



UNIVERSITY OF LEEDS

This is a repository copy of *The hierarchical emergence of worm-like chain behaviour from globular domain polymer chains*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/152139/>

Version: Accepted Version

Article:

Hanson, BS orcid.org/0000-0002-6079-4506, Head, D orcid.org/0000-0003-0216-6787 and Dougan, L orcid.org/0000-0002-2620-5827 (2019) The hierarchical emergence of worm-like chain behaviour from globular domain polymer chains. *Soft Matter*, 15 (43). pp. 8778-8789. ISSN 1744-6848

<https://doi.org/10.1039/C9SM01656B>

© The Royal Society of Chemistry 2019. This is an author produced version of an article published in *Soft Matter*. Uploaded in accordance with the publisher's self-archiving policy.

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.



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

Cite this: DOI: 00.0000/xxxxxxxxxx

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The hierarchical emergence of worm-like chain behaviour from globular domain polymer chains

Benjamin S. Hanson¹, David Head², and Lorna Dougan^{1,3}

Biological organisms make use of hierarchically organised structures to modulate mechanical behaviour across multiple lengthscales, allowing microscopic objects to generate macroscopic effects. Within these structural hierarchies, the resultant physical behaviour of the entire system is determined not only by the intrinsic mechanical properties of constituent subunits, but also by their organisation in three-dimensional space. When these subunits are polyproteins, colloidal chains or other globular domain polymers, the Kratky-Porod model is often assumed for the individual subunits. Hence, it is implicitly asserted that the polymeric object has an intrinsic parameter, the persistence length, that defines its flexibility. However, the persistence lengths extracted from experiment vary, and are often relatively small. Through a series of simulations on polymer chains formed of globular subunits, we show that the persistence length itself is a hierarchical structural property, related not only to the intrinsic mechanical properties of the underlying monomeric subunits, but emerging due to the organisation of inhomogeneous geometry along the polymer contour.

1 Introduction

Many naturally occurring proteins form polymers in order to perform their biological function, with examples such as fibrinogen, actin, tubulin¹ and the large set of titin subdomains². Due to their shape, intrinsic mechanical properties and local physical interactions with one another, each of these monomeric subunits polymerise into structures with very different structural architectures and emergent mechanical properties. Taking the eukaryotic cytoskeletal proteins as an example, G-actin monomers polymerise linearly to form long F-actin microfilaments. These relatively thin filaments, with a diameter between 7nm and 9nm, then go on to form bundles, cellular transport networks, and part of the muscle structure³. On the other hand, tubulin polymerises linearly to form protofilaments but also radially to form microtubules. With a diameter of ~ 25 nm, these systems are much larger than actin filaments whilst also being hollow with respect to protein material³. Microtubules are then the higher order structure which goes on to form cellular networks, cilia, and the core flagellar axonemal structure involved in the beating of sperm tails⁴. Even with these diverse emergent architectures and applications, each of these globular domain polyproteins fall under the

single umbrella term ‘polymer’. In a similar vein, colloidal chains can also be considered as globular domain polymers. These systems are formed of mostly spherical subunits which may be connected by linker domains, local electrostatic interactions or even dispersive interactions such as van der Waals forces⁵. In addition to the intrinsic mechanical properties of the colloids themselves and the localised steric interactions between them, the properties of colloidal chains can also be determined by functionalised electrostatic interactions^{6–8}.

Of current interest in the study of these types of polymeric object is their rational design, namely, how we can design polymeric structures with specific and predictable biophysical properties. These may be the end products in and of themselves, such as elastomeric polyproteins⁹ or renewable biopolymers¹⁰, or they may be required for subsequent use in the rational design of more complex systems, such as biopolymer solutions¹¹ and networks^{12,13}, protein hydrogels^{14,15} and even mimicking complex biological systems such as cilia and flagella⁸. In each of these applications, both experimental and simulation studies have shown that the microscopic mechanical properties of these polymeric objects have a significant effect on the resulting hierarchical properties, whether it is macroscopic elasticity or network connectivity. Using a variety of experimental techniques^{6,7,16–18} we can reliably create globular domain polymeric objects with controlled subunit geometries. To enable the specific rational design of these polymers using these methods, or the hierarchical sys-

¹ School of Physics & Astronomy, University of Leeds, Leeds, UK
E-mail: b.s.hanson@leeds.ac.uk / l.dougan@leeds.ac.uk

² School of Computing, University of Leeds, Leeds, UK

³ Astbury Center for Structural Molecular Biology, University of Leeds, Leeds, UK

† Electronic Supplementary Information (ESI) available: See DOI: 10.5518/709

tems formed from them, we must first understand how the properties of the polymers themselves are dependent on their own underlying structures.

The most commonly applied physical model applied to all types of polymer is the Kratky-Porod model, colloquially known as the worm-like chain (WLC) model¹⁹. The WLC model assumes a continuous contour along the polymeric object, with length L_c , along which there is an energy density proportional to the square of the curvature at each point. The Hamiltonian function for this model, H_{WLC} , is

$$H_{WLC} = \frac{1}{2}k_B T \int_0^{L_c} L_p \left(\frac{\partial^2 \vec{r}}{\partial s^2} \right)^2 ds \quad (1)$$

where s is a scalar parameter specifying a point along the continuous contour of the polymer, \vec{r} is the position vector at s , and $k_B T$ is the Boltzmann energy factor. Eq.1 results in the emergence of a bending resistance throughout the polymer, characterised by an exponential decay in correlation between the tangent vectors at any two points along the contour

$$\langle \vec{r}(s) \cdot \vec{r}(s') \rangle = \exp\left(-\frac{|s' - s|}{L_p}\right) \quad (2)$$

where $\vec{r}(s)$ is the tangent unit vector at s . We now see that the value L_p from Eq.1 appears as the decay constant associated with this exponential decorrelation. Hence, L_p is known as the persistence length. At constant temperature, which we will assume throughout this article, the WLC model assumes L_p to be an intrinsic property of a given polymer. Indeed, there are many expansions to the WLC²⁰ accounting for chain contour extension, additional interactions, externally applied forces and others, and within each of these models the persistence length remains an intrinsic property of the polymer under consideration. In the context of globular domain polymers, however, this may not be the case.

We can generalise the physical structures of polyproteins and colloidal chains, viewing them both as specific type of polymer with an inhomogeneity in the aspect ratio along the contour i.e. alternation between thin domains (linker domains) and thick domains (globular subunits). From a theoretical perspective, this inhomogeneity leads to a lack of clarity in how we should define the contour of the object and hence, what contour length the persistence length relates to. Similarly from an empirical perspective, in the case of experimental extraction of these L_c and L_p parameters, it is again unclear what physical structure the contour length refers to, if any, and hence what contour the persistence length is relative to. This dependence of L_p on L_c was implicitly noted by Kellermeyer in a study of the biopolymer titin²¹, who noted the ‘apparent’ persistence length changed as certain titin subdomains unfolded under experimental conditions. Further, Fu *et al.* introduced the idea of a ‘natural length’ for a polymer²², termed L_n , representing the end-to-end distance of the polymer at zero external force. Yet if the lengths of the polymeric structure change as a function of applied force, this leaves the contour length L_c acting as a fitting parameter with a somewhat loose physical interpretation, rather than explicitly defining the polymeric structure as we might assume. More explicit examples of

the disjunction between pure polymer physics and its application to globular domain polymers can be seen in some of the collective research on titin.

Titin is a large, cable-like structure resulting from the polymerisation of 244 individual globular domains and is responsible for muscle elasticity²³. Two of these domains, the I27 and PEVK domains, have been extensively studied and exhibit mechanical behaviour relevant to this work. PEVK is known to be relatively flexible, unfolding so easily that it was defined by Linke *et al.* as an ‘entropic spring’²⁴. On the other hand, I27 is significantly more thermodynamically stable²⁵, and hence more rigid whilst in the folded state.

The elasticity of the PEVK domain of titin was investigated by Sarkar *et al.* via a single molecule force spectroscopy atomic force microscopy (SMFS) study on titin co-polymers, also termed hetero-polyproteins, formed of alternating subdomains of I27 and PEVK in the form $[[PEVK]_3I27]_4$ ²⁶. The force-extension relationship of this process results in a characteristic unfolding curve that corresponds to an analytic expression derived from the WLC model²⁰, albeit an approximate relation, and from this Sarkar *et al.* were able to extract a value for L_p . They measured the persistence lengths of three different $[[PEVK]_3I27]_4$ polyproteins, each corresponding to a PEVK domain formed from the expression of a different exon of the human titan gene, exons 120, 161 and 184. These persistence lengths were measured to be $0.89 \pm 0.42\text{nm}$, $0.92 \pm 0.38\text{nm}$ and $0.98 \pm 0.4\text{nm}$ respectively, which are on the order of 2 amino acid lengths²⁷. Carrion-Vazquez *et al.* also used SMFS to measure the persistence length of $[I27]_8$ polyproteins. Although there were linker domains present between the I27 domains, there were no flexible PEVK domains²⁸. They measured the persistence length of this system as $0.39 \pm 0.07\text{nm}$, even less than Sarkar *et al.* measured the unfolded PEVK domain to be.

On the other hand, Li *et al.* made measurements using single molecule electron microscopy²⁹, which differs from SMFS in that it does not apply high tensile strain or spatial boundary conditions to the polymer. By extracting an end-to-end distribution from the electron micrographs, a distribution which also has an analytic form derived from the WLC model, they were able to measure the persistence length of $[I27]_{12}$ as $9.8 \pm 0.6\text{nm}$, significantly greater than the disordered PEVK regions and, more importantly, greater than the L_p value of the $[I27]_8$ polyprotein measured using SMFS. This suggests that the persistence lengths of polyproteins measured by SMFS differ from those measured in the absence of applied force.

