Electronic Supplementary Material (ESI) for Journal of Materials Chemistry C.

Equilibration and thermal reversibility in mixtures of model OPV small-molecules and polymers

AM Higgins, P Gutfreund, V Italia and EL Hynes

Supplementary Information

1 Materials and Methods

Sample list

Except for samples H and I, all samples had Bis-PCBM bottom layers that were spin-coated at 2.2 x 10^3 revolutions per minute (2.2k RPM) from chlorobenzene solutions (2.5 wt% Bis-PCBM). Samples H and I had bottom layers that were spin-coated at 2.7 k RPM from 2.8 wt% solutions of Bis-PCBM in chlorobenzene. The top layers were spin-coated onto freshly-cleaved mica using the spin speeds and solutions given in table S1.

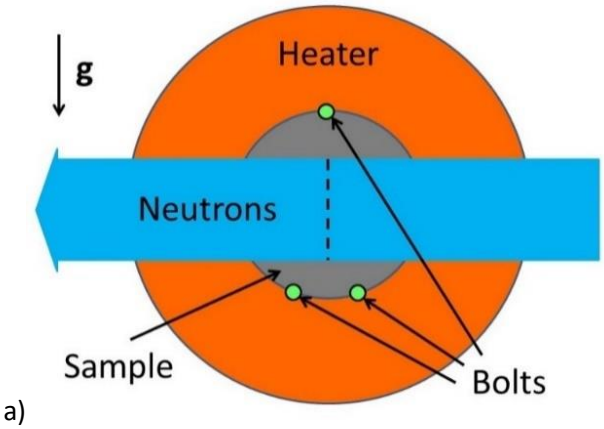
Batch 1	**Top layer solutions and spin speeds**
Sample A	1.5 wt% 300k PS in toluene at 2.4k RPM
Sample B	"
Sample C	"
Sample D	1.6 wt% solution (solutes; 34 wt% Bis-PCBM/66 wt% 300k PS) in chlorobenzene @ 2.5k RPM
Sample E	"
Sample F	2 wt% 5k PS in toluene at 2k RPM
Sample G	2.3 wt% solution (solutes; 33 wt% Bis-PCBM/67 wt% 5k PS) in chlorobenzene @ 3k RPM
Sample H	2.1 wt% 2k PS in toluene at 2k RPM
Sample I	"
Batch 2	
Sample 1	1.6 wt% solution (solutes; 35 wt% Bis-PCBM/65 wt% 300k PS) in chlorobenzene @ 2.5k RPM
Sample 2	"
Sample 3	1.6 wt% solution (solutes; 27 wt% Bis-PCBM/73 wt% 300k PS) in chlorobenzene @ 2.5k RPM
Sample 4	1.6 wt% solution (solutes; 14 wt% Bis-PCBM/86 wt% 300k PS) in chlorobenzene @ 2.5k RPM

Table S1; Solutions and spin speeds used in fabricating the top layers in batches 1 and 2.

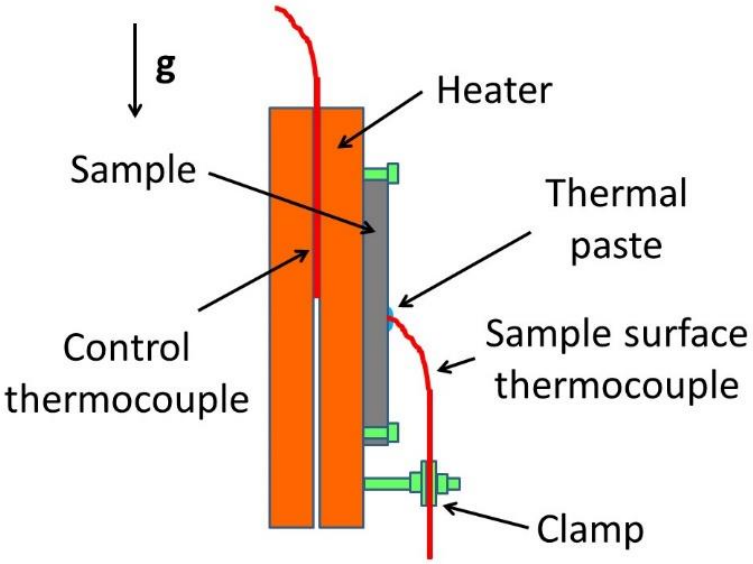
Thermal annealing

Set-up and calibration - batch 1; The sample heater was placed within a vacuum chamber with quartz windows, that sits in the neutron beam. For the samples in batch 1 the chamber was pumped down to a pressure of 10^{-4} mbar before heating the samples to set-point temperatures of between 80 °C and 200 °C. The samples were bolted onto the heater using three bolts (placed above and below the beam footprint, as shown in figure S1a). The tightness of the bolts was carefully adjusted to minimise the chance of bending the samples. Temperature calibration was performed for three different tightnesses of bolt (loose, medium & tight), although only loose and medium were employed during the NR experiments. Temperature calibration was performed using a blank silicon sample with a blob of thermal paste on the

surface. Fine gauge k-type thermocouple wire (with an accuracy of ± 2.2 °C, from Omega, USA) was then clamped in such a position that it was pressing onto the calibration sample surface, as shown in figure S1b. This was connected to a k-type thermocouple feedthrough in one of the vacuum chamber ports. Figure S2 shows the stabilised sample surface temperatures during heating and cooling (plus a third 'repeat measurement', following unmounting and re-mounting of the heater unit and re-attachment of the calibration sample, for some temperatures), for different bolt tightness, as a function of the set-point temperature. Averaging the stabilised sample surface temperatures at each set-point temperature gives the calibration offsets shown in table S2. Figure S3 shows the temporal behaviour of the sample surface temperature following steps up and down in the set-point temperature (mostly 10 °C steps). For all set-point temperatures of 120 °C and above the sample surface temperature is completely stabilised within 6 minutes, following a 10 °C step (up or down) in the set-point temperature.



a)



b)

Figure S1; Schematic diagrams of *in-situ* heating and temperature calibration set-ups (not to scale). a) The *in-situ* heating set-up (face on view) with a sample in place. b) The sample surface temperature calibration set-up (edge on view) using a dummy sample and thermocouple. The direction of gravity is indicated by g in both diagrams.

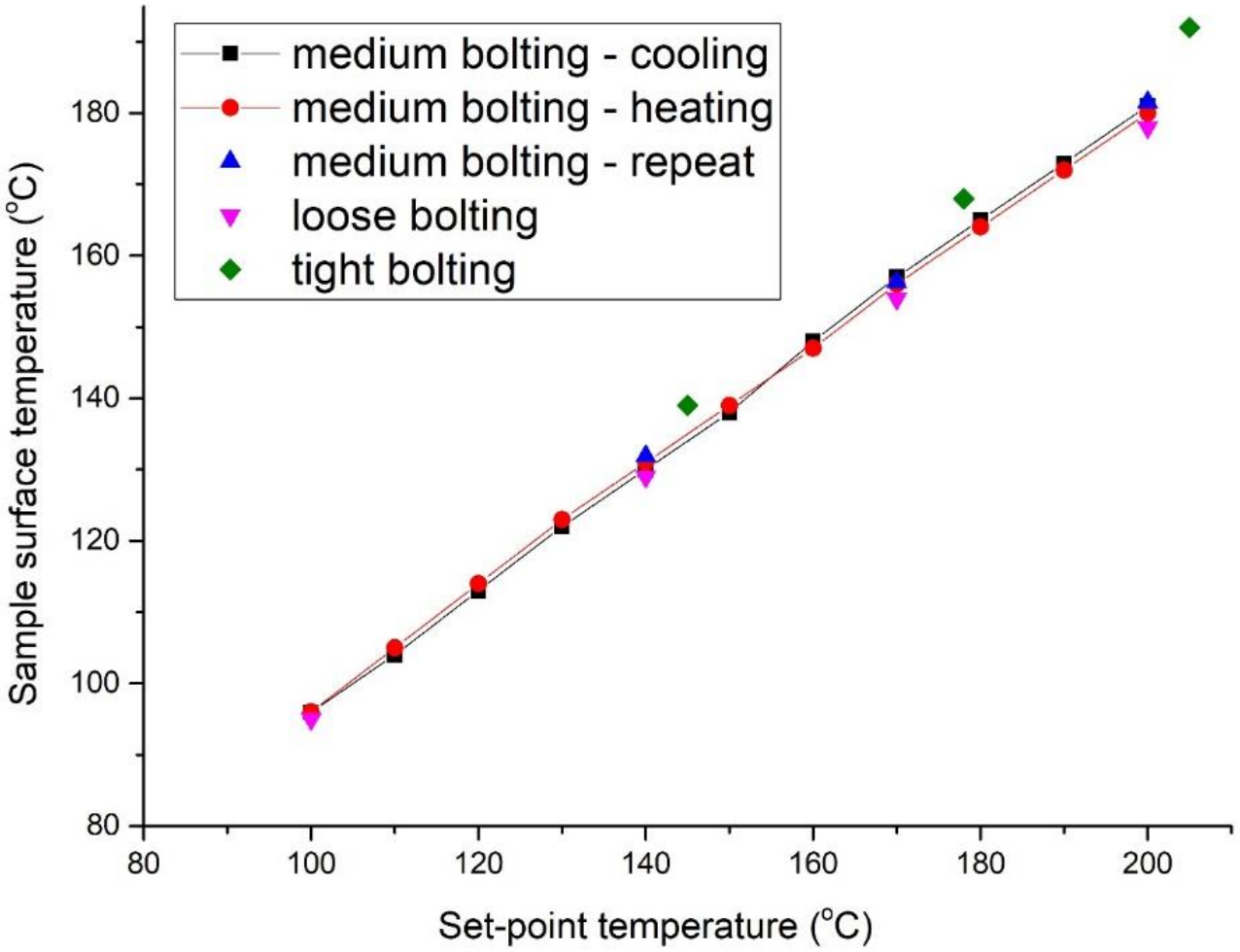


Figure S2; Offset between sample surface temperature and heater set-point for batch 1 (D17; 2021) The plot shows the stabilised sample surface temperature versus the set-point temperature for the calibration sample. Measurements were taken using three different tightnesses of bolting; tight (which was not used during the NR experiments), medium (the tightness used for almost all of the samples), and loose (used for one sample; sample F). The measurements for medium bolting, taken during cooling and heating of the samples are the mean sample surface temperatures between 6 and 8 minutes after changing the set-point to the given temperature (reducing or increasing in 10 °C steps). The measurements labelled 'medium bolting repeat' were performed following complete detachment of the calibration sample from the heater and re-attachment (the medium bolting calibrations and the repeats were actually performed several weeks apart; one before and one after the NR experiments). The repeat measurements for medium bolting were taken 15 minutes after changing the set-point to the given values.

