UNIVERSITY OF LEEDS

This is a repository copy of *Inducing protein aggregation by extensional flow*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/114410/

Version: Accepted Version

Article:
Dobson, J, Kumar, A orcid.org/0000-0001-8391-6812, Willis, LF orcid.org/0000-0001-6616-3716 et al. (8 more authors) (2017) Inducing protein aggregation by extensional flow. Proceedings of the National Academy of Sciences, 114 (18). pp. 4673-4678. ISSN 0027-8424

https://doi.org/10.1073/pnas.1702724114

This article is protected by copyright. All rights reserved. This is an author produced version of a paper published in the Proceedings of the National Academy of Sciences. Uploaded in accordance with the publisher's self-archiving policy. In order to comply with the publisher requirements the University does not require the author to sign a non-exclusive licence for this paper.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Inducing protein aggregation by extensional flow

John Dobson[1]*, Amit Kumar[2,3]*, Leon F. Willis[2,3]*, Roman Tuma[2,3], Daniel R. Higazi[4#], Richard Turner[4], David C. Lowe[4], Alison E. Ashcroft[2,3], Sheena E. Radford[2,3], Nikil Kapur[1†], and David J. Brockwell[2,3†]

[1] School of Mechanical Engineering, [2]School of Molecular and Cellular Biology, Faculty of Biological Science, [3]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK and [4] Medimmune Ltd, Granta Park, Cambridge, CB21 6GH, UK

* These authors contributed equally to this work

Current address: Ipsen Biopharm Ltd UK, Wrexham Industrial Estate, Wrexham, LL13 9UF

†Correspondence should be addressed to:

Nikil Kapur: e-mail: n.kapur@leeds.ac.uk Phone: +44(0)113 343 2152

Or

David Brockwell: e-mail: d.j.brockwell@ leeds.ac.uk Phone: +44(0)113 343 7821

Abstract

Relative to other extrinsic factors, the effects of hydrodynamic flow fields on protein stability and conformation remain poorly understood. Flow-induced protein remodelling and/or aggregation is observed both in Nature and during the large-scale industrial manufacture of proteins. Despite its ubiquity, the relationships between the type and magnitude of hydrodynamic flow, a protein's structure and stability and the resultant aggregation propensity are unclear. Here, we assess the effects of a defined and quantified flow-field dominated by extensional flow on the aggregation of bovine serum albumin (BSA), β_2-microglobulin (β_2m), granulocyte colony stimulating factor (G-CSF) and three monoclonal antibodies (mAbs). We show that the device induces protein aggregation after exposure to an extensional flow field for 0.36-1.8 ms, at concentrations as low as 0.5 mg mL^{-1}. In addition, we reveal that the extent of aggregation depends on the applied strain rate and the concentration, structural scaffold and sequence of the protein. Finally we demonstrate the *in situ* labelling of a buried cysteine residue in BSA during extensional stress. Together, these data indicate that an extensional flow readily unfolds thermodynamically and kinetically stable proteins, exposing previously sequestered sequences whose aggregation propensity determines the probability and extent of aggregation.

Significance of the work

Proteins are inherently sensitive to environmental factors that include hydrodynamic flow. Flow-induced protein remodelling is utilized *in vivo* and can also trigger the aggregation of therapeutic proteins during manufacture. Currently, the relative importance of shear- and extensional hydrodynamic flow fields to aggregation remains unclear. Here we develop a flow device that subjects proteins to a defined and quantified flow field that is dominated by extensional flow. We show that extensional flow is crucial to induce the aggregation of globular proteins and that flow-induced aggregation is dependent on both protein structure and sequence. These observations rationalize the diverse effects of hydrodynamic flow on protein structure and aggregation propensity seen in both Nature and in protein manufacture.

Introduction

Proteins are dynamic and metastable and consequently have conformations that are highly sensitive to the environment (1). Over the last 50 years the effect of changes in temperature, pH and the concentration of kosmatropic/chaotropic agents on the conformational energy landscape of proteins has become well understood (1). This, in turn, has allowed a link to be established between the partial or full unfolding of proteins and their propensity to aggregate (2). The force applied onto a protein as a consequence of hydrodynamic flow has also been observed to trigger protein aggregation and has fundamental (3), medical (4) and industrial relevance, especially in the manufacture of bio-pharmaceuticals (5-8). While a wealth of studies have been performed (7, 9-13), no consensus has emerged on the ability of hydrodynamic flow to induce protein aggregation (7, 14, 15). This is due to the wide variety of proteins used (ranging from lysozyme, BSA and alcohol dehydrogenase to IgGs), differences in the type of flow field generated (e.g. shear, extensional or mixtures of these) and to the presence or absence of an interface (16). A shearing flow field (Fig. 1A, top) is caused by a gradient in velocity perpendicular to the direction of travel and is characterized by the shear rate (s^{-1}). This results in a weak rotating motion of a protein alongside translation in the direction of the flow. An extensional flow field (Fig. 1A, bottom) is generated by a gradient in velocity in the direction of travel and is characterized by the strain rate (s^{-1}). A protein in this type of flow would experience an extensional force between the front (faster flow) and the rear (slower flow), potentially leading to elongation of the molecule as directly observed for a single DNA molecule (17, 18). The majority of protein aggregation studies to date have considered shear flows within capillaries (3) or through using viscometric type devices (10, 16). On the whole, these studies show that globular proteins are generally resistant to shear flow in the absence of an interface (3, 14, 19). By contrast, Simon *et al.* (11) showed increased aggregation of BSA with increasing extensional flow. Many operations within bio-pharmaceutical

manufacture such as filtration, filling and pumping (5, 20) also create extensional flow fields. This observation together with the importance of extensional flow fields to thrombosis (21) and spider silk spinning (22) suggests a link between extensional flow and protein aggregation.

To assess the relative importance of extension and shear to flow-induced aggregation, we have developed a low-volume flow device, characterized using Computational Fluid Dynamics (CFD), which uses a rapid constriction to generate an extensional flow field followed by flow within a capillary that generates a shearing flow. Using our device, it is possible to de-convolute the effects of shearing and extensional flow fields. We demonstrate that extensional flow can trigger the aggregation of BSA and that the extent of aggregation is dependent on the total exposure time, strain rate and protein concentration. We also show that the aggregation of a range of globular, natively folded proteins (β_2m, G-CSF and three mAbs) under extensional flow is diverse and is particularly damaging to therapeutic proteins (G-CSF and mAbs) under conditions analogous to those encountered during their manufacture. Finally, we show directly that the device triggers the aggregation of BSA by inducing partial unfolding and that the extent of aggregation is strain-rate and protein concentration dependent, suggesting that aggregation occurs by interaction of partially unfolded proteins whose population is induced by extensional flow.

Results

Design and computational characterization of extensional flow device

Extensional flow can be generated by methods including a cross-slot (23-25), a four-mill-roller (11), opposed jets (26), or by introduction of a constriction in a pipe while maintaining laminar flow conditions (18, 27). The simplicity of the latter method led to the design of a reciprocating flow device driven by a linear actuator comprising two gas-tight 1 mL syringes (bore diameter 4.61 mm) connected *via* a glass capillary (75 mm long, inner diameter 0.3 mm) using compression fittings (Fig. 1B and 1C). The rapid reduction in tube diameter at the square-edged constriction (~ 15:1) produces a 238-fold increase in linear velocity at each syringe:capillary connection.

