

## Electronic Supplementary Information

### Morphology, energetics and growth kinetics of diphenylalanine fibres

Phillip Mark Rodger,<sup>a</sup> Caroline Montgomery,<sup>a</sup> Giovanni Costantini,<sup>a</sup> and Alison Rodger<sup>a,b</sup>

<sup>a</sup>*Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK.*

<sup>b</sup>*Department of Molecular Sciences, Macquarie University, NSW, 2109, Australia.*

#### 1. Root-mean-square deviation

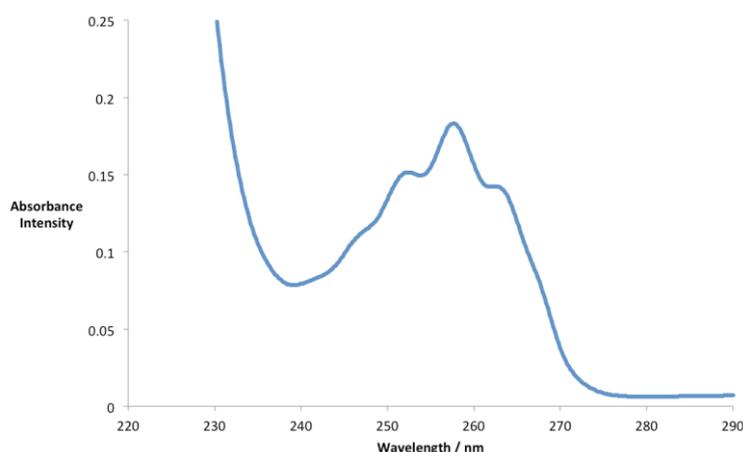
The root-mean-square deviation (RMSD) was calculated using

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} (r_i(t) - r_i(t_0))^2}{N_{atoms}}}$$

where  $r_i(t_0)$  refers to the position of the diphenylalanine (FF) atoms at the start of the simulation and  $r_i(t)$  to the position at time  $t$ . At each time step,  $t$ , the RMSD was minimised through a rotation and translation fit. The optimised RMSD was taken as a measurement of the physical stability of a structure.

#### 2. Absorbance of diphenylalanine

The peaks seen in the absorbance spectrum of FF are due to the  $\pi$ - $\pi^*$  transitions from the phenyl chromophore (Figure S1).



**Figure S1.** Absorbance signal for FF in water (2 mg/mL).

### 3. Linear Dichroism

Linear dichroism (LD) is a polarized light spectroscopy technique which relies on the orientation of the molecules that are being probed. Two beams of perpendicularly linearly polarised light are shone at the sample, and the difference in absorbance of each of these polarisations is measured

$$LD = A_{parallel} - A_{perpendicular}$$

Net alignment of molecules in the sample is essential for a non-zero LD signal because the electronic transition moments will only absorb light when they are aligned with the electric field of the radiation. The relationship between the LD signal and the orientation of the molecules is given by

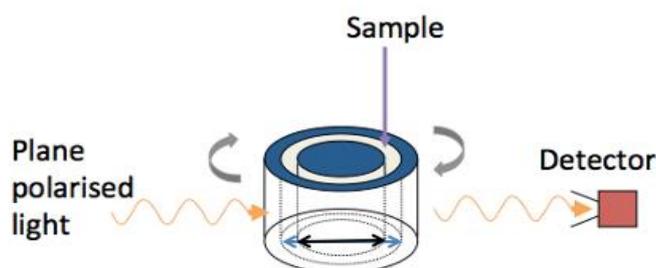
$$LD = \frac{3}{2}SA_{isotropic}(3\cos^2\alpha - 1)$$

where  $A_{isotropic}$  is the absorbance of an unoriented sample,  $S$  is the orientation parameter ( $S = 1$  for an oriented sample and  $S = 0$  for an isotropic one), and  $\alpha$  is the angle between the transition moment and the alignment (fibre) axis.

In our case the sample is oriented by Couette flow, with the outer capillary, which contains the liquid sample, rotating and a stationary rod rigidly suspended in the middle of the capillary (Figure S2). The rotating capillary has a 3.0 mm internal diameter and spins at 3,000 rpm, while the stationary rod suspended in the middle has a 2.5 mm outer diameter.<sup>1, 2</sup> The instrument geometry is such that one beam of light is parallel to the orientation of the sample and the other is perpendicular to it. Molecular structures with significant aspect ratios become aligned in the flow, and their absorbance signals show an LD signal.