Heater set-point temperature (°C)	100	110	120	130	140	150	160	170	180	190	200
Average sample surface temperature (°C)	96	105	114	123	131	139	148	156	165	173	181

Table S2; Average offset between sample surface temperature (for medium bolting) and heater set-point for batch 1 (D17; 2021).

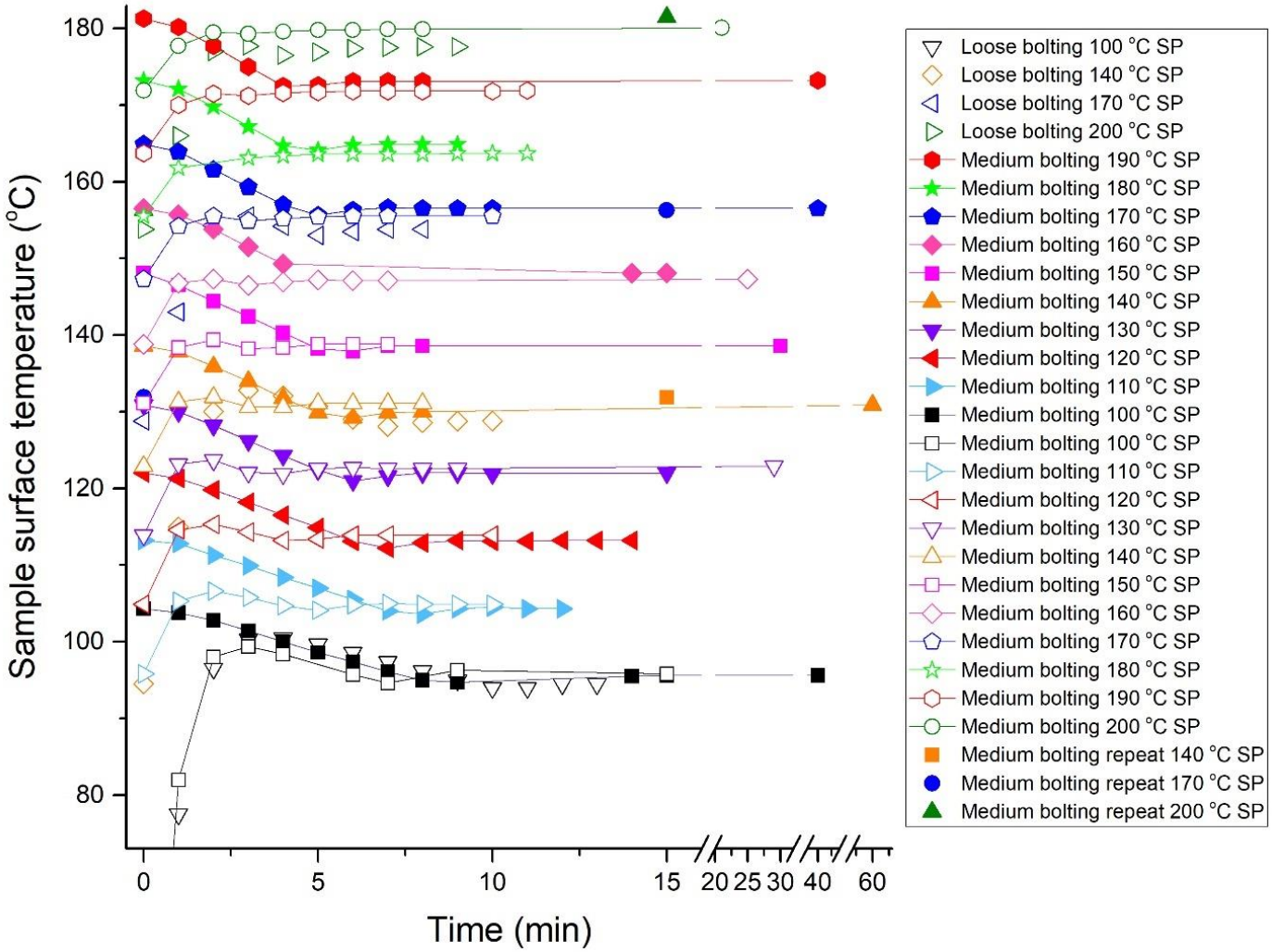


Figure S3; Sample surface temperature as a function of time, at various heater set-point (SP) temperatures for the vacuum chamber set-up used to measure batch 1. The heater temperature was changed to the given set-point at time zero in all cases.

Set-up and calibration - batch 2; The same vacuum chamber and heater set-up used for batch 1 was used for batch 2, except that the chamber was pumped down to a pressure of between 2 x 10^{-2} and 2 x 10^{-1} mbar. Calibration measurements were performed in the same way as described for batch 1, for different degrees of bolting tightness. Sample surface temperatures were stabilised within approximately 6 minutes following step changes in set-point temperatures of up to 30 °C. Figure S4 shows the sample surface temperature against the heater set-point temperature for three different tightness bolts. Fitting a straight line, y = mx + c, where y is the sample temperature, x is the set-point temperature, m is the gradient and c the intercept on the y-axis, to the data for each bolting tightness, gives the fit parameters shown in table S3. For intermediate bolting (used during the NR experiments) this gives sample surface temperatures of 99 °C, 147 °C and 157 °C, at set-point temperatures of 100 °C, 150 °C and 160 °C respectively.

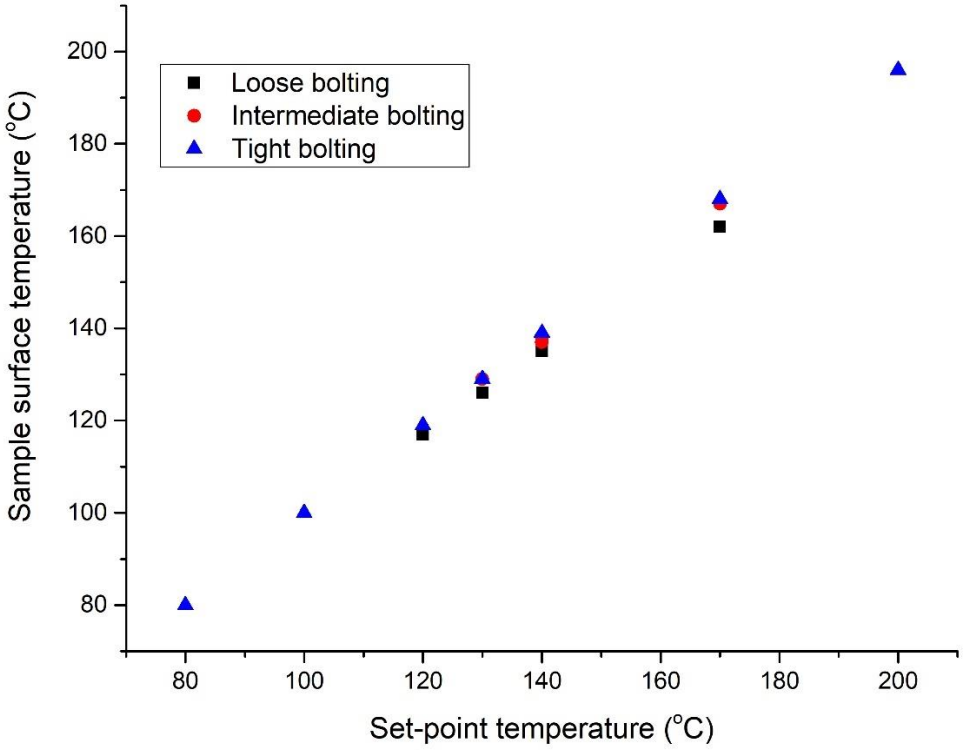


Figure S4; Offset between sample surface temperature and heater set-point for batch 2 (D17; 2019) The plot shows the stabilised sample surface temperature versus the set-point temperature for the calibration sample. Measurements were taken using three different tightnesses of bolting; intermediate bolting was used during the NR measurements.

	m	c
Loose bolting	0.9	9.2
Intermediate bolting	0.96	3.4
Tight bolting	0.97	3.0

Table S3; Linear fit parameters (y = mx + c; see text for details) for the data with the three different tightness bolts in figure S4.

The batch 1 samples (D17;2021) were annealed using a better vacuum than the batch 2 samples (D17;2019) to prevent any degradation of the materials when heated to elevated temperature for extended periods of time. We have previously performed FTIR-ATR measurements on PCBM samples annealed under rotary pump vacuum,[1] and found no evidence of the characteristic oxidation peaks that can emerge (by examining the absorption peak at 1737 cm^{-1} [2]). We repeated these measurements, but at higher temperatures, and for longer annealing periods, using rotary pump vacuum (similar to that achieved during annealing of batch 2 samples). We prepared thick films of Bis-PCBM by drop-casting onto silicon wafers. Again, no evidence of any degradation was found (see figure S5).

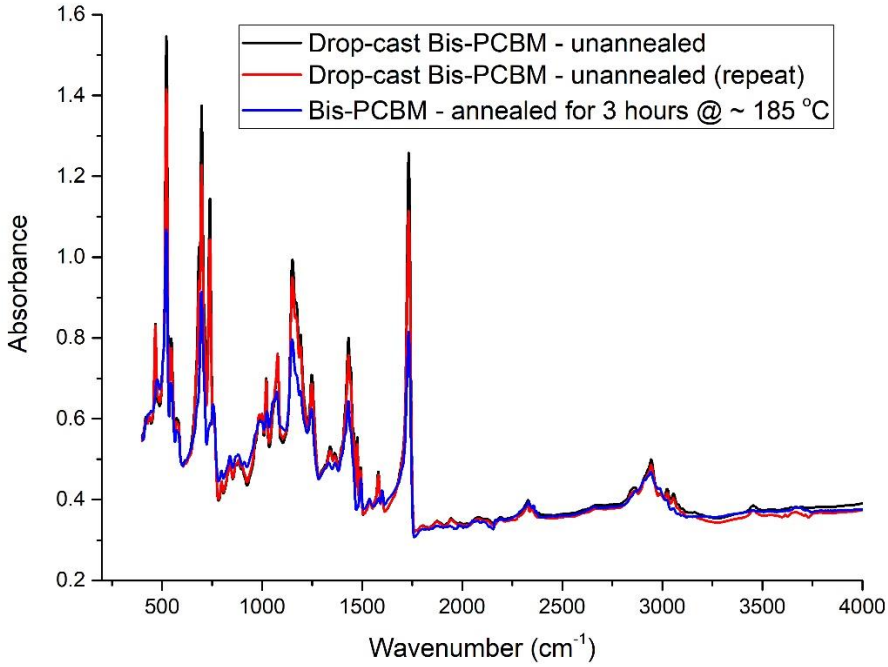


Figure S5; Fourier transform infra-red (FTIR) spectroscopy on a Bis-PCBM sample, before and after thermal annealing for three hours, in a vacuum oven (Binder, Germany) pumped down with a rotary vacuum pump (RV5, Edwards, UK) at a set-point of 195 °C (sample surface temperature of ~185 °C). The red and blue spectra have each been scaled to approximately match the black spectrum at wavenumbers above 2000 cm^{-1}.

