How flexible is a protein: simple estimates using FRET microscopy

Shourjya Sanyal[1], David F. Coker[1,2] and Donal MacKernan[1]

[1] School of Physics and Complex Adaptive Systems Laboratory, University College Dublin, Ireland. [2] Department of Chemistry, Boston University, Boston, USA

*Correspondence: shourjya.sanyal@ucdconnect.ie,donal.mackernan@ucd.ie

S1 Mapping L to N

To map the parameter L used in our model of the flexible linker, with experiment where the linker length is proportional to the total number of residues/beads N, it is necessary to relate N to L. This is done by simply plotting the mean square end to end distance of the chromophore $\langle R^2 \rangle$ of the model as a function of L, as shown in Fig.S1. An excellent fit to the data is given by

$$\langle R^2 \rangle = d(R_0, D_0) \times L^2 + e(R_0, D_0) \tag{1}$$

Here the quantities $d(R_0, D_0) = 243.06$ and $e(R_0, D_0) = 285.30$ Å^2. The difference between the mean square end to end distances of the chromophore and the unperturbed polymer chain for our models can be approximated as D_0^2, giving us

$$\langle R^2 \rangle - D_0^2 = d(R_0, D_0) \times L^2 + e(R_0, D_0) \tag{2}$$

S2 Models

0.1 Gaussian chain model

In the gaussian chain model (GCM) every bond vector itself is gaussian distributed, and $P^{linker}(r)$ can be computed using,

$$P^{linker}(r) = 4\pi r^2 \times \left\{ \frac{3}{2\pi C_\infty N b_0^2} \right\}^{3/2} exp \left\{ -\frac{3r^2}{2C_\infty N b_0^2} \right\} \tag{3}$$

GCM is an good approximation for an ideally flexible linker, whose end-to-end distribution is also gaussian. Using GCM we can directly compute the $P^{linker}(r)$ using Eq.3 with the predicted value of C_∞ and the number of residues N, as shown in Fig.2 and Fig.S3 (Right Panel, Dashed line).

0.2 Bead and spring linker model

In the case of the bead and spring linker model the macro particles are joined by a discrete chain made of N point-like beads at position $r_i (i = 1, ..., N)$. These point particles are connected to the macro particles and among themselves by harmonic springs, and interact with a simple isotropic pair potential[1].

$$\begin{aligned} V_\ell = V_{chain} = \tfrac{1}{2} k_s [|r_1 - R_1| - b_1]^2 + \textstyle\sum_{i=1}^{N-1} \tfrac{1}{2} k_s [|r_{i+1} - r_i| - b_0]^2 \\ + \tfrac{1}{2} k_s [|R_2 - r_n| - b_1]^2 \end{aligned} \tag{4}$$

