducibly different for the two polymorphs at the high salt concentration of 250 mg/ml NaCl, see Figure S9. The $\gamma$ nucleation time distribution initially has a very fast relative nucleation rate which
slows down over time in all of the runs, while the $\alpha$ nucleation time distribution initially has a very slow relative nucleation rate which speeds up over time giving an ‘s-shaped’ curve. This is well illustrated by the median nucleation times of the subpopulations of $\alpha$ nucleating wells and $\gamma$ nucleating wells for each run. The median $\alpha$ nucleation times for the three runs are 77 h, 102 h and 81 h, while median $\gamma$ nucleation times for those runs are 13.5 h, 14 h and 34 h respectively. We can therefore say, in addition to the fraction of nucleation that corresponds to each polymorph being reproducible, the relative change in the effective nucleation rates of the two polymorphs is also reproducible.

![Graph](a)

**Figure S10.** Plots of the sum of the two CIF functions, $I_\alpha + I_\gamma$, i.e., the fraction of wells where crystallisation has occurred, is plotted as a function of time. This is at 250 mg/ml salt. (a) The red triangles are the average $I_\alpha + I_\gamma$, for all three isothermal runs, and the purple crosses are is the average of two runs that are at 21 °C for the first 18 h, after which the microplate is maintained at 30.4 °C for the following 48 h. (b) Here we have plotted the individual runs of the systems that were averaged to obtain the data in (a). This is two runs heated to 30°C after 18 h (runs 4 and 5), and the isothermal individual runs (runs 1, 2 and 3).

D. Time-dependent supersaturation increases polymorph purity

In Figure S10(a), we have plotted the sum of the two CIFs, $I_\alpha + I_\gamma$ for both isothermal experiments, and experiments where the temperature is increased from room temperature (close to 21 °C) to 30.4 °C, after 18 hours. All runs are at 250 mg/ml NaCl. Note that for the experiments warmed to 30.4 °C, nucleation is almost completely stopped. In Figure S10(b), we show the individual runs that make up Figure S10(a). We can see the trends observed in the individual runs are the same as we observe in the datasets where all runs under the same conditions are combined.
Figure S11. Polymorph composition as a function of time. The composition is the fraction of the wells where crystallisation has occurred, that contain the $\gamma$ polymorph. The lines start at the point when five nucleation events have occurred. Each curve is one run, and the colour indicates the salt concentration. Different runs at the same concentration are distinguished by being solid, dashed and dotted.

E. Individual purity vs times

We have looked at how polymorph composition varies over time. Here we show that the runs that make up the datasets plotted in Figure 4 follow the same trends. The runs are plotted in Figure S11 and data is shown in Table S4.

VI. ADDITIONAL RESULTS ON GROWTH RATES AND CRYSTAL HABITS

A. The effect of growth rates on the error in nucleation time measurements

Our measurements for nucleation times are only accurate if the time for nucleation, i.e., for the crystal to cross the nucleation barrier and start growing irreversibly, is much larger than the time taken for the crystal to grow from just past the barrier, to a size large enough to be visible.

We have plotted the sizes of 10 $\alpha$ and 10 $\gamma$ crystals, at 90 mg/ml NaCl in Figure S12. Data for 250 mg/ml NaCl are in Figure 5 of the main text. Note that growth rates vary with salt concentration. The crystals grow faster at 90 mg/ml of salt than 250 mg/ml. We focus
Table S4. Results of individual runs, for the fraction of $\gamma$ at early times, and at the end.

<table>
<thead>
<tr>
<th>NaCl conc / mg/ml</th>
<th>Fraction $\gamma$ at $n_{\geq 10}$</th>
<th>Final $\gamma$ fraction</th>
<th>$n_{\geq 10}$</th>
<th>$t_{n_{\geq 10}}$ / h</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.0</td>
<td>0.0</td>
<td>10</td>
<td>0.16</td>
</tr>
<tr>
<td>60</td>
<td>0.083</td>
<td>0.086</td>
<td>12</td>
<td>0.33</td>
</tr>
<tr>
<td>90</td>
<td>0.30</td>
<td>0.10</td>
<td>10</td>
<td>1.16</td>
</tr>
<tr>
<td>90</td>
<td>0.15</td>
<td>0.099</td>
<td>13</td>
<td>1.0</td>
</tr>
<tr>
<td>150</td>
<td>0.45</td>
<td>0.13</td>
<td>11</td>
<td>2.5</td>
</tr>
<tr>
<td>150</td>
<td>0.18</td>
<td>0.17</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>200</td>
<td>0.50</td>
<td>0.50</td>
<td>10</td>
<td>34</td>
</tr>
<tr>
<td>200</td>
<td>0.40</td>
<td>0.63</td>
<td>10</td>
<td>106</td>
</tr>
<tr>
<td>250</td>
<td>1.0</td>
<td>0.47</td>
<td>10</td>
<td>2.66</td>
</tr>
<tr>
<td>250</td>
<td>1.0</td>
<td>0.60</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>250</td>
<td>1.0</td>
<td>0.72</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

The second column is the fraction of crystals in the $\gamma$ polymorph, at a time, $t_{n_{\geq 10}}$, early in the experiment. $t_{n_{\geq 10}}$ is the earliest observation time at which we have 10 or more nucleation events; the precise number of nucleation events, $n_{\geq 10}$, is in column four. The third column is the fraction of the $\gamma$ polymorph at the end of the experiment.