Finally we look at the work of Huber *et al.*, who performed a study of the PEVK domain in isolation using Förster resonance energy transfer (FRET)³⁰, a method that can also be used to measure an end-to-end distance distribution, in this case of a single PEVK domain. Using two synthetic PEVK sequences, one with 11 residues and another with 21, they were able to extract persistence lengths of $0.63 \pm 0.01\text{nm}$ and $0.48 \pm 0.02\text{nm}$ respectively. Within error, these correspond to the measurements of Sarkar *et al.*

From this series of experiments we see that the persistence length of a polyprotein formed of alternating PEVK/I27 domains is relatively small²⁶ and approximately equivalent to that

of an independent PEVK domain³⁰, indicating that the rigid I27 monomers do not contribute significantly to the persistence length in these cases. However, a polyprotein formed of only I27 domains measured using electron microscopy (at zero applied tensile force) has a persistence length approximately 25 times greater than a polyprotein formed of the same monomeric subunits under SMFS conditions^{28,29}. This indicates that applied force somehow affects the measured persistence length, even though additions to the WLC model Hamiltonian attempt to account for this³¹. These observations indicate some discrepancies between idealised polymer physics and applications of polymer physics to globular domain polymers, and the difference of potentially orders of magnitude between the apparent and true persistence lengths.

For general SMFS experiments on polyproteins, it is often the case that the WLC model is used simply as a fitting model to validate the measurement of single protein properties, such as the peak forces and the peak-to-peak distances of the SMFS unfolding curves. Nevertheless, many SMFS experiments observe persistence lengths of around the length of a single amino acid: in titin^{27,28,32,33}, general polyproteins^{34–36}, and other structural motifs^{37–40}, and this magnitude of persistence length corresponds well to that of individual, flexible amino acid chains^{41,42}. Yet we have seen from the work of Li *et al.* that in the absence of applied tensile force, the persistence length can be significantly larger. Biophysical systems which incorporate these types of polymeric object as a structural subunit are currently being investigated^{14,15,43}, and within these systems it is likely that the subunits are under conditions similar to SMFS experiments. In addition, current theoretical models for the viscoelastic properties of homogeneous semiflexible polymers in solution have recently been shown to be insufficient^{11,12}. Schuldt *et al.* observed that the resultant shear modulus of a polymer network solution, formed of tunable DNA nanotubes, has a non-trivial dependence on both the persistence length and polymer concentration that contradicts current theoretical frameworks¹². Tassieri then argues that the current theoretical understanding of the persistence length with respect to emergent hierarchical properties is insufficient for a variety of potential systems¹¹. Thus, as we attempt to rationally design these hierarchical systems, it is important to revisit the WLC model itself so we can employ previously extracted WLC parameters appropriately to make new predictions about their resultant mechanical properties.

With these ideas in mind, we have performed a series of simulations of single globular domain polymers which aim to investigate how the underlying subunit geometries, mechanical properties and externally applied forces affect the persistence length of the polymeric object (see Figure 1). Via simulations that mimic the electron microscopy experiments of Li *et al.*²⁹, we study the emergence of persistence length in the absence of any external forces other than the statistical effects of temperature. With further simulations designed to mimic SMFS spectroscopy experiments, we investigate how the contour and persistence lengths implicitly measured from force-extension curves relate to the previous explicit calculations in the absence of force.

2 Methods & Simulations

2.1 Globular Domain Polymer Simulations

Standard polymer simulation techniques cannot be used to gain insight into how the persistence length of a globular domain polymer emerges as a function of the underlying subunit structure. These techniques often treat polymeric properties as defined parameters rather than emergent, measurable properties as we require. Instead, we have used BioNet, a new software package in active development designed to simulate biological network structures. Similarly to LAAMPS⁴⁴, BioNet allows us to design biophysical systems from various structural building blocks and perform a variety of simulation protocols on them. To best represent the polymers studied in this work, our models consisted of 10 spherical beads connected explicitly at the surface by linear springs as shown in Figure 3b.

The spheres in our model represent the globular subunits. They are geometrically rigid, but are associated with a soft core pair-pair steric repulsion of the form

$$U_{st} = \begin{cases} \frac{1}{2}k_{st} \left(\left(\frac{V_o}{A_c} \right)^2 + (r - 2R)^2 \right) & r < 2R \\ 0 & r \geq 2R \end{cases} \quad (3)$$

where r is the center to center distance between two spheres, R is the radius of each sphere, V_o is the volumetric overlap between the two spheres⁴⁵, A_c is the cross-sectional area of each sphere, and k_{st} is an associated elastic constant. This interaction has been tested and shown to be fully conservative, and, because it is a volumetric interaction with $U_{st} \propto r^6$, this potential can be said to represent the intrinsic stiffness of the globular subunits. Although alternative interaction potentials are available, such as Weeks-Chandler-Anderson⁴⁶ or Hertzian⁴⁷ potentials, our interaction potential has been written in this form so that the stiffness k_{st} is directly comparable with the linear elastic stiffnesses of the neighbouring linker domains, whilst leaving the functional form proportional to the volumetric overlap. For completeness however, we have performed a subset of equivalent simulations using these different potential forms, the results of which are discussed in Supplementary Section 5.2.

To aid in simulation efficiency, our linker domains are represented by simple Hookean springs. As such they can be thought of as end-to-end distance fluctuations of a Gaussian chain, representing the underlying linker domain, or simply as a linearisation of the local forcefield. If required, the combination of equilibrium length and stiffness in this domain can be converted into a contour length and a Kuhn length and hence, a persistence length. Increases in linear stiffness whilst keeping the equilibrium length constant can be loosely thought of as an increase of the persistence length for the underlying linker domain.

A final point of importance is that our beads are not simply points with ‘effective radii’. The linkers are connected to the beads explicitly at their surface, giving our beads real three dimensional structure. Due to this explicit surface connectivity, on each sphere the axes of rotation for each connected spring are different and hence an effective bending energy will emerge due to steric interactions. This is the core feature of our model from

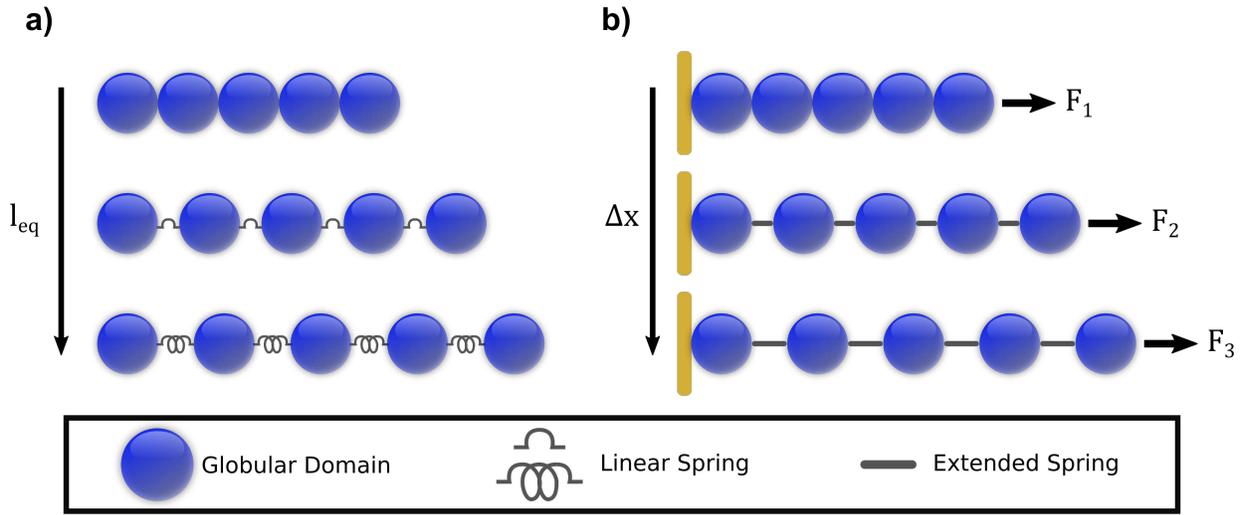


Fig. 1 A schematic overview of the systems we are considering in this work. a) A series of globular domain polymers with an increasing linker domain length, l_{eq} , between the subunits and b) a globular domain polymer under SMFS conditions. Increasing the applied force, F , increases the overall extension in the polymer, Δx , and hence can be used to force a globular domain polyprotein into the same geometric configuration as the polyproteins in a).

which we expect to see the emergence of persistence lengths on the order of the length of the polymer itself (see Figure 2). For completeness, rotational degrees of freedom were modelled for each sphere, with dynamics calibrated to correspond with the theoretically expected rotational diffusion.

Once each of our polymer models had been built, we ran simulations of each single polymer model using a Brownian dynamics protocol. Each simulation ran with a simulation timestep a factor of 10 lower than the smallest relaxation time within the system and for a simulation time a factor of 100 longer than the theoretical Rouse time for the equivalent Gaussian chain¹⁹, an approximation of the longest relaxation time within these systems. This ensures the simulation trajectories have both numerical accuracy and dynamical convergence. The timescales were determined by a background viscous medium with viscosity $\mu = 1\text{MPa}\cdot\text{ns}$, similar to water⁴⁸. This provided a localised Stokes drag on each sphere rather than a full hydrodynamic effect. Specific simulation protocols for each investigation will be described in Section 3.

2.2 Molecular Dynamics Simulations

To calibrate the springs within our coarse-grained model, we performed our own study into the persistence length of linker domains used in polyproteins via molecular dynamics simulations on a set of theoretically disordered amino acid chains⁴⁹. Our sequences were built into physical chains using the ‘Build Structure’ utility within UCSF Chimera⁵⁰. Each structure was assembled using the Dunbrack rotamer libraries⁵¹ and initialised into helical structures for simplicity.

Subsequent dynamical equilibrations and simulations were performed using the Gromacs 2018 and 2019 Molecular Dynamics packages⁵². Each single amino acid chain was immersed in explicit water and simulated using the ‘AMBER99SB-ILDN’ force-field. Each simulation completed a different amount of real time dynamics, ranging from $2\mu\text{s}$ and $7\mu\text{s}$.