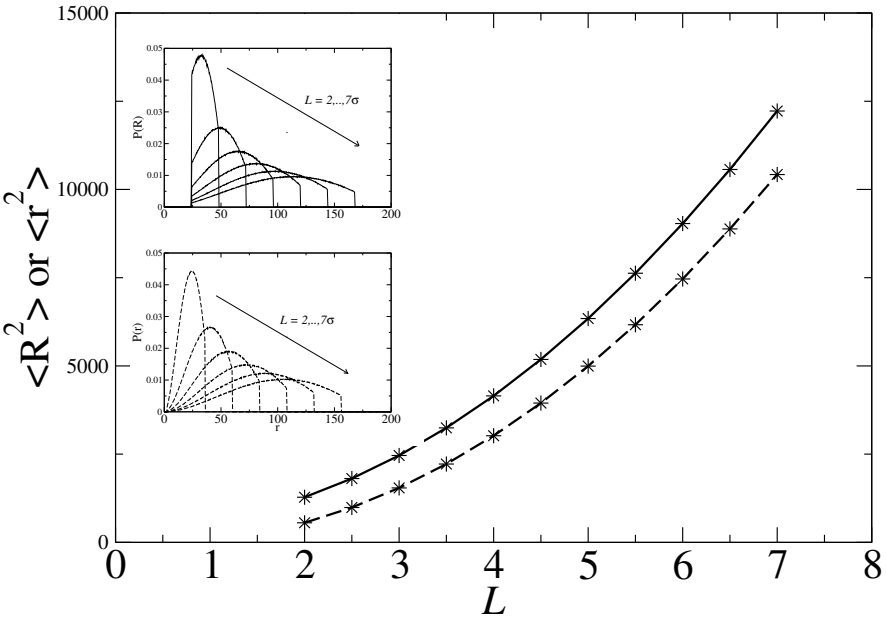

Figure S1: Mean square displacement of the macroparticles $\langle R^2 \rangle$ (Full Black Line) and the isolated linker $\langle r^2 \rangle$ (Dashed Black Line) as a function of L. (Inset) End to end distribution of the of the macro particle $P^{macro}(R)$ and of the polymer in absence of the macro particle $P^{linker}(r)$ for L=2,. . . ,7 σ

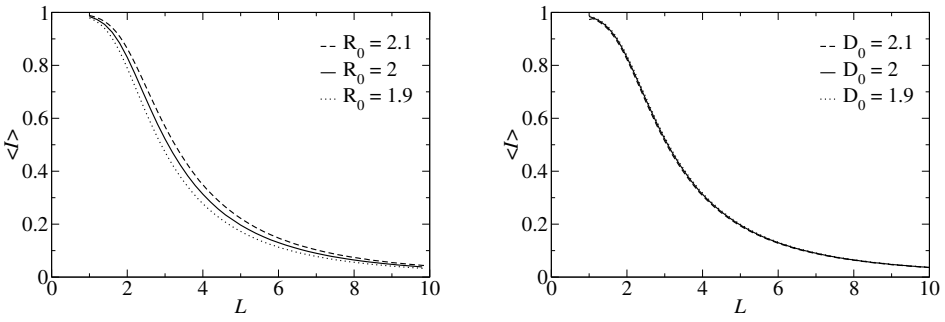

Figure S2: FRET efficiency $\langle I \rangle$ as a function of length L and computed through Monte Carlo simulation, as a function of varying (A) Föster radius R_0 and (B) FP diameter D_0. All distances are measured in units of σ.

The equilibrium distance b_1 between the macro particle and the bead is defined as sum of radius of the macro particle and the bead i.e. 13.9 Å . The value of k_s was chosen to reproduce the experimentally

observed values of FRET efficiency $\langle I \rangle$ for a given linker system.[2] Using this model we can then compute the corresponding $P^{macro}(R)$ and $P^{linker}(r)$ for a given number of residues N, as shown in Fig.2 and Fig.S3 (Left/Right Panel, Full line).

0.3 Monte Carlo Simulation

We use the Monte Carlo simulation approach introduced by Metropolis et al.[3,4] to estimate the statistical properties of each model. Below, $\vec{r}^n$ denotes the position of the n macro-particles (in our models $n = 2$). The sampling procedure comprises:

1. Choose a macro-particle position $\vec{r}^n$ at random in the available configuration space and calculate $V(\vec{r}^n)$;
2. Propose a random displacement $\vec{r}^n$ to $\vec{r}^{n'} = \vec{r}^n + \Delta$; and calculate $V(r'^n)$;
3. Accept the proposed move from $\vec{r}^n$ to $\vec{r}^{n'}$ with probability $min\left(1, \exp(-\beta\left[V(\vec{r}^{n'}) - V(\vec{r}^n)\right)\right.$.

0.4 MARTINI Simulations of Peptides in Water

In order to further validate the end to end distribution for such linker systems in solvent, we have simulated representative flexible linkers using Gromacs (version 4.6.5) with the the MARTINI 2.2 polarised coarse grained force field. The reduction of the number of degrees of freedom and the use of shorter-range potential functions makes the force-field computationally very efficient, allowing for a reduction of the simulation time by ∼2-3 orders of magnitude compared to fully atomistic molecular mechanics force fields (AMBER, CHARMM etc). Thereby allowing us to compute fundamental statistical properties of this intrinsically disordered linkers, using our available computational resources for the large size of the periodic box and simulation lengths up to a μs required for these calculations.