Computational Fluid Dynamics (CFD) was then used to characterize the type, magnitude and timescale of the hydrodynamic forces generated by this device (Methods and SI Appendix). The two-dimensional axisymmetric domain used to represent the physical flow device is shown in Fig. 1D, together with the contraction geometry from which it is derived (note: only one contraction was modelled). As the velocity and strain rate experienced by the fluid depends on its initial radial position, the centre-line strain rate values are reported. At a plunger speed of 8 mms^{-1}, CFD analysis (Fig. 1D*ii-iv* and SI Appendix, Fig. S1) shows that distal to the constriction, the mean flow velocity is constant (8 mm s^{-1} and 1.9 m s^{-1} at the extremity of each pipe), but increases rapidly (*i.e.* accelerates) over 2 mm in the vicinity of the constriction (blue line, Fig. 1D*iv*). These data show that as a protein transits a constriction, it will experience high strain effects due to hydrodynamic extensional flow. Under these conditions the residence time of a particle within the syringe barrels, the capillary and the constriction is approximately 5, 40×10^{-3} and 18×10^{-6} s (SI Appendix, Fig. S1B) respectively. Repeating CFD using the same geometry but at various inlet velocities (i.e. the velocity at which the plunger moves down

the syringe barrel) revealed that the length of the extensional region is independent of the plunger velocity, but that the maximum strain rate (along the centre-line) is directly dependent on plunger velocity (SI Appendix, Fig. S1A). Additionally, the exposure time of fluid to the high strain regime decreases exponentially with plunger velocity (SI Appendix, Fig. S1B). Finally, the flow exponent for this device was calculated as n= 0.905 (SI Appendix) which is consistent with a dominant laminar flow regime across the contraction. For simplicity, as the geometry of the device is fixed throughout this study, we report the plunger velocity and number of passes, allowing recovery of the fundamental fluid mechanical parameters by reference to SI Appendix, Table S1 and Fig. S1A.

Defined flow fields can aggregate BSA

The all α-helical, 583-residue protein BSA (Fig. 2A) was used for our initial studies as it has well-characterized intrinsic aggregation pathways (28-30) and its behaviour under shear and extensional flow-fields has been investigated previously (11, 31). To assess whether our extensional flow device can induce protein aggregation, 500 μL of gel-filtered mono-disperse BSA (Methods) at a concentration of 1, 2, 5 or 10 mg mL^{-1} was passed through the capillary 500, 1000, 1500 or 2000 times at a plunger velocity of 8 mm s^{-1} (equivalent to total exposure times to extensional flow of 55, 109, 164 and 218 ms, respectively, at fixed centre-line strain- (11750 s^{-1}) and shear- (52000 s^{-1}) rates (Methods and SI Appendix). The concentration of soluble protein was then quantified after ultra-centrifugation (Fig. 2B) and any aggregates present in the unclarified sample visualized using transmission electron microscopy (TEM) (Fig. 2C). These data show that the designed device can induce protein aggregation, with the amount of insoluble material increasing with increasing pass number in a protein concentration-dependent manner (Fig. 2B, black to magenta bars). TEM images (Fig. 2C, SI

Appendix, Fig. S2) revealed that the aggregates which form as a result of extensional flow are amorphous in nature and show greater compaction and demarcation with increasing pass number.

Our extensional flow device comprises two constrictions connected by a capillary. Consequently, the observed aggregation could be induced, at least in part, by the shear flow present in the glass connector capillary. To investigate this possibility, 5 mg mL^{-1} BSA was subjected to 1000 or 100 passes at 8 and 16 mm s^{-1}respectively through the same extensional flow device fitted with a half-length (37.5 mm) connecting capillary, reducing the exposure time of the protein to shear by half. The yield of insoluble material was found to be unaffected (Fig. 2D, SI Appendix, Fig. S3), identifying extensional flow rather than shear flow as a key trigger of aggregation for BSA.

Dissecting the dispersity of extensional flow-induced aggregates

To determine how the dispersity of BSA changes with increasing pass number, Nanoparticle Tracking Analysis (NTA, Fig. 2E and S4) and Dynamic Light Scattering (DLS), (SI Appendix Fig. S5 and S6) were used to analyze samples subjected to between 20 and 2000 passes at 8 mm s^{-1}. NTA visualizes particles with hydrodynamic diameters of 10-2000 nm, allowing the sizing and numeration of polydisperse colloidal solutions (32). Hydrodynamic diameter frequency histograms for BSA solutions (5 mg mL^{-1}) show an increase in particle size (ranging from 40 nm to > 1 μm) with increasing pass number (SI Appendix, Fig. S4) while unstressed BSA, or BSA stressed for fewer than 50 passes, yielded no detectable particles (monomeric BSA has a hydrodynamic radius (R_h) of ~3.5 nm, see below). The total number of aggregates was also found to increase with increasing pass number and with increasing BSA concentration

(Fig. 2E). Furthermore, the width of the error bars also increases with the number of passes, suggesting an increase in sample dispersity.

These data were then corroborated using DLS. Comparison of regularization plots (14, 33) for BSA before and after 2000 passes, showed a reduction in intensity for monomeric BSA (R_h = 3.2 ± 0.9 nm) and the appearance of a range of particles with R_h values > 100 nm (SI Appendix, Fig. S5, Table S2). Whilst accurate determination of the size and relative amount of each particle type in polydisperse solutions is difficult by DLS (34), polydispersity can, nonetheless, be assessed qualitatively by calculation of the z-average radius and polydispersity index (PDI). The z-average radius (the average size of a particle within a disperse solution based on the averaged intensity of all species within the solution) is obtained directly from the raw autocorrelation data using cumulants analysis (SI Appendix). The PDI (where values of ~ 0.1 and > 0.6 reveal monodisperse and highly polydisperse species, respectively) is calculated from the z-average radius (SI Appendix) and is used to assess polydispersity (35). Comparing the z-average radii (SI Appendix, Fig. S6, Table S3) and PDI (SI Appendix, Table S3) for 5 mg mL^{-1} BSA before (3.5 ± 0.1 nm and ~0.1) and after (87.9 ± 66.3 nm and >0.6) 2000 passes through the device clearly demonstrates that exposure to extensional flow has induced aggregation. The polydispersity of various concentrations of BSA (1, 2, 5 and 10 mg mL^{-1}) after 0 – 2000 passes assessed by regularization (SI Appendix, Table S2), z-average radii (SI Appendix, Fig. S6, Table S3) and PDI (SI Appendix, Table S3) all showed that the mean hydrodynamic radius increases with increasing pass number and that aggregates appear at lower pass numbers upon increasing BSA concentration. Together, the soluble protein assay, NTA and DLS provide evidence for extensional flow-induced aggregation and indicate that 50 passes are required to observe any visible aggregates for BSA at concentrations ≥ 1 mg mL^{-1}.