Figure S12. Plot of the size of a crystal, as a function of time. The NaCl concentrations is 90 mg/ml. Data is shown for 10 $\alpha$ and 10 $\gamma$ crystals. The sizes of $\alpha$ crystals are shown as light blue (needles) and black (non-needles) lines-and-points. The sizes of $\gamma$ crystals are shown as red (needles) and green (non-needles) lines-and-points. For each crystal $t = 0$ is defined as the time of the first image in which there is a visible crystal.

on the $\gamma$ crystals for estimating nucleation time errors as their growth is slightly slower than that of the $\alpha$ crystals. Here the growth rate for the $\gamma$ crystals is around 2 mm/h. This means the error in our nucleation time measurements should be less than 10 minutes in most
cases. This is a small error for all but the very shortest nucleation times.

For a salt concentration of 250 mg/ml, we plotted the sizes of 10 α and 10 γ crystals, as a function of time, in Figure 5 of the main text. Most but not all of the γ crystals are growing at rate of around 0.2 mm/h. This implies that our measured nucleation times are on average just over 30 mins too long, i.e., we first see a γ crystal about 30 mins after it nucleated as a microscopic nucleus. A few α crystals are also growing at around 0.2 mm/h, but most are initially growing of order 10 times faster. Our runs at 250 mg/ml salt are of 168 h duration, so except for nucleation at early times, a 30 mins error is a relatively small error.

Table S5. Mean glycine sizes at initial detection and at two subsequent times, for both α and γ polymorphs.

<table>
<thead>
<tr>
<th>NaCl conc / mg/ml</th>
<th>Time / mins</th>
<th>Size/ mm</th>
<th>Interquartile range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean α γ</td>
<td>α γ</td>
</tr>
<tr>
<td>90</td>
<td>0</td>
<td>1.6 ± 0.5 0.6 ± 0.1</td>
<td>3.2 0.69</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.5 ± 0.7 1.2 ± 0.3</td>
<td>4.0 1.5</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>4.3 ± 0.5 3.0 ± 0.2</td>
<td>3.0 0.56</td>
</tr>
<tr>
<td>250</td>
<td>0</td>
<td>0.5 ± 0.1 0.17 ± 0.02</td>
<td>0.74 0.11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.9 ± 0.2 0.23 ± 0.01</td>
<td>1.2 0.025</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1.8 ± 0.3 0.9 ± 0.1</td>
<td>2.3 0.23</td>
</tr>
</tbody>
</table>

We estimate the uncertainties of the mean values with error estimates that are the standard deviation of the measured sizes, divided by the square root of the number of crystals measured (10 crystals of each polymorph at each concentration). We measure the width of the distribution of sizes of the crystals via the interquartile range.

B. Variation in growth rates between crystals

In Table S5, we present crystal sizes at three times, from the same data set as used in Figure 5 and Figure S12. In this table, the interquartile range (IQR) is $Q_3 - Q_1$, where $Q_1$ is the first quartile of the distribution of crystal sizes, i.e., the size at which 25% of the crystals are smaller and 75% are larger, and $Q_3$ is the size where 75% are smaller and 25% are larger. The IQR is a convenient measure of the spread in crystal sizes as it is relatively insensitive to outliers (unlike the standard deviation), and we have outliers in the crystal size, see Figure S12. It is clear that the growth rates of both polymorphs vary widely between one crystal and another. This is especially true for the α polymorph where at early times the IQR is
larger than the mean.

![Graphs showing crystal size vs. nucleation time](image)

Figure S13. Here we plot the size of crystals 20 mins after they were initially detected, as a function of the time at which they nucleated. In (a) the salt concentration is 250 mg/ml, and in (b) it is 90 mg/ml.

C. Growth rate and nucleation time

In the main text we showed that needle-like crystals tend to have faster growth rates. We have also showed that crystals that nucleate at later times are more likely to be needle-like. It follows that the crystals with longer nucleation times generally have faster growth rates. As we have discussed, the growth rate of crystals is difficult to quantify for our data because the growth rate changes with time. Here we plot the size of a crystal 20 mins after it is initially detected against nucleation time. This is shown in Figure S13. We see that at both salt concentrations, the earliest nucleating crystals are small at 20 mins.
Figure S14. For the γ polymorph, we plot the fraction of crystals with a needle-like habit, as a function of the total number of wells that contain crystals of that polymorph.

D. Nucleation time and crystal habit

For γ crystals at high salt concentrations, we observe the same correlation between crystal habit and nucleation time, as we did for the α polymorph (Figure 7). The data for the γ polymorph are shown in Figure S14. We see the same trend as for the α polymorph, with non-needle-like crystals nucleating at early times, and needle-like crystals at late times, although the fraction of needles is lower for the γ polymorph. At low salt concentrations, there are too few γ crystals to make clear statements. For example, at 60 mg/ml and 90 mg/ml there are only six and twelve γ nucleation events in total, respectively.

REFERENCES