Thermal annealing protocols - Batch 1. Two main types of annealing protocols were used for batch 1; examples of each are shown in figure S6. Firstly, full NR curves were measured at 80 °C. For the samples with top layers that were initially pure PS, the samples were typically then heated straight to a maximum value. After a waiting time of around 20 minutes for this first temperature step, a full NR measurement (measurement at both incident angles, with a total acquisition time of 10 minutes) began. Stepwise changes in the heater set-point of 10 °C were then carried out; cooling to a minimum, and then heating back up to the maximum temperature. At each temperature, measurements of a full NR curve began after a waiting time of approximately 6 minutes. For some samples, two full NR curves were measured at three different set-points during cooling (for example at set-points of 190 °C, 160 °C and 140 °C for sample C, and at set-points of 170 °C, 140 °C and 120 °C for sample H). The reproducibility of the fit parameters extracted from such repeat measurements confirmed that the waiting times that were used were sufficient to allow (re)equilibration of the sample after each temperature step. Samples were then cooled to 80 °C and a final NR curve was measured. For samples with top layers that were initially PS/Bis-PCBM blends, an isothermal annealing step (typically lasting several hours) was included in the protocol prior to the temperature cycling.

Two samples with pure PS top layers were annealed in a different way, involving larger temperature steps, but with four full NR measurements consecutively at each temperature; i) following heating at a set-point of 190 °C (sample surface @ 173 °C), sample A (300k PS top layer initially) was heated at a set-point of 150 °C (sample surface @ 139 °C) and ii) following heating at 170 °C (sample surface @ 156 °C), Sample I (2k PS top layer initially) was heated at a set-point of 120 °C (sample surface @ 114 °C).

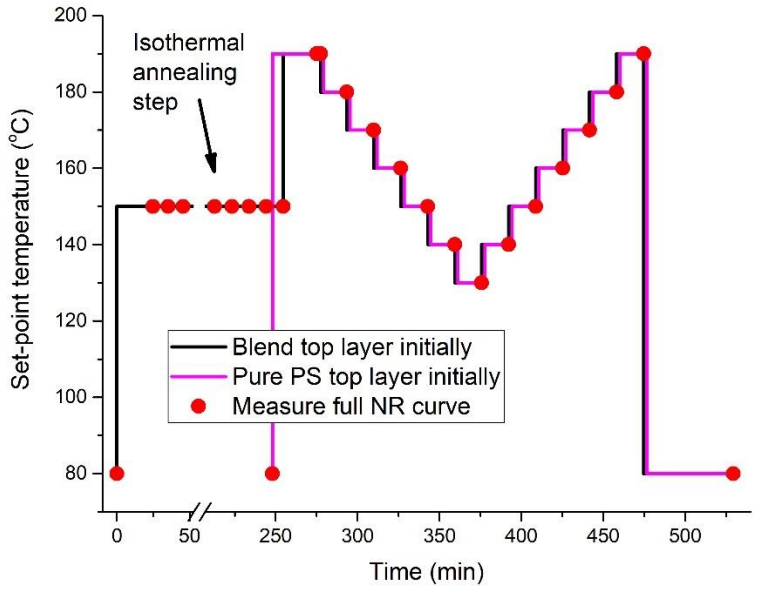


Figure S6; Typical annealing protocols for samples with pure PS and blend top layers initially. The times of the full NR curves are those at the end of each full measurement. The measurement times shown for the sample with a blend top layer initially, correspond to those for sample D.

Thermal annealing protocols - Batch 2. Samples (which all had 300k PS/Bis-PCBM blend top layers initially) were first heated to a set-point of 100 °C and then a full NR curve was measured. Samples were then heated to set-points of either 150 °C or 160 °C (see table S4 for details). Consecutive full NR measurements were then taken at these elevated temperatures.

Sample	**Annealing/NR measurement protocol;** sample surface temperatures (with heater set-points (SP) in brackets).
Sample 1	Measured at 99 °C (SP 100 °C) initially, then at 147 °C (SP 150 °C) 4 times, 157 °C (SP 160 °C) 5 times, and 147 °C 2 times (corresponding to the 12 data points for sample 1 in figures 5b and S10.
Sample 2-4	Measured at 99 °C (SP 100 °C) initially, then at 157 °C (SP 160 °C) for all remaining times.

Table S4; Annealing protocols for batch 2 samples.

NB; Throughout the text in the main paper, unless stated otherwise, given temperatures are the sample surface temperatures, rather than heater set-point temperatures (for both batches).

Data Reduction and Analysis

The vast majority of samples were reduced using the standard 'incoherent' setting within Cosmos, which integrates specularly reflected neutrons from the detector map in a Cartesian fashion in scattering angle-wavelength (θ–λ) space. Two samples (sample F in batch 1 and sample 4 in batch 2), which had broader specular reflections in the detector map were, however, reduced using the 'coherent' setting within Cosmos, which integrates within the specularly reflected region along contours of constant q_z (momentum transfer in the direction normal to the sample). As discussed below, these different methodologies gave very similar NR curves, fits, and extracted fit parameters.

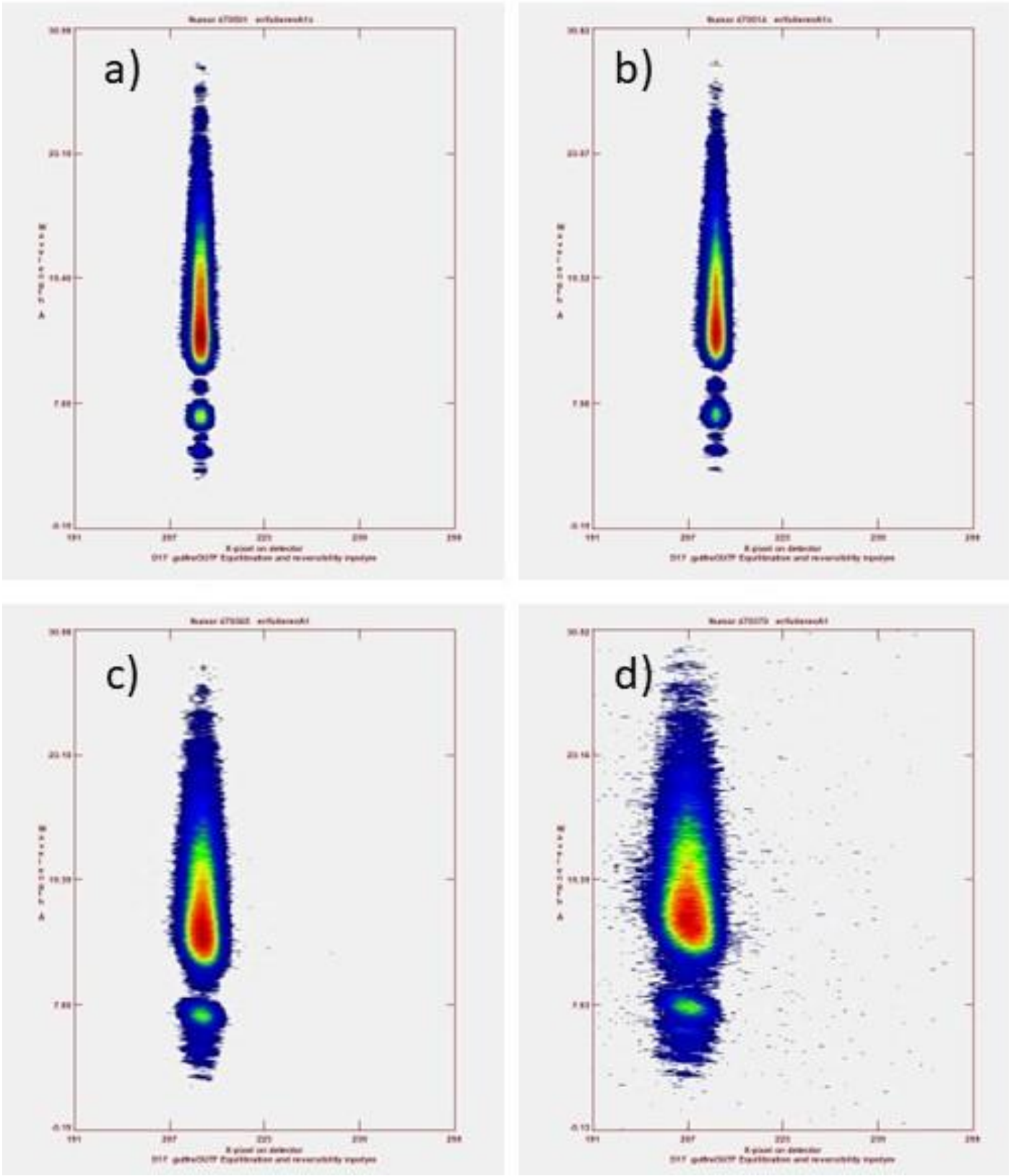


Figure S7; Detector maps (with incident angles of 0.8°) comparing broadened specular peaks (which occurred for 2 samples) to the other 11 samples. a) Sample 3 at a set-point of 100 °C. b) Sample 3 at a set-point of 160 °C. c) Sample 4 at a set-point of 100 °C. d) Sample 4 at a set-point of 160 °C. All 4 panels plot detector pixel on the x-axis (zoomed in to only shows pixels 191-255) and neutron wavelength (Å) on the y-axis (with a scale from approximately 0 up to 31 Å). Integration of each detector map along the y-axis and projection onto the x-axis gives a specular peak with a full-width at half-maximum approximately as follows; a) 2 pixels, b) 2 pixels, c) 4 pixels and d) 7 pixels.