2.3 Theoretical Background

To aid us in our analysis we will define some useful quantities in advance. For a polymer chain in the absence of any steric forces or self-interaction, a simple model to use is the freely-jointed chain model, a specific instance of the more general Gaussian chain model. Our structure, as shown in Figure 3b, is a slight extension of the standard freely-jointed chain. In addition to the freely rotating linker domains we have an additional rigid globular component of radius R with explicit surface binding sites defined, and the linkers are Hookean springs with associated stiffness k whose length can thermally fluctuate about an equilibrium length l_{eq} . If we define the end-to-end distance, E , as the distance between the centers of the two terminal beads, the expected squared end-to-end distance for a system of N beads can be analytically calculated from an analysis of the end-to-end vector, \vec{E} . The details involved in this derivation can be found in Supplementary Section 5.3. Here, we simply quote the resulting equation

$$\langle E^2 \rangle = (N-1) \left(\frac{k_B T}{k} + l_{eq}^2 + 2 \left(2 - \frac{1}{N-1} \right) R^2 \right). \quad (4)$$

We can see that in this model, the contribution of the radius R to the overall end-to-end distance is dependent upon the length of the chain itself. However, as N increases this effect quickly becomes negligible and we can define an effective segment length l'_{eq} as

$$l'_{eq} = \sqrt{l_{eq}^2 + (2R)^2}. \quad (5)$$

It is this length about which the chain as a whole undergoes thermal fluctuations that increase its expected end-to-end distance. In the linear elastic linker domains these thermal fluctuations will have a characteristic magnitude $l_{rms} = \sqrt{k_B T/k}$, which is the root mean square value as calculated from the equipartition theorem (see Supplementary Section 5.3). From here we can define a di-

dimensionless variable, \hat{l}'_{rms} , which we will refer to as the relative fluctuation magnitude

$$\begin{aligned}\hat{l}'_{rms} &= \frac{l_{rms}}{l'_{eq}} \\ &= \frac{\frac{l_{rms}}{2R}}{\sqrt{\left(\frac{l_{eq}}{2R}\right)^2 + 1}}.\end{aligned}\quad (6)$$

This parameter measures how large the thermal fluctuations are relative to the equilibrium length of the effective segment defined in Eq.5. As we may expect, \hat{l}'_{rms} can be expressed in terms of the specific relative geometries within the system, which are

$$\hat{l}_{eq} = \frac{l_{eq}}{2R} \quad (7)$$

$$\hat{l}'_{rms} = \frac{l_{rms}}{2R} \quad (8)$$

where \hat{l}_{eq} and \hat{l}'_{rms} are, respectively, the sizes of the equilibrium length and fluctuation magnitudes of the linker domain relative to the diameter of the globular domain. As Eq.6 encompasses both Eq.7 and Eq.8, it follows that \hat{l}'_{rms} is an appropriate dimensionless quantity to measure the relative size of the thermal fluctuations with respect to the length of the structural subunits within the chain. We now consider how these parameters can affect the steric interaction.

In our model Eq.4 is the limit as the local steric interaction strength $k_{st} \rightarrow 0$. For the non-zero case we can imagine the situation as shown in Figure 2, where θ_M is the maximum possible angle before steric interactions begin to occur. Under the approximation that the fluctuating angle θ is symmetric for neighbouring beads, then it can be shown that

$$\cos \theta_M = 1 - (\hat{l}_{eq} + \hat{l}'_{rms}). \quad (9)$$

Therefore, because the relative rotational motion is inhibited by the steric interaction which is itself dependent on the relative geometries in the system, we would expect a persistence length to arise within the polymer as a function of those same relative geometries, in addition to the intrinsic stiffnesses of the globular and linker domains.

From our discussion here, we conclude that the natural dimensionless parameters to use in the study of the emergence of persistence length in globular domain polymers are \hat{l}_{eq} and \hat{l}'_{rms} . Varying \hat{l}_{eq} will allow us to alter the relative equilibrium sizes of the globular and linker domains, and varying \hat{l}'_{rms} will allow us to alter the size of the fluctuations in the linker domain relative to the effective fluctuating segment of the chain defined by Eq.5. For the remainder of this work, keep in mind that our notation is such that \hat{l}_{eq} and \hat{l}'_{rms} are normalised by $2R$, whereas \hat{l}'_{rms} is normalised by l'_{eq} i.e. the inverted comma denotes a different type of normalisation.

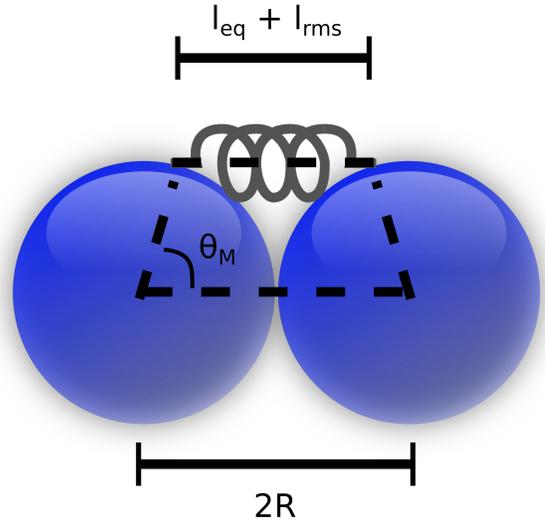


Fig. 2 The emergence of bending rigidity due to local steric interactions. The angle θ_M is a geometric property dependent upon the relative sizes of the bead radius, R , and the sum of the equilibrium length, l_{eq} and fluctuation magnitude, l_{rms} , of the associated linker domain. When combined with a non-zero stiffness in the globular domain, k_{st} , the result will be an effective bending potential.

3 Results

3.1 The persistence length of amino acid chains is comparable to the size of single amino acids, and small compared to globular proteins

To investigate the intrinsic stiffnesses we might expect from our linker domains, we performed molecular dynamics simulations of theoretically disordered amino acid chains, containing small side chains and zero net charge, which are commonly used as linker domains⁴⁹. From these simulations we were able to calculate persistence lengths using a discretised form of Eq.2, which can be approximated by defining the contour as a piecewise function defined from a set of discrete points s_i , and hence can be rewritten as

$$\langle \vec{r}(s_i) \cdot \vec{r}(s_j) \rangle = \exp\left(-\frac{|s_j - s_i|}{L_p}\right) \quad (10)$$

where the set of values s_i now explicitly define the contour. An obvious choice for these discrete points in our amino acid chains is the set of C_α backbone atoms, but it must be emphasised for the purpose of this work that this is by no means the only choice, and the persistence lengths calculated will reflect this choice. Defining each segment vector $\vec{l}_i = \vec{r}(s_{i+1}) - \vec{r}(s_i)$, we subsequently define the tangent vector $\vec{t}(s_i)$ as

$$\vec{t}(s_i) = \frac{\vec{l}_i}{|\vec{l}_i|} \quad (11)$$

or, in words, $\vec{t}(s_i)$ is the normalised difference in position vectors between neighbouring C_α atoms i and $i+1$. As the length of our chain, and each segment, may vary over time due to thermal fluctuations, we chose to define each our overall contour length of the chain, L_c , as

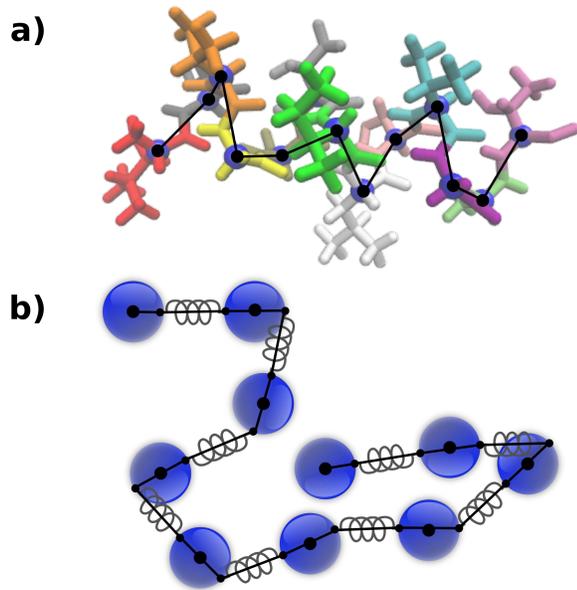


Fig. 3 The two types of model studied in this work, with the defined contours superimposed. **a)** The contour as defined for an amino acid chain, piecewise between the C_α atoms and **b)** The contour defined for our globular domain polymers, piecewise between the effective linker connections to each globular subunit.

$$L_c = \sum_{i=1}^{N-1} \langle |\vec{l}_i| \rangle. \quad (12)$$

In words, our contour length is defined as the average contour length, formed as a summation of the average segment lengths and where the averages are taken over the full set of simulation frames. This is shown in Figure 3a.

Here it is important to note that each different length of contour, if thought of as an inextensible worm-like chain, could in principle be assigned its own ‘intrinsic’ persistence length. What we measure experimentally and in this simulation environment is the average effect of each of these systems as the chain entropically transitions between different contour lengths under thermal fluctuations. Nevertheless, using these properties we were able to fit Eq.2 to the simulation trajectory as shown in Figure 4, and calculate the persistence lengths of amino acid chains as shown in Table 1.

Model	L_p (nm)	L_p (Amino acids)
G[P] ₅ C	0.60 ± 0.02	1.80 ± 0.07
[LSVGATI] ₂	0.54 ± 0.06	1.49 ± 0.17
[GSS] ₄	0.37 ± 0.04	1.05 ± 0.12

Table 1 The calculated persistence lengths for a set of amino acid chains, shown in SI units and in terms of the number of amino acid lengths. Both columns are therefore equivalent.

We can immediately see that whilst dynamics of the [LSVGATI]₂ and [GSS]₄ chains correspond relatively well to the WLC model, the G[P]₅C chain dynamics do not. For G[P]₅C, a relatively strong correlation exists around the center of the chain indicating some

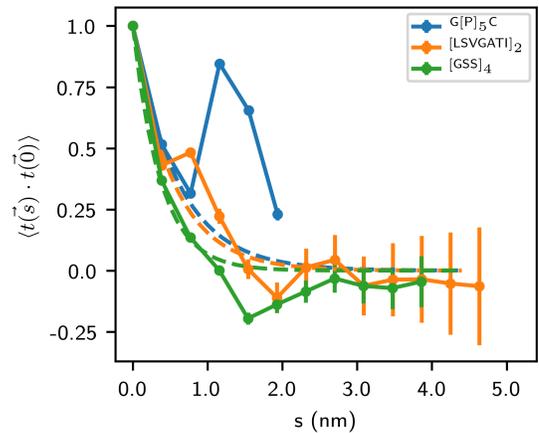


Fig. 4 The correlations between tangent vectors along the contour of 3 different amino acid chains.

form of self-interaction. The source of this interaction is out of scope for this work, but we have included this simulation for completeness. We are still able to fit the Eq.2 to the first 3 C_α atoms of the G[P]₅C chain, although the specificity of this fit is questionable. Nevertheless, this is the value quoted in Table 1.