Investigating early events in BSA aggregation

The absence of aggregates prior to 50 passes suggests that BSA must either display a history effect (some threshold of 'damage' must be attained before aggregation occurs) or that NTA and DLS are insensitive to the (presumably) low concentration of aggregate present at low pass numbers. To address the latter possibility fluorescence correlation spectroscopy (FCS) was used to measure the effect of pass number on the diffusion time of a 1 – 10 mg mL^{-1} BSA solution doped with 1 % (*v/v*) Alexa 488-labelled BSA (SI Appendix). In agreement with the solubility, NTA and DLS data above, the autocorrelation functions for BSA solutions exposed to > 50 passes deviate from that expected for a mono-disperse monomeric species (SI Appendix, Fig. S7 and Table S4) and that the aggregates increase in size with increasing pass number. By contrast with NTA and DLS, however, FCS is able to detect the presence of smaller oligomers after as few as 10 passes at 5 mg mL^{-1} (R_h increases ~3-fold). After 20 or more passes, correlation functions could not be fitted to a single component model but fitted to a two component model (SI Appendix). These data show that the R_h of the larger species increases with increasing pass number (23.3, 34.9 and 52.0 nm after 20, 50 and 100 passes, SI Appendix, Tables S4). In accord with the observations above, increasing protein concentration was found to lead to the detection of polydispersity after fewer passes (SI Appendix, Table S4), as expected for an aggregation reaction of high molecular order.

Aggregation of other proteins

The effects of extensional flow on the aggregation behaviour of a range of other proteins that differ in size, secondary structure content and topology (SI Appendix, Table S5) were next examined to assess whether the behaviour of BSA is typical for globular, folded proteins. Each protein was subjected to 20 or 100 passes at 8 mm s^{-1} and the resulting aggregation quantified by UV-spectrophotometry and the dispersity characterized by NTA and DLS. β_2m (100 residues, R_h = 2.3 nm) was selected as fluid flow has been implicated previously in the

aggregation of the protein into amyloid fibrils in the joints of patients undergoing long-term dialysis (36). 5 mg mL^{-1} β_2m was found to be more sensitive than BSA to extensional flow as assessed by the pelleting assay (2 and 10 % was insoluble after 20 or 100 passes, respectively, compared with 1 % and 1.5 % of BSA, Fig. 3A). Indeed, NTA and DLS measurements (SI Appendix, Fig. S8) detected aggregates of β_2m after 100 passes. Visualization of these aggregates by TEM revealed short needle-like fibrils (Fig. 3B). The behavior of β_2m contrasts starkly with that of the C3 variant of G-CSF (37). This 175-residue, four-helical all α-protein was found to be extremely sensitive to the effects of extensional flow. After only 20 passes of 0.5 mg mL^{-1} GCSF-C3 through the device (a ten-fold reduction in concentration relative to BSA and β_2m), heterogeneous amorphous aggregates as large 8 μm were observed by NTA and DLS (PDI > 0.6, SI Appendix, Fig. S9) with 20 % of GCSF C3 rendered insoluble (Fig. 3A and 3C), increasing to 40 % after 100 passes. Finally, three model IgG biopharmaceuticals (MEDI1912_WFL (38), MEDI1912_STT (38) and mAb1) were subjected to extensional flow. These IgGs were chosen as they have known, but different, aggregation propensities (38). MEDI1912_WFL and MEDI1912_STT differ by only six residues (WFL substituted by STT in CDR1 (W and F) and CDR2 (L) in each V_H domain), yet the former IgG has poor pharmacokinetic and biophysical properties, while its rationally engineered variant displays the same pM affinity for its target, but low self-association and enhanced serum persistence(38). The third mAb (mAb1) was chosen as it has low sequence identity to MEDI1912_WFL and displays aggregation behaviour in line with a typical 'bioprocessible' IgG (e.g. greater than 95 % monomeric purity post protein A purification with a degradation rate of less than 2 % monomer loss per year in solution by HPLC-SEC) (38). All three IgGs (at a concentration of 0.5 mg mL^{-1}) were found to be sensitive to the effects of extensional flow; aggregates were detected by NTA after 20 passes for mAb1 and MEDI1912_WFL (SI Appendix, Fig. S10A and S11A) and after 100 passes for MEDI1912_STT (SI Appendix, Fig. S12A). Notably, these

three proteins display markedly different sensitivity to extensional flow, despite their structural similarity. The aggregation-prone MEDI1912_WFL was so sensitive to extensional flow that only ~45 and ~15 % of protein remained in solution after 20 and 100 passes, respectively (Fig. 3A). This sensitivity is remarkable given the low concentration of protein used (0.5 mg mL^{-1}). By contrast, mAb1 yielded significantly less insoluble material (~15 and ~25 % after 20 and 100 passes) with MEDI1912_STT exhibiting still less susceptibility (~2 and ~5% insoluble material after 20 and 100 passes, respectively (Fig. 3A)). These data show that bio-pharmaceuticals with diverse structures (G-CSF C3 is all-α and mAbs are all-β, SI Appendix, Table S5) are prone to extensional flow-induced aggregation and, surprisingly, that the aggregation propensity of IgGs that differ only at three positions in the CDR loops of each V_H domain show remarkably different responses to hydrodynamic flow.

What drives aggregation?

Previous studies have suggested that hydrodynamic forces can induce conformational changes in proteins (10, 39) but how these changes result in aggregation remained unclear. To assess the mechanism of extensional flow-induced unfolding, 5 mg mL^{-1} BSA was subjected to 10, 20, 50 or 100 passes at 8 mm s^{-1} in the presence of 5 mM 5-[2-(Iodoacetamido)ethylamino]naphthalene-1-sulfonic acid (IAEDANS, a sulfhydryl reactive fluorophore). In the native state, all but one of the 35 cysteine residues of BSA form disulfide bridges and the only free cysteine residue available for labelling (Cys34) is buried (13.8 $Å^2$ solvent accessible surface area, Fig. 2A) and recalcitrant to labelling (Fig. 3F, lane 1). Exposure to extensional flow, however, renders BSA sensitive to labelling, the extent of which increases with increasing pass number (quantified in Fig. 3F, bottom). To determine whether the change in solvent accessibility of Cys34 occurs during extensional flow rather than conformational changes upon aggregation, BSA was subjected to 10, 20, 50 and 100 passes,

allowed to relax for up to ten minutes (see Methods) and then incubated with IAEDANS for a time equivalent to that of the extensional flow experiment (100 passes ~ 10 minutes, see Methods). No labelling was evident under these conditions (SI Appendix, Fig. S13). To determine the effect of the unusually dense disulfide network present in BSA, which may limit the extent of flow-induced unfolding, the labelling experiments were repeated by applying the extensional flow in the presence of 0.5 mM TCEP (Fig. 3F). The data show an approximate two-fold increase in labelling, together with a 2-fold increase in insoluble material produced in the presence of TCEP (SI Appendix, Fig. S14) when BSA is subjected to extensional flow. Interestingly, when IAEDANS was added ten minutes after stressing the protein in the presence of 0.5 mM TCEP, BSA was again labelled (SI Appendix, Fig. S13) indicating that breakage of the disulphide bridges under extensional flow in the presence of reductant yields aggregates comprising unfolded monomers with sulfhydryl groups exposed to the solvent, by contrast to the aggregates formed in the absence of reductant. These data accord with experimental and theoretical investigations using proteins, DNA, organic polymers and coarse-grained models which suggest that flow applies a stretching force to molecules along the flow field that is proportional to the strain rate. As the hydrodynamic drag of the protein increases as it unfolds, the already destabilized protein has a greater susceptibility to unfold further and interact with other proteins (40-44) which increases its size.

