



Probing the oligomeric re-assembling of bacterial fimbriae in vitro: a small-angle X-ray scattering and analytical ultracentrifugation study

Alexandra S. Solovyova¹  · Daniel T. Peters²  · Gema Dura^{3,4}  · Helen Waller²  · Jeremy H. Lakey²  · David A. Fulton³ 

Received: 22 October 2020 / Revised: 14 April 2021 / Accepted: 20 April 2021 / Published online: 4 May 2021
© The Author(s) 2021

Abstract

Capsular antigen fragment 1 (Caf1) is an oligomeric protein consisting of 15 kDa monomeric subunits that are non-covalently linked through exceptionally strong and kinetically inert interactions into a linear polymer chain. It has been shown that after its thermal depolymerisation into unfolded monomeric subunits, Caf1 is able to efficiently repolymerise in vitro to reform its polymeric structure. However, little is known about the nature of the repolymerisation process. An improved understanding of this process will lead to the development of methods to better control the lengths of the repolymerised species, and ultimately, to better design of the properties of Caf1-based materials. Here we utilize small-angle X-ray scattering to estimate the size of Caf1 polymers during the first 24 h of the re-polymerisation process. Analytical ultracentrifugation measurements were also used to investigate the process post-24 h, where the rate of repolymerisation becomes considerably slower. Results show that in vitro polymerisation proceeds in a linear manner with no evidence observed for the formation of a lateral polymer network or uncontrolled aggregates. The rate of Caf1 in vitro repolymerisation was found to be concentration-dependent. Importantly, the rate of polymer growth was found to be relatively fast over the first few hours, before continuing at a dramatically slower rate. This observation is not consistent with the previously proposed step-growth mechanism of in vitro polymerisation of Caf1, where a linear increase in polymer length would be expected with time. We speculate how our observations may support the idea that the polymerisation process may be occurring at the ends of the chains with monomers adding sequentially. Our findings will contribute towards the development of new biomaterials for 3D cell culture and bio-printing.

Keywords Oligomeric growth · Repolymerisation · Mass increment · Flexibility

Introduction

Capsular antigen fragment 1 (Caf1) is an oligomeric protein consisting of 15 kDa monomer subunits that are non-covalently linked into a linear polymer chain (Fig. 1a). The protein belongs to the chaperone–usher class of bacterial pilins (Waksman and Hultgren 2009; Zav'yalov et al. 2010), which polymerise at the surface of *Yersinia pestis* bacteria to form long chains [up to 2 μm in length (Soliakov et al. 2010)] that protect the bacteria from the host immune system (Du et al. 2002). Each subunit possesses an incomplete immunoglobulin-like fold, which must be completed through donation of an N-terminal donor strand from a second subunit into the hydrophobic acceptor cleft, and by linking subunits together in this way, a polymer is formed (Zavialov et al. 2003). The association constant (K_A) of the subunit interaction is estimated to be at least 10^{14} M^{-1} , with dissociation times in the order of billions of years (Yu et al. 2012; Zavialov et al.

Special Issue: COST Action CA15126, MOBIEU: Between atom and cell.

✉ Alexandra S. Solovyova
alexandra.solovyova@newcastle.ac.uk

✉ David A. Fulton
david.fulton@newcastle.ac.uk

¹ Proteome and Protein Analysis, Medical School, Newcastle University, Framlington Place, Newcastle-upon-Tyne NE2 4HH, UK

² Biosciences Institute, Medical School, Newcastle University, Framlington Place, Newcastle-upon-Tyne NE2 4HH, UK

³ Chemistry, School of Natural and Environmental Sciences, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

⁴ Present Address: Departamento de Química Inorgánica, Orgánica y Bioquímica, Universidad de Castilla-La Mancha, Facultad de Ciencias y Tecnologías Químicas-IRICA, Avda. C. J. Cela, 10, 13071 Ciudad Real, Spain

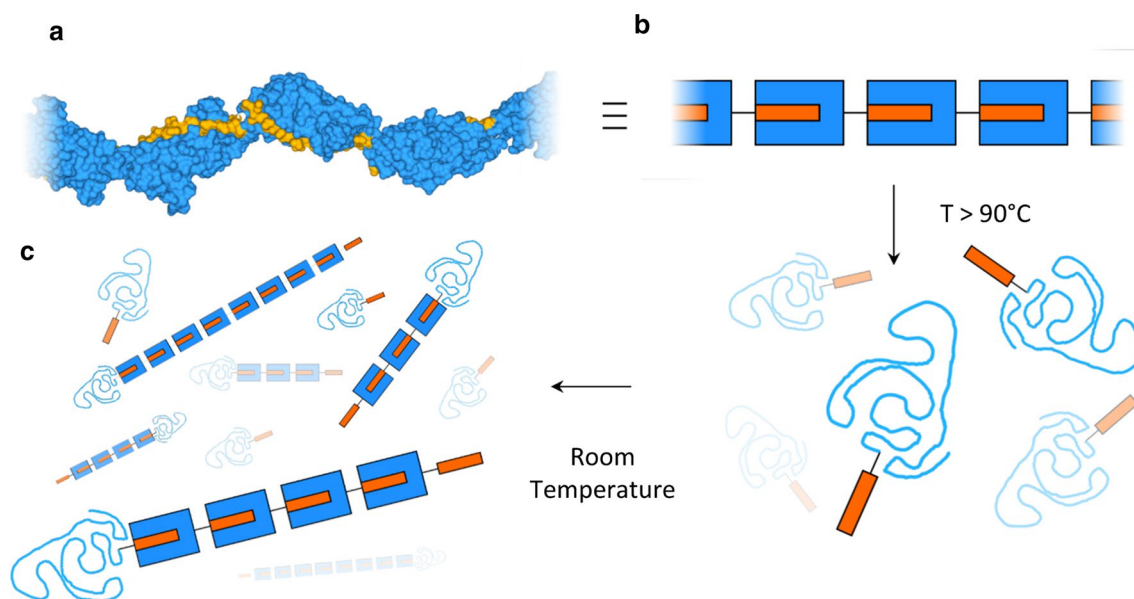


Fig. 1 Structure and reversible thermal unfolding of Caf1 polymers. **a** Model of a segment of Caf1 polymer (generated from PDB entry 1P5U). The N-terminal donor strands (coloured orange) are inserted into the acceptor cleft of adjacent subunits. Protein structures were visualised using the CCP4MG molecular graphics package (McNicholas et al. 2011). **b** Cartoon depicting the reversible thermal unfolding of Caf1 polymers. When heated above the subunit melting tempera-

ture ($T_m > 86\text{ }^\circ\text{C}$) the subunits (blue) thermally unfold with concomitant decomplexation of the N-terminal donor strands (orange), causing Caf1 to depolymerize into its unfolded monomeric form. When the subunits are incubated at room temperature they subsequently refold with donor strand recomplexation and concomitant repolymerisation, with a fraction of unfolded monomer still present (c)

2005), and the remarkable strength of these kinetically inert inter-subunit linkages bestows the Caf1 polymer with excellent stability towards extremes of pH and proteases (Chalton et al. 2006; Ulusu et al. 2017).

We recently reported (Dura et al. 2020); however, that, despite their high stability, the inter-subunit linkages can be completely dissociated by utilizing thermal protein unfolding (Fig. 1b). At its melting temperature of $86\text{ }^\circ\text{C}$ (Chalton et al. 2006; Zavialov et al. 2005), the subunits of Caf1 unfold with concomitant decomplexation of the N-terminal donor strands, causing complete depolymerisation of the Caf1 chain. We speculate that heat perturbs the tertiary structure of the acceptor to the extent that it is no longer complementary to the N-terminal donor strand, resulting in decomplexation. Surprisingly, this thermal unfolding-depolymerisation process is reversible, and upon cooling the subunits were shown to refold with concomitant recomplexation of the N-terminal donor strand, reforming the polymer chain (Fig. 1c). Throughout this process, a large initial fraction of uncomplemented (and hence disordered) monomer is present, decreasing over time as the subunits refold and reassemble into oligomers, but still present in small amounts even at longer time scales (Dura et al. 2020). Remarkably, Caf1's capacity for reversible thermal unfolding is maintained even when extensively cross-linked to form a hydrogel network and endows their hydrogels with the capacity to melt and then reform, i.e., undergo a reversible gel–sol

transition. This feature of Caf1 presents new material possibilities, and we were able to successfully demonstrate hydrogel reshaping, welding and blending. It was even possible to encapsulate live cells inside the hydrogel network, thus avoiding the use of reactive chemical cross linkers or expensive biorthogonal chemistries (Dura et al. 2020).

Macroscopic material properties depend critically on polymer molecular weight, and thus to further exploit the potential of the reversible unfolding feature of the Caf1 polymer in material development, it is important to better understand the nature and kinetics of the repolymerisation process. Indeed, our previous work (Dura et al. 2020) showed that the repolymerised form of Caf1 was of shorter lengths than the original bacteria-synthesized material, and an improved understanding of the repolymerisation process is likely to lead to the development of approaches to control of the lengths of the repolymerised forms of Caf1.

Insight to the kinetics of the Caf1 refolding process can be obtained from circular dichroism (CD) spectroscopy, which allows for the real-time observation of changes in protein tertiary structure during refolding. However, CD measurements do not provide direct information as to the changes in polymer molecular weight occurring during repolymerisation, the key information required to understanding any polymerisation process. In principle, the evolution of polymer molecular weight with time might be observable by size exclusion chromatography, but it is difficult to obtain quantitative molecular weight

data for very large species. Transmission electron microscopy (TEM) can reveal the shape and length of the polymers but cannot make bulk, time resolved quantitative measurements. The highly complementary techniques of small-angle X-ray scattering (SAXS) and analytical ultracentrifugation (AUC) are widely used approaches for the study of biopolymers (Byron and Vestergaard 2015; Perkins et al. 2011) as they allow solute mass determination and shape characterisation either in terms of radius of gyration or frictional ratio/hydrodynamic radius, respectively. Furthermore, the almost instantaneous SAXS measurements available at modern synchrotron beamlines enable high resolution time-resolved studies (Franke and Svergun 2020). The SAXS method is sensitive to the mass and shape of the scattering molecule, allowing effective estimation of solute mass changes during the initial (fast) growth of Caf1 chains during repolymerisation. These measurements can be complemented with the determination of sedimentation velocity (SV) by AUC, which is a hydrodynamic method to estimate mass and shape of macromolecules. The changes in particle molecular weight and radius of gyration revealed uniquely by SAXS/AUC can thus be used to build a new understanding of rapidly polymerising biomolecules.