Figure S7 shows the broader specularly reflected peaks for sample 4 (batch 2) in comparison to the typical specular reflections from other samples measured in this study. We were not able to eliminate this broader reflection, despite

adjusting the bolting of these samples to the heater block. For sample F the width of the specular peak narrowed and then broadened again during thermal cycling, while for sample 4 it broadened after being heated from a set-point of 100 °C to a set-point of 160 °C, and the position of the specular reflection shifted slightly on the detector (see figures S7c and d). It is possible, therefore, that the broadness of the specular peak in these two samples is a combination of an intrinsically less flat sample, combined with potential strain due to clamping of the sample (which can change with temperature). It was therefore decided to reduce the data from these two samples using two different methodologies; i) standard 'incoherent' data reduction in which the specular reflection in the detector map shown in figure S7 is integrated over the scattering angle, θ, at constant wavelength, and ii) 'coherent' data reduction, in which (to accommodate slightly bent samples which present a range of incident angles to the incoming neutron beam) the specular peak is integrated along lines of constant q_z.[3] The NR curves and the extracted fit parameters are very similar for these two reduction methodologies for both samples. The extracted fit parameters for sample 4 are shown in figure S8. As can be seen, the data is robust with-respect-to the choice of data reduction methodology.

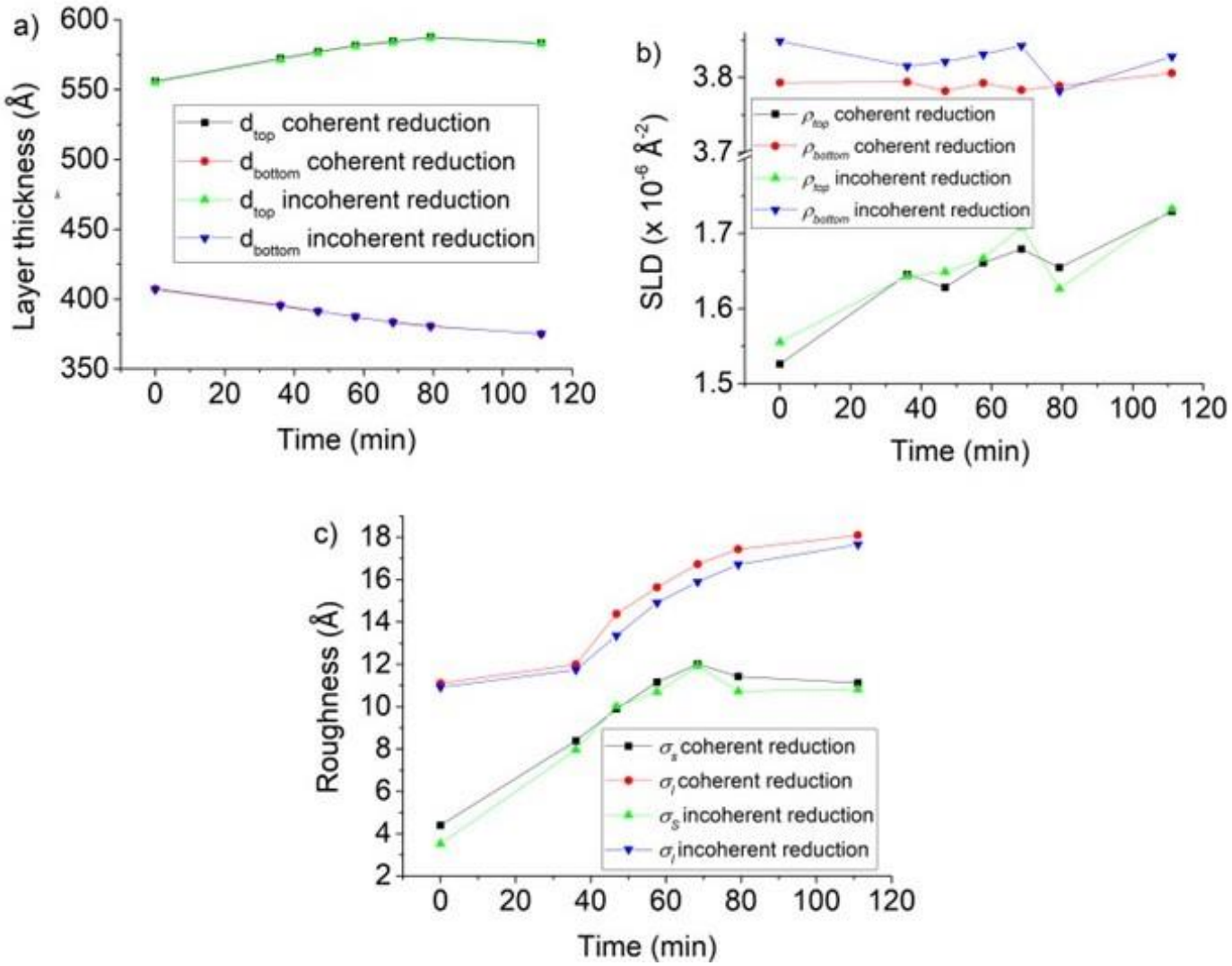


Figure S8; Comparison of coherent and incoherent data reduction methods for sample 4. Fit parameters versus time for a) layer thicknesses, d_{top} and d_{bottom}, b) SLDs, ρ_{top} and ρ_{bottom} and c) surface/interface roughnesses, σ_s and σ_i. NB; The coherent SLD parameters in b) are the same as for sample 4 plotted in figure 5b.

Goodness-of-fit, χ^2, parameters for samples A-I and 1-4 are given in table S5.

Sample	χ^2 (N_D)
A	277-517 (247-251)
B	251-406 (247-252)
C	283-379 (248-253)
D	217-452 (244-251)
E	309-428 (247-252)
F	218-659 (183-185)
G	281-689 (247-254)
H	269-504 (248-255)
I	318-624 (251-253)
1	501 (199), 2641 (179), 2031 (181), 1521 (179), 1356 (183), 635 (173), 449 (181), 285 (172), 296 (176), 299 (180), 291 (184), 262 (175)
2	643 (164), 2334 (163), 1088 (163), 805 (164), 622 (163), 557 (162)
3	710 (269), 1047 (273), 881 (271)
4	283 (154), 295 (139), 265 (141), 220 (136), 259 (138), 254 (139)

Table S5; Goodness-of-fit, χ^2, parameters for the (Levenburg-Marquardt/differential evolution) fits to NR measurements on samples A-I and 1-4. The χ^2 values given for samples A-I (batch 1) represent the entire range of NR measurements on these samples (i.e. at 80 ^oC before and after annealing and during thermal cycling, and also during isothermal annealing of samples D, E and G). The N_D parameters, given in brackets, are the number of data points in the reflectivity curves (again representing the entire range of NR measurements on these samples). For samples 1-4 (batch 2) the χ^2 values (and N_D values in brackets) are stated for every fit; these are given in chronological order (i.e. corresponding to the data points in sequence in figure 5b).

2 Supplementary Information for Figure 5

The roughness fit parameters that accompany the batch 1 SLDs in figure 5a are shown in figure S9b (the SLDs from 5a are reproduced in S9a, with a new combined legend).

The thickness parameters that accompany the batch 2 SLD data in figure 5b are shown in figure S10. This shows that the three samples with mean top layer compositions initially in the two-phase region of the phase-diagram (samples 1-3), experience thinning of the top layer and thickening of the bottom layer, on annealing, as fullerene moves from the top layer into the bottom layer. In contrast, sample 4, which has a mean top layer composition in the single-phase region of the phase-diagram shows the opposite behaviour, as mass transfer of fullerene from the bottom to the top layer occurs.

The lines in figure 5b) are fits of type $y = A + (B-A)\, e^{-t/\tau}$, where y represents the SLD, t is the annealing time, and A, B and τ are constants. In the fits of the 4 samples shown in figure 5b, the parameters A and τ are shared between all 4 samples, but B is allowed to take a different value for each sample. The fitted B parameters for samples 1-4 are 2.07, 2.06, 1.88 and 1.56 x 10^{-6} $Å^{-2}$ respectively. The fitted A and τ parameters are 1.73 x 10^{-6} $Å^{-2}$ and 95 minutes respectively.

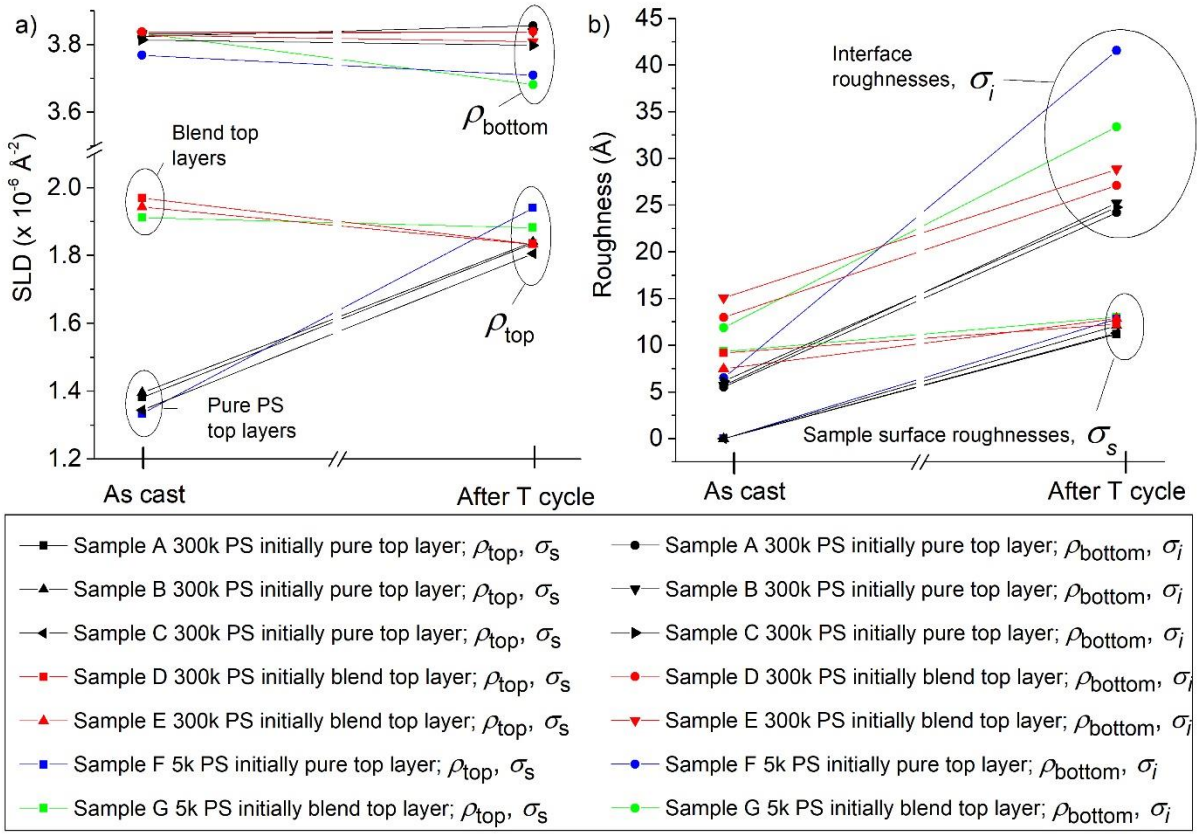


Figure S9; This figure accompanies figure 5a. a) A duplicate of the SLD fit parameter data shown in figure 5a (with different colours and symbols identifying each sample, and here matching those in b)). b) The accompanying surface and interface roughness fit parameters. The legend applies to a) and b).

