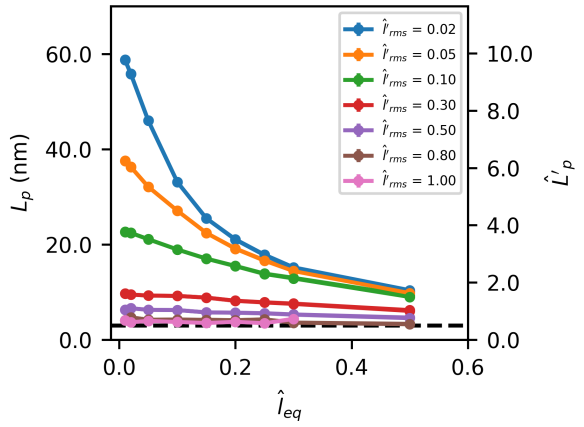
5 Supplementary Information

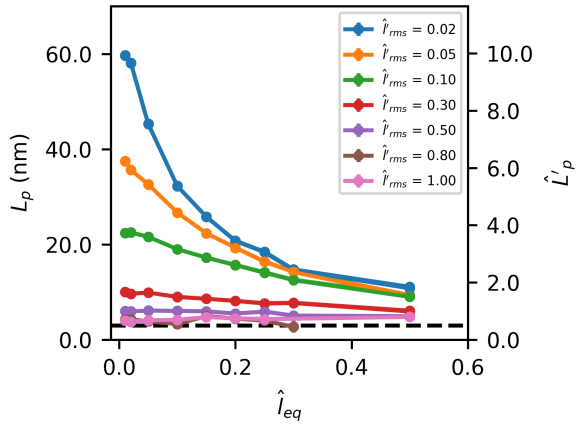
5.1 Data Access & Reproducibility

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)

Figure S1. 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 S1 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 S1a shows the results of using a Hertzian contact force, F_H, given by Liu *et al.* as[47,57]

$$F_H = \begin{cases} \frac{4}{3} C E^* \varepsilon^n R_c^2 & \varepsilon \geq 0 \\ 0 & \varepsilon > 0 \end{cases} \tag{17}$$

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

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

This piecewise empirical function is intended to capture the nonlinearity 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 S1a shows this to be a suitable approach, as the trends are almost exactly equivalent to those shown in Figure 5.

Figure S1b shows the results using a Weeks-Chandler-Anderson potential[46], U_{WCA}, a modification to the standard Lennard-Jones potential with functional form

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

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 S1a and S1b 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 S1a and S1b, 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

$$\vec{E} = \sum_{i=1}^{N-1} \vec{r_i'} + \vec{l}_{i,i+1} + \vec{r}_{i+1} \tag{20}$$

where $\vec{r_i'}$ is the vector from the center of the ith globular domain to the surface binding site, $\vec{l}_{i,i+1}$ is the vector from the ith 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

$$\begin{aligned}\langle E^2 \rangle = \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} \langle \vec{r_i'}\cdot\vec{r_j'} + \vec{l}_{i,i+1}\cdot\vec{l}_{j,j+1} + \vec{r}_{i+1}\cdot\vec{r}_{j+1} \\ + \vec{r_i'}\cdot\left(\vec{l}_{j,j+1}+\vec{r}_{j+1}\right) \\ + \vec{l}_{i,i+1}\cdot\left(\vec{r_j'}+\vec{r}_{j+1}\right) \\ + \vec{r}_{i+1}\cdot\left(\vec{r_j'}+\vec{l}_{j,j+1}\right)\rangle.\end{aligned} \tag{21}$$

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

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

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

$$\langle l^2\rangle = l_{eq}^2 + \frac{k_BT}{k} \tag{23}$$

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

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

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

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

which is equivalent to Eq.4.