In this work we use SAXS to characterise the early, relatively rapid growth of Caf1 repolymerisation in solution. We then use AUC to investigate the behavior of the system in the latter stage of the repolymerisation, where the process is proceeding at a very slow rate. Methods to describe the structure and dynamic behavior of such protein polymers in solution are still evolving, and here we try to overcome these limitations by employing a number of different models to analyse the data. The results suggest that Caf1 repolymerisation proceeds in a linear fashion, with the absence of any lateral assembling, and with kinetics that are concentration-dependent. Importantly, over the first few hours (0–2 h) the polymer length was observed to grow relatively quickly, before continuing at a dramatically slowed rate. We discuss how this does not support our previously proposed in vitro polymerisation by a step-growth mechanism, and we speculate how these measurements hint at the possibility of a process more akin to chain growth, where one subunit at a time is added to the polymer chain. This analysis is the first of its kind for a bacterial fimbriae and our findings will help inform the use of SAXS and AUC in the in vitro repolymerisation of other protein-based polymeric species and, more broadly, the application of Caf1 polymers in the development of new biomaterials.

Materials and methods

Caf1 overexpression and purification

Caf1 was produced as described previously (Dura et al. 2020, 2018). Briefly, BL21(DE3) *E. coli* cells (NEB) were

transformed with the pT7-COPΔR plasmid and grown in Terrific Broth (TB) media for 22 h at 35 °C. Cells were pelleted and Caf1 protein was extracted from the supernatant and flocculent layer using tangential flow filtration and gel filtration with a Captocore column (Cytiva). Samples were dialysed against nanopure water, lyophilised and resolubilised in phosphate buffered saline (Dulbecco PBS).

Small-angle X-ray scattering (SAXS)

Sample preparation and repolymerisation procedure

Native Caf1^{WT} samples were used in our time-course SAXS measurements, and these were obtained as described above. The initial stock solution of the lyophilised protein was prepared at 13 mg/mL in PBS and the experimental sample concentrations (1.0, 2.0, 5.0, 10 mg/mL), subjected to repolymerisation were obtained by dilution of the initial stock in PBS.

The native Caf1^{WT} samples were heated to 90 °C for 5 min to achieve complete unfolding and monomerization then cooled on ice for one minute or left at room temperature for the remaining sample handling time. In general, a fast cooling method (i.e., on ice) was used to trigger the repolymerisation process, unless slow cooling (at room temperature) is specifically mentioned. The ice cooling procedure was expected to provide more efficient repolymerisation of Caf1 after thermal denaturation and was easier to standardise, while the room temperature cooling was considered to be less efficient for the repolymerisation process. Additionally, the fast cooling method was used in our previous study (Dura et al. 2020), so results obtained with this method are comparable to previously published data. Samples were centrifuged for 30 s at 20,238×g and placed into the Arinax sample handling robot (the bioSAXS robot, designed at EMBL) to continue repolymerisation process at 20 °C followed by timed SAXS measurements.

Some of the Caf1^{WT} samples used in our SAXS experiments went through the refolding process in-house over 24 h before the SAXS experiments. These samples were prepared using the fast cooling method described above. The concentration of these samples was 13 mg/mL during the repolymerisation process.

SAXS measurements

The small-angle X-ray scattering (SAXS) experiments were carried out at the beamline B21, Diamond Light Source (Didcot, UK) using a monochromatic X-ray beam ($\lambda = 0.9466 \text{ \AA}$) at the electron energy of 13.08 keV. The beamline was equipped with an EIGER 4M detector (DeCTRIS, Switzerland) and the sample to detector distance was 2.696 m, covering a momentum transfer range of

$0.0038 < q < 0.44 \text{ \AA}^{-1}$, where $q = (4\pi\sin\theta)/\lambda$ and 2θ is the scattering angle. The beam was focused on the detector, the beam size was 1 mm (horizontal) \times 0.5 mm (vertical) at the sample position and 0.008 mm (horizontal) \times 0.07 mm (vertical) at the focal point. The protein sample (50 μ L) was injected into a quartz flow-through capillary with a diameter of 1.6 mm and a thickness of 10 μ m. The experimental temperature was kept at 20 $^{\circ}$ C.

The time-course SAXS measurements for each Caf1 concentration series were carried out by sequentially collected scattering curves of the same thermally denatured Caf1 specimen. The Caf1 material required to complete a single concentration time-course series was placed into the Eppendorf-holding rack of bioSAXS sample changer and delivered by the robotic system in standard aliquots (50 μ L) to a quartz cell capillary inside the robot for X-ray scattering measurements. The measurements were carried out in the batch mode. During the initial rapid stages of Caf1 oligomerization (within the first hour after the native polymer monomerization) the time interval between consecutively collected scattering curves was governed by the time required to clean the sample-containing capillary and fill it with the next sample aliquot (1.5 min) inside the bioSAXS robot and safely exit the beamline hutch (7.5 min) before starting the data collection time course and obtaining the first data point. Thus, the first timepoint data were collected 7.5 min after the sample was heated and cooled and all consecutive measurements were made each 1.5 min during the first hour of data collection. These time intervals were determined experimentally before starting the actual time-course measurements. The same PBS solution was used in all measurements for all concentration series, the blank scans were recorded before and after first hour of uninterrupted sample injections in batch mode and, also, the buffer blanks were measured before and after the sample at later stages of the time course (3 h apart). Overall 30 blank scans were collected.

SAXS data analysis

Raw SAXS 2-D images were processed with the DAWN package (<https://dawnsci.org/>), the processing pipeline available at the beamline to produce normalised, integrated 1D un-subtracted SAXS curves. Caf1 behaviour under the beam was carefully monitored for radiation damage and the optimal number of frames for averaging was determined as 10 frames with 1 s exposure per measurement. The background subtraction, averaging of the data and determination of the scattering parameters in both reciprocal and real space along with molecular mass were executed in the program ScÅtter (<http://www.bioisis.net>). The form-factor modelling of the scattering data ($I(q)$)

was carried out using the SasView package (<http://www.sasview.org/>). SAXS contrast for Caf1 and buffer solution was calculated using the SASSIE contrast calculator module (Sarachan et al. 2013). The polydispersity distributions were not included in the data analysis, thus all fitted parameters should be considered as mean values for the whole system. Thus, the molecular mass of the species estimated in our experiments was a weight average number $\langle M \rangle_w$ and the radii of gyration obtained from slope of the scattering curve were the higher momentum median values (also known as z -average) $\langle R_g \rangle_z$ (Glatter et al. 1982).

The SoMo–SAS simulation module (Brookes et al. 2013) incorporated into the package SoMo (Brookes et al. 2010a, b; Rai et al. 2005) was utilised to calculate $P(r)$ distributions for different oligomeric models of Caf1. The coordinates for the models were generated by the program Swiss PDB Viewer (Guex et al. 1999; Guex and Peitsch 1997) using the atomic coordinates for Caf1 dimer (Zavialov et al. 2003) (PDB accession code 1P5U).

Radii of gyration and mass determination

- (i) The radius of gyration $\langle R_g \rangle_z$ for every scattering curve (i.e., a single timepoint measurement) was estimated using the Guinier approximation (Guinier et al. 1955) which allows the evaluation of $\langle R_g \rangle_z$ at low q values ($q \times \langle R_g \rangle_z < 1.3$) and could be determined from linear approximation in the following form: $\text{Ln}I(q) \cong \text{Ln}I(0) - \frac{R_g^2}{3}q^2$, where $I(q)$ is the scattering intensity; $I(0)$ is forward scattering intensity. Furthermore, the Guinier analysis could be enhanced using the Guinier peak analysis (GPA) procedure (Putnam 2016) which represents a modified form of Guinier approximation by multiplying both sides of the Guinier approximation by q as follows: $\text{Ln}[qI(q)] = \text{Ln}[I(0)] + \frac{\text{Ln}(q^2)}{2} - \frac{R_g^2}{3}q^2$. The dimensionless version of GPA ($qR_g^2I(q)/I(0)$ versus $(qR_g)^2$) is implemented in the program ScÅtter. The GPA characteristic peak position allows the validation of R_g values in complex systems when the Guinier region of scattering curve is represented by a small number of data points.

Alternatively, the value of $\langle R_g \rangle_z$ could be evaluated from indirect Fourier transform (IFT) $P(r)$ distance distribution

$$\text{function: } \quad \langle R_g \rangle_z^2 = \frac{\int_0^{D_{\max}} r^2 P(r) dr}{2 \int_0^{D_{\max}} P(r) dr}, \quad \text{where}$$

$P(r) = 2r^2 \int_0^\infty I(q) \frac{\sin(2\pi qr)}{2\pi qr} dq$, with r denoting a distance within the scattering object, and D_{\max} denoting the maximal distance of the scattering object (Glatter et al. 1982). In contrast with

the Guinier approximation, the $\langle R_g \rangle_z$ determined from IFT is calculated from the whole scattering curve.