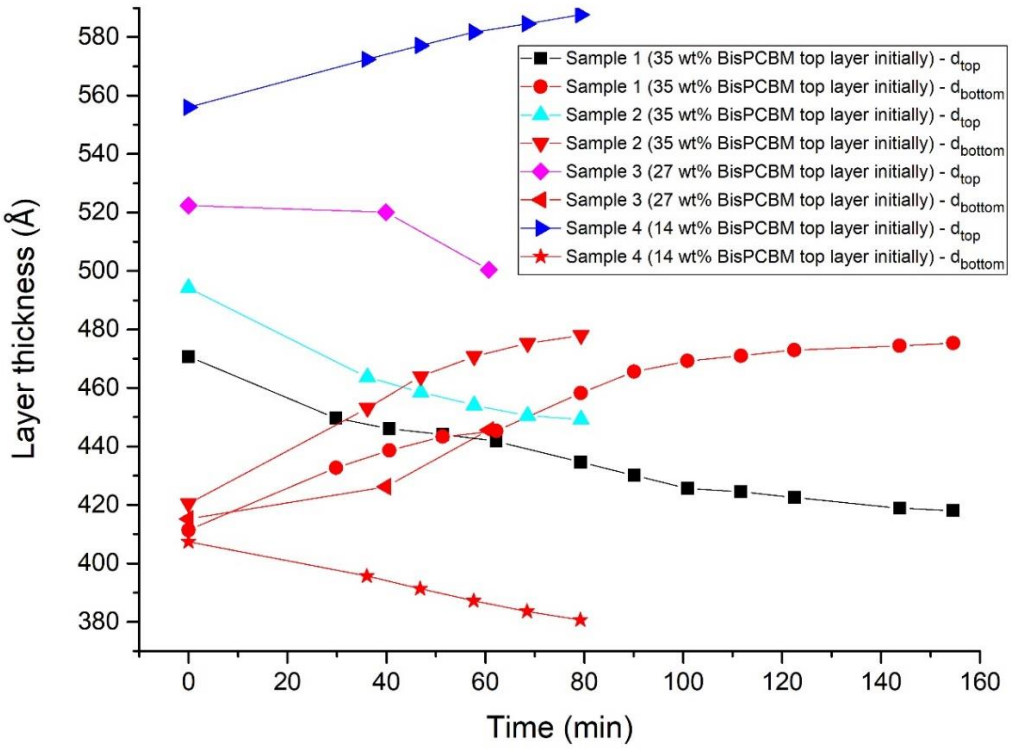


Figure S10; Layer thickness fit parameters that accompany the SLD fit parameters for the batch 2 samples in figure 5b. Samples 1-3 have top layers with mean compositions that are initially in the two-phase region of the phase-diagram. Sample 4 has a top layer composition that is initially in the one-phase region of the phase-diagram. All time zero measurements were done at a sample surface temperature of 99 °C. Subsequent measurements were done at 157 °C for samples 2-4, but at a mixture of 147 °C and 157 °C for sample 1 (see table S4 for details).

3 Supplementary Information for Figure 6

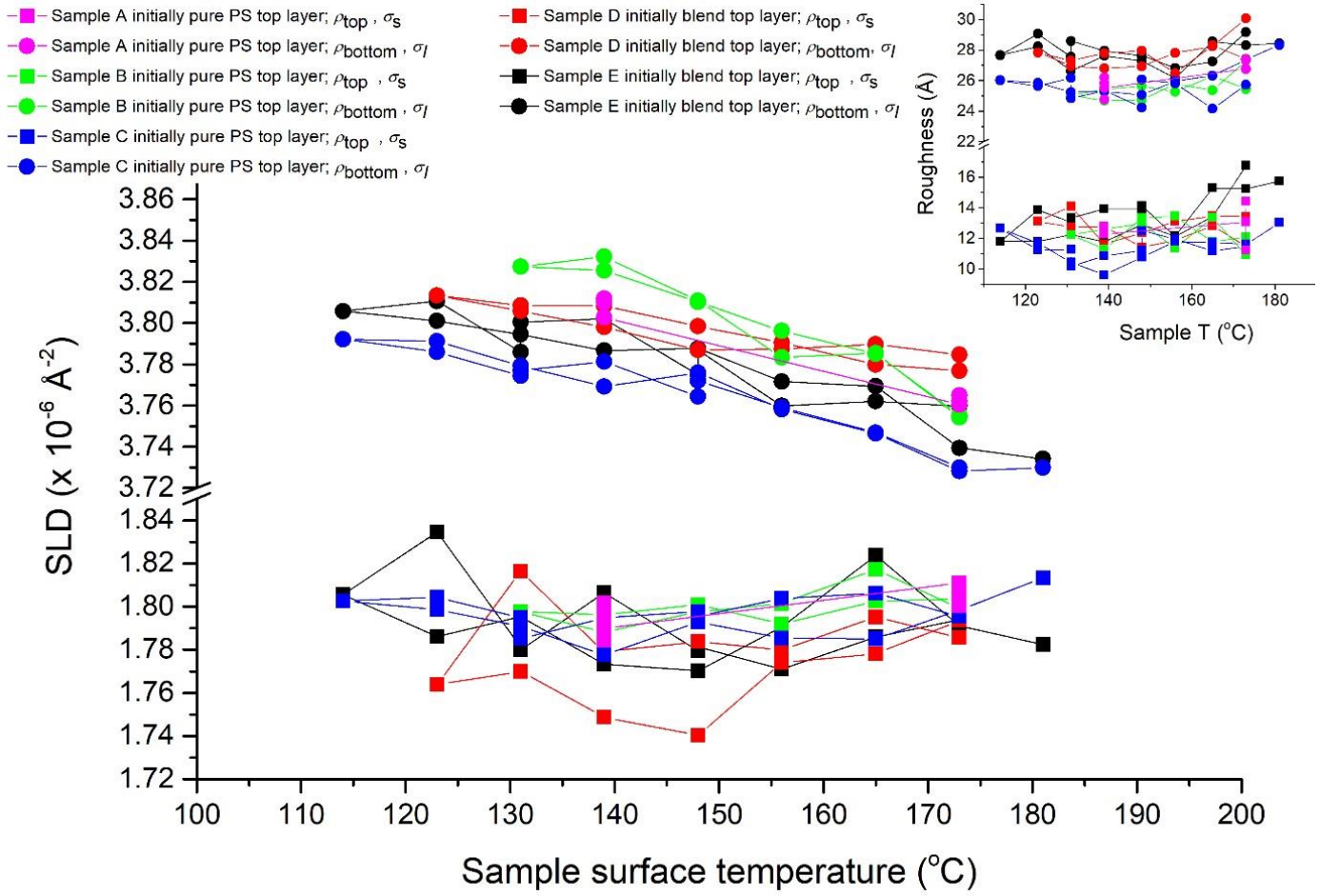


Figure S11; SLD fit parameters and (inset) surface/interface roughness fit parameters that accompany the thickness fit parameters shown in figure 6. The legend applies to both the main figure and the inset.

Thermal expansion data for a single Bis-PCBM layer on silicon is shown in figure S12. This involves fitting a single uniform layer to NR measurements on a sample heated between 80 °C and 180 °C (set-points). The SLD goes down and the thickness goes up with temperature, with a change in gradient above around 130-140 °C (most evident in the thickness data), close to the reported glass transition temperature of Bis-PCBM.[4] Linear fits to the data in figures S12 a) and b), for temperatures above 135 °C, gives thermal expansion coefficients of approximately 1.7 x 10^{-4} K^{-1} and 2.9 x 10^{-4} K^{-1} respectively (in each case, this is for fits of three sets of data combined; i) angle 1 measurements during heating, ii) angle 1 measurements during cooling and iii) angle 1 and angle 2 measurements combined).