Each of the values of L_p shown in Table 1 are relatively small compared to the size of the globular domains of polyproteins, the folded proteins themselves. By calculating the average size of an amino acid in a similar manner to Eq.12 we are also able to show that these persistence lengths are on the order of just 1-2 amino acids lengths as shown in the second column. This implies that disordered linker domains used within polyproteins are likely the least rigid components in terms of bending. Additionally, we find that these values of L_p are similar in magnitude to each of the experiments discussed in Section 1, with the exception of the electron microscopy experiments of Li *et al.* This indicates that the PEVK domain of titin behaves similarly to an unfolded protein, but perhaps more importantly, it indicates that the SMFS experiments of [I27]₈ by Carrion-Vazquez *et al.* are likely measuring the persistence lengths of the linker domains connecting the I27 domains, and not the persistence length of the polyprotein as a whole²⁸. Although mentioned implicitly throughout the literature, we emphasise that this result has strong implications for the mechanical characterisation of polyproteins and other globular domain polymers. The following sections will investigate the persistence length of these globular domain polymers in the absence of force.

3.2 The persistence length of a globular domain polymer is dependent on the relative sizes of the linker domain and the globular subunits

Now we have seen the order of magnitude of persistence length that can emerge from disordered amino acid chains, we investigate the how these chains affect the persistence length of a general globular domain polymer when embedded within them as linker domains. As stated in Section 2.1, our polymers are repre-

sented as soft core globular domains connected explicitly at the surface by linker domains (see Figure 3). We first investigate the effect of the relative geometries within the system on the persistence length by performing a series of dynamic simulations with varying values of \hat{l}_{eq} and \hat{l}'_{rms} , together with a constant value of k_{st} . From the resulting trajectories we calculate the persistence lengths using the same explicit method as in Section 3.1, using Eq.10, but we now define our contour as shown in Figure 3b, with associated an contour length

$$L_c = \sum_{i=1}^{N-1} 2R + \langle |\vec{l}_i| \rangle. \quad (13)$$

Our new definition of L_c for these systems now accounts for the rigid globular subunits. However, the structural inhomogeneity along the defined contour leads to a non-trivial decay in correlation along the contour. Our method for extracting unique persistence lengths from these correlation curves in accordance with Eq.10 is addressed in Supplementary Section 5.4. Figure 5 shows the persistence lengths extracted from these simulations.

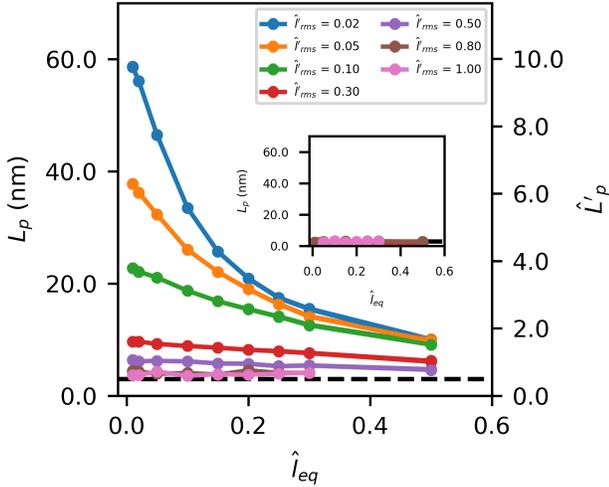


Fig. 5 The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of relative fluctuation magnitude values, \hat{l}'_{rms} , and in the presence and absence of the steric interaction, as represented by the globular domain stiffness k_{st} . *Main* $k_{st} = 192.31$ pN.nm and *Inset* $k_{st} = 0$ pN.nm

From Figure 5 we can see that as the relative linker size \hat{l}_{eq} increases, there appears to be convergence in the persistence length L_p towards an asymptotic value for each value of the relative fluctuation magnitude \hat{l}'_{rms} . We will denote this asymptote as $L_{p,\infty}$. Given our definition of the contour, which begins at the centre of the first subunit, our hypothesis is that as $\hat{l}_{eq} \rightarrow \infty$, $L_p \sim R$. This theoretical asymptote is shown in Figure 5 as a black dashed line. As the globular domains get further apart from one another due to longer linker domains, local steric interactions will occur less often. With only a freely jointed connection between the linker domains and the globular subunits in our model, a reduction in local steric interactions will subsequently lead to the reduction in correlation between the tangents of neighbouring contour seg-

ments until, at infinite separation, there is no correlation at all except within each of the discrete contour segments. For our definition of the contour, this discrete change in correlation will occur at the surface of the first globular subunit, a distance R from the beginning of the contour. This is validated by the inset graph of Figure 5, where because the steric interaction has effectively been switched off, there is no source of correlation at any value of \hat{l}_{eq} and so we observe the same, small persistence length regardless of the polymeric geometry.

If we instead consider reductions in the relative linker size \hat{l}_{eq} in Figure 5, we observe an increase in L_p for most of the smaller values of \hat{l}'_{rms} . Due to the increase in the rate of steric interactions as the globular subunits get closer to one another, there is an emergence of an effective bending resistance along the defined contour as per Eq.9, leading to correlations of the tangent vectors between contour segments. Reducing \hat{l}_{eq} , then, although having no effect on the intrinsic stiffness of the linker domain, leads to a larger bending resistance via an increase in the rate of local steric interactions. Again, in the inset graph of Figure 5, the complete lack of steric interaction effectively switches of this bending resistance entirely by reducing the intrinsic stiffness of the globular domain to zero. Thus we have shown that an increase in the average separation between the globular domains in globular domain polymers decreases the persistence length of the polymer. Figure 11 in Supplementary Section 5.5 shows an intermediate case between the two graphs in Figure 5, showing the same convergence behaviour but with smaller values of L_p at small \hat{l}_{eq} .

We will consider the exact effect of the steric interaction in Section 3.4, but first we notice that Figures 5 and 11 clearly show that the increase in L_p with a reduction in \hat{l}_{eq} is much more pronounced for smaller values of \hat{l}'_{rms} . Hence, L_p must also be dependent upon the intrinsic flexibility of the linker domain.

3.3 The persistence length of a globular domain polymer is affected by the flexibility of the linker domains

From Eq.6, \hat{l}'_{rms} is a dimensionless value representing the flexibility of the linker domain. As the magnitude of the thermal fluctuations increases, the effective length of the domain, and hence the entire chain, increases as per Eq.4.

Figure 5 shows that as the relative fluctuation magnitude \hat{l}'_{rms} decreases, the persistence length L_p increases for each value of the relative linker size \hat{l}_{eq} . This indicates that as the thermal fluctuations comprising \hat{l}'_{rms} decrease, there is a reduction in the effective length of the linker domain, promoting a higher rate of steric interactions as per Eq.9.

An additional point of note is that although in the absence of steric interactions the thermal fluctuations will be symmetric, such that the average length of the linker domain $\langle l_i \rangle = l_{eq}$, the inclusion of steric interactions will tend to bias the fluctuations towards increasing the overall contour length for smaller values of \hat{l}_{eq} . Hence, the reduction in \hat{l}'_{rms} generates a greater increase in the persistence length for smaller values of \hat{l}_{eq} . We can conclude that decreasing \hat{l}'_{rms} corresponds to an increase in L_p not only because the linker is intrinsically stiffer, but because the reduction in magnitude of thermal fluctuations leads to a reduction in the

contour length and hence promotes a higher rate of local steric interactions. Thus we have shown that in addition to increases in the average linker length reducing the overall persistence length, increases in the fluctuations about this average also reduces the persistence length.

From the differences between the traces at each value of \hat{l}'_{rms} in Figures 5 and 11, we can see that at all values of \hat{l}_{eq} , the magnitude of the increase in L_p due to a reduction in \hat{l}'_{rms} is itself continuously dependent upon the stiffness of the globular domain k_{st} .

3.4 The persistence length of a globular domain polymer is dependent on the stiffness of the globular domain

We have seen that an increase in the rate of local steric interactions between globular subunits causes the emergence of an effective bending resistance along the polymer contour, leading to larger persistence lengths. The strength of these interactions, then, should further modulate this effect. This can be seen in Figure 6, where we have performed simulations with varying values of \hat{l}_{eq} and k_{st} with a constant value of \hat{l}'_{rms}

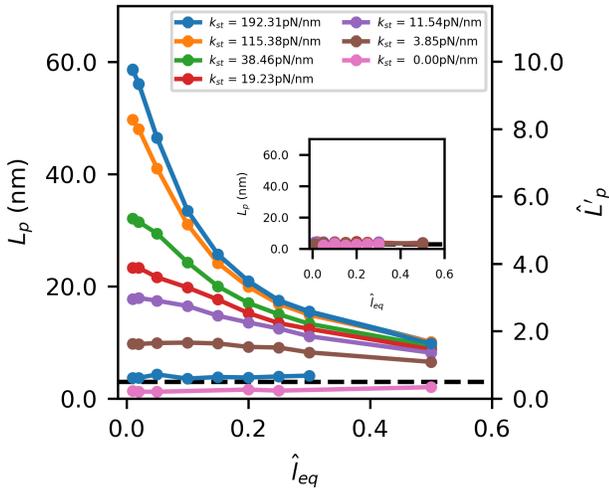


Fig. 6 The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of globular domain stiffness values, k_{st} , and at two different values of the relative fluctuation magnitude, \hat{l}'_{rms} . Main $\hat{l}'_{rms} = 0.02$ and Inset $\hat{l}'_{rms} = 1.0$

With a similar form to Figure 5, Figure 6 shows that as we increase the globular domain stiffness k_{st} , we observe an increase in the persistence length L_p . This effect is significantly more pronounced at smaller values of the relative linker size \hat{l}_{eq} . However, we now see from the inset graph in Figure 6 that for relatively large values of the relative fluctuation magnitude \hat{l}'_{rms} , no additional bending resistance emerges for any values of \hat{l}_{eq} or k_{st} . This indicates that the linkers are so flexible that the effective increase in contour length prevents steric interactions from occurring altogether. This is not equivalent to setting $k_{st} = 0$, however, but does comprise a reduction to the freely-jointed chain model as given by Eq.4 or, had we included explicit amino acid chains, a reduction to a homogeneous worm-like chain model. Thus we have shown

that if steric interactions can occur between the globular domains within the polymer, then a reduction in the strength of the steric interaction will reduce the overall persistence length. Figure 12 in Supplementary Section 5.5 shows an intermediate case between the two graphs in Figure 6.