Highly asymmetric objects, such as rods, can be characterised by the cross-sectional radius of gyration (R_c) along with the general value of R_g , describing the object as a whole. R_c characterises the shape of object in the cross section, and could be obtained from the intermediate Guinier region of the scattering curve (Glatter et al. 1982) in the following form: $I(q) = \frac{I(0)}{q} \exp\left[-\frac{q^2 R_c^2}{2}\right]$. This equation could be linearized as $\ln[qI(q)]$ versus q^2 and the slope of the straight line is proportional to R_c . In our experiments the linear approximation was considered in the q -range of 0.0055–0.06 Å⁻¹, satisfying the requirement: $\frac{2\pi}{L} < qR_c < 1$, where L is the length of Caf1 dimer (116 Å), the smallest folded oligomeric form of Caf1, roughly satisfying the requirement for an elongated particle $D_{\max}/d_{\text{crosssection}} > 2.5$ (Glatter et al 1982) (the axial ratios for the dimer model are: $X:Z = 3$, $X:Y = 2.4$). In our experiments the cross-sectional radius of gyration was determined as a z -average value. Objects with a circular cross section, R_c can be converted to the geometrical radius (R) as $R = \sqrt{2}R_c$.

The Z -average value of the radius of gyration could also be evaluated if the scattering curve is modelled with a known form factor. Thus, if a fractal model (Teixeira 1988) is used, the radius of gyration could be evaluated as $\langle R_g \rangle_z^2 = \frac{D(D+1)\xi^2}{2}$, where D is fractal dimension and ξ is the correlation length (the average distance between entanglement points in a flexible polymeric chain) of the polymeric chain (Teixeira 1988).

In the case of the flexible cylinder model, the radius of gyration could be estimated as: $\langle R_g \rangle_z^2 = \frac{a^2 N}{6}$, where a is the Kühn length (statistical length or twice the persistence length) and N is the number of subunits (Flory and Volkenstein 1969). In this present work all form factors were modelled using the SasView package (<https://www.sasview.org/>).

The weight-averaged mass of re-polymerising Caf1 for every timepoint was estimated using a volume-of-correlation (V_c) invariant (Rambo and Tainer 2013) in the following form: $M = \left[\frac{1}{c} \frac{V_c^2}{R_g}\right]^{1/k}$, where the empirical constants are $c = 0.1231$ and, in the case of proteins; $k = 1$. Volume-of-correlation could be expressed as $V_c = \frac{I(0)}{\int qI(q)dq}$, therefore, $= \frac{1}{c} \times \frac{I(0)^2}{R_g} \times \frac{1}{[\int qI(q)dq]^2}$. The experimentally determined values for the molecular weight were the weight average numbers.

Determination of V_c and, consecutively, the mass was performed using the program ScÅtter. The average number of subunits in oligomeric chain could be determined as $\langle N \rangle_w = \langle M_w^{\text{exp}} \rangle / M_0$, where M_0 is known mass of the monomer.

Analytical ultracentrifugation sedimentation velocity (AUC-SV)

Sample preparation and refolding procedure

Solutions of post-denaturation Caf1 at concentrations of 1, 2, 5 and 7.5 mg/mL were prepared using lyophilised Caf1 in PBS. Monomeric Caf1 was obtained by heating the sample to 90 °C for 5 min. The repolymerisation process was triggered by cooling the samples at 4 °C. The repolymerised material was examined within the next 72 h after monomerisation followed by repeated measurement of the same sample stored at 4 °C. The repolymerised samples were diluted with relevant buffer prior to AUC experiments to concentrations between 0.2 and 0.4 mg/mL.

Data collection and analysis

Sedimentation velocity (SV) experiments were carried out in a Beckman Coulter Palo Alto, CA, USA ProteomeLab XL-I analytical ultracentrifuge using interference optics. All AUC runs were carried out at a rotation speed of 20,000 rpm at 20 °C using an 8-hole An-Ti50 rotor and double-sector aluminium-Epon centrepieces. The sample volume was 400 µL and the sample concentration was in range of approximately 0.2–0.4 mg/mL. Every repolymerised Caf1 sample was examined at two concentrations, at least, to ensure the reproducibility of the results. Caf1 partial specific volumes \bar{v} were calculated from the protein amino acid sequence, using the program SEDNTERP (Laue et al. 1992). The density and viscosity of the experimental buffer (Dulbecco's PBS) was also calculated using SEDNTERP. Sedimentation velocity profiles were treated using both the size-distribution $c(s)$ and size-and-shape $c(s, f/f_0)$ models implemented in the program SEDFIT (Brown and Schuck 2006; Schuck 2000). The resulting distributions were also converted to standard conditions ($s_{20,w}$). To determine the mass of each species, the $c(s)$ distribution was converted either to a $c(M)$ distribution or $c(M, f/f_0)$, respectively. Each peak on the distribution plot was integrated to obtain the weight-averaged values for sedimentation coefficient and molecular mass.

Results

The early stages of Caf1 repolymerisation can be observed by SAXS

Samples of native-Caf1 [the prefix native indicates this sample was obtained directly from the bacterial biosynthesis, consisting of chains averaging ~ 100 subunits (Soliakov et al. 2010)] at a selection of concentrations (1.0, 2.0, 5.0, and 10 mg/mL) were thermally unfolded and then cooled

to allow refolding with concomitant repolymerisation. SAXS data were collected for each sample at the timepoints shown in Fig. 2a. An illustrative example of a set of scattering curves obtained at 5.0 mg/mL in reciprocal space is shown in Fig. 2b. The same dataset was converted to the total scattering intensity plot $q \times I(q)$ versus q (Fig. 2c) demonstrating the increase in total scattering intensity with time, particularly during the first hour of repolymerisation. The $q \times I(q)$ estimate for a hypothetical Gaussian coil simulating scattering of disordered Caf1 monomer ($R_g = 41.3 \text{ \AA}$, 5 mg/mL) at zero time was made to show the relatively low contribution that results from the disordered monomer (Fig. 2c). The generation of scattering curve for hypothetical Gaussian coil is described in details in Supplementary Materials and illustrated in Fig S1a, b.

Intramolecular $P(r)$ distance distribution

Scattering curves obtained in reciprocal space $I(q)$ could be converted to real space using indirect Fourier transformation (IFT) and represented in the form of a distance distribution function $P(r)$ (Fig. 3a). $P(r)$ distributions were calculated using ScÅtter (<https://www.bioisis.net>). The Legendre method was used for indirect Fourier transform as this was most suitable for elongated molecules. For transformation,

q_{\min} was set to 0.0067, 0.0085 and 0.011 for 5 mg/mL, 2 mg/mL and 1 mg/mL, respectively, and kept fixed through the time course. The value of q_{\max} for each dataset was set at the point when total scattering intensity was zero ($q \times I(q) = 0$), and was also evaluated in ScÅtter.

The shape of the $P(r)$ functions resulting from the repolymerisation time course suggests the formation of continuously elongating species with two clearly visible common peaks observed at shorter distances throughout the whole time course. The first peak (with a maximum at $\sim 20 \text{ \AA}$) represents distances within the $\sim 60 \times 30 \text{ \AA}$ monomeric unit. The following peak (at $\sim 44 \text{ \AA}$) is relevant to the correlation distances to the second neighbouring subunit. As the repolymerisation of Caf1 proceeds, this peak shifts slightly towards longer distances, as would be expected if inter subunit correlations increase. Correlations to more distant subunits are poorly resolved on account of the inter-subunit flexibility which results in the smoothing of the $P(r)$ function. These findings agree with published results for a model flexible multimeric system, albeit significantly smaller than Caf1 (Bernado 2010). The observed irregular ripples of $P(r)$ at larger distances are probably the result of oversampling of the scattering data in the indirect Fourier transform as a consequence of the high dispersion of the scattering at high angles. Theoretical $P(r)$ distributions calculated for

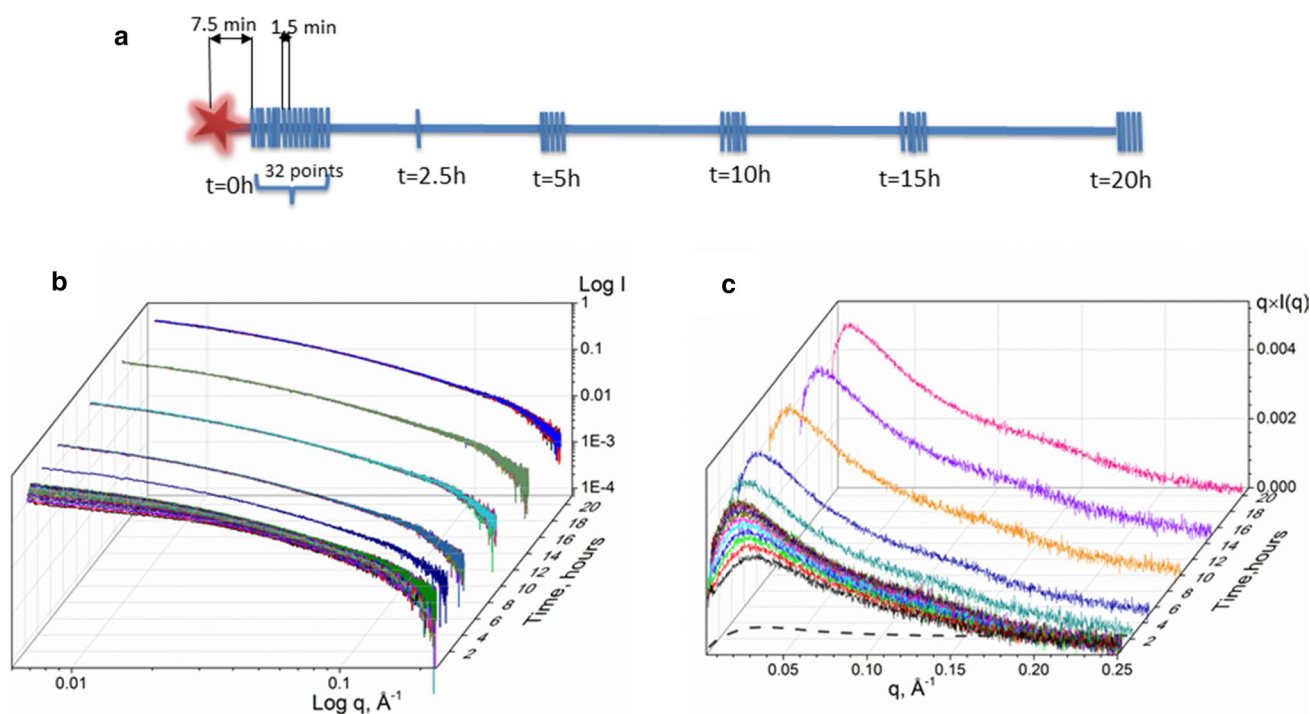


Fig. 2 Experimental design of SAXS measurements for the Caf1 post-denaturation repolymerisation time course (a); a set of resulting scattering curves (the sample concentration was 5 mg/mL) in reciprocal space (b); and selected scattering curves from the same set represented as total scattering intensity versus scattering angle (c). These

curves are shown together with the hypothetical total scattering intensity from a Gaussian coil ($R_g = 41.3 \text{ \AA}$, $c = 5 \text{ mg/mL}$) (dashed line) which simulates the scattering of a disordered Caf1 monomer at zero time