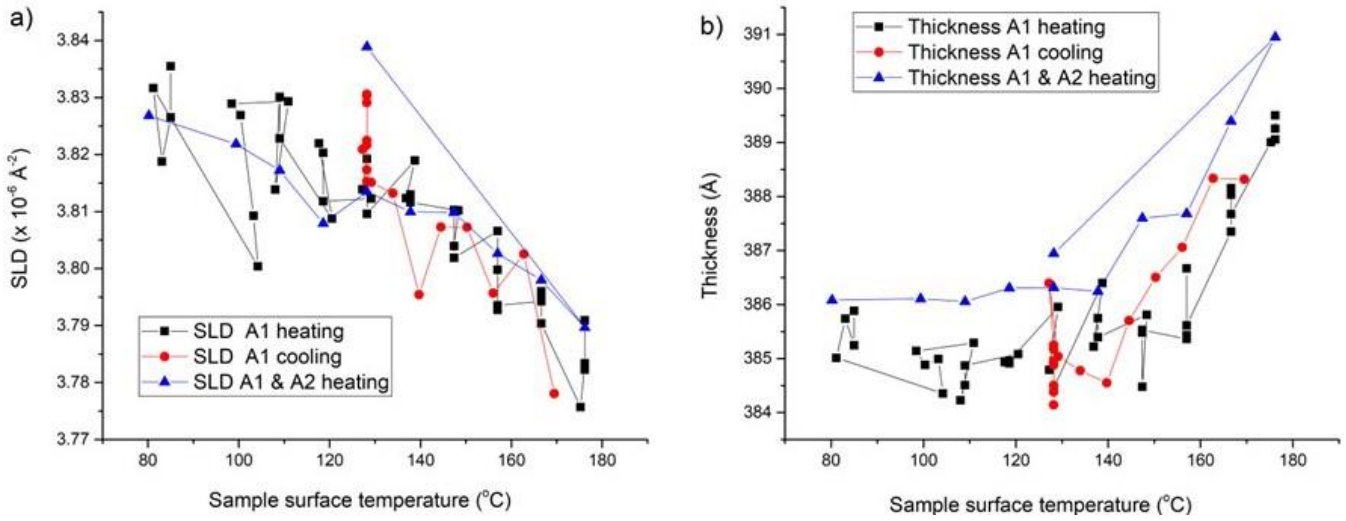


Figure S12; Single layer Bis-PCBM data (measured as part of batch 2 on D17 in 2019). This sample was heated at a set-point (SP) temperature of 80 °C and then from SPs of 100 °C to 180 °C in 10 °C steps. Following each change in SP temperature at 100 °C and above, 3 consecutive neutron reflectivity (NR) measurements (2 minutes acquisition time each) were taken at an incident angle of 0.8° (A1). This was then followed by a full NR measurement at each SP, using incident angles of 0.8° (A1) and 3° (A2). During cooling to 130 °C (SP), repeated measurements at 0.8° (A1) were performed. The sample surface temperatures during the NR measurements were calculated from the heater temperatures, using the fit to the intermediate bolting data for the batch 2 data (see figure S4 and table S3).

4 Supplementary Information for Figure 8

Bilayer and spline fits for the two samples containing 2k PS (samples H and I) are shown in figures S13 and S15 below. The fit parameters as a function of temperature are then shown in figures S14 and S16 for samples H and I respectively. Given the low contrast between the two layers shown in the SLD profiles, the relatively low thickness of the bottom layer and the appearance of only a single obvious fringe periodicity in the NR curves, single layer fits were also attempted. However, these had a significantly higher value of the goodness-of-fit, χ^2, parameter, in comparison to the bilayer and spline fits, and did not correctly capture the size of the fringes (in particular, the minima, were not fitted well).

With regard to the fit parameters for samples H and I, figure S17 shows that there is significant variability in the buried interface roughness parameter at both high and low annealing temperatures, and the values of these parameters should therefore be treated with particular caution. With regard to the other fit parameters, one potentially significant similarity between the 2k PS data and the data from 5k and 300k PS samples is that again we see the bottom layer thickness decrease with temperature, but the top layer thickness increase (see figures S14a and S16a), again indicative of greater miscibility between the components at higher temperatures. There is however a subtle difference in the behaviour of the fitted layer SLDs with temperature for 2k PS (figure 8). If the data from figure 8 for sample H is replotted on its own (see figure S14b), a *reduction* in the SLD of *both* layers with temperature is evident. This is also found at the two measured temperatures for sample I (see figure S16), and is in contrast to the behaviour found for 5k PS samples (figures 7c and 7f) in which the top layer SLD, ρ_{top}, increases with temperature. Given the caveat regarding the perturbation of the SLD profiles due to the relatively low thickness of the bottom layers in these samples, these findings are rather tentative, and further investigations would be required to confirm (or otherwise) whether such thermal behaviour is reliable for 2k PS/Bis-PCBM systems.

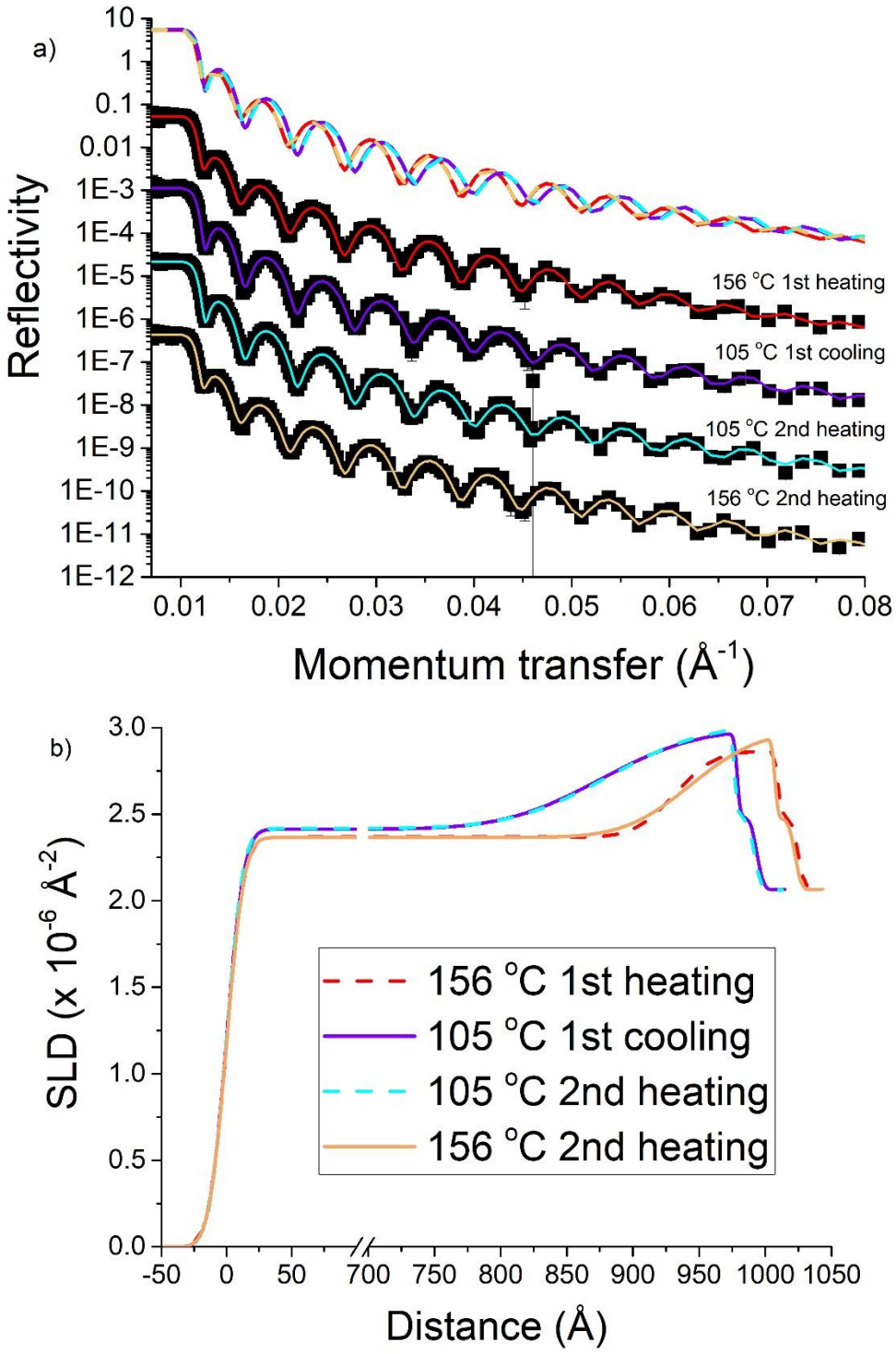


Figure S13; Selected NR data and fits for sample H (initially pure 2k PS top layer). After initial measurement at 80 °C, this sample underwent heating to a sample surface temperature of 156 °C, then stepwise cooling to 96 °C, and then a second heating to 165 °C. Finally, the samples were cooled slowly to 80 °C. The plots show data for two measurements at 105 °C and two at 156 °C. a) NR data and fits (the curves are offset vertically for clarity). To enable comparison the fit curves are reproduced at the top, with no vertical offset. b) SLD profiles (bilayer fits) corresponding to the fits in a).