The extensional force or tension experienced along a protein represented by two globular domains with a diameter of 2.34 nm connected by a 44.3 nm linker to form a dumb-bell (3) is of the order of 10 fN for plunger velocities up to 20 mm s^{-1} (SI Appendix). This is 2 to 3 orders of magnitude lower than that required to mechanically unfold a protein using atomic force microscopy over a similar timescale (45) suggesting that a protein cannot be globally unfolded. Instead of the force exerted, the global energy requirements for a protein-sized fluid packet (3.5 nm radius) to pass through either the extensional flow-dominated acceleration region or

the shear-dominated capillary region was calculated (SI Appendix). This analysis (Fig. 4A) shows that even the global energy available (~2.7 and 1.1 k_BT at 8 mm s^{-1} for the extensional- and shear- dominated regions, respectively is not sufficient to completely unfold the protein, noting that only a very small proportion of this energy will be absorbed into the structure. The global energy requirement of the shear-dominated capillary region is similar to that of the extensional flow region, despite the latter having been shown to be responsible for the aggregation of BSA. A more appropriate parameter to consider may thus be the rate of energy transfer to the fluid packet $E/(k_BT\ t)$ (s^{-1}) where t is the time to transit the capillary or the extensional flow region (Fig. 4B) since this better represents how quickly energy is added into the protein solution, noting that any perturbed structure will also dissipate energy at a finite rate. Scaling by t also removes the dependence of capillary length on (E/k_BT) for the shear-flow as observed in this study. This analysis shows that the rate of energy transfer by extensional flow is orders larger than that for shear.

To investigate the effect of plunger velocity on the extent of aggregation, the insoluble material generated after subjecting 5 mg mL^{-1} BSA to 100 passes at 2, 4, 8, 10, 12, 14 and 16 mm s^{-1} was quantified (Fig. 4C). At velocities below 10 mm s^{-1}, BSA aggregation was found to be independent of strain rate yielding only ~1-2 % pelletable material. By contrast, at plunger velocities between 10-16 mm s^{-1}, formation of insoluble protein increased with strain rate, so that at 16 mm s^{-1} ~15 % of BSA was rendered insoluble after only 100 passes. Importantly, the degree of aggregation was found to be similar for full- and half-length capillaries at 8 and 16 mm s^{-1} (Fig. 2D), in accord with the simulations shown in Fig. 4B. These data suggest that at lower extensional strain rates, the transient unfolding force exerted onto the protein is insufficient to trigger full unfolding, or that this force triggers partial unfolding close to equilibrium (hydrodynamic force countered by folding energy). At higher strain rates,

however, partial unfolding of BSA (the extent of which is limited by the disulfide network) triggers aggregation.

Discussion

While prior work has demonstrated that hydrodynamic flow can induce the unfolding of supercoiled plasmid DNA (20, 27), polymers (26), von Willebrand factor (21, 46), and other proteins (11), the relative ability of shear and extensional flow to induce aggregation for these different systems has remained unclear. In order to address this question, we designed a device to generate an extensional flow field that would subject natively folded globular proteins to high and well defined strain rates. Using this device, we demonstrated that extensional flow has the ability to induce the aggregation of BSA. This, together with previously published studies, suggests that while shear and extensional flow fields can both induce aggregation (11, 31), their ability to do so is protein dependent. For example, both spider silk and Von Willebrand factor have been observed to undergo shear-induced remodelling (which nonetheless are exposed to mixed shear / extensional flows *in vivo* (22, 46)). As both of these proteins are evolved to respond to low levels of hydrodynamic force it may be that their response to shear is atypical for globular, stably folded proteins. The latter proteins are relatively insensitive to shear flow where the presence of an interface is often required to induce aggregation (19).

Repeating our experiments on a variety of proteins demonstrated that the extent of aggregation caused by extensional flow depends on the structure, topology, concentration and precise sequence of the protein. In addition to delineating these determinants, we have shown using *in situ* cysteine labelling that extensional flow can induce conformational remodelling. The theoretical considerations and data discussed above (Fig. 4), suggest that extensional flow can catalyse the partial/full unfolding of proteins. A critical rate of energy transfer must, however, be reached to allow the unfolding barrier to be traversed during exposure to the flow force. Superficially, hydrodynamic forced unfolding is similar to mechanical unfolding of single protein molecules using optical tweezers or the atomic force microscope. These single-

molecule forced unfolding studies have shown that mechanical strength is related to the ability of regions local to the points of force application to resist extension by contrast with traditional measures of stability such as thermal or chemical denaturation (45, 47). If flow-induced aggregation occurs from a partially or fully unfolded state, then the threshold strain rate (i.e. that required to bring about exposure of an aggregation-prone region) will be protein dependent. As a consequence, natively folded globular proteins will be generally recalcitrant to shear flow, whereas inherently extensible unstructured proteins are not. After the initial partial unfolding step, the likelihood of two (or more) unfolded molecules interacting productively is dependent on the affinity of the exposed aggregation-prone regions, the protein concentration and the rate at which the protein regains its native structure, rationalizing the diverse sensitivity observed for the highly homologous IgG pair (MEDI1912_WFL and STT, Fig. 3A). Furthermore, as both the unfolding and aggregation steps are likely to dependent on factors such as pH, temperature and ionic strength, even the same protein may display different extensional flow behavior in different environments.

In summary, we have shown the utility of characterizing the behavior and dispersity of protein solutions subjected to well-defined hydrodynamic flows in order to de-convolve the effects of shear, extensional flow, protein topology and sequence on their unfolding and aggregation properties. The results have revealed the sensitivity of proteins to unfolding and consequent aggregation under extensional flow in a manner dependent on the protein sequence and structure. The approach adopted will aid the rational re-design of protein sequences that are more robust to bio-processing and help to understand how flow has been utilized by nature in biological processes as diverse as silk spinning and blood clotting.

Methods

Characterization of flow geometry using computational fluid dynamics (CFD): CFD (using the general finite element simulation package, Comsol Multiphysics) was used to visualise and quantify the flow field generated by the extensional flow device. This allowed the velocity, strain rate and exposure time, amongst other parameters, to be calculated. A description of the CFD model, along with details of how to obtain the strain rate is given in the SI Appendix.

Extensional flow apparatus and experiments

Two 1 mL gas-tight syringes with inner bore diameter of 4.61 mm (Hamilton Syringes model 1001 RN Valco SYR) were modified to take a glass capillary tube of inner diameter 0.3 mm with a compression fitting (Hamilton Syringes RN 1 mm) producing an abrupt contraction with diameter ratio ~15:1 producing a 238-fold increase in velocity. Protein solutions were stressed for a defined number of passes at a given plunger velocity, then the rig stopped, dissembled and the solution expelled slowly from the syringe. Control samples were incubated at ambient temperature for the duration of a given stress experiment (e.g. 10 passes at a plunger velocity of 8 mm s^{-1} takes 1 minute to complete). See SI Appendix. All experiments were performed at least twice unless otherwise stated.