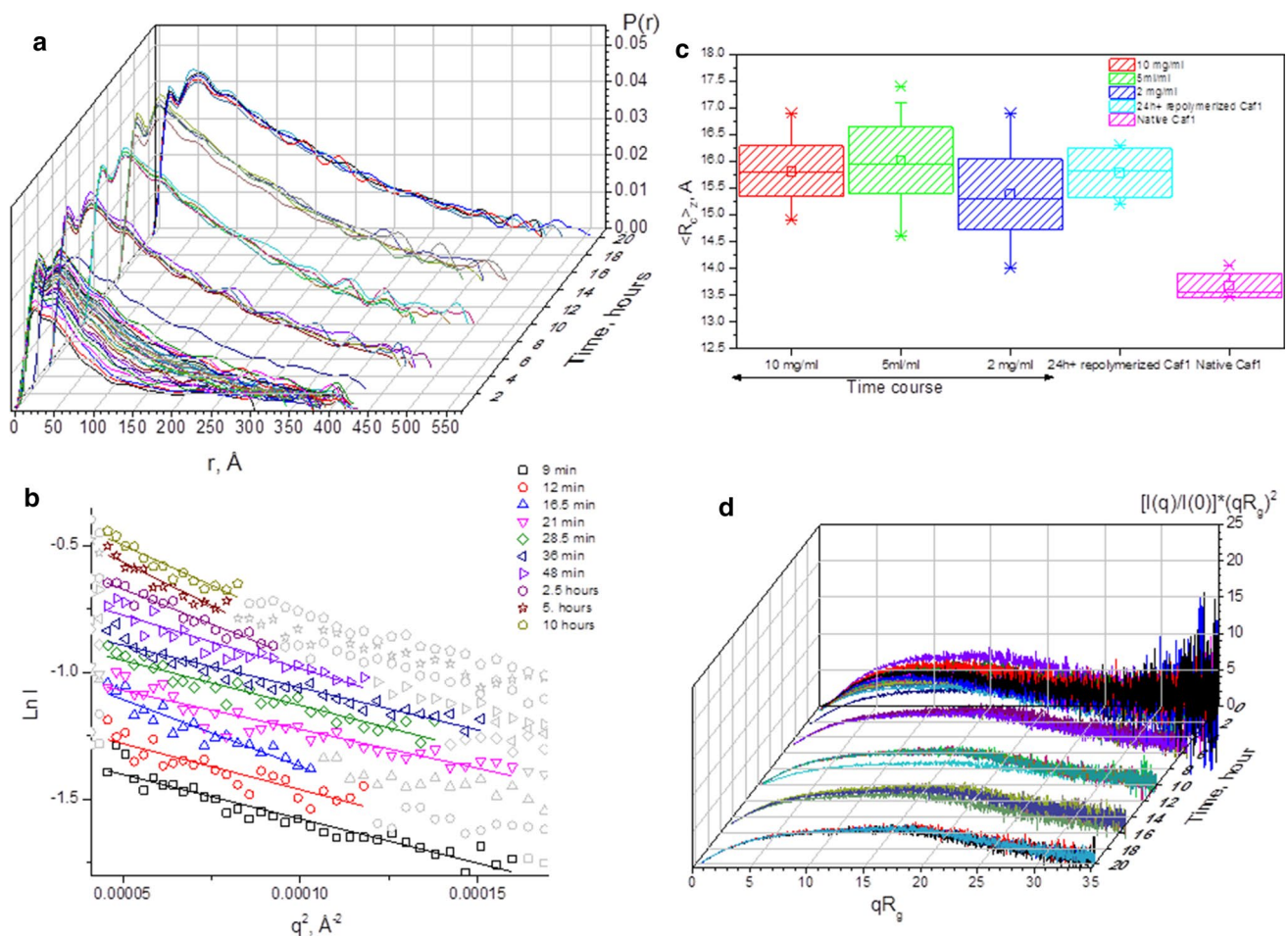


Fig. 3 SAXS standard plots. **a** $P(r)$ distributions show an increase in scattering particles' maximal distance and similar positions for two main peaks on the distribution during the time course (the sample concentration was 5 mg/mL). **b** Guinier plot for the time-course sample at the concentration of 5 mg/mL. **c** Values of cross-sectional

radius of gyration $\langle R_{g,z} \rangle$ do not substantially change with concentration and time for the repolymerizing sample. A smaller value of R_g was obtained for native Caf1^{WT}. **d** Shape of a dimensionless Kratky plot suggests multimeric scattering particles elongating with time (shown for the sample concentration of 5 mg/mL)

the examples of multimeric linear rigid Caf1 structures (shown in Fig. S2) show overall similarity with experimentally obtained $P(r)$ distributions. Thus, these observations indicate that the Caf1 subunits likely repolymerise to form growing flexible oligomeric species.

Guinier analysis

The Guinier plot is one of the important standard plots in SAXS data analysis which allows the determination of the radius of gyration (R_g) of the scattering object (Glatter et al. 1982). The characteristic linear approximation of the scattering curve could be made in the low- q region, allowing determination of the general value of R_g . TEM images of samples of repolymerized Caf1 (Dura et al. 2020), suggest that the species formed could be polydisperse and contain variable numbers of Caf1 monomers, therefore, only a

z -average value of radius of gyration ($\langle R_g \rangle_z$), characterising the whole system could be determined (Glatter et al. 1982). The linear approximation $\text{Ln}I(q)$ versus q used the same starting point (q_{\min}) as the IFT above (q_{\min} was set to 0.0067, 0.0085 and 0.011 for 5 mg/mL, 2 mg/mL and 1 mg/mL, respectively—the linear approximation for the 5 mg/mL time-course set are shown in Fig. 3b). All Guinier approximations were run through the Guinier peak analysis procedure to ensure that the region selected for linear approximation fits the shape of the Guinier peak (Putnam 2016). Several illustrative examples of the Guinier peak analysis are shown in Fig. S3.

The TEM images obtained in our earlier work (Dura et al. 2020) indicated the formation of elongated structures, thus the determination of $\langle R_{g,z} \rangle$ has to be included in the characterization of repolymerising Caf1. The Guinier plot in the so-called “intermediate- q region” ($\text{Ln}(qI(q))$ versus q^2)

(Glatter et al. 1982) at each timepoint in the repolymerisation process (Fig. S4a), along with the native Caf1^{WT} (Fig. S4b) allows the estimation of the cross-sectional radius of gyration ($\langle R_c \rangle_z$), a highly informative parameter for elongated multimeric molecules (Sund et al. 1969). The $\langle R_c \rangle_z$ value for repolymerised Caf1 was found to be 15.8 ± 0.2 Å and was observed to remain constant during the time course for all concentrations examined for repolymerised samples (Figs. 3c and S4c). This observation suggests that the growing polymer remains linear in nature and suggests no lateral association. Analysis of a sample of native-Caf1^{WT}, on the other hand, revealed a slightly smaller cross section (13.7 ± 0.06 Å), which translates into a geometrical radius of 19.2 Å (Fig. 3c). The hydrodynamic model built on the basis of Caf1 monomer atomic coordinates (PDB accession code 1P5U (Zavialov et al. 2003), reveals a prolate ellipsoid with the dimensions of $57.3 \times 38.3 \times 37.3$ Å (hydration included); therefore, the geometrical radius (R) and R_c for the model can be estimated to be 18 – 19 Å and 13 Å, respectively. This is in good agreement with our experimentally determined value of $\langle R_c \rangle_z$ (13.3 Å) for native Caf1^{WT} (the relationship between R and R_c for the circular cross section can be found in the Materials and Methods section). Thus, repolymerised Caf1 samples demonstrate an excess of 3 Å in $\langle R_c \rangle_z$ values compared to native Caf1^{WT} chains (Fig. 3c). This excess was also observed in the samples repolymerising for the period over 24 h. We believe that the presence of disordered monomer and/or propensity to further oligomeric growth may cause this difference.