The initial coordinates of this unstructured polymer were generated using Antechamber. We then simulated the isolated linker segement in square box of side 100 Å for the smallest system with 25 residues, 150 Å for the system with 41 and 52 residues, and 200 Å for the system with 65 and 84 residues. The number of solvent molecules ranged from ∼6000 water molecules for the smallest box size to ∼55000 water molecules for the largest box. For the 52 residue system where we also simulated the linker attached with the FPs we have use a box of size 200 Å with ∼55000 water molecules. The energy of the system was minimised through steepest descent. This was followed by a relaxation step of 10 ns where the whole system was heated to 300K. A step size of 1 femtoseconds was used throughout. Then an isotropic barostat was switched on and constant pressure (1 atm) dynamics was maintained using the Berendsen coupling for 10 ns. Production runs were performed in the NPT ensemble to capture dynamical and equilibrium properties for 1000 ns i.e 1 μs with Particle Mesh Ewald for long-range electrostatics interaction. Then for each of the MD simulations we compute the corresponding $P^{linker}(r)$, as shown in Fig.2 and Fig.S3 (Right Panel, Dotted line). For the 52 residue system where we also simulated the linker attached with the FPs we compute the corresponding $P^{macro}(R)$, as shown in Fig.2 (Left Panel, Dotted line).

Table S1: Comparison of the five different flexible linkers.

Linker Type	N	Predicted C_∞	$L_{Eff\ Pot}$	$\langle R^2\rangle_{Eff\ Pot}$	$\langle R^2\rangle_{Bead}$	$\langle r^2\rangle_{Eff\ Pot}$	$\langle r^2\rangle_{GCM}$	$\langle r^2\rangle_{Bead}$	$\langle r^2\rangle_{MARTINI}$
$(SAGG)_{13}$	52	1.3	2.55	1784.41	1883.27	976.134	967.693	1069.8	1162.74
$(SAGG)_{21}$	84	1.3	2.97	2477.22	2572.43	1573.84	1554.64	1633.9	1919.3
$(GGSGGS)_4$	41	2.1	2.76	2093.65	2315.63	1243.03	1227.28	1405.14	1277.94
$(GGSGGS)_8$	65	2.1	3.24	2932.09	3166.73	1952.39	1948.32	2159.13	1906.13
$(GGGGS)_5$	25	2.6	2.51	1738.09	1905.24	938.594	929.019	1070.31	1161.95

Linker Type	N	$\langle I\rangle_{Eff\ Pot}$	$\langle I\rangle_{Bead}$	$\langle I\rangle_{MARTINI}$
$(SAGG)_{13}$	52	0.6778	0.6767	0.6726
$(SAGG)_{21}$	84	0.5188	0.5318	
$(GGSGGS)_4$	41	0.5990	05905	
$(GGSGGS)_8$	65	0.4352	0.4440	
$(GGGGS)_5$	25	0.6714	0.6907	

References

[1] G. Strob, *The Physics of Polymers*, BerlinHeidelberg: Springer-Verlag, 2007.

[2] S. Sanyal, D. MacKernan and D. F. Coker, *Journal of Physics : Conference Series.*, 2015, **640**, year.

[3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.

[4] D. Frenkel and B. Smit., *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, Inc., 1996.

[5] N. Komatsua, K. Aoki, M. Yamadac, H. Yukinagac, Y. Fujitac, Y. Kamioka, and M. Matsuda, *Mol. Biol. Cell.*, 2011, **22**, 4647–4656.

[6] T. H. Evers, E. M. v. Dongen, A. C. Faesen, E. W. Meijer and M. Merkx, *Biochemistry*, 2006, **45**, 13183–13192.

[7] G. Li, Z. Huang, C. Zhang, B. Dong, R. Guo, H. Yue, L. Yan and X. Xing, *Appl. Microbiol. Biotechnol.*, 2015, **100**, 215–225.