Protein preparation

Bovine serum albumin (BSA, Sigma Aldrich) was purified by gel filtration chromatography using a Superdex 200 (26/60) gel filtration column (GE Healthcare) equilibrated with 25 mM ammonium acetate buffer, pH 5.1 and stored in aliquots at -20°C. Prior to stressing experiments, the protein was concentrated using a centrifugal concentrator with a 30 kDa cut-off filter (Merck Millipore). After filtration through a 0.22 μm membrane (Merck Millipore),

the concentration was determined by UV spectroscopy (SI Appendix, Table S5) and adjusted as necessary. GCSF-C3 (37) was over-expressed in BL21[DE3]pLysS cells transformed with a pET23a_GCSF_C3 vector and purified as described in SI Appendix. Extensional flow experiments with GCSF C3 were performed in filtered (0.22 μm) and de-gassed 25 mM sodium phosphate, 25 mM sodium acetate buffer, pH 7.0. β_2m was purified as described (48) and extensional flow experiments performed in filtered (0.22 μm) and de-gassed 25 mM sodium phosphate buffer, pH 7.2. Antibodies were provided by MedImmune Ltd (Cambridge). Antibodies were prepared by dialyzing into 0.22 μm filtered and de-gassed 150 mM ammonium acetate buffer, pH 6.0, diluting prior to stressing experiments as appropriate.

Insoluble protein pelleting assay

After stressing for the desired number of passes, the apparatus was dissembled and 200 μL of protein solution ultracentrifuged using a Beckmann Coulter Optima TLX Ultracentrifuge equipped with a TLA100 Rotor at 30,000 rpm for 30 minutes at 4 °C. 150 μL of supernatant was then removed and diluted to 2 mL (BSA) or 250 μL (all other proteins) in 6 M guanidine hydrochloride (Gdn HCl) 25 mM TrisHCl buffer, pH 6. The pellet and remaining supernatant were diluted in the same buffer to 2 mL (BSA) or 250 μL (all other proteins) and incubated overnight. The amount of protein in the pellet was then calculated by measuring the protein concentration of this solution, the supernatant after ultracentrifugation and the protein solution in the absence of extensional flow using UV-visible spectroscopy (see SI Appendix, Table S6 for extinction coefficients). This procedure was performed in duplicate.

Biophysical characterization of polydispersity

Experimental procedures for DLS, NTA, TEM and FCS are described in the SI Appendix.

IAEDANS (5-naphthalene-1-sulfonic acid) labelling of BSA.

5 mg mL^{-1} BSA solution (25 mM ammonium acetate pH 5.1) was mixed with 5 mM IAEDANS and stressed for 0 – 100 passes at a plunger velocity of 8 mm s^{-1} (strain rate = 11750 s^{-1}). TCEP at 0.5 mM was added to the tube before stressing as required. In another experiment a5 mg mL^{-1} BSA was stressed for 0 – 100 passes in the presence or absence of TCEP (this protein was left for the same length of time as the extensional flow experiment as above). Subsequently, this protein was mixed with 5 mM of IAEDANS and incubated for the same time as the protein was stressed for in the presence of IAEDANS above. The IAEDANS labelling was quenched with SDS PAGE loading buffer containing 200 mM DTT. The diluted samples (~100 µg) were then analyzed by SDS PAGE (using a 12 % w/v (37.5:1 acrylamide:bis-acrylamide) gel). Fluorescent bands in the gel were excited by UV light provided by a UV-trans illuminator (Syngene Gel documentation). The intensities of the fluorescent bands were analyzed with the Gene Tool software supplied with the instrument. The gel was then stained with Coomassie Brilliant Blue.

Acknowledgements

We thank Prof Joanne Tipper and Dr Saurabh Lal for their help with NTA and Mr David Sharples for access to the ultracentrifuge facility. JD and AK are co-funded funded by Medimmune Ltd and the University of Leeds. LFW is funded by the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies, UK (EP/I033270/1) and SER and DJB acknowledge funding by the ERC (FP7/2007-2013 Grant agreement 32240). NK holds a Chair in Pharmaceutical Processing at the University of Leeds funded by GSK and the Royal Academy of Engineering. The DLS instrument was funded by the MRC (G0900958), the EM by the Wellcome Trust (108466/Z/15/Z) and the FCS apparatus partly funded by the

ERC (as above). We thank Professor Peter Olmsted for many insightful discussions at the inception of this work.

References

1. Jahn TR & Radford SE (2005) The Yin and Yang of protein folding. *FEBS J.* 272:5962-5970.
2. Tipping KW, van Oosten-Hawle P, Hewitt EW, & Radford SE (2015) Amyloid Fibres: inert end-stage aggregates or key players in disease? *Trends Biochem. Sci.* 40:719-727.
3. Jaspe J & Hagen SJ (2006) Do protein molecules unfold in a simple shear flow? *Biophys. J.* 91:3415-3424.
4. Arora D, Behr M, & Pasquali M (2004) A tensor-based measure for estimating blood damage. *Artif. Organs.* 28:1002-1015.
5. Cromwell ME, Hilario E, & Jacobson F (2006) Protein aggregation and bioprocessing. *AAPS J.* 8:E572-E579.
6. Wang W, Nema S, & Teagarden D (2010) Protein aggregation--pathways and influencing factors. *Int. J. Pharm.* 390:89-99.
7. Thomas CR & Geer D (2011) Effects of shear on proteins in solution. *Biotechnol. Lett.* 33:443-456.
8. Tolbert W & Prior C (1989) *Perfusion Culture, in Advanced Research on Animal Cell Technology* (Springer Netherlands, Springer Netherlands) pp 119-145.
9. Rathore N & Rajan RS (2008) Current perspectives on stability of protein drug products during formulation, fill and finish operations. *Biotechnol. Prog.* 24:504-514.

10. Charm SE & Wong BL (1981) Shear effects on enzymes. *Enzyme Microb. Technol.* 3:111-118.
11. Simon S, Krause HJ, Weber C, & Peukert W (2011) Physical degradation of proteins in well-defined fluid flows studied within a four-roll apparatus. *Biotechnol. Bioeng.* 108:2914-2922.
12. Bekard IB, Asimakis P, Bertolini J, & Dunstan DE (2011) The effects of shear flow on protein structure and function. *Biopolymers* 95:733-745.
13. Szymczak P & Cieplak M (2011) Hydrodynamic effects in proteins. *J. Phys. Condens. Matter* 23:033102.
14. Tirrell M & Middleman S (1975) Shear modification of enzyme-kinetics. *Biotechnol. Bioeng.* 17:299-303.
15. Thomas CR, Nienow AW, & Dunnill P (1979) Action of shear on enzymes: studies with alcohol dehydrogenase. *Biotechnol. Bioeng.* 21:2263-2278.
16. Brückl L, Schröder T, Scheler S, Hahn R, & Sonderegger C (2016) The effect of shear on the structural conformation of rhGH and IgG1 in free solution. *J.Pharm.Sci.*105:1810-1818.
17. Atkins ED & Taylor MA (1992) Elongational flow studies on DNA in aqueous solution and stress-induced scission of the double helix. *Biopolymers* 32:911-923.
18. Larson JW*, et al.* (2006) Single DNA molecule stretching in sudden mixed shear and elongational microflows. *Lab chip* 6:1187-1199.
19. Bee JS*, et al.* (2009) Response of a concentrated monoclonal antibody formulation to high shear. *Biotechnol. Bioeng.* 103:936-943.
20. Zhang H*, et al.* (2007) Prediction of shear damage of plasmid DNA in pump and centrifuge operations using an ultra scale-down device. *Biotechnol. Prog.* 23:858-865.