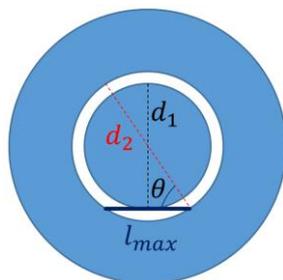
A background spectrum is taken with no spinning (as no signal is expected from an isotropic sample) and subtracted from the spinning spectrum in order to create an LD spectrum. In a time-series (*e.g.* during kinetics studies) the capillary spinning was switched off between measurements at every time point in order to acquire a reference background spectrum. Macros were created within the Jasco J-815 Spectra Manger software in order to complete the sequences needed for the variable temperature and kinetics measurements. Moreover, an electronic control box was built in order to switch the spinning on and off between measurements. However, as the fibres grew larger, an LD signal was observed even in the absence of Couette flow, indicating that some fibres were probably stuck inside the cell. Therefore, the background measurement at the start of the experiment, before any fibres had formed, was used for plotting all spectra and the change in background was used to indicate when the fibres had become too large for the experimental set-up.

Special care had to be taken when cleaning the LD capillary and rod, as the fibres stuck to them and could have provided nucleation sites in successive experiments. Capillaries with a Teflon stopper at the base were manufactured for this work to facilitate cleaning. Between experiments, the components were soaked in a 6 M HNO<sub>3</sub> bath overnight to dissolve any residual FF.



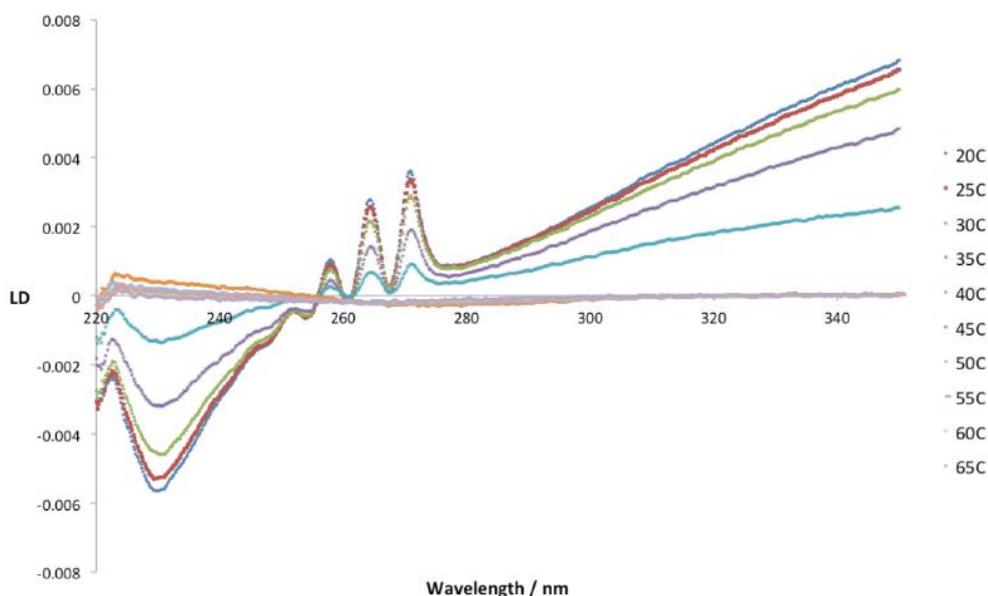
**Figure S2.** Schematic setup of a linear dichroism experiment. Two perpendicular beams of linearly polarized light are shone at a sample which is in a spinning capillary with a stationary rod in the middle. The difference in absorbance of these two polarisations is measured by the detector. The sample fits in the 250  $\mu$ m gap between the quartz rod and quartz capillary (quartz indicated in blue).

Some aspects of the specific experimental setup used in this work have important consequences for the interpretation of the data. Firstly, small fibres are invisible to this technique because they are not long enough to be able to align in shear flow. There is also a maximum length, which is essentially the diagonal across the gap between capillary and rod. The capillary has an inner diameter of  $d_2 = 3.0$  mm and the rod has a diameter  $d_1 = 2.5$  mm so that when fibres grow longer than  $l_{max}$  they get stuck and can no longer align properly in the flow (see Figure S3). Therefore, the signal that we get from them after this point is no longer comparable to the signal from when they were shorter. In particular from Figure S3 it can be seen that  $l_{max} = \frac{d_1}{\tan[\sin^{-1}(d_1/d_2)]} = 1.66$  mm.



**Figure S3.** Schematic representation of the maximum length that can be reached by growing fibres before they get stuck in the gap between capillary and rod in the LD Couette cell.