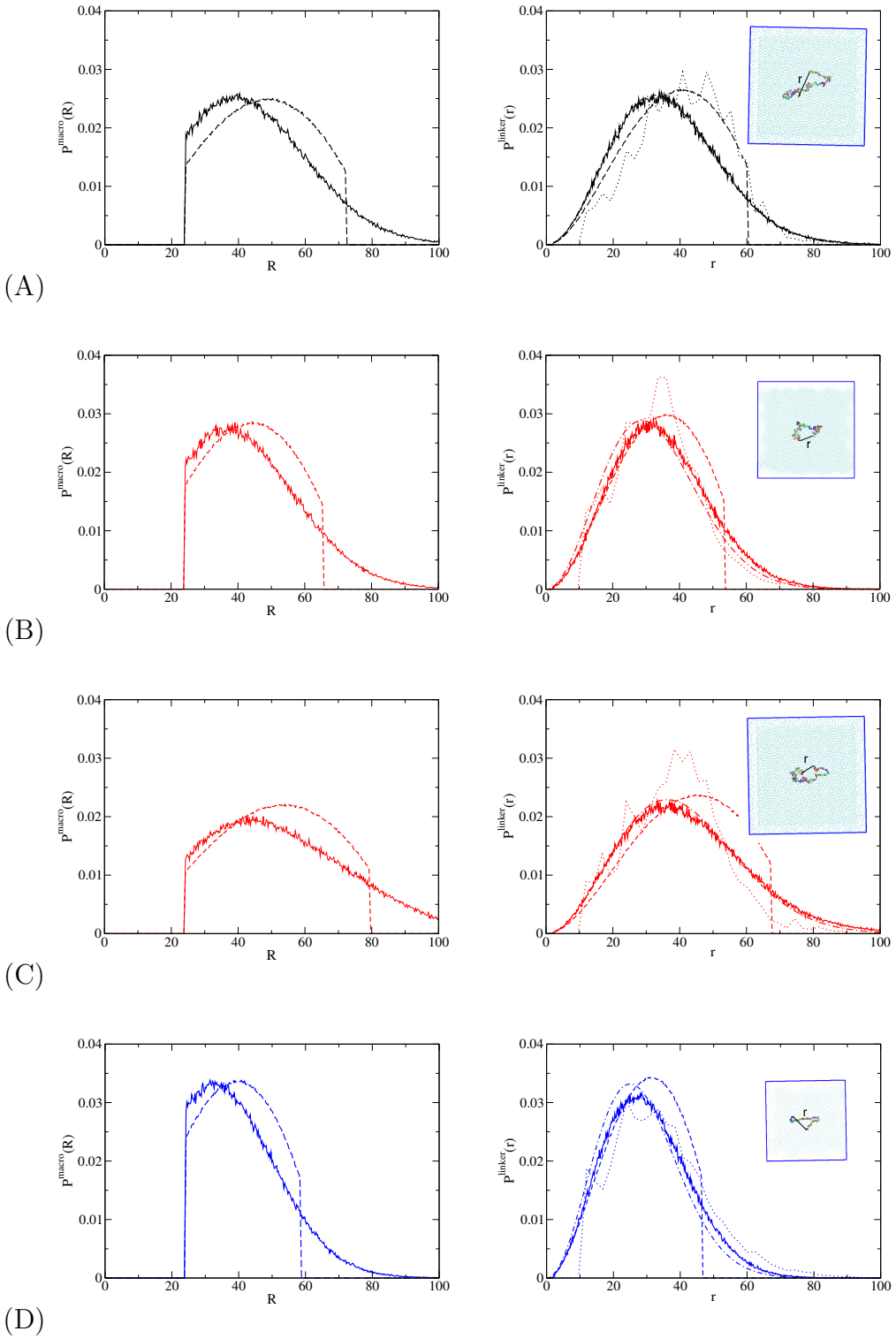

Figure S3: End to end distribution of the macroparticles $P^{macro}(R)$ (Left) and the isolated linker $P^{linker}(r)$ (Right) for (A) $(SAGG)_{21}$ [5], (B) $(GGSGGS)_4$ [6], (C)5$(GGSGGS)_8$ [6], and (D) $(GGGGS)_5$ [7], (distances are measured in units of Angstroms) computed using (1) Effective Potential (Dashed Line), (2) GCM (Dot and dashed line), (3) Bead and spring (Full line) and (4) MARTINI (Dotted line). (Inset Right) The linkers modelled using MARTINI.