21. Springer TA (2014) von Willebrand factor, Jedi knight of the bloodstream. *Blood* 124:1412-14125.

22. Rammensee S, Slotta U, Scheibel T, & Bausch AR (2008) Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. USA* 105:6590-6595.

23. Bae YB*, et al.* (2016) Microfluidic assessment of mechanical cell damage by extensional stress. *Lab chip* 16:96-103.

24. Perkins TT, Smith DE, & Chu S (1997) Single polymer dynamics in an elongational flow. *Science* 276:2016-2021.

25. Haward SJ, Oliveira MS, Alves MA, & McKinley GH (2012) Optimized cross-slot flow geometry for microfluidic extensional rheometry. *Phys. Rev. Lett.* 109:128301.

26. Odell JA, Muller AJ, Narh KA, & Keller A (1990) Degradation of Polymer-Solutions in Extensional Flows. *Macromolecules* 23:3092-3103.

27. Meacle FJ*, et al.* (2007) Degradation of supercoiled plasmid DNA within a capillary device. *Biotech. Bioeng.* 97:1148-1157.

28. Bhattacharya A, Prajapati R, Chatterjee S, & Mukherjee TK (2014) Concentration-dependent reversible self-oligomerization of serum albumins through intermolecular beta-sheet formation. *Langmuir* 30:14894-14904.

29. Bhattacharya M, Jain N, Bhasne K, Kumari V, & Mukhopadhyay S (2011) pH-induced conformational isomerization of bovine serum albumin studied by extrinsic and intrinsic protein fluorescence. *J. Fluoresc.* 21:1083-1090.

30. Bhattacharya M, Jain N, & Mukhopadhyay S (2011) Insights into the mechanism of aggregation and fibril formation from bovine serum albumin. *J. Phys. Chem. B* 115:4195-4205.

31. Bekard IB*, et al.* (2012) Bovine serum albumin unfolds in Couette flow. *Soft Matter* 8:385-389.

32. Filipe V, Hawe A, & Jiskoot W (2010) Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm. Res.* 27:796-810.
33. Hanlon AD, Larkin MI, & Reddick RM (2010) Free-solution, label-free protein-protein interactions characterized by dynamic light scattering. *Biophys. J.* 98:297-304.
34. Hassan PA, Rana S, & Verma G (2015) Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering. *Langmuir* 31:3-12.
35. Roger V, Cottet H, & Cipelletti L (2016) A new robust estimator of polydispersity from dynamic light scattering data. *Anal. Chem.* 88:2630-2636.
36. Mangione PP*, et al.* (2013) Structure, folding dynamics, and amyloidogenesis of D76N beta2-microglobulin: roles of shear flow, hydrophobic surfaces, and alpha-crystallin. *J. Biol. Chem.* 288:30917-30930.
37. Buchanan A*, et al.* (2012) Improved drug-like properties of therapeutic proteins by directed evolution. *Protein Eng. Des. Sel.* 25:631-638.
38. Dobson CL*, et al.* (2016) Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance *in vivo*. *Sci. Rep.* 6:38644.
39. Ashton L, Dusting J, Imomoh E, Balabani S, & Blanch EW (2009) Shear-induced unfolding of lysozyme monitored *in situ*. *Biophys. J.* 96(10):4231-4236.
40. Agarwal US (2000) Effect of initial conformation, flow strength, and hydrodynamic interaction on polymer molecules in extensional flows. *J. Chem. Phys.* 113:3397-3403.
41. Agarwal US, Bhargava R, & Mashelkar RA (1998) Brownian dynamics simulation of a polymer molecule in solution under elongational flow. *J. Chem. Phys.* 108:1610-1617.

42. De Gennes PG (1974) Coilstretch transition of dilute flexible polymers under ultrahigh velocity gradients. *J. Chem. Phys.* 60:5030-5042.

43. Szymczak P & Cieplak M (2006) Stretching of proteins in a uniform flow. *J. Chem. Phys.* 125:164903-164908.

44. Szymczak P & Cieplak M (2007) Proteins in a shear flow. *J. Chem. Phys.* 127:155106-155107.

45. Chen Y, Radford SE, & Brockwell DJ (2015) Force-induced remodelling of proteins and their complexes. *Curr. Opin. Struct. Biol.* 30:89-99.

46. Sing CE & Alexander-Katz A (2010) Elongational flow induces the unfolding of von Willebrand factor at physiological flow rates. *Biophys J.* 98:L35-L37.

47. Crampton N & Brockwell DJ (2010) Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.* 20:508-517.

48. Tipping KW*, et al.* (2015) pH-induced molecular shedding drives the formation of amyloid fibril-derived oligomers. *Proc. Natl. Acad. Sci. USA* 112:5691-5696.

Figure Legends

Fig. 1. Design of extensional flow apparatus and validation of the generated flow field using CFD. (A) The differences between shear- (top) and extensional flow (bottom). Solid black arrows indicate velocity, dashed lines show streamlines indicative of flow direction. The red dot in both diagrams represents a protein as a sphere in the flow. The curved red arrow represents rotation due to shear. Straight red arrows indicate the relative velocity of the protein, which differs before and after the contraction in the extensional flow. (B) Schematic of the extensional flow apparatus showing two syringes connected by a single capillary. (C) Image of the extensional flow device. (D) (i) A 3D schematic of the contraction geometry where the barrel of the syringe meets the capillary (dotted red line shows the location of the contraction). The 2D axis-symmetric approximation used for CFD analysis is superimposed in blue. CFD results of the extensional flow region showing the flow velocity (ii) and strain rate (iii) profiles for a typical flow with a plunger velocity (average inlet velocity) of 8 mms^{-1} (centre line strain rate = 11750 s^{-1}). (iv) Velocity and strain rate along a streamline located on the axis of symmetry at a plunger velocity of 8 mms^{-1}.

Fig. 2. Aggregation of BSA is induced by an extensional flow field. (A) Structure (top, PDB: 3V03 and topology diagram (bottom) of BSA colored by domain. Disulfide bridges are shown as blue lines and loops as grey lines. The free cysteine is shown as an open green circle highlighted by the red arrow (top) and a closed green circle (bottom). (B) Bar graph showing % BSA remaining in solution after 500-2000 passes at 8 mm s^{-1} at a protein concentration of 1 (black), 2 (red), 5 (blue) or 10 (magenta) mg mL^{-1}. (C) TEM images of 5 mg mL^{-1} BSA after 0 (top) and 2000 passes (bottom). The grids were imaged at 10000× magnification and the scale bar represents 500 nm. (D) Halving the exposure time to shear flow using a 37.5 mm (Half) instead of a 75 mm (Full) connecting capillary has no effect on the extent of aggregation of 5 mg mL^{-1} BSA after 1000 passes at 8 mm s^{-1} or 100 passes at 16 mm s^{-1}. (E) Total number of 10-2000 nm particles tracked by NTA in 1, 2, 5 and 10 mg mL^{-1} BSA solutions after 50-2000 passes at 8 mm s^{-1}. Error bars represent the error from two independent experiments.