5.4 Alternate persistence length definitions

Figure S2 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 S2a and S2b 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

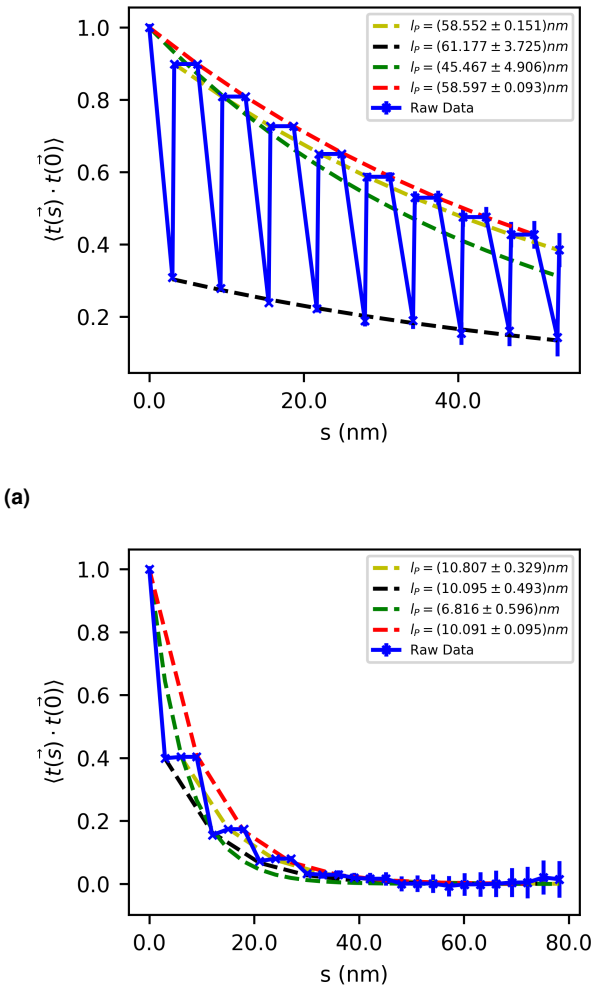


Figure S2. The tangent correlations at each point along the contour of a globular domain polymer with $\hat{l}'_{rms} = 0.02$, $k_{st} = 192.31$ pN.nm and for two different values of $\hat{l}_{eq}$. We show multiple possible fits of Eq.10 to the data. See text for an explanation of the different coloured 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 S2a and Figure S2b 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

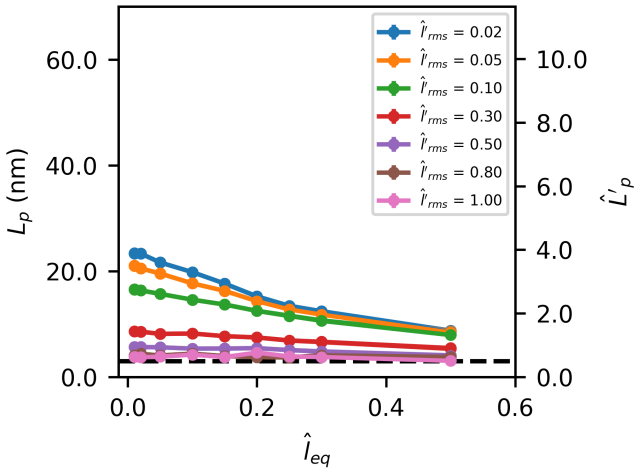


Figure S3. The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of relative fluctuation magnitude values, $\hat{l}'_{rms}$, and at an intermediate value of the globular domain stiffness, $k_{st} = 19.23$ pN.nm

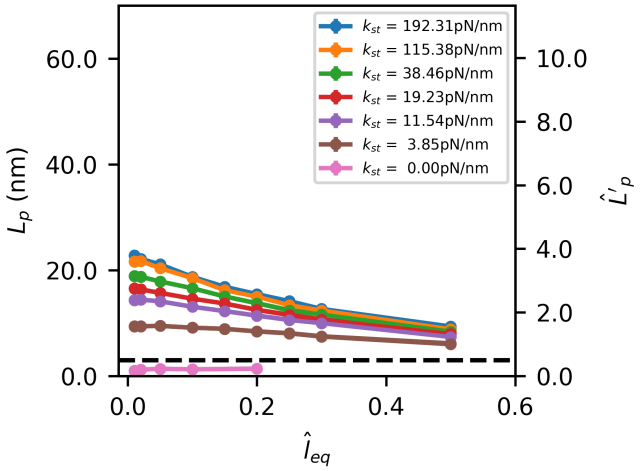


Figure S4. The effect of the linker equilibrium length on the persistence length of a globular domain polymer, for a variety of globular domain stiffness values, k_{st}, and at an intermediate value of the relative fluctuation magnitude, $\hat{l}'_{rms} = 0.1$.