Dimensionless Kratky plot

The scattering curves obtained at each timepoint during Caf1 repolymerisation were processed to obtain the relevant dimensionless Kratky plots ($(qR_g)^2 \times I(q)/I(0)$ versus qR_g) to assess the cumulative conformation of the polymeric chains (Durand et al. 2010). The dimensionless Kratky plot for a time-course set of curves (shown in Figs. 3d and S5 for the 5 mg/mL sample concentration series) is neither exactly parabolic nor hyperbolic, the shapes characteristic of folded/globular particles or random walk polypeptide chains, respectively (Receveur-Brechot and Durand 2012). Instead, it reveals an asymmetric broad peak extending towards high qR_g values with time. This suggests that growing flexible Caf1 polymers elongates with time and in agreement with the results of Bernado's theoretical work (Bernado 2010). The dimensionless Kratky plot for the data collected at the earlier stages of the time course, up until the 10th–15th timepoint (depending on the sample concentration), shows a slight elevation above zero on qR_g -axis (shown in Fig. S5, top panel). This could be caused by some amount of free disordered monomer that remains at the earlier stages of repolymerisation time course and as the repolymerisation

progresses, the folded polymeric chains start to dominate on the plot in form of fully converged curves (Fig. S5, bottom panel).

Evolution of Caf1 chain sizes during repolymerisation

As described, SAXS can provide unique information about the Caf1 repolymerisation process by allowing us to observe how the radius of gyration of the whole system changes during repolymerisation. Both approaches to determine radius of gyration (from Guinier approximation and IFT) gave very similar values for $\langle R \rangle_z$ (Fig. S6). We observed a concentration dependency, with the growth rate in $\langle R_g \rangle_z$ increasing with Caf1 concentration (Fig. 4a). The rate of growth in $\langle R_g \rangle_z$ values decreases as the repolymerisation progresses. At the end of the data collection (20 h) the rate of increase was observed to be very slow.

The solute weight-averaged mass could be determined using a volume-of-correlation invariant obtained from scattering intensities (Rambo and Tainer 2013) and, finally, the weight-averaged number of subunits in polymeric chain ($\langle N \rangle_w$) as $\langle N \rangle_w = \langle M^{\text{exp}} \rangle_w / M_0$, where M_0 is the mass of the monomer.

The increase in $\langle N \rangle_w$ versus time consists of a steep exponential section during the early phase (0–2.5 h) of the repolymerisation process and at the later phase (> 2.5 h) growth continues but at a significantly lower rate similarly to the changes in $\langle R_g \rangle_z$ (shown in Fig. 4b for a repolymerising sample at 5 mg/mL).

Form factor modelling

In our earlier work, the native Caf1 oligomer was captured on the TEM images as long chains which appear as a random walk polymers (Soliakov et al. 2010) forming wiggles and loops. Such behavior is suggestive of the extremely flexible nature of these bio-filaments. The form-factor modelling approach to analyze the small-angle scattering data for hydrogel-forming systems was demonstrated earlier (Draper et al. 2017; Fuentes-Caparrós et al. 2019). These developments could be relevant for Caf1 polymers to some extent, owing to Caf1's ability to form a hydrogel (Dura et al. 2018). Thus, for the next step in this study we tried to fit Caf1 scattering curves to a number of form-factor models in the attempt to describe the scattering behavior of Caf1 chains analytically.

In general terms, small-angle scattering of a particle (i.e., the scattering intensity function $I(q)$) in solution can be defined by the particle's concentration (volume fraction), its contrast in solution and its form-factor scattering intensity function. In many cases, scattering curves produced by biomolecules and their ensembles in solution can

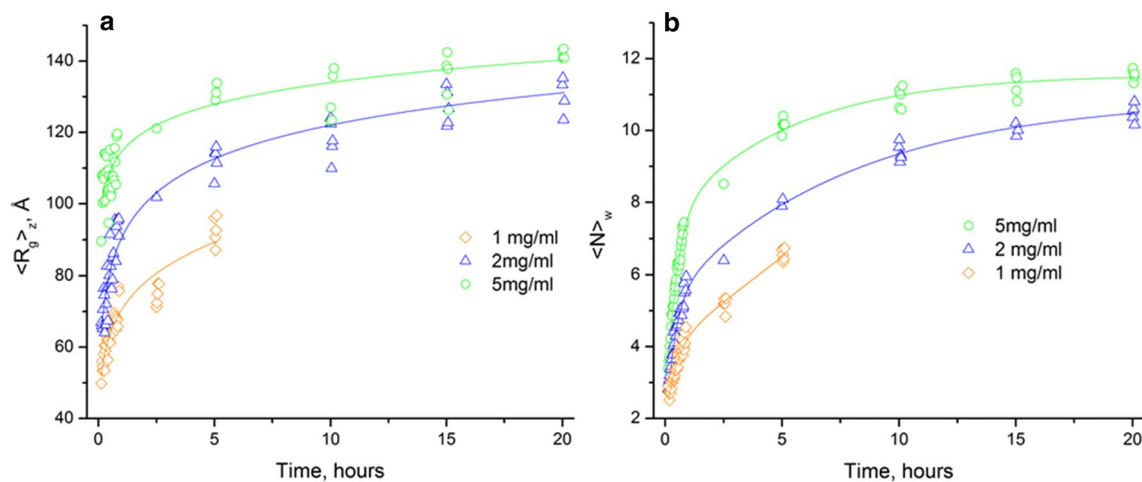


Fig. 4 Increase in z-average values of radius of gyration (**a**) and in weight-averaged number of chain subunits (**b**) of repolymerised Caf1 during the time course. The concentration of repolymerised Caf1 affects the rate of increase. $\langle R_g \rangle_z$ values were obtained from $P(r)$ dis-

tribution and the average number of subunits ($\langle N \rangle_w$) in the repolymerised chain was calculated on the basis of $\langle M \rangle_w$ values evaluated using the Guinier approximation. The samples at 1 mg/mL were examined for the fast stage only

be successfully fitted to known form factors calculated for geometrical bodies (Pedersen 1997). In the first instance we analysed the scattering curves of the native Caf1^{WT} using a flexible cylinder model (Pedersen and Schurtenberger 1996) which satisfactorily fitted our data [Fig. 5 bottom panel, (a)]. The length of about 3000 Å and radius of approximately 15 Å were in reasonable agreement with published earlier TEM findings for native Caf1^{WT} polymeric chains (Soliakov et al. 2010). The fitted value for the flexible cylinder radius is also in agreement with a hydrated Caf1 monomer model based on its high resolution structure (Zavialov et al. 2005). The value for the statistical length (Kühn length) of a native Caf1^{WT} oligomeric chain was found to be 170 Å, which is approximately three monomer lengths. This model fit the experimental data almost comprehensively except at very high sample concentrations (10 mg/mL). The model for a polymer with excluded volume (Hammouda 1993) fits the experimental data at this sample concentration slightly better due to accounting for the concentration effects in solution.

The repolymerisation process of Caf1 is a dynamic process since the shape of growing oligomers (and, consequently, the form-factor models describing them) may vary depending on the sample concentration, the length of the growing oligomer and the rate of growth. Therefore, several form-factor models were examined to analyze the time-course data. An illustrative example of the form factor model fits is shown in Figs. 5b and S7 representing the datafit and/or residuals by the following models (i) a flexible cylinder model, which was used to analyse the native Caf1^{WT} samples, (ii) a rigid cylinder model, which could be applicable for the earlier stages of growth when the chains are still short (about the statistical length of the polymer), (iii) a fractal

model and (iv) a polymer with excluded volume model. The two latter models are shape-independent and are applicable to a flexible polymer (Capp et al. 2014). On the same note Caf1 growing oligomers in vitro may be modelled as a fractal system, where each Caf1 monomer joining the chain can be considered as a repeat building block. As the chain grows from monomer into a larger oligomer, its general geometrical features change, which can be described quantitatively in terms of fractal dimension (D). Folded globular proteins typically exhibit a value of $D \sim 2.8$ (Isogai and Itoh 1984), i.e., a 3D shape, while the behaviour of a long oligomer chain can be represented as a line (i.e., 1D shape).

A flexible polymeric chain is expected to have fractal dimensions varying between 2 and 1 as it entangles and wriggles (Dewey 1998). We anticipated that as the Caf1 chain grows in length during repolymerisation, D would decrease towards unity, the value anticipated for a flexible indefinitely thin cylinder ($L \gg d$). A plot of D versus time for repolymerisation at a selection of concentrations (2.0, 5.0, 10 mg/mL) is presented in Fig. 6. At the first timepoint after the initiation of repolymerisation, the observed fractal dimension of the system was observed to be in the range of 2.25 suggesting the presence of low order oligomers, such as trimers or tetramers, for example. As the repolymerisation proceeds, D was observed to decrease rapidly then essentially plateau, tending towards a value of $D \sim 1.72$, detected for a Caf1 sample repolymerised more than 24 h in advance. The latter seemingly low value for D could be explained by the fact that this sample was repolymerising at much higher concentration (13 mg/mL) which results in the formation of a longer polymeric chain compared to the repolymerised Caf1 samples studied in the time course. Overall, these