3.5 The persistence length of a globular domain polymer measured under SMFS conditions is not equivalent to that of the free polymer

In previous sections we have been able to explicitly calculate persistence lengths from the definition given in Eq.10, using the ‘true’ contour of the system defined in advance. These native conditions are similar to the electron microscopy experiments of Li *et al.* and so can result in similarly high persistence lengths such as those measured experimentally for the [127]₁₂ polyprotein²⁹. SMFS, however, changes the relative geometries of the globular domain polymer through the action of an applied force, which imparts an overall extension to the polymer.

To investigate how this process affects the values of both L_p and L_c we performed an additional series of globular domain polymer simulations with constraints representing the experimental conditions of SMFS force spectroscopy. To accomplish this, we first aligned the polymer along the x -axis. During each simulation a constant force \vec{F} was applied in the positive x direction to the final node in the chain to simulate the action of the AFM tip, whilst all forces on the base node were set to zero to simulate surface adhesion.

To calculate the persistence length in each of our simulations, we used the equation from the original work of Marko and Siggia³¹,

$$F = \frac{k_B T}{L_p} \left(\frac{1}{4} \left(\left(1 - \frac{\Delta x}{L_c} \right)^{-2} - 1 \right) + \frac{\Delta x}{L_c} \right) \quad (14)$$

where F is the applied force, and Δx the extension in the x direction only, as would be measured by the changing height of the AFM tip. We note that for an ideal polymer i.e. one in which there is a continuous aspect ratio along the contour, we would expect that the value of L_p extracted from Eq.14 would be approximately equal to that calculated from Eq.2. However, as we are studying globular domain polymers, we cannot assume that these values will be the same due to the constantly changing geometry ratios. Therefore, we make the distinction that the values of L_p measured in the remainder of this section are apparent persistence lengths, and will be strongly dependent upon the specific conditions of applied force.

As with earlier sections, we will rewrite Eq.14 with dimensionless parameters for a more general analysis. In the absence of thermal fluctuations, the application of a force to our system would simply result in $F = k\Delta l$ for each linker domain, where $\Delta l = l - l_{eq}$. Thus, we can define our force in terms of the length-scale Δl . Additional normalisation by $2R$ allows us to rewrite Eq.14 as

$$\Delta \hat{l} = \hat{l}'_{rms} \frac{l'_{rms}}{L_p} \left(\frac{1}{4} \left(\left(1 - \frac{\Delta x}{L_c} \right)^{-2} - 1 \right) + \frac{\Delta x}{L_c} \right) \quad (15)$$

where \hat{l}_{rms} and l_{rms} are as defined in Section 2.3 and

$$\begin{aligned}\Delta\hat{l} &= \frac{\Delta l}{2R} \\ &= \frac{F}{2Rk}\end{aligned}$$

We performed a series of single polymer simulations using a value of $\hat{l}_{eq} = 0.01$ only, and for a range of values of \hat{l}'_{rms} and k_{st} . Each simulation applied a different value of $\Delta\hat{l}$ and from each we calculated the extension as

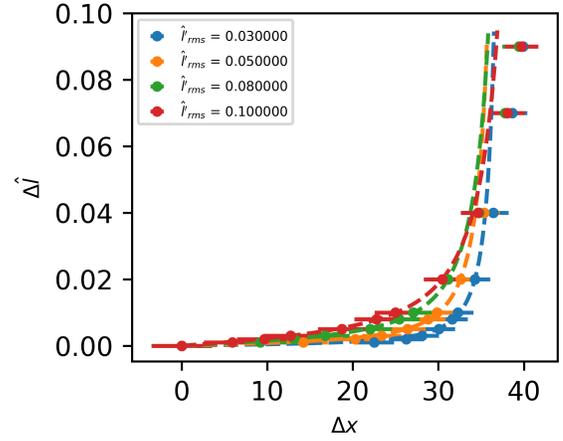
$$\Delta x = \langle x_N - x_1 \rangle \quad (16)$$

where x_1 and x_N are the x coordinates of the initial and final nodes respectively, and the average is again taken over the set of simulation frames. Hence, for each value of \hat{l}'_{rms} and k_{st} , we were able to extract a force-extension curve that is equivalent to a single peak of a characteristic force-extension curve from SMFS experiments using AFM. Examples of the application of Eq.15 to these curves are shown in Figure 7, where both L_p and L_c were left as variable fitting parameters.

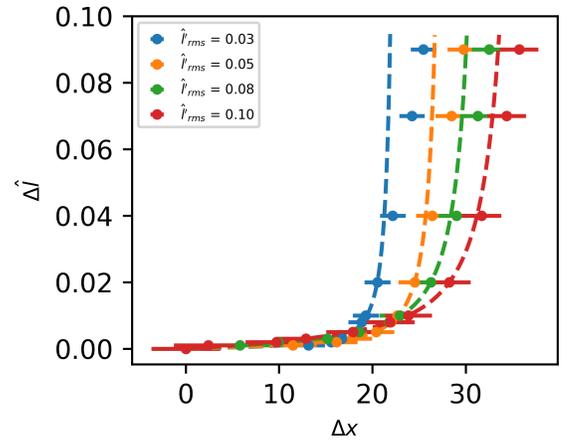
Just as in standard SMFS applications of the WLC model, Eq.15 does not fit well at large applied forces. Due to our parameter normalisation, we were able to standardise our fits of Eq.15 to the first 8 data points in each curve as shown in Figure 7, which gave the best fit in terms of minimising the fitting parameter covariances whilst deviating from the data at high strain. The effect of both k_{st} and \hat{l}'_{rms} on the effective contour and persistence lengths measured from the force-extension curves are made clearer in Figure 8, where we plot the extracted values of L_c and L_p as functions of these mechanical parameters.

The first thing to recognise from Figure 8a is that neither the values of L_c or L_p are the same as those defined or measured for the equivalent free polymers (see Sections 3.2 and 3.4). This implies that the persistence length of globular domain polymers is in some way dependent upon the environment and the applied constraints, specifically the forces applied to them, and hence the apparent and true persistence lengths are not equivalent for globular domain polymers. We also see that the values of L_p are not as small as the asymptotic values seen for large values of \hat{l}'_{rms} in simulations of free polymers ($L_p \sim R$). This implies that the SMFS force spectroscopy technique, by varying the relative geometries within the system as it applies increasing force, measures some form of average of each of the true values of L_p corresponding to each individual extension.

From a geometrical perspective and with respect to Figure 3b, we would expect any combination of values of \hat{l}_{eq} and \hat{l}'_{rms} to explicitly define our contour, independently of the protein stiffness. Indeed, as we increase \hat{l}'_{rms} we also see an increase in L_c . This may be expected given that \hat{l}'_{rms} acts to increase the average squared length of a fluctuating chain as per Eq.4. However, Figure 8a clearly shows a large reduction in our measured contour length L_c as we increase k_{st} . This implies something akin to the Kuhn length appearing for these globular domain polymers, in that the effective segment lengths in our polymeric system are



(a)



(b)

Fig. 7 The application of the WLC model to the force-extension behaviour of a globular domain polymer with $\hat{l}_{eq} = 0.01$, for a variety of relative fluctuation magnitude values, \hat{l}'_{rms} , and at two different values of the globular domain stiffness, k_{st} . a) $k_{st} = 19.23$ pN.nm b) $k_{st} = 192.31$ pN.nm

being reduced from the true polymeric backbone as we reduce the protein stiffness k_{st} and allow the globular subunits to overlap. Therefore, even though in the unstrained state the explicit definition of L_c and calculation of L_p via Eq.2 works well, when SMFS constraints are applied the mathematical contour required to fit the WLC model via Eq.15 is significantly shorter and does not correspond to the backbone of the polymer itself. Hence, the contour length L_c is also an apparent value.

Aside from the contour length, we see that the measured persistence length L_p increases with k_{st} as expected. However, compared with the free polymer simulations the effect of k_{st} on L_p is much less pronounced. This is true even at low values of \hat{l}'_{rms} when, contrary to the free polymer calculations, k_{st} has almost no effect on L_p .

Although our measurements of L_c and L_p appear unusual in isolation, given our results for the true values from the free polymer simulations, Figure 8b shows how the ratio of these two val-

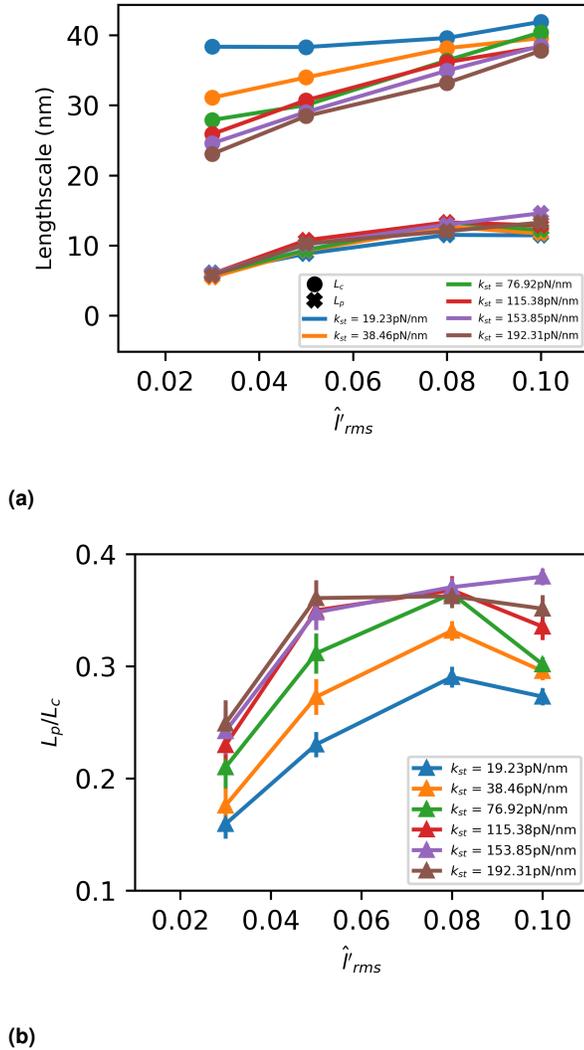


Fig. 8 The effect of the SMFS force spectroscopy procedure on the measurement of both the contour length L_c and the persistence length L_p of a globular domain polymer with $\hat{l}_{eq} = 0.01$, for a variety of values of \hat{l}_{rms} and k_{st} . a) The individual L_c and L_p values and b) the ratio of L_p and L_c values