However, when the fibres are short, LD can tell us all about the time frame of the kinetics of the fibre assembly. It is also possible to infer information about the onset of the assembly from LD data. Although it is not possible to see the initial small assemblies, it is possible to detect fibres at a fixed point after this has happened (Figure S4). The fibres have to be a set length for the signal to appear, and this size will not change between experiments. Therefore, the variation in onset time that we see with linear dichroism is real and comes from the fact that the nucleation is a random process.



**Figure S4.** LD signal of FF fibres forming from a 2 mg/ml FF monomer solution as the reaction mixture is cooled from 70°C to 20°C at a rate of 1 °C/min. Initially there is no signal, but as the fibres start to assemble peaks appear in the 230 nm and 250–270 nm regions. No baseline correction has been implemented. The change in baseline as a function of time is due to the increase in light scattering as the fibres become larger.

#### 4. LD at 70 °C for 4 Hours

The data in Figure S5 show the LD signal at 70 °C over 4 hours. No peaks are visible, indicating that there are no fibres present. Therefore, fibres do not form at 70 °C for at least 4 hours.

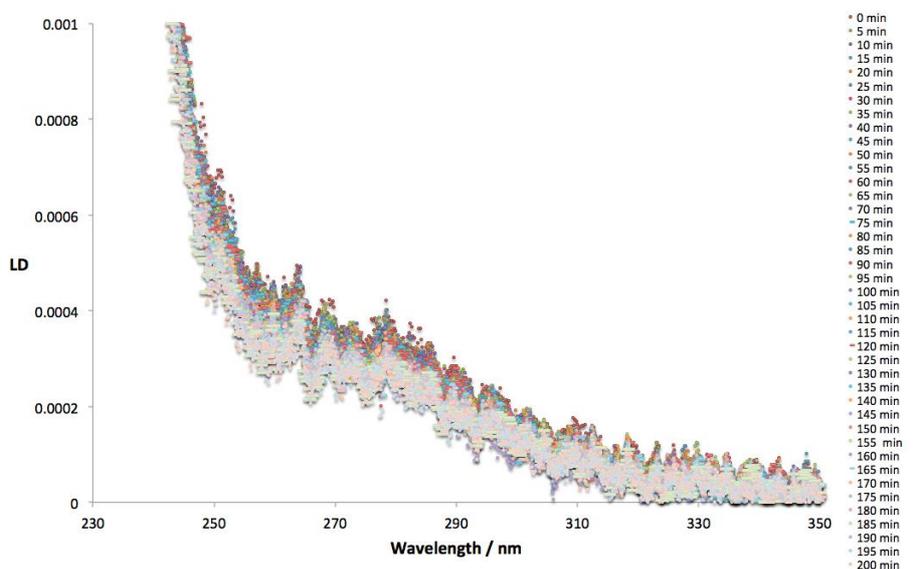


Figure S5. Absence of the aromatic LD signal of FF (2 mg/mL in water) at 70 °C over 4 hours.

#### 5. LD fit

The wavelength scan LD data of Figure S6 shows a positive LD absorbance signal due to the FF — between about 260 and 280 nm — overlaid on a negative trending scattering signal. An example of the fit (pink curve) made to the LD data (red curve) to remove the scattering is shown in Figure S6.

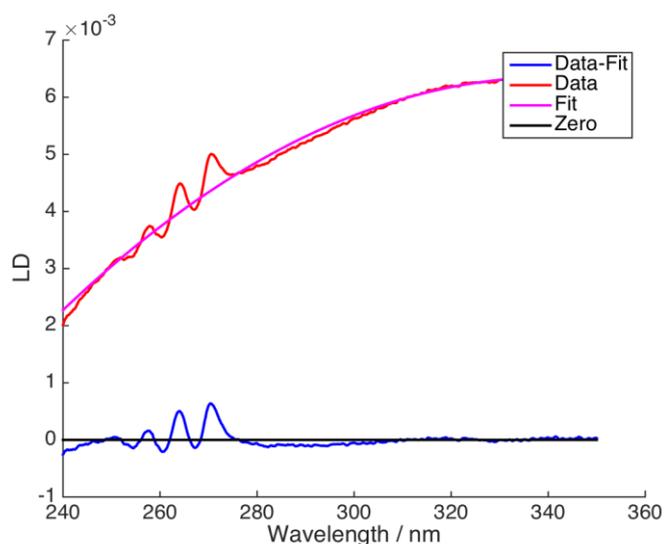
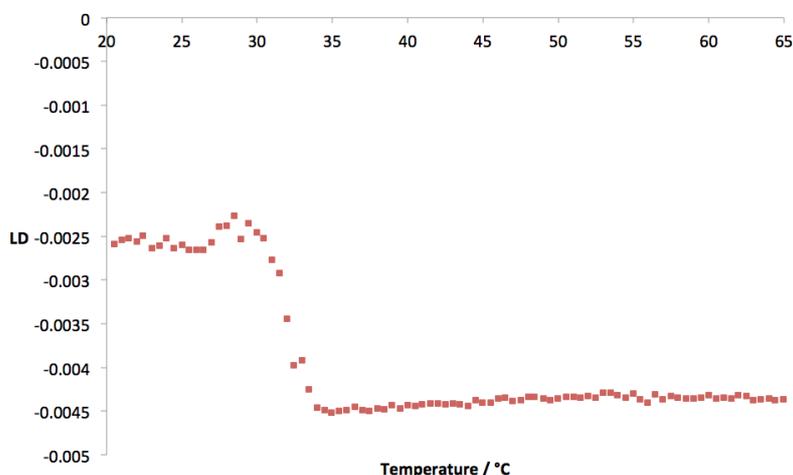