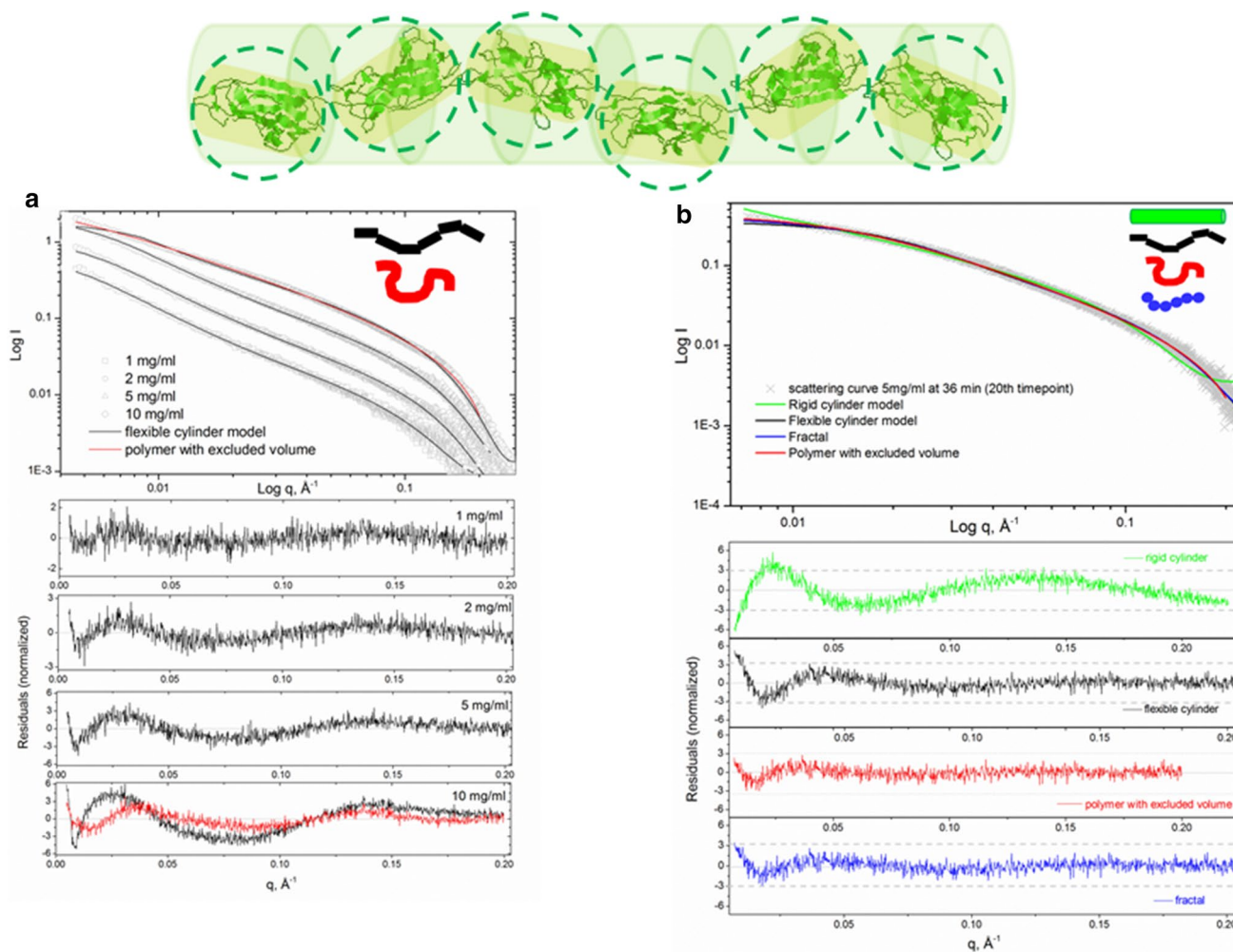


Fig. 5 Linear model of CafI oligomer (cartoon in green-top panel) represented as various form-factor models: rigid cylinder (whole oligomer), flexible cylinder (every monomer or group of monomers in the chain is considered as fragments of flexible cylinder), and shape-independent mass fractal (spherical subunit monomers) and polymer with excluded volume—a basic model for scattering of various polymeric chains. The bottom panel shows an illustrative example of the

experimental scattering curve fit to these models. The native CafI^{WT} modelled as a flexible cylinder (black line) at 1 mg/mL, 2 mg/mL, 5 mg/mL and 10 mg/mL and a polymer with excluded volume (red line) at 10 mg mg/mL (a). Repolymerising CafI at 36 min post-denaturation, (sample concentration 5 mg/mL) modelled as a rigid cylinder (green line), flexible cylinder (black line) fractal (blue line) and a polymer with excluded volume (red line) (b)

observations indicate that CafI repolymerizes to reform a linear species with no lateral self-association, i.e., its repolymerized form is identical to its original bacteria-synthesized form (Soliakov et al. 2010).

The fitted values of fractal dimension (between 2.3 and 1.7) (Fig. 6) for in vitro CafI polymerisation suggest that the polymer with excluded volume model could be also applied (Hammouda 1993). From the perspective of the $\langle R_g \rangle_z$ values characterising the entire system, this model is quite similar in essence to the fractal model. The $\langle R_g \rangle_z$ value derived from the polymer with excluded volume model is a parameter that is directly fitted to the scattering curve, while in the fractal model, it has to be calculated. Overall, all models give similar values for $\langle R_g \rangle_z$, which are comparable to those obtained in the Guinier approximation/IFT (Fig. S7), with

the exception of the $\langle R_g \rangle_z$ values derived from the parameters of the rigid cylinder—those are in the agreement with others at early stages of repolymerisation only.

The residual plots (Fig. 5b, bottom panel, and Fig. S7) show that the flexible cylinder and both shape independent models possess a similar distribution of residuals, although the polymer with excluded volume model gives the smallest deviations. The rigid cylinder model gives a reasonable fit only at very early stages of repolymerisation (as shown in Fig. S7, top panel). A noticeable systematic deviation at very low scattering angles ($q < 0.01$) was present for all form-factor model fits, particularly for the scattering curves collected in first 10–20 min of the time course. Hence, it may be concluded that neither model gives an ideal fit to the data in this q -range. This outcome could be explained by a

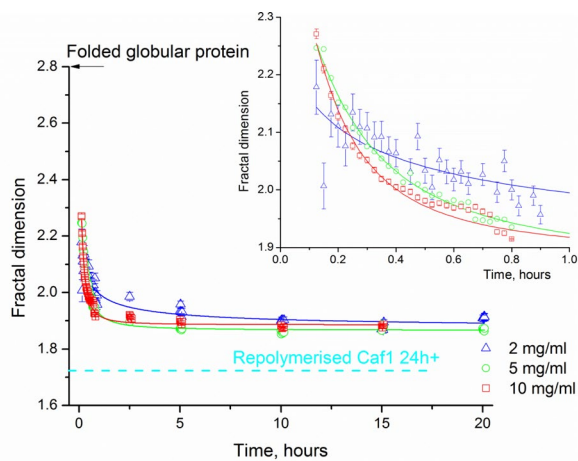


Fig. 6 Changes in the geometry of a repolymerized Caf1 chain in terms of fractal dimension. Fractal dimension drops very quickly during the first hour of repolymerisation suggesting the formation of polymers in the form of wiggly lines $D \cong 1.9$, ultimately approaching the D value of 1.72 detected for the Caf1 sample repolymerised for more than 24 h in advance (the concentration of repolymerising sample was 13 mg/mL, whereas this sample concentration in SAXS experiment was 6.75 mg/mL). The changes in fractal dimension during first hour of repolymerisation are shown in the insert. The data obtained for the low concentration (2 mg/mL) repolymerized material, which polymerizes slower than at the higher concentrations, demonstrates slower linearization, probably on account of a relatively large amount of disordered monomeric material being present during the polymerisation

contribution of a small number of large species present in the rapidly oligomerising Caf1 solution at the beginning of the repolymerisation process, when the majority of the material is still rather small. As the repolymerisation progresses, this effect seemingly diminishes as the majority of the material acquires longer polymeric chains, thereby decreasing the degree of polydispersity in solution. It seems that the level of polydispersity in solution is dependent on the method used to initiate Caf1 repolymerisation, i.e., it is characteristic for thermally denatured Caf1 samples cooled on ice and then subjected to repolymerisation. In a contrast, the scattering curves produced by the slow repolymerising method (i.e., where the repolymerisation process was triggered at room temperature) are much better fitted by the form-factor models over the low q -range (Fig. S9). The latter samples demonstrate significantly slower time-dependent increase in mass and radius of gyration (Fig. S10).

The latter phase of Caf1 repolymerisation can be observed by AUC

The SAXS studies indicated that after an initial fast growth phase, the rate of growth of Caf1 chains becomes significantly slower. Analytical ultracentrifugation sedimentation velocity can provide information about the chain lengths of

repolymerized Caf1 in this latter slow growth phase, i.e., at significantly longer experimental time scales in comparison with SAXS experiments. Samples of native-Caf1^{WT} at 2.0, 5.0 and 7.5 mg/mL were thermally unfolded, cooled to 4 °C and left to repolymerize for 72 h (3 days) and then analysed by AUC. The solute distribution in an ultracentrifuge cell can be represented in the form of 1D $c(s)$ or 2D $c(s,*)$ (Brown and Schuck 2006; Schuck 2000). The latter is applicable for particularly complex heterogeneous systems, where $*$ could represent properties of the biomolecule, such as ff_0 , M , D , R_S (Brown and Schuck 2006; Chaton and Herr 2015). Both 1D and 2D size-distribution models were applied to analyze sedimentation velocity profiles (the experimental data and fit are shown in Fig. S11a–d). The distribution of solute masses is shown in Fig. 7a (the corresponding distributions of sedimentation coefficients is shown in Supplementary material, Fig. S12a). The concentration of repolymerising sample seems to be an important factor, since the mass distributions broadened and shifted towards higher values as the concentration of repolymerised sample increased from 1.0 mg/mL to 7.5 mg/mL. The observed increase in mass was accompanied by frictional ratio changes from 1.8 to 2.9 in the 1D distribution suggesting an overall elongation of the molecules as the refolding concentration increases. To minimize the contribution of highly asymmetric shapes that result in the non-ideal behavior of the solute, it was important to dilute the solutions immediately prior to the experiment. Thus, the actual sample concentration in sedimentation velocity experiments was typically about 0.2–0.4 mg/mL; however, due to the extremely slow rates of chain growth, the sample dilution effect was unlikely to affect the observed masses. More detailed insight on the nature of repolymerized Caf1 was gained by moving from 1D distributions, where the values of ff_0 are average numbers for all observed species, to 2D distributions, where the parameters mass and shape were fitted simultaneously (Figs. 7b, S12b, d). This approach demonstrates clearly that the higher masses that are detected are characterized by higher ff_0 values, i.e., the chain becomes longer as the sample concentration increases. Thus, the largest detected species ranged from ~148.2 kDa at the Caf1 repolymerisation concentration of 1 mg/mL, to ~576.3 kDa at repolymerisation concentration of 7.5 mg/mL. Interestingly, the analysis of samples stored for 60 days show a more populated profile with a strong peak (Fig. 7c) at mass ~600 kDa (16% of total material) and less resolved peaks at 1MDa (8% of total material), indicating further repolymerisation beyond 3 days, when the maximal detected mass was ~400 kDa (20%) (marked in Fig. 7c with red arrows). The species with the mass of ~200 kDa were detected in both cases (marked with a yellow line) at 23% and 26% of the total, respectively. These observations indicate that Caf1 repolymerisation slowly continues beyond the initial 20 h