Fig. 3. Quantification of flow-induced aggregation of different proteins and the mechanism of extensional flow-induced aggregation of BSA. (A) Bar graph showing percentage of protein remaining in solution after 0, 20 or 100 passes at a plunger velocity of 8 mm s^{-1}. The protein concentrations used are shown on the top. (B) – (E) TEM images of β_2m, GCSF-C3, mAb1 and MEDI1912-WFL after 100 passes. The grids were imaged at 10000× magnification (scale bar = 500 nm). (F) Top: visualization by Coomassie Brilliant Blue staining (upper image) and fluorescence (lower image) of 5 mg mL^{-1} IAEDANS-labelled BSA resolved on a 12 % SDS-PAGE gel. Bottom: quantification of fluorescence in lanes 1-10 of SDS-PAGE gel. Lane 1: BSA (in presence of IAEDANS after 0 passes); lanes 2 – 5: BSA stressed for the indicated number of passes in the presence of IAEDANS, lane 6: BSA in the presence of IAEDANS and 0.5 mM TCEP after 0 passes; lanes 7 – 10: BSA stressed for the indicated number of passes in the presence of IAEDANS and 0.5 mM TCEP. Plunger velocity was 8 mm s^{-1} (strain rate = 11750 s^{-1}). Error bars represent the error from two independent experiments.

Fig. 4. Energy distribution in different regions of the extensional flow device. (A) Average energy dissipated within the extensional region (red line) or within the shear region (black line) per pass for a protein with a diameter of 4 nm as a function of plunger velocity. (B) Average rate of energy dissipation within the extensional region (red line) and within the shear region (black line) per representative protein volume as a function of plunger velocity. Data calculated by CFD using parameters described in SI Appendix, Table S1. (C) Percentage of insoluble material of 5 mg mL^{-1} BSA samples stressed for 100 passes at the plunger velocities indicated (strain rates = 3184 - 23421 s^{-1} for plunger velocities of 2 – 16 mm s^{-1}).

Figure 1

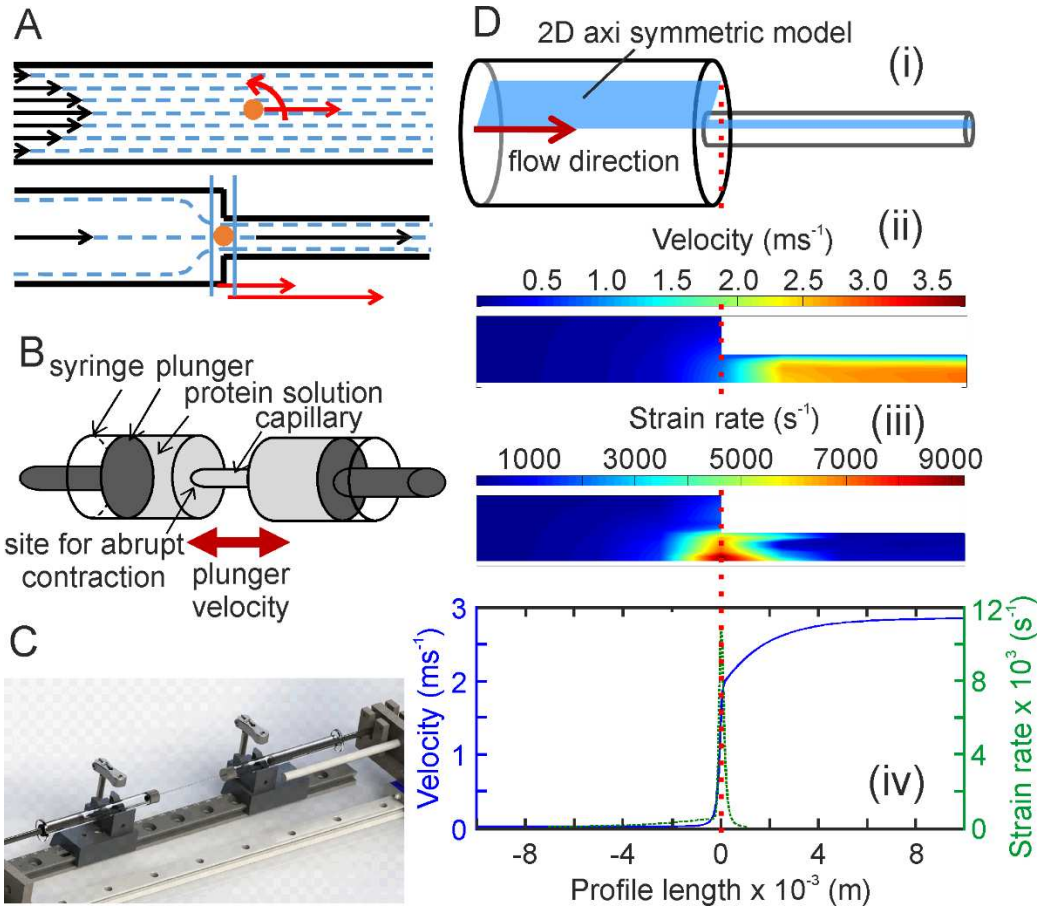

A
B
syringe plunger
protein solution
capillary
site for abrupt contraction
plunger velocity
C
D
2D axi symmetric model
(i)
flow direction
Velocity (ms-1)
(ii)
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Strain rate (s-1)
(iii)
1000 3000 5000 7000 9000
Velocity (ms-1)
Strain rate x 10^3 (s-1)
(iv)
Profile length x 10-3 (m)

Figure 2:

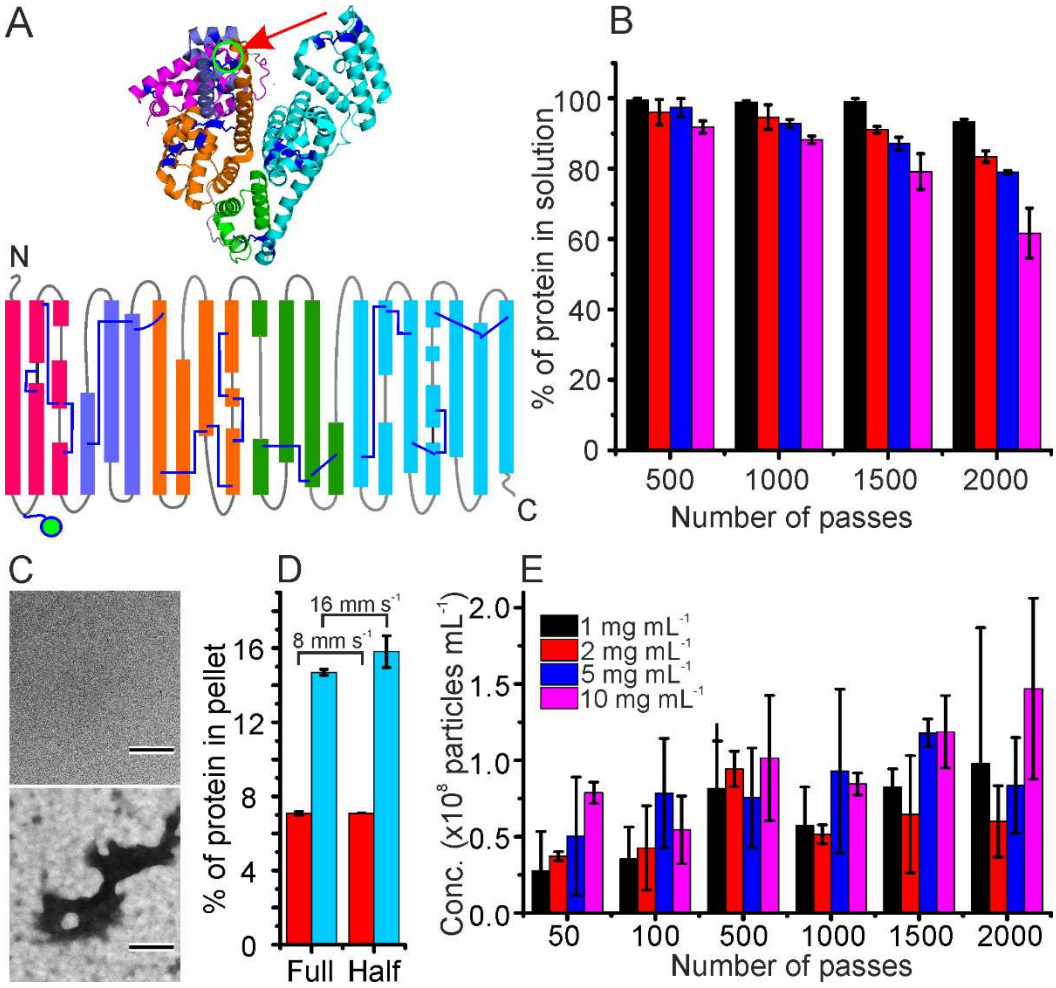

A
N
C
B
% of protein in solution
100
80
60
40
20
0
500
1000
1500
2000
Number of passes
C
D
16 mm s-1
8 mm s-1
% of protein in pellet
16
12
8
4
0
Full
Half
E
1 mg mL-1
2 mg mL-1
5 mg mL-1
10 mg mL-1
Conc. (x10^8 particles mL-1)
2.0
1.5
1.0
0.5
0.0
50
100
500
1000
1500
2000
Number of passes

Figure 3

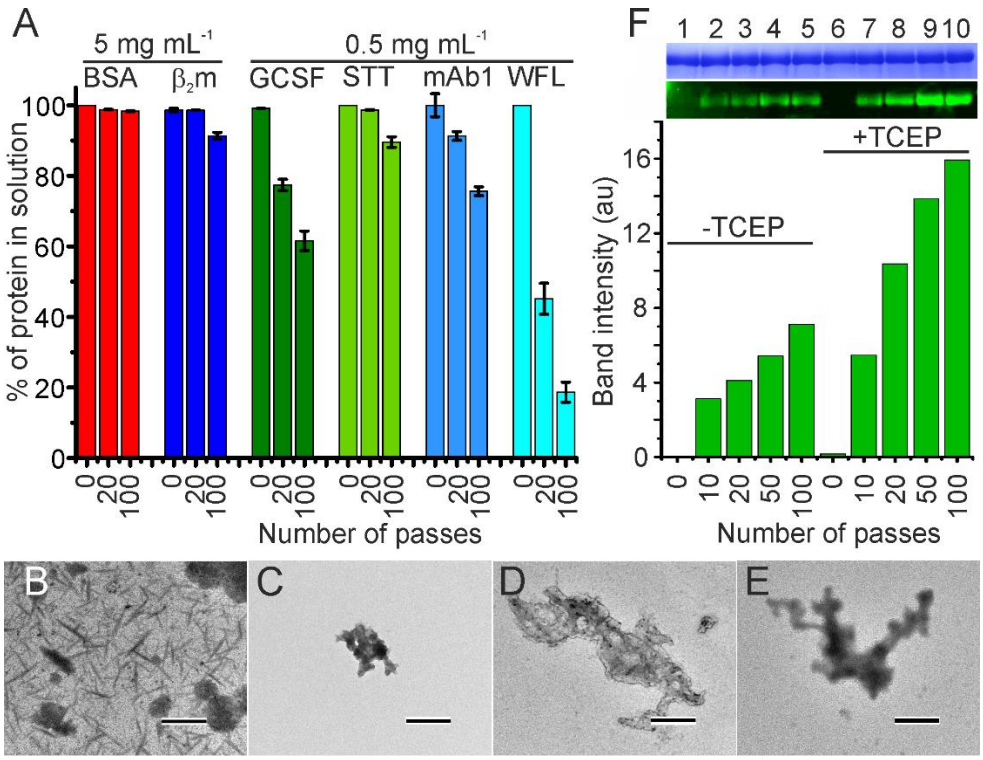
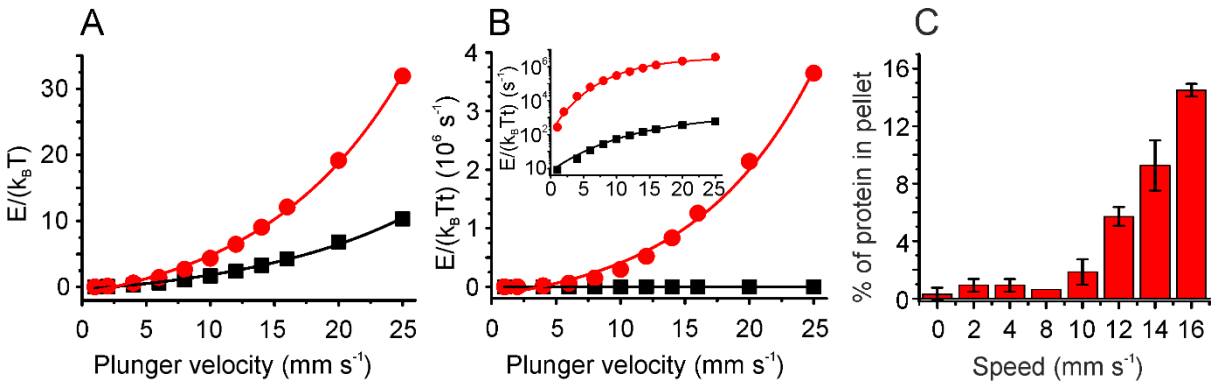

A
5 mg mL-1
0.5 mg mL-1
BSA
β2m
GCSF
STT
mAb1
WFL
% of protein in solution
Number of passes
F
1 2 3 4 5 6 7 8 9 10
+TCEP
-TCEP
Band intensity (au)
Number of passes
B
C
D
E

Figure 4


A
B
C
E/(k_BT)
Plunger velocity (mm s^-1)
E/(k_BTt) (10^6 s^-1)
E/(k_BTt) (s^-1)
% of protein in pellet
Speed (mm s^-1)