Figure S6. Example of LD fit at 20 °C from the thermal stability measurements of FF in water (2 mg/mL).

## 6. Cooling at 0.5 °C/min, single wavelength measurement

Figure S7 shows the LD of a solution of FF as the temperature was cooled from 65°C to 20°C in 0.5 °C /min decrements. Figure 3c in the main paper displays the results obtained at a cooling rate of 1 °C/min.



**Figure S7.** LD signal at 258 nm as the sample is cooled from 65 °C to 20 °C at 0.5 °C/min.

## 7. Derivation of the cumulative distribution function of multiple nucleation events

Nucleation is a stochastic process and in particular has been shown to be Poisson-type. The Poisson distribution describes the probability of a discrete random variable  $M$  that a number of nucleation events,  $m$ , occurs in a given time interval,  $t$ , where  $\lambda$  is the rate of nucleation.

$$P(M = m) = \frac{(\lambda t)^m e^{-\lambda t}}{m!} \quad (\text{S1})$$

In order to explore the times associated with this Poisson process, we consider time as a continuous random variable  $X$ . The cumulative distribution function (CDF),  $F(t)$ , is described as follows.

$$F(t) = P(X \leq t) \quad (\text{S2})$$

$$F(t) = 1 - P(X > t) \quad (\text{S3})$$

The CDF is equal to the probability that at least one event has occurred up to and including the time interval  $t$ . We can find this probability,  $P(X > t)$ , by setting the number of nucleation events,  $M$ , equal to zero in the Poisson distribution (Equation (S1)).

$$P(X > t) = P(M = 0) \quad (\text{S4})$$

$$P(M = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \quad (\text{S5})$$

$$P(M = 0) = e^{-\lambda t} \quad (\text{S6})$$

This can now be substituted into the CDF (Equation (S3)).

$$F(t) = 1 - e^{-\lambda t} \quad (S7)$$

This function describes the cumulative distribution of the probability for the time taken for a nucleation event to occur.

However, in experiments, it may be the case that multiple nucleation events are needed in order to produce a density of fibres large enough to be detectable, so this must be taken into account in deriving the CDF. Equation (S7), *i.e.* the probability that at least one nucleation event is occurring within time  $t$ , was obtained by subtracting the probability of 0 nucleation events occurring from unity. The probability that up to  $n$  events happen within time  $t$ , *i.e.* the probability that up to  $n$  nuclei exist at time  $t$ , is given by the sum of the probabilities that 0, 1, 2, ...,  $n$  events happen.

$$P(\text{up to } n \text{ nuclei exist}) = \sum_{i=0}^n \frac{(\lambda t)^i e^{-\lambda t}}{i!} \quad (S8)$$

This can be substituted into Equation (S3) to produce the CDF of nuclei required in order to create detectable fibres.

$$F(t) = 1 - \sum_{i=0}^n \frac{(\lambda t)^i e^{-\lambda t}}{i!} = 1 - \frac{\Gamma(n+1, \lambda t)}{n!} \quad (S9)$$

where  $\Gamma$  is the upper incomplete gamma function.