ues changes. We now see that with only a single exception, at each value of \hat{l}_{rms} an increase in k_{st} increases the effective persistence length when measured with respect to the effective contour length. We can also see that as \hat{l}_{rms} itself increases, the relative size of L_p with respect to L_c appears to grow initially, and then begins to fall again. As \hat{l}_{rms} acts as an effective increase to the total length of the spring, then as this value initially increases, the effective contour length also increases, and the effective persistence length increases with it. However, just as in the free polymer simulations, by increasing \hat{l}_{rms} we also reduce the frequency of steric interactions, which contributes to a reduction in the persistence length as per Eq.9. As we begin to move into the large force regime of Eq.15, where steric interactions can no longer occur in our polymers due to the rigid boundary conditions, the effect of the increases in contour length on the persistence length are overtaken by the lack of steric interaction, thus reducing the

persistence length again. Hence, we can see that the apparent persistence length can only be correctly interpreted when measured relative to the apparent contour. However, we also see that whilst our very small initial value of $\hat{l}_{eq} = 0.01$ would correspond to a semi-flexible polymer with true values of L_c and L_p measured from a free polymer using Eq.10 (see Section 3.4), even for our stiffest globular domains in Figure 8b, the polymer remains in the fully flexible state for all values of \hat{l}_{rms} . This indicates the the application of tensile force to a globular domain polymer, and the subsequent change in internal geometry, significantly reduces its apparent flexibility.

4 Conclusions

From this simulation study we have seen that the persistence length of globular domain polymers, whether they are colloidal chains, polyproteins or some other inhomogeneous polymer, is strongly dependent upon the relative geometries along the polymeric contour. This dependence causes the true persistence length, intrinsic to the free polymer, and apparent persistence length, specific to an experimental setup, to diverge from one another. If the linker domain is long compared with the size of the globular domains, or if the thermal fluctuations are relatively large, then the intrinsic persistence length will be dominated by the stiffness of the linker domains themselves. From the intrinsic persistence lengths of amino acid chains referenced in Section 1 and verified in Section 3.1, it is likely that these polymers would be classed as flexible. On the other hand, if the linker domain is short compared with the size of the globular domains, and the thermal fluctuations are sufficiently small, then the persistence length will be dominated by the stiffness of the globular domains. Given the relatively high intrinsic stiffness of proteins⁵³ and their low aspect ratio, it is likely that only this structural geometry will enable globular domain polyproteins to move into the semi-flexible or rigid regimes. Finally, we have shown that the alteration of internal geometry via SMFS means that the apparent persistence lengths of globular domain polymers measured via this technique, although qualitatively useful, cannot be said to represent the true stiffness of the polymer in any situation other than that specific experiment.

Our work is in qualitative agreement with Hsu *et al.*, who were able to show via a thorough simulation study that although the WLC model fits well to so-called ‘bottle-brush’ polymers, due to their complex side chain geometry the resulting L_c and L_p values do not have clear a physical meaning⁵⁴. They found that the persistence length, although clearly defined with respect to a contour, was actually proportional to the contour length itself, which should not be the case for a WLC system. Our work is complimentary, showing that although the WLC can be applied to the SMFS force-extension curves of globular domain polymers, the emergent persistence lengths and contour lengths do not accurately reflect the ‘intrinsic’ stiffness of object. It is true that persistence length is dependent on external factors such as temperature and uniform hydrophobicity⁵⁵, which provide something akin to a delocalised restoring force equivalent to a modification in the intrinsic stiffness of the object. What we have shown is that that the stiffness of an entire globular domain polymer is inescapably de-

pendent upon *localised* interactions with the environment, which act to alter the relative internal geometries and therefore reduce the frequency and strength of the local interactions from which ‘intrinsic’ stiffness arises⁵⁴.

The effect of local geometries on the persistence length of a long molecule was alluded to by Landau and Lifshitz in their seminal series of textbooks⁵⁶, and we now see the relevance of these insights. The hierarchical emergence of worm-like chain behaviour in globular domain polyproteins provides a route for exploring the rational design of biological systems. Based on the work of Da Silva *et al.*¹⁴, Shmilovich *et al.* recently developed a model to describe the force/extension behaviour of hydrogels formed of polyproteins of protein L⁴³. Using rigid rods to model the polyproteins, and via the inclusion of unfolding kinetics for the individual protein subunits, their model predicts the emergence of a non-linear viscoelastic response as force is applied. It would be of great interest to observe how the effective change in persistence length of these polyproteins via the application of force to the hydrogel would effect the emergent properties of the system.

We have shown that through careful consideration of the relative geometries of the linker and globular domains, the microscopic mechanics of a hierarchical system can be manipulated by orders of magnitude. We hope this study will provide a useful empirical reference in the design of hierarchical biological systems which incorporate globular domain polymers.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors would like to extend thanks to the group of Professor Lorna Dougan at the University of Leeds, who provided useful experimental insight into this theoretical work. Thanks also to Dr David Brockwell for proofreads and clarification on nomenclature.

The computational experiments in this work were undertaken on ARC3, part of the High Performance Computing facilities at the University of Leeds, UK

The project was supported by a grant from the Engineering and Physical Sciences Research Council (EPSRC) (EP/P020088X/1) to L. Dougan.

Notes and references

- 1 R. H. Pritchard, Y. Y. Shery Huang and E. M. Terentjev, *Soft Matter*, 2014, **10**, 1864–1884.
- 2 W. A. Linke, *Annual Review of Physiology*, 2017, **80**, 389–411.
- 3 H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh and A. Amon, *Molecular Cell Biology*, WH Freeman and Company, New York, 7th edn, 2013.
- 4 M. E. Porter and W. S. Sale, *Journal of Cell Biology*, 2000, **151**, 37–42.
- 5 H. R. Vutukuri, A. F. Demirörs, B. Peng, P. D. Van Oostrum, A. Imhof and A. Van Blaaderen, *Angewandte Chemie - International Edition*, 2012, **51**, 11249–11253.
- 6 B. Bharti, G. H. Findenegg and O. D. Velev, *Scientific Reports*, 2012, **2**, 1–5.
- 7 Z. Rozynek, M. Han, F. Dutka, P. Garstecki, A. Józefczak and E. Luijten, *Nature Communications*, 2017, **8**, 4–8.
- 8 S. Gonzalez and R. Soto, *New Journal of Physics*, 2018, **20**, 053014.
- 9 Y. Cao and H. Li, *Nature Materials*, 2007, **6**, 109–114.
- 10 A. L. Holmberg, K. H. Reno, R. P. Wool and T. H. Epps, *Soft Matter*, 2014, **10**, 7405–7424.
- 11 M. Tassieri, *Macromolecules*, 2017, **50**, 5611–5618.
- 12 C. Schuldt, J. Schnauß, T. Händler, M. Glaser, J. Lorenz, T. Golde, J. A. Käs and D. M. Smith, *Physical Review Letters*, 2016, **117**, 1–6.
- 13 Y. Zhang, E. P. Debenedictis and S. Keten, *Soft Matter*, 2019, **15**, 3807–3816.
- 14 M. A. Da Silva, S. Lenton, M. Hughes, D. J. Brockwell and L. Dougan, *Biomacromolecules*, 2017, **18**, 636–646.
- 15 J. Wu, P. Li, C. Dong, H. Jiang, B. Xue, X. Gao, M. Qin, W. Wang, Bin Chen and Y. Cao, *Nature Communications*, 2018, **9**, 1–11.
- 16 D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nature Methods*, 2009, **6**, 343–5.
- 17 E. Weber, R. Gruetzner, S. Werner, C. Engler and S. Marillonnet, *PLoS ONE*, 2011, **6**, year.
- 18 P. Vladimir, J. L. Ong, R. B. Kucera, B. W. Langhorst, K. Bilotti, J. M. Pryor, E. J. Cantor, B. Canton, T. F. Knight, T. C. Evans and G. Lohman, *bioRxiv*, 2018, 322297.
- 19 M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, 1988.
- 20 M. L. Hughes and L. Dougan, *Reports on Progress in Physics*, 2016, **79**, year.
- 21 M. S. Kellermayer, C. Bustamante and H. L. Granzier, *Biochimica et Biophysica Acta - Bioenergetics*, 2003, **1604**, 105–114.
- 22 Y. B. Fu, Y. R. Liu, P. Y. Wang and P. Xie, *Journal of Polymer Science, Part B: Polymer Physics*, 2018, **56**, 297–307.
- 23 J. Hsin, J. Strümpfer, E. H. Lee and K. Schulten, *Annual Review of Biophysics*, 2011, **40**, 187–203.
- 24 W. A. Linke, M. Kulke, H. Li, S. Fujita-Becker, C. Neagoe, D. J. Manstein, M. Gautel and J. M. Fernandez, *Journal of Structural Biology*, 2002, **137**, 194–205.
- 25 M. H. Abolbashari and S. Ameli, *Scientia Iranica*, 2012, **19**, 1526–1533.
- 26 A. Sarkar, S. Caamano and J. M. Fernandez, *Journal of Biological Chemistry*, 2005, **280**, 6261–6264.
- 27 S. R. K. Ainaravapu, J. Brujić, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li and J. M. Fernandez, *Biophysical Journal*, 2007, **92**, 225–233.
- 28 M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci.*, 1999, **96**, 3694–3699.
- 29 H. Li, A. F. Oberhauser, M. Carrion-vazquez, J. G. Kerkvliet, P. E. Marszalek and J. M. Fernandez, *Nature*, 2002, **418**, 998–1002.