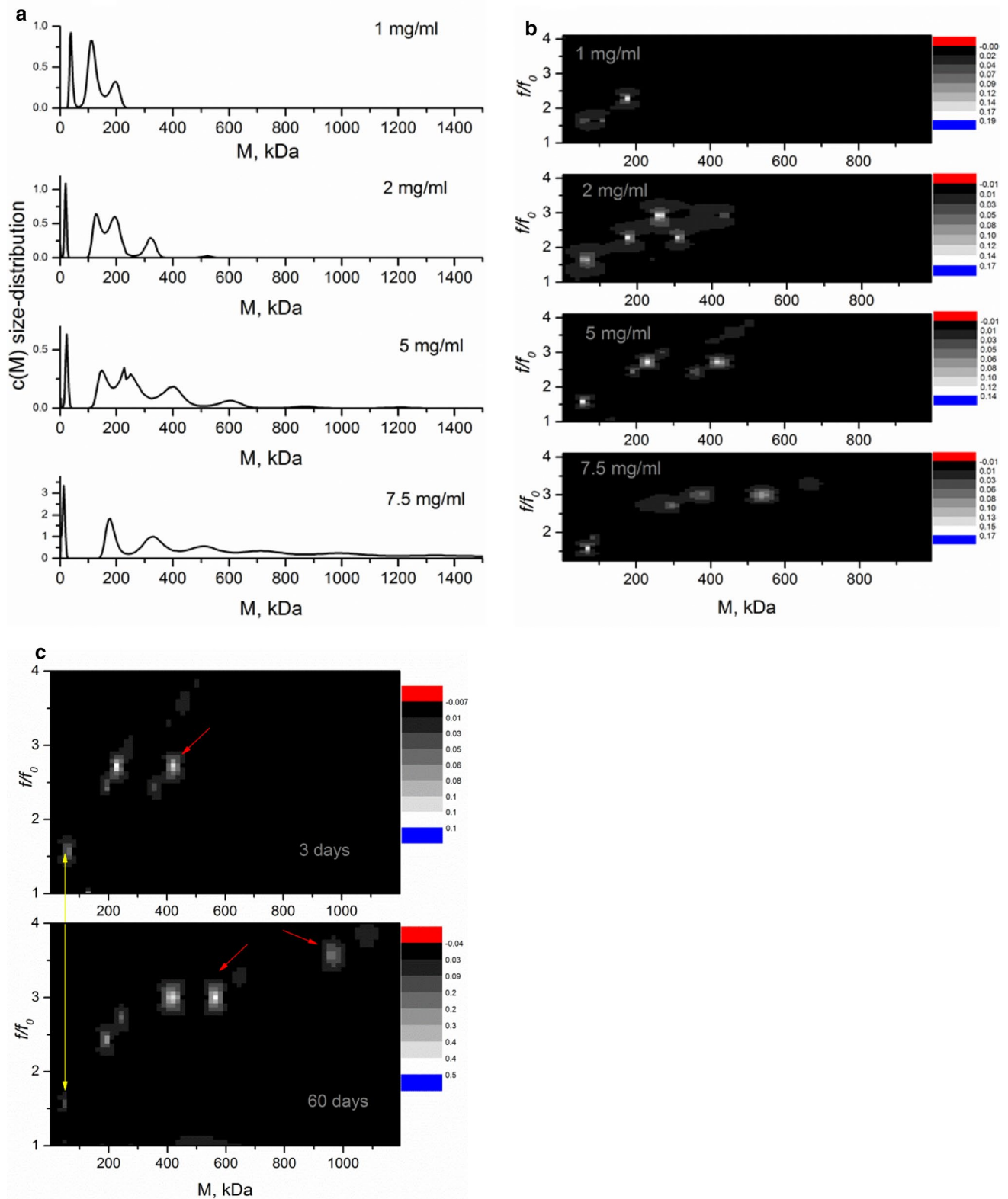


Fig. 7 Repolymerised Caf1 samples analysed by sedimentation velocity analytical ultracentrifugation experiments (the rotation speed was 20,000 rpm) modelled as 1D $c(M)$ (**a**) and 2D $c(M, f/f_0)$ (**b**) size distributions. Long storage time result in the appearance of heavier and

longer species (marked with red arrows), while smaller species still exists (marked with yellow arrows) as seen from the 2D distribution $c(M, f/f_0)$ (**c**)

window of the SAXS experiments, and suggests the terminal subunits of Caf1 oligomers remain ‘active’ and are capable of forming linkages with other monomers or chains, even at long reaction times. In all experiments, a fraction of small Caf1 oligomers (dimeric or trimeric size) were also detected alongside the larger polymers.

Discussion

Repolymerized Caf1 oligomers are considerably shorter in length than the native form of Caf1, but exhibit some variety in mass and shape. Therefore, analysis methods based on statistical conformation of the whole ensemble are particularly useful as shown in (Capp et al. 2014). However, most SAXS studies of oligomeric flexible systems were based on preliminary knowledge of the number of subunits within the oligomer as in (Bernado 2010, Capp et al. 2014) that help to enable modelling of the scattering intensity resulting from stochastic movements of the oligomeric chain, as implemented in EOM (Bernadó et al. 2007) or SASSIE (Perkins et al. 2011). An attempt to describe the indefinite repolymerization of Caf1 subunits in terms of mass fractals is probably the most justified approach, since this model is totally shape-independent. To the best of our knowledge, this model has not yet been applied to regular oligomeric growth. The fractal model only describes the conformational dynamics of the chain and does not provide any information about actual sizes of particles, however, combining this analysis with another approaches, such as Guinier/IFT, allows for a more complete characterisation of the system.

A key advantage of SAXS experiments in the study of the repolymerization of Caf1 is its capacity to passively observe the almost instantaneous determination of solute mass and shape; however, the measurements are limited by q -window of the camera length distances in bioSAXS instrument. We were thus able to monitor effectively the repolymerization for approximately the first 10 h before some oligomers likely became too long to be detected by the q -window of the measurement. Conveniently, it was possible to detect further growth on a much slower experimental time scale in sedimentation velocity experiment and trace the small oligomeric forms of Caf1 which could be “hiding” behind the dominant signal coming from larger oligomers in the SAXS experiments.

In contrast to SAXS, where detected masses are the weight averaged, analytical ultracentrifugation can, in principle, detect the mass for each individual oligomeric form. The method cannot describe the mass distribution of oligomeric species forming during first few hours of the repolymerisation, but it is effective approach for slow maturing oligomeric forms.

SAXS and AUC together provide experimental data that indicate how the Caf1 polymer grows with time during the repolymerization process. The experimental evidence obtained suggests that Caf1 repolymerization occurs in a linear fashion, with no evidence of lateral association or aggregation. This result is in agreement with published earlier TEM images of native and repolymerised Caf1 (Soliakov et al. 2010; Dura et al. 2020) which reported no observation of Caf1 chain branching. The rate of repolymerisation, as determined by SAXS, appears to show a concentration dependency (in the range 1.0–5 mg/mL), which is in agreement with earlier CD studies (Dura et al. 2020). Irrespective of the model used to fit the SAXS data, the oligomeric growth shown in Fig. 4b displays two key features: a relatively fast exponential growth over the initial stage (0–2.5 h) and then a much slower linear growth over the remaining duration of the experiment. The donor–acceptor nature of the Caf1 monomer unit suggests that its *in vitro* polymerization likely proceeds by a step growth mechanism, where the monomers react together in a stepwise fashion to form shorter oligomers that continue to react to form longer oligomers. However, with classical step growth polymerizations a linear growth of molecular weight with time can be expected (Stille 1981). Thus, the observations made by SAXS and AUC are not consistent with the idea of Caf1 polymerizing by step-growth process. We thus considered the possibility that the initial exponential growth in molecular weight may arise on account of the operation of an isodesmic polymerization mechanism (Zhao and Moore 2003). An isodesmic polymerization can be considered the supramolecular analogue of a classical step growth polymerization in which monomers reversibly interact in a stepwise fashion to form short oligomers that then continue to react to form longer oligomers. However, an isodesmic polymerization would be expected to reach equilibrium (i.e., molecular weight stops growing and becomes constant), but that does not appear to be the case as the AUC experiments indicate observed small increases in molecular weight occurring, even after several days, suggesting an isodesmic mechanism also is not operational. The kinetic plot in Fig. 4b is also not consistent with nucleation growth (Zhao and Moore 2003), a well-known protein polymerization mechanism in which there is a relatively slow nucleation phase followed by rapid polymerization, eliminating the possibility of this mechanistic pathway.

In vivo, native-Caf1^{WT} polymer biogenesis follows the conserved chaperone–usher pathway used by Gram-negative bacteria to construct a variety of surface polymeric structures (Waksman and Hultgren 2009; Zav’yalov et al. 2010). This process involves the sequential addition of subunits onto the growing end of the polymer chain, and can be considered to resemble a chain-growth polymerization mechanism (McGrath 1981), where monomer units add one at a time onto the end of a growing chain. The evolution of

molecular weight with time observed in our *in vitro* experiments does appear to present similarities with that observed for chain growth polymerizations. Early in a chain growth process when monomer is present in a large excess in relation to chain ends, molecular weight increases quickly. When the concentration of monomer decreases to become similar to that of active chain ends, the rate of molecular weight growth slows down, and then becomes increasingly slower as the concentration of monomer continues to decrease. It is tempting to think that such a process might be operating in the *in vitro* polymerization process. This, however, would require a plausible mechanism as to how Caf1 subunits can add one at a time onto the end of a chain in the absence of chaperone and usher proteins that are required to mediate the *in vivo* polymerization observed in Gram-negative bacteria, which is beyond the scope of this current work. Experiments ongoing in our laboratories hope to shed further light on the mechanism of *in vitro* Caf1 polymerization.