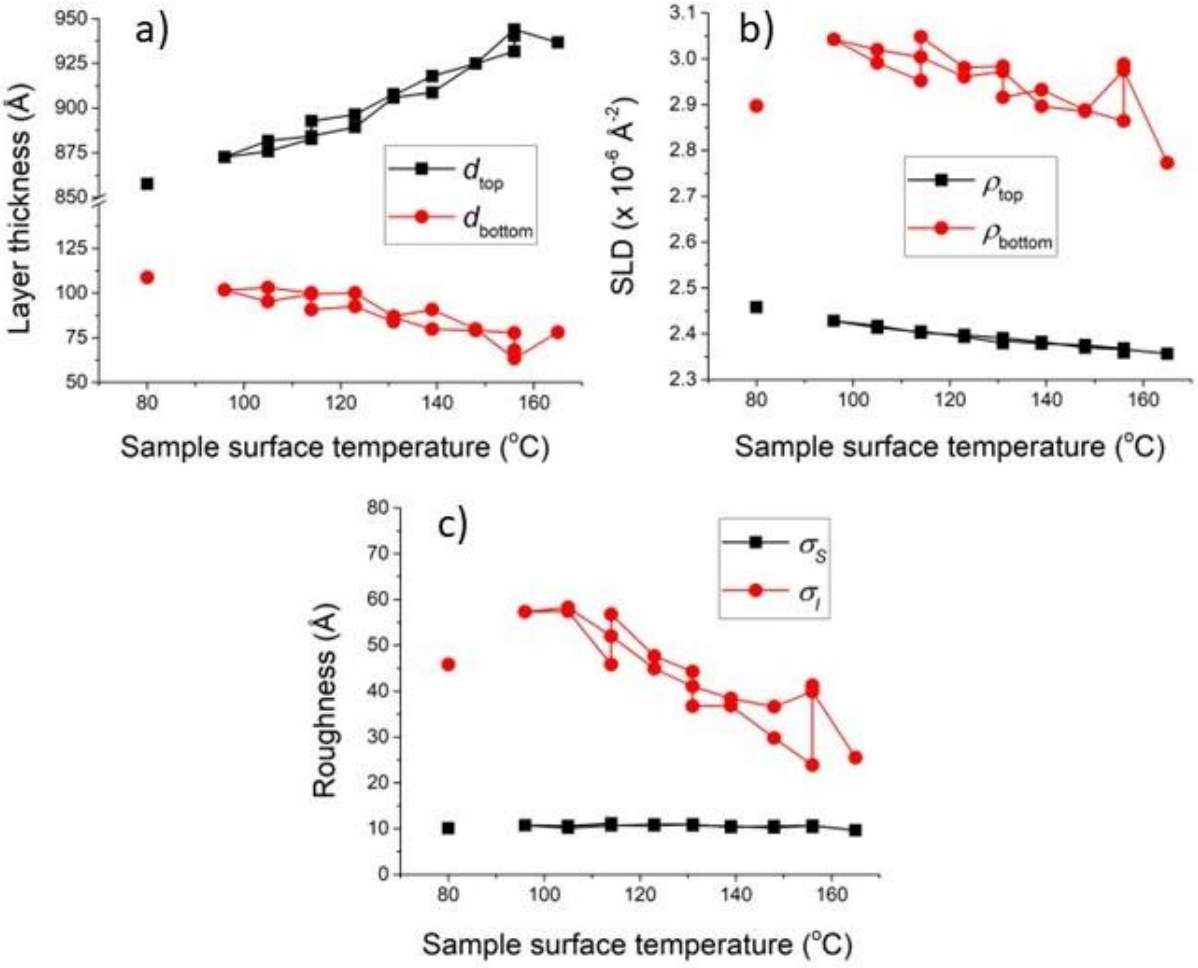


Figure S14; Fit parameters as a function of temperature, following equilibration, for sample H (initially pure 2k PS top layer). a) Layer thicknesses, d_{top} and d_{bottom}. b) Scattering length densities (SLDs), ρ_{top} and ρ_{bottom}. c) Sample surface roughness σ_S and interfacial roughness σ_I. The data in b) is the same as in figure 8 (plotted alongside data from samples A-G and I), but plotted on a different scale, with the x and y axes swapped. The data in c) is plotted in figure S17 below, alongside data from samples A-G and I. The data points at 80 °C represent measurements after final cooling to 80 °C.

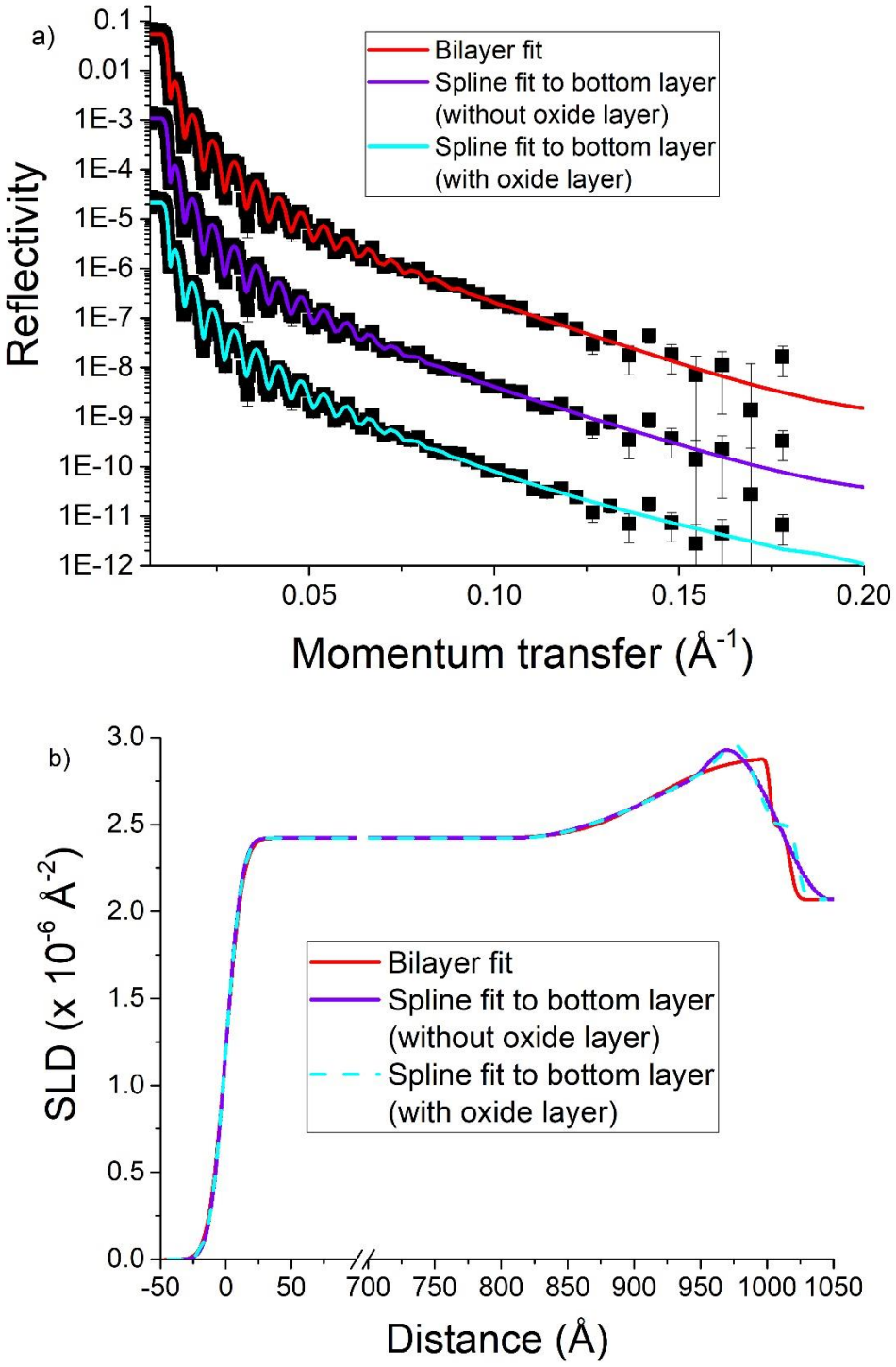


Figure S15; Comparison of a bilayer fit, and fits in which the bottom layer was represented by a spline[5] for sample I (initially pure 2k PS top layer) at a sample surface temperature of 156 °C. a) NR data (plotted 3 times with vertical offsets) and fits. b) Comparison of SLD profiles corresponding to the fits in a). The spline layers had 2 nodes. The position and SLD of the nodes, and the thickness of the bottom layer was adjustable (along with the thickness, SLD and roughness of the top layer), giving a total of eight adjustable parameters in the spline fits.

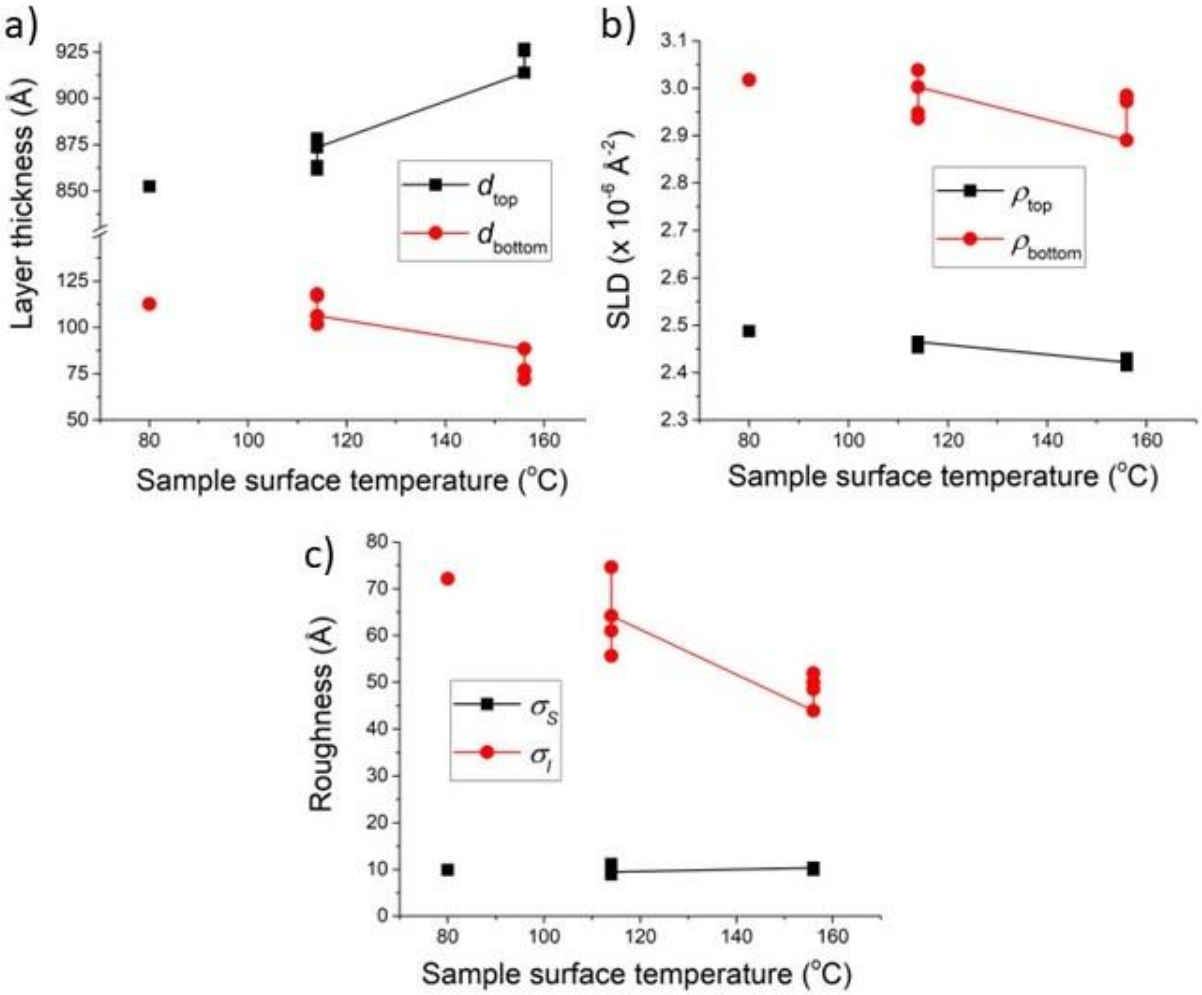


Figure S16; Fit parameters (bilayer fits) as a function of temperature, following equilibration, for sample I (initially pure 2k PS top layer – this is a duplicate of sample H in figure S14, but with 4 consecutive full NR measurements performed at sample surface temperatures of 156 °C and then 114 °C). a) Layer thicknesses, d_{top} and d_{bottom}. b) Scattering length densities (SLDs), ρ_{top} and ρ_{bottom}. c) Sample surface roughness σ_S and interfacial roughness σ_I. The axes in a)-c) are all plotted on the same scales as in figure S14. The data in b) is the same as in figure 8 (plotted alongside data from samples A-G and I), but plotted on a different scale, with the x and y axes swapped. The data in c) is plotted in figure S17 below, alongside data from samples A-G and I. The data points at 80 °C represent measurements after final cooling to 80 °C.