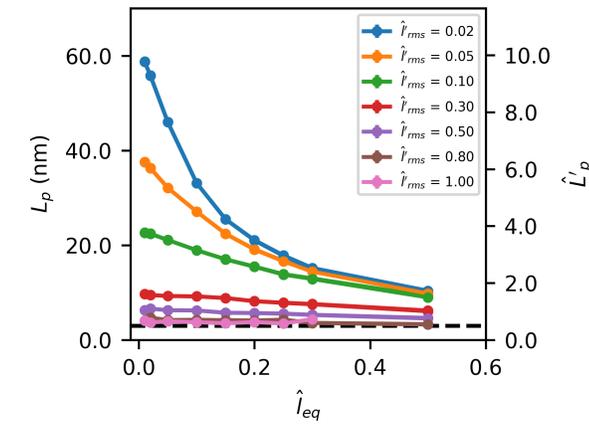
- 30 T. Huber, L. Grama, C. Hetényi, G. Schay, L. Fülöp, B. Penke and M. S. Kellermayer, *Biophysical Journal*, 2012, **103**, 1480–1489.
- 31 J. F. Marko and E. D. Siggia, *Macromolecules*, 1995, **28**, 8759–8770.
- 32 J. A. Rivas-Pardo, E. C. Eckels, I. Popa, P. Kosuri, W. A. Linke and J. M. Fernández, *Cell Reports*, 2016, **14**, 1339–1347.
- 33 J. Oroz, M. Bruix, D. V. Laurents, A. Galera-Prat, J. Schönfelder, F. J. Cañada and M. Carrión-Vázquez, *Structure*, 2016, **24**, 606–616.
- 34 Z. T. Yew, M. Schlierf, M. Rief and E. Paci, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2010, **81**, 1–4.
- 35 S. Lv, D. M. Dudek, Y. Cao, M. M. Balamurali, J. Gosline and H. Li, *Nature*, 2010, **465**, 69–73.
- 36 M. Jahn, K. Tych, H. Girstmair, M. Steinmaßl, T. Hugel, J. Buchner and M. Rief, *Structure*, 2018, **26**, 96–105.e4.
- 37 A. Sarkar, S. Caamano and J. M. Fernandez, *Biophysical Journal*, 2007, **92**, 36–38.
- 38 M. Wolny, M. Batchelor, P. J. Knight, E. Paci, L. Dougan and M. Peckham, *Journal of Biological Chemistry*, 2014, **289**, 27825–27835.
- 39 A. Alemany, B. Rey-Serra, S. Frutos, C. Cecconi and F. Ritort, *Biophysical Journal*, 2016, **110**, 63–74.
- 40 K. M. Tych, M. Jahn, F. Gegenfurtner, V. K. Hechtel, J. Buchner, T. Hugel and M. Rief, *Journal of Physical Chemistry B*, 2018, **122**, 11373–11380.
- 41 T. H. Evers, E. M. W. M. Van Dongen, A. C. Faesen, E. W. Meijer and M. Merckx, *Biochemistry*, 2006, **45**, 13183–13192.
- 42 H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels and B. Schuler, *Proceedings of the National Academy of Sciences*, 2012, **109**, 16155–16160.
- 43 K. Shmilovich and I. Popa, *Physical Review Letters*, 2018, **121**, 168101.
- 44 S. Plimpton, *Journal of Computational Physics*, 1993, **117**, 1–19.
- 45 E. W. Weisstein, *Sphere-Sphere Intersection*, <http://mathworld.wolfram.com/Sphere-SphereIntersection.html>.
- 46 D. M. Heyes and H. Okumura, *Molecular Simulation*, 2006, **32**, 45–50.
- 47 K. K. Liu, D. R. Williams and B. J. Briscoe, *Journal of Physics D: Applied Physics*, 1998, **31**, 294–303.
- 48 J. R. Rumble, *CRC handbook of chemistry and physics.*, CRC Press LLC, 99th edn, 2018.
- 49 M. Van Rosmalen, M. Krom and M. Merckx, *Biochemistry*, 2017, **56**, 6565–6574.
- 50 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of computational chemistry*, 2004, **25**, 1605–12.
- 51 M. V. Shapovalov and R. L. Dunbrack, *Structure*, 2011, **19**, 844–858.
- 52 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.
- 53 J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton.*, Sinauer Associates, Sunderland (Massachusetts), 2001.
- 54 H.-P. Hsu, W. Paul and K. Binder, *Polymer Science Series C*, 2013, **55**, 845–865.
- 55 S. J. Haward, P. R. Shewry, J. Marsh, M. J. Miles and T. J. McMaster, *Microscopy Research and Technique*, 2011, **74**, 170–176.
- 56 L. Landau and E. Lifshitz, *Course of Theoretical Physics: Statistical Physics Part 1*, Pergamon Press Ltd, Oxford, 3rd edn, 1980, pp. 396–400.
- 57 T. Liu, F. Khabaz, R. T. Bonnecaze and M. Cloitre, *Soft Matter*, 2018, **14**, 7064–7074.

5 Supplementary Information

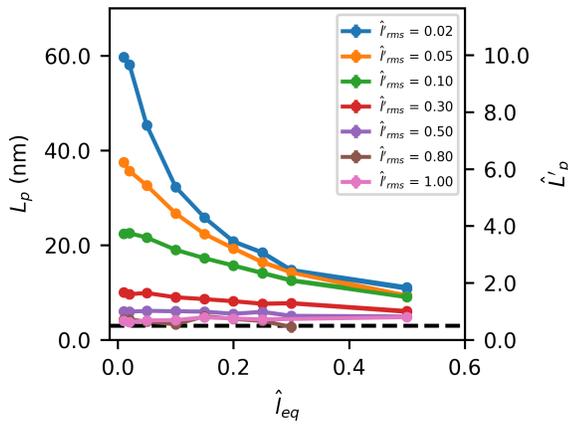
5.1 Data Access & Reproducibility

All of our globular domain polymer simulations were performed using BioNet, a software package in development at the University of Leeds. This software is available for download from the Bitbucket repository <https://bitbucket.org/GokuBH/proteinhydrogelsoftware/>. Although the core software is in active development for additional applications and features, for reproducibility purposes the branch *PolyProteinSims* remains unchanged since this work was performed.

All data and graphs used in this work, as well as additional movies which comprise the range of globular domain polymer flexibilities observed in this work, can be found at <https://doi.org/10.5518/709>.



(a)



(b)

Fig. 9 The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of relative fluctuation magnitude values, \hat{l}_{rms} , and using two different functional forms for the steric interaction. a) The Hertzian interaction potential given by Eq.17, and b) the Weeks-Chandler-Anderson potential given by Eq.19.

5.2 Alternate forms of steric interaction

We previously mentioned that the functional form of Eq.3 was chosen to give an energy proportional to the volume overlap, thus modelling volumetric compression together with a linear stiffness, k_{st} , which is directly comparable to the linear stiffness of the linker domain. However, alternate functional forms are available. Figure 9 shows the results of a range of single globular domain polymer simulations equivalent to those in Section 3.2 but using alternative steric interaction potentials.

Figure 9a shows the results of using a Hertzian contact force, F_H , given by Liu *et al.* as^{47,57}

$$F_H = \begin{cases} \frac{4}{3}CE^*\varepsilon^n R_c^2 & \varepsilon \geq 0 \\ 0 & \varepsilon < 0 \end{cases} \quad (17)$$

where R_c is the contact radius, ε is a dimensionless overlap distance, E^* is an elastic contact modulus and C and n are empirical constants which depend on ε as

$$\begin{aligned} 0 < \varepsilon \leq 0.1 & \Rightarrow C = 1, n = 1.5 \\ 0.1 < \varepsilon \leq 0.2 & \Rightarrow C = 31.62, n = 3 \\ \varepsilon > 0.2 & \Rightarrow C = 790.6, n = 5 \end{aligned} \quad (18)$$

This piecewise empirical function is intended to capture the non-linearity involved in volumetric compression. Specific to our simulations, where each bead has the same radius R , it can be shown that $R_c = R/2$ and $\varepsilon = 2(2 - (r/R))$ for $r < 2R$, and $\varepsilon = 0$ otherwise. To obtain a value of E^* that best reflects our original simulations, we took the derivative of Eq.17 with respect to radial distance and equated this with a constant value of $k_{st} = 192.31$ pN.nm from Figure 5. This leads to a variable contact modulus, but one which reflects the constant stiffness used in our main simulations. Figure 9a shows this to be a suitable approach, as the trends are almost exactly equivalent to those shown in Figure 5.

Figure 9b shows the results using a Weeks-Chandler-Anderson potential⁴⁶, U_{WCA} , a modification to the standard Lennard-Jones potential with functional form

$$U_{WCA} = \begin{cases} E_m \left(\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 + 1 \right) & r \leq r_0 \\ 0 & r > r_0 \end{cases} \quad (19)$$

where r_0 is the cutoff distance, and E_m the energy minimum of the equivalent Lennard-Jones potential. In our case, we require the cutoff to be at the surface of the spherical domains and so $r_0 = 2R$. This potential is significantly steeper than Eq.3 or Eq.17, and so we chose a value of $E_m = 10$ pN.nm in order to keep our simulations numerically stable.

That Figures 9a and 9b are so similar to the main simulations shown in Figure 5 indicates that each form of steric interaction is interchangeable for the calculations we have been performing. With regards to the Hertzian potential representation of contact forces, we have not subjected the subunits to any significant compression, only thermal collisions. Hence, as we are at high values of k_{st} in Figures 9a and 9b, it is likely that we remain in the linear regime of steric interaction in our simulations. This justifies

our linear parameterisation in this section, but also supports our geometric interpretation of the hierarchical emergence of persistence length in this work i.e. it is not the functional form of steric interaction that matters, merely the rate of occurrence due to geometric factors.

Additional software branches, *PolyProteinSimsHertz* and *PolyProteinSimsWCA*, were created with these alternate steric interaction forms and are also available for download (see Section 5.1).