Conclusion

In this work we have demonstrated how SAXS and AUC can be utilized to study the *in vitro* repolymerization of Caf1. SAXS data could be used to determine the evolution of polymer growth with time, and reaffirms the concentration dependency of the rate of repolymerization. SAXS data analysis of a 20-h time course was based on a combination of both shape-dependent and shape-independent approaches giving an insight on fast initial phase of Caf1 repolymerization. The AUC method was employed to observe evidence of the process at longer time scales, where there continued to be a very slow growth of the oligomers. A key finding of this work was that the observed evolution of polymer molecular weight with time, where weight grows rapidly (0–2.5 h) and then far more slowly, was not consistent with that of classical step growth or isodesmic polymerizations. Intriguingly, the evolution of molecular weight might be better explained by a chain growth mechanism, whereby subunits attach one at a time onto the end of a chain. Such a polymerization process may bear similarities to the conserved chaperone–usher pathway used *in vivo*, where the polymerization process is mediated by chaperone and usher proteins to sequentially attach subunits one at a time onto the C-terminus of the growing polymer.

This work shines new light on the Caf1 repolymerization mechanism, which was previously thought to involve irreversible association of subunits and proceed through a conventional step-growth mechanism. This improved understanding of the Caf1 repolymerization process will allow the development of methods to better control polymer growth, and ultimately, better control the properties of Caf1-based materials.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00249-021-01543-3>.

Acknowledgements This work was funded by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) award number BB/M018318/1 to JHL, to support the translation, development and commercialisation of innovative Industrial Biotechnology processes. Authors are thankful to Diamond Light Source (DLS) for SAXS beamtime (proposal SM24306, beamline B21) granted for this study to ASS. Special thanks to Drs Katsuaki Inoue and Nathan Cowieson (B21, DLS) for help in SAXS data collection and primary data treatment. The authors would like to thank Dr Benjamin R Horrocks (Newcastle University) and Dr Mariya S Antyufeyeva (V.N.Karazin Kharkiv National University) for fruitful and insightful discussions.

Author contributions The study was designed by ASS and DAF; sample preparation was made by HW, DTP and GD. Experiments were performed by ASS, DTP and GD. Data were analysed by ASS. ASS, DAF, DTP and JHL wrote the paper. All authors read and approved the final manuscript.

Funding Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) award number BB/M018318/1.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bernado P (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J* 39:769–780
- Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664
- Brookes E, Demeler B, Rocco M (2010a) Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. *Macromol Biosci* 10:746–753
- Brookes E, Demeler B, Rosano C, Rocco M (2010b) The implementation of SOMO (SOLUTION MODeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur Biophys J* 39:423–435
- Brookes E, Perez J, Cardinali B, Profumo A, Vachette P, Rocco M (2013) Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. *J Appl Crystallogr* 46:1823–1833
- Brown PH, Schuck P (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys J* 90:4651–4661

- Byron O, Vestergaard B (2015) Protein-protein interactions: a supra-structural phenomenon demanding trans-disciplinary biophysical approaches. *Curr Opin Struct Biol* 35:76–86
- Capp JA, Hagarman A, Richardson DC, Oas TG (2014) The statistical conformation of a highly flexible protein: small-angle X-ray scattering of *S. aureus* protein A. *Structure* 22:1184–1195
- Chalton DA, Musson JA, Flick-Smith H, Walker N, McGregor A, Lamb HK, Williamson ED, Miller J, Robinson JH, Lakey JH (2006) Immunogenicity of a *Yersinia pestis* vaccine antigen monomerized by circular permutation. *Infect Immun* 74:6624–6631
- Chaton CT, Herr AB (2015) Chapter eight—elucidating complicated assembling systems in biology using size-and-shape analysis of sedimentation velocity data. In: Cole JL (ed) *Methods in enzymology*, vol 562. Academic Press, pp 187–204
- Dewey TG (1998) *Fractals in molecular biophysics*. Oxford University Press
- Draper ER, Su H, Brasnett C, Poole RJ, Rogers S, Cui H, Seddon A, Adams DJ (2017) Opening a can of Worm(-like Micelle)s: the effect of temperature of solutions of functionalized dipeptides. *Angew Chem Int Ed* 56:10467–10470
- Du Y, Rosqvist R, Forsberg A (2002) Role of fraction 1 antigen of *Yersinia pestis* in inhibition of phagocytosis. *Infect Immun* 70:1453–1460
- Dura G, Waller H, Gentile P, Lakey JH, Fulton DA (2018) Tuneable hydrogels of CafI protein fibers. *Mater Sci Eng C Mater Biol Appl* 93:88–95
- Dura G, Peters DP, Waller H, Yemm AI, Perkins ND, Ferreira AM, Crespo-Cuadrado M, Lakey JH, Fulton DA (2020) A thermally reformable protein polymer. *Chem* 6:3132–3152
- Durand D, Vives C, Cannella D, Perez J, Pebay-Peyroula E, Vachette P, Fieschi F (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol* 169:45–53
- Flory PJ, Volkenstein M (1969) Statistical mechanics of chain molecules. *Biopolymers* 8:699–700
- Franke D, Svergun DI (2020) Synchrotron small-angle X-ray scattering on biological macromolecules in solution. In: Jaeschke EJ, Khan S, Schneider JR, Hastings JB (eds) *Synchrotron light sources and free-electron lasers: accelerator physics, instrumentation and science applications*. Springer International Publishing, Cham, pp 1645–1672
- Fuentes-Caparrós AM, de Paula G-F, Dietrich B, Wilson C, Brasnett C, Seddon A, Adams DJ (2019) Annealing multicomponent supra-molecular gels. *Nanoscale* 11:3275–3280
- Glatter O, Kratky O, Kratky HC (1982) *Small angle X-ray scattering*. Academic Press
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
- Guex N, Diemand A, Peitsch MC (1999) Protein modelling for all. *Trends Biochem Sci* 24:364–367
- Guinier A, Fournet G, Walker CB, Yudowitch KL (1955) *Small-angle scattering of X-rays*. Wiley
- Hammouda B (1993) SANS from homogeneous polymer mixtures: a unified overview. *Polymer characteristics*. Springer, Berlin, pp 87–133
- Isogai Y, Itoh T (1984) Fractal analysis of tertiary structure of protein molecule. *J Phys Soc Jpn* 53:2162–2171
- Laue TM, Shah BD, Ridgeway TM, Pelletier S (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: *Analytical ultracentrifugation in biochemistry and polymer science*. Redwood Press Ltd, Melksham, pp 90–125
- McGrath JE (1981) Chain reaction polymerization. *J Chem Educ* 58:844
- McNicholas S, Potterton E, Wilson KS, Noble MEM (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr Sect D* 67:386–394
- Pedersen JS (1997) Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv Colloid Interface Sci* 70:171–210
- Pedersen JS, Schurtenberger P (1996) Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules* 29:7602–7612
- Perkins SJ, Nan R, Li K, Khan S, Abe Y (2011) Analytical ultracentrifugation combined with X-ray and neutron scattering: experiment and modelling. *Methods* 54:181–199
- Putnam CD (2016) Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *J Appl Crystallogr* 49:1412–1419
- Rai N, Nollmann M, Spotorno B, Tassara G, Byron O, Rocco M (2005) SOMO (SOLUTION MOdeler) differences between X-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* 13:723–734
- Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496:477–481
- Receveur-Brechot V, Durand D (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* 13:20
- Sarachan KL, Curtis JE, Krueger S (2013) Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution. *J Appl Crystallogr* 46:1889–1893
- Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619
- Soliakov A, Harris JR, Watkinson A, Lakey JH (2010) The structure of *Yersinia pestis* CafI polymer in free and adjuvant bound states. *Vaccine* 28:5746–5754
- Stille JK (1981) Step-growth polymerization. *J Chem Educ* 58:862
- Sund H, Pilz I, Herbst M (1969) Studies of glutamate dehydrogenase. 5. The X-ray small-angle investigation of beef liver glutamate dehydrogenase. *Eur J Biochem* 7:517–525
- Teixeira J (1988) Small-angle scattering by fractal systems. *J Appl Crystallogr* 21:781–785
- Ulusu Y, Dura G, Waller H, Benning MJ, Fulton DA, Lakey JH, Peters DT (2017) Thermal stability and rheological properties of the “non-stick” CafI biomaterial. *Biomed Mater* 12:051001
- Waksman G, Hultgren SJ (2009) Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* 7:765–774
- Yu XD, Fooks LJ, Moslehi-Mohebi E, Tischenko VM, Askarieh G, Knight SD, Macintyre S, Zavialov AV (2012) Large is fast, small is tight: determinants of speed and affinity in subunit capture by a periplasmic chaperone. *J Mol Biol* 417:294–308
- Zavialov AV, Berglund J, Pudney AF, Fooks LJ, Ibrahim TM, MacIntyre S, Knight SD (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113:587–596
- Zavialov AV, Tischenko VM, Fooks LJ, Brandsdal BO, Aqvist J, Zav’yalov VP, Macintyre S, Knight SD (2005) Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem J* 389:685–694
- Zav’yalov V, Zavialov A, Zav’yalova G, Korpela T (2010) Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS Microbiol Rev* 34:317–378
- Zhao D, Moore JS (2003) Nucleation-elongation: a mechanism for cooperative supra-molecular polymerization. *Org Biomol Chem* 1:3471–3491