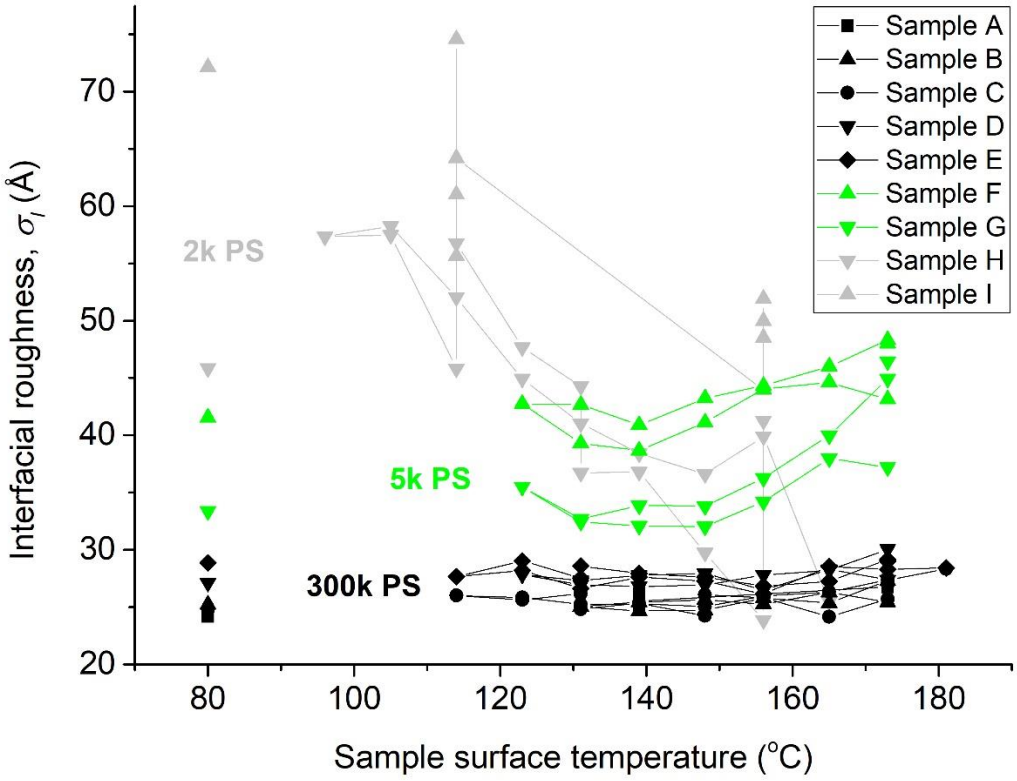


Figure S17 Summary of equilibrated interfacial roughness parameter σ_I as a function of MW and temperature. The plot shows roughness parameters obtained during *in-situ* thermal cycling, following equilibration (measured as the temperature is cycled from high-to-low-to-high, and finally after cooling to 80 °C). The plot combines the (bilayer) fit parameters for 2k, 5k and 300k PS/Bis-PCBM bilayers shown in figures S14c, S16c, 7e, 7f and S11 (inset).

5 Post Annealing Microscopy

Following annealing, all samples were imaged using optical microscopy (using a Nikon Eclipse microscope with a x50 objective). Samples A-E and 1-4, that contained 300k PS (both those with top layers that were initially pure PS, and those that were initially blends) were very uniform at the end of annealing, with little discernible lateral morphology (see figures S18 a and b). In contrast, both samples containing 5k PS show a morphology containing undulations on a lateral lengthscale of order a few tens of micrometres (figure S18 c and d). The 2k PS samples also exhibit some lateral morphology (see figures S18 e and f). Atomic force microscopy (AFM) imaging of these samples (measured in tapping mode using either a Veeco Dimension or a JPK Nanowizard 3) reveals root-mean-square (rms) roughness of around 1-2nm (measured over image sizes of 50 x 50 μm), for 300k, 5k and 2k PS samples. This is of the same order as the surface roughness fit parameters obtained from NR. The origin of the observed undulations is therefore not clear. They could be associated with the early stages of dewetting of one of the layers within the sample. This could potentially be the cause of the fit parameters at high temperatures immediately before final cooling to 80 °C not reproducing the fit parameters extracted at the same temperatures on first heating for the 5k PS samples (see the data points at the highest annealing temperatures of 173 °C in figures 7 b, d, e and f). It is, however, clear that (non-reversible) dewetting has not affected the behaviour at temperatures below 173 °C, where reversible behaviour is observed. Samples also contain linear features arising from folds (an example of which is shown in figure S18 f) or occasional splits in the top layer during floating/drying, or arising from terraces in the mica during cleaving. These linear defects cover only a small fraction of the area of the samples, and are therefore not expected to contribute to the measured NR from the samples.

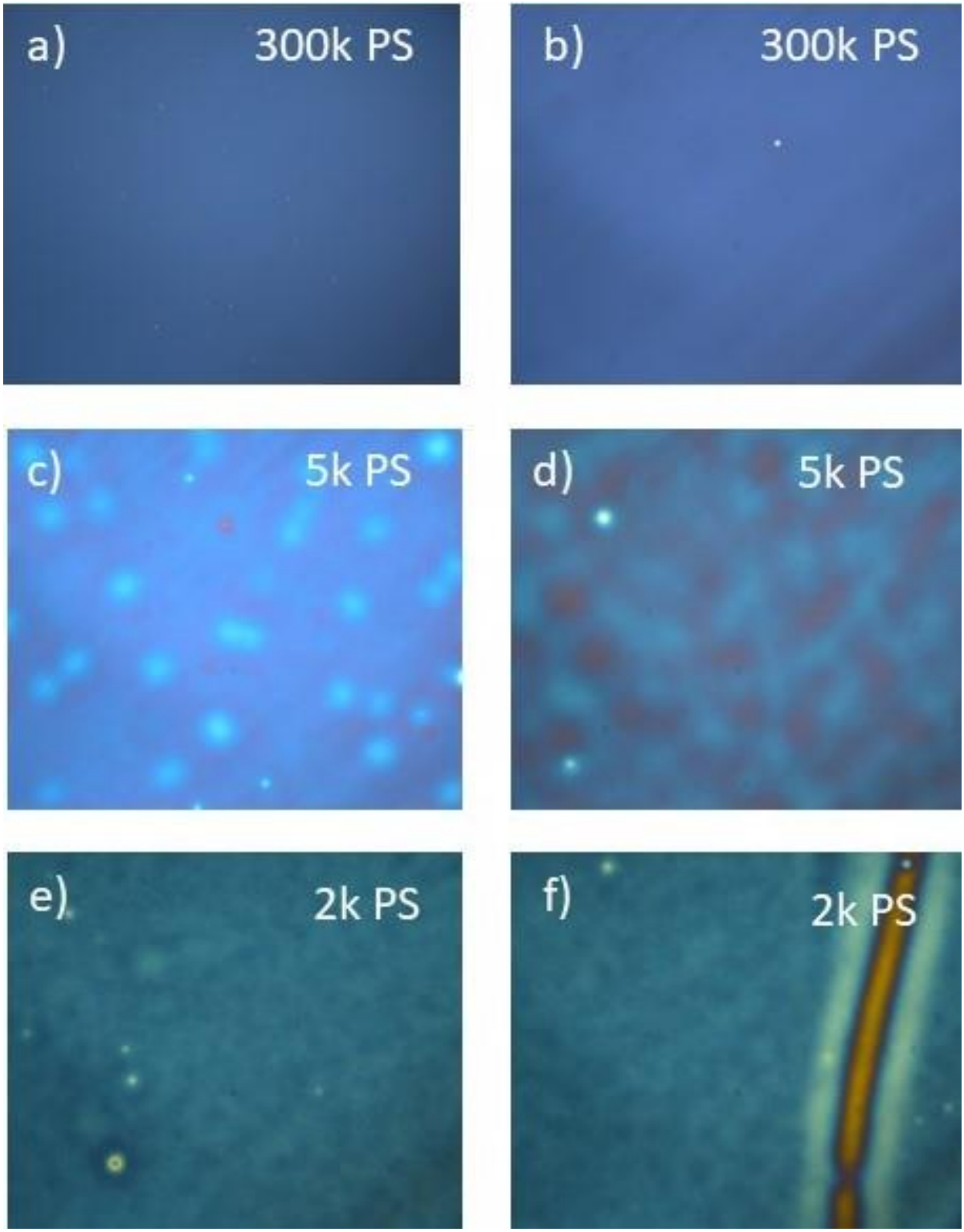


Figure S18; Optical micrographs of samples after annealing. a) Sample D. b) Sample D. c) Sample G. d) Sample F. e) Sample I. f) Sample I. All image sizes are 200 μm x 240 μm, except for a) which is 1 mm x 1.2 mm.

References for Supplementary Information

1 Hynes, E. L. *et al.* Interfacial width and phase equilibrium in polymer-fullerene thin-films. *Communications Physics* **2**, doi:10.1038/s42005-019-0211-z (2019).

2 Speller, E. M. *et al.* Impact of Aggregation on the Photochemistry of Fullerene Films: Correlating Stability to Triplet Exciton Kinetics. *Acs Applied Materials & Interfaces* **9**, 22739-22747, doi:10.1021/acsami.7b03298 (2017).

3 Gutfreund, P. *et al.* Towards generalized data reduction on a chopper-based time-of-flight neutron reflectometer. *J. Appl. Cryst.* **51**, 606-615 (2018).
4 Leman, D. *et al.* In Situ Characterization of Polymer–Fullerene Bilayer Stability. *Macromolecules* **48**, 383-392, doi:10.1021/ma5021227 (2015).
5 Nelson, A. R. J. & Prescott, S. W. refnx: neutron and X-ray reflectometry analysis in Python. *J Appl Crystallogr* **52**, 193-200, doi:10.1107/S1600576718017296 (2019).