5.3 Derivation of the extended freely-jointed chain model

For a globular domain polymer comprised of N globular domains and following the contour defined in Figure 3b, the end-to-end distance vector, \vec{E} , can be written as

$$\vec{E} = \sum_{i=1}^{N-1} \vec{r}_i + \vec{l}_{i,i+1} + \vec{r}_{i+1} \quad (20)$$

where \vec{r}_i is the vector from the center of the i th globular domain to the surface binding site, $\vec{l}_{i,i+1}$ is the vector from the i th globular domain surface binding site to the $i+1$ th, and \vec{r}_{i+1} is the vector from the $i+1$ th globular domain surface binding site to the center of the $i+1$ th globular domain (see Figure 3). Using Eq.20, we can calculate the inner product of \vec{E} with itself, which can be expanded as

$$\begin{aligned} \langle E^2 \rangle &= \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \langle \vec{r}_i \cdot \vec{r}_j + \vec{l}_{i,i+1} \cdot \vec{l}_{j,j+1} + \vec{r}_{i+1} \cdot \vec{r}_{j+1} \\ &\quad + \vec{r}_i \cdot (\vec{l}_{j,j+1} + \vec{r}_{j+1}) \\ &\quad + \vec{l}_{i,i+1} \cdot (\vec{r}_j + \vec{r}_{j+1}) \\ &\quad + \vec{r}_{i+1} \cdot (\vec{r}_j + \vec{l}_{j,j+1}) \rangle. \end{aligned} \quad (21)$$

In this freely-jointed chain model, the absence of steric interactions between globular domains means that any term involving an inner product between neighbouring subunits will reduce to zero in the ergodic limit. However, we note that it is these very terms that cause the emergence of a persistence length in globular domain polymers once a steric interaction is included. Nevertheless, neglecting these terms from Eq.21 and simplifying gives

$$\langle E^2 \rangle = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} 2R^2 \delta_{i,j} + \langle l^2 \rangle + R^2 (\delta_{i,j+1} + \delta_{i+1,j}) \quad (22)$$

where R is the radius of each globular domain, and l is the length of each linker domain. The delta functions have emerged from the ensemble averages taken between subunits, which can only correlate with themselves in a freely-jointed chain model.

To calculate $\langle l^2 \rangle$, we recognise that $l = l_{eq} + \Delta l$, where l_{eq} is the equilibrium length of the linker domain, and Δl are fluctuations about that length due to thermal noise. Applying the equipartition theorem, we find that

$$\langle l^2 \rangle = l_{eq}^2 + \frac{k_B T}{k} \quad (23)$$

where k is the linker stiffness. Now, introducing Eq.23 to Eq.22, we can take summations over the delta functions, resulting in

$$\langle E^2 \rangle = (N-1) \left(\frac{k_B T}{k} + l_{eq}^2 + 2R^2 \right) + 2(N-2)R^2. \quad (24)$$

For the final term in Eq.24 we have taken into account the boundary conditions implied by $\delta_{i,j+1}$ and $\delta_{i+1,j}$, where the upper and lower limits of the two sums must be taken into account for polymers of finite length. Finally, factorising Eq.24 yields

$$\langle E^2 \rangle = (N-1) \left(\frac{k_B T}{k} + l_{eq}^2 + 2 \left(2 - \frac{1}{N-1} \right) R^2 \right) \quad (25)$$

which is equivalent to Eq.4.

5.4 Alternate persistence length definitions

Figure 10 shows the raw correlation traces between all tangent vectors along the contour as defined in Section 3.2 and shown as black discs in Figure 3. Thus, the centre-to-centre vector between adjacent globules consists of three vectors, corresponding to (i) globule centre to start of linker, (ii) start of linker to end of linker, and (iii) end of linker to the next globule centre. We can see that the graph is highly discontinuous every third data point. With reference to the starting position of our defined contour, we see that the significantly lower set of points correspond to the correlations with the linker domain vectors, and that the pairs of equally correlated points correspond to the rigid globular domain vectors. This shows very clearly the geometric inhomogeneity along the defined contour for globular domain polymers.

Although the steric interaction potential we have used allows a variable degree of overlap between the globular domains, it does not model explicit compression and so the linker-globular domain connection sites (N and C termini) within each globular domain remain in a constant orientation with respect to the globular domain throughout the simulation. Thus, these rigid body vectors remain perfectly correlated throughout the simulation. However, the linker domains do not exhibit any steric interactions with their neighbours throughout the simulations, and so their correlations are relatively small. In fact, it is likely that the only reason they are non-zero is due to the interactions of the globular domains on either side. It is interesting, therefore, that the linker domain vectors, representing the end-to-end distance of the underlying amino acid chains, can have such low correlations with respect to the initial vector amidst globular domains that are significantly more correlated due to the linker geometry itself.

In Figures 10a and 10b we can see four of the possible fits of Eq.10 to the data. In the main text the highest fit, coloured in red, was employed, as this represents the correlations along the contour emerging from interactions between the globular domains, the focus of this paper. Due to the perfect correlation within globular domains, the second fit, coloured in yellow, is exactly equivalent. The third fit, coloured in black, represents the correlations between the linker domains. Although these correlations are low with respect to the initial vector, the persistence of the correlations along the chain is significant, resulting in a persistence length of the same magnitude as the globular domains

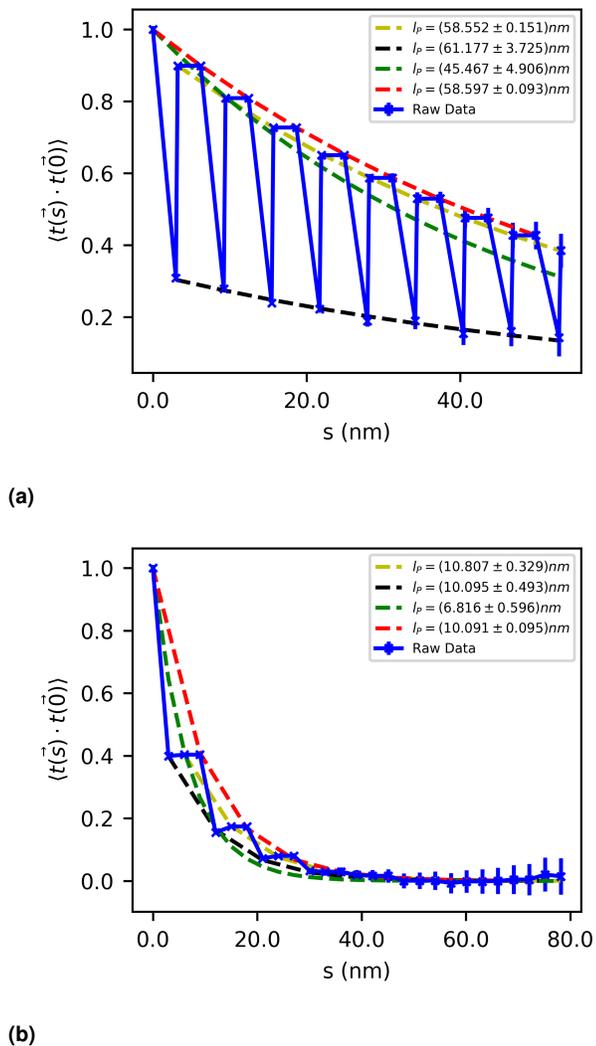


Fig. 10 The tangent correlations at each point along the contour of a globular domain polymer with $\hat{l}_{rms} = 0.02$, $k_{st} = 192.31$ pN.nm and for two different values of \hat{l}_{eq} . We show multiple possible fits of Eq.10 to the data. See text for an explanation of the different coloured fits. a) $\hat{l}_{eq} = 0.01$ b) $\hat{l}_{eq} = 0.5$

themselves. The final fit, coloured in green, represents a form of average between the correlations of the linker and globular domains. It is possible that this fit more accurately represents the effective persistence lengths measured in Section 3.5, or the value that would be measured via end-to-end measurements such as in FRET experiments, which take into account the resultant flexibilities of all components of the polymer chain. However, for the purposes of this work, it is sufficient to observe that alternate fitting models exist for different discretisations of the polymer chain, but all generate a similar prediction for the persistence length.

We note that if we had defined our contour to start at another point, say, the N or C termini of the first globular subunit, our correlation graphs would look slightly different. However, the similarities between the persistence lengths extracted from only the linker domains (black curves), or only the globular domains

(red curves), are similar enough in both Figure 10a and Figure 10b that we can infer the same emergent patterns would occur as a function of the local geometries and steric interactions.

5.5 Graphs of the intermediate regimes of hierarchical emergence of persistence length in globular domain polymers

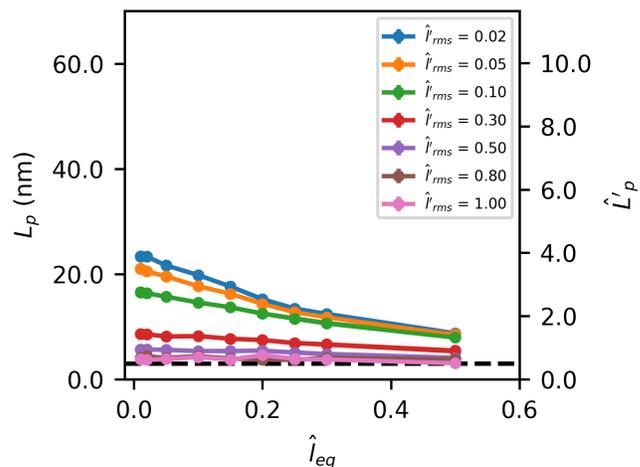


Fig. 11 The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of relative fluctuation magnitude values, \hat{l}_{rms} , and at an intermediate value of the globular domain stiffness, $k_{st} = 19.23$ pN.nm

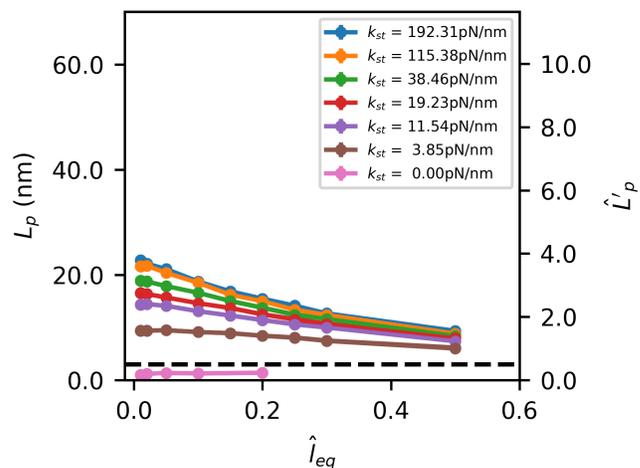


Fig. 12 The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of globular domain stiffness values, k_{st} , and at an intermediate value of the relative fluctuation magnitude, $\hat{l}_{rms} = 0.1$.