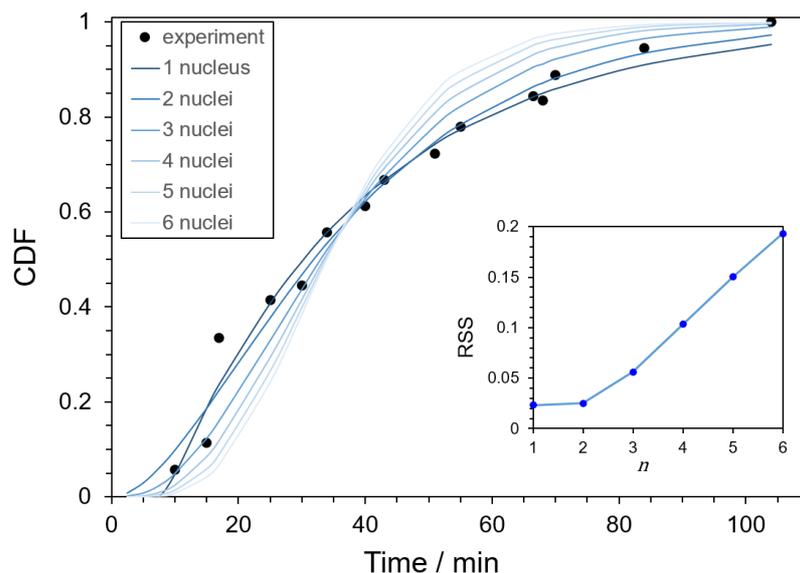
It was noted by Jiang and Horst<sup>3</sup> that, in experiments, it is not possible to measure the time taken for one (or more) nucleation events to occur, but instead, the time it takes to *detect* that one (or more) nucleation events have occurred. In other words, the crystals must grow large enough to be detected and, therefore, a growth period,  $t_g$ , must be introduced into the CDF equation. The total time,  $t$ , is the sum of the time for the nucleation event(s) to occur,  $x$ , and the growth time,  $t_g$ .

$$t = x + t_g \quad (S10)$$

Substituting this into Equation (S9) produces the modified CDF

$$F(t) = 1 - \sum_{i=0}^n \frac{(\lambda(t-t_g))^i e^{-\lambda(t-t_g)}}{i!} = 1 - \frac{\Gamma(n+1, \lambda(t-t_g))}{n!} \quad (S11)$$

This function should be compared with the fraction of experiments in which fibres have been detected within time  $t$ , shown in Figure 4c. Figure S8 shows the comparison between the experimental data and a series of CDF for increasing values of  $n$ , each of which has been individually optimised in terms of  $\lambda$  and  $t_g$ .

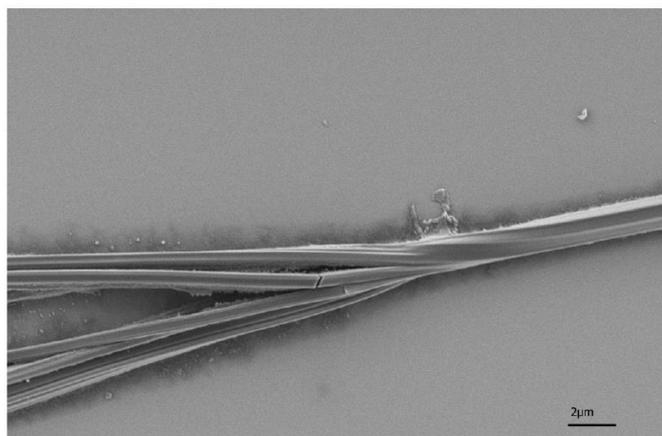


**Figure S8.** Comparison between the CDFs in Equation (S11) (continuous lines) and the experimental fraction of events which have nucleated versus time (filled black circles). The different lines correspond to different numbers of nucleation events  $n$ , and each of them has been optimised in terms of  $\lambda$  and  $t_g$ . Inset: residual sum of squares (RSS) for the fits of Equation (S11) to the experimental data for  $n = 1$  to 6.

The best fits in Figure S8 are obtained for  $n = 1$  ( $t_g = 8$  min and  $\lambda = 0.03$  min<sup>-1</sup>) and  $n = 2$  ( $t_g = 0$  min and  $\lambda = 0.05$  min<sup>-1</sup>); the corresponding CDF functions approximate the experimental data equally well with residual sum of squares (RSS) values of 0.23 and 0.25, respectively. All higher values of  $n$  have an optimal growth period  $t_g = 0$  min but result in worse fits with the experimental data, as demonstrated by the increasing RSS values (inset in Figure S8).

## 8. Secondary nucleation

Secondary nucleation events are evident as illustrated in Figure S9.



**Figure S9.** SEM image of a fibre showing branching and breaking of fibres.

## 9. References

1. R. Marrington, T. R. Dafforn, D. J. Halsall, M. Hicks and A. Rodger, *Analyst*, 2005, **130**, 1608-1616.
2. R. Marrington, T. R. Dafforn, D. J. Halsall and A. Rodger, *Biophysical Journal*, 2004, **87**, 2002-2012.
3. S. Jiang and J. H. ter Horst, *Crystal Growth and Design*, 2011, **11**, 256-261.