Volume 1 | Number 1 | Jan 2013 | Pages 1–100

Molecular Biosystems

www.rsc.org/molecularbiosystems

THE BIOLOGY OF PLAGUE

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

ARTICLE TYPE

How flexible is a protein: simple estimates using FRET microscopy [†]

Shourjya Sanyal,[*a] David F. Coker,[b‡] and Donal Mackernan[a]

Abstract. Flexible proteins are frequently used to link sub-units of larger complexes in various contexts, for instance, in the construction of umimolecular sensors used in FRET microscopy, and fusion proteins. How flexible such linkers are can be an important question in the overall design of the complex, and yet sometimes suprisingly difficult to establish. Such difficulties can arise because the actual flexibility of a protein depends significantly on its interactions with the solvent, and when the local environment is a subcellular compartment, even the conditions of the solvent, may not be known. In this communication we propose a simple numerical procedure through which the flexibility of such proteins can be extracted from FRET based microscopy data.

Flexibile proteins frequently link components within biomolecular complexes, such as unimolecular Föster resonance energy transfer (FRET) probes, and fusion proteins.[1–4] One common example of the former comprises a pair of fluorescent proteins connected by a long flexible linker protein, flanked at either end by a ligand binding and a sensor domain.[5,6] As genetically encoded sensors, their location in cellular organelles can be selected, and used to locally measure the concentration of target ligands/analytes of interest, through optical microscopy. Likewise, fusion proteins using multiple protein segments joined by flexible linkers,[7,8] find a wide range of application including designing fragment based immunoassays,[9–12] bifunctional enzymes,[13] affinity purification,[14] protein stabilisation,[15] and drug transport.[16]

The flexibility of a linker depends not only on its length and constituent residues but also importantly on the nature of the solvent. This is due to the fact that the effective charges and polarisation of residues depends on the solvent, which can be difficult to ascertain within a subcellular compartment or organelle.[5,17] A good estimator of flexibility of the peptide linker is given by the characteristic ratio $C_N = \langle r^2 \rangle / N b_0^2$, and its limiting value C_∞. Here, $\langle r^2 \rangle$ is the root mean square end-to-end distance of the isolated polymer chain, N is the number of residues, and $b_0 = 3.8$ Å is the distance between the adjacent C_α carbons.[18] In a non-cellular context, Brant et al. experimentally measured the characteristic ratio C_∞ of a given poly-peptide using dilute solutions in an ideal / θ-solvent.[19,20] Small-angle X-ray scattering has also been used to estimate C_∞.[21]

Theoretical models have been employed to compute the C_∞ of the rigid homo-polymer linkers as early as 1960's by Schimmel et al. by predicting the rotational freedom of the residues involved.[18] The models show the characteristic ratio C_N of polyproline scales as N (i.e. like a rigid rod), whereas for polyalanine and polyglycine, C_N converges to a constant value of 9.27 and 2.16 respectively.[22,23] Various polymer chain models like the Gaussian chain model (GCM) by Flory,[24] Worm like chain (WLC) by Kratky-Porod,[25] and self-avoiding chain model have been used in literature to estimate C_∞.[26] In such cases, the solvent is treated at most implicitly. Recent work has revealed that the measured flexibility of "flexible" proteins/polymers is often not as great as the predictions of such theoretical models due in part to the inadequate treatment of solvent interactions with the peptide. Deficiencies in the most frequently used forcefields in molecular dynamics, and corresponding parameterizations of mesoscale simulations often leads to the over-estimation of flexibility for various peptides in solvents.[27]

Measurements of C_∞ in subcellular compartments and organelles pose particular challenges both to experiment and detailed molecular simulation, as the precise solvent conditions of the linker are often unknown. Evers et al. have tackled this challenge by measuring the flexibility of protein linkers in a protein buffer solution, using fluorescent proteins (FPs) attached to either end of the linker of interest, and effectively measuring their FRET efficiency as a function of increasing linker length.[28] In particular,

[a] School of Physics, University College Dublin, Ireland. Tel: 353 894537082; E-mail: shourjya.sanyal@ucdconnect.ie, donal.mackernan@ucd.ie
[b] Department of Chemisrty, Boston University, USA.
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

they also showed that the experimental FRET efficiency could be reproduced through a theoretical approach using Gaussian Chain or wormlike chain models of the flexible linker, combined with a geometrical correction accounting for the distances between the centres of the FP's and the end point of the chains. It is feasible that an analytical expression for the dependence of C_∞ on FRET efficiency and the number of residues in the flexible chain can be extracted. The principle objective of the work reported here is to present such an expression, obtained through a simple theoretical model whose salient properties are detailed using Monte Carlo simulation.

Typically, the FRET efficiency I, the fraction of energy transfer events per donor excitation event, falls off quickly with distance between the FPs, and can be approximated by the expression,

$$I(R) = \frac{1}{1 + (R/R_0)^6} \tag{1}$$

with the Föster Radius, $R_0 \sim 4 - 5$ nm, giving the distance at which the energy transfer efficiency is 50% and R is the distance between the macro-particles. The most common method to estimate the average FRET efficiency, $\langle I \rangle$, is through the ratio of the emission intensity of the acceptor FP to that of the donor (also called the FRET sensitised donor-to-acceptor intensity ratio or simply the emission ratio), as measured in experiments by a method known as ratio-metric FRET[29].

In our approach, the fluorescent proteins are modelled as two non-overlapping spherical macro particles of diameter σ e.g. σ = 24 Å (which corresponds to the minimum inter-chomophore distance between two GFP proteins). In the absence of any explicit attraction between the fluorescent proteins, the FRET probe interactions can be modelled using an effective potential $V(R) = V_s(R) + V_l(R)$. Where V_s ensures that the spherical macro-particles cannot overlap ($V_s(R) = \infty$ if $R < \sigma$ and is zero otherwise). The linker part of the interaction is a simple isotropic pair potential with the form.

$$V_l = \begin{cases} \infty & \text{if } R > L/2 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Geometrically, this can be visualised as two non-overlapping macro-particles free to move inside a sphere of diameter L. In our Monte Carlo model, we fix the Föster radius at $R_0 = 48$ Å, and the FP diameter $D_0 = 24$ Å. All the experimental system we choose to model have values of R_0 and D_0 similar to these typical values used in our simulations.

Using the Boltzmann distribution, $P^{macro}(R)$, the equilibrium averages of observables such as the mean square displacement $\langle R^2 \rangle$, or the FRET efficiency, $\langle I(R) \rangle \equiv \langle I \rangle$, are easily estimated through the Monte Carlo simulation approach, as detailed in section 2 of the supplementary information.[30] The estimated average FRET efficiency is plotted as a function of the parameter L in the insert of Fig. 1, and its dependence can be accurately fit by the functional form

$$\langle I \rangle = \frac{1}{1 + \left(\frac{L - a(R_0, D_0)}{b(R_0, D_0)} \right)^{c(R_0, D_0)}} \tag{3}$$

The values of the parameters that provide the best fit to our Monte Carlo calculations are $a(R_0, D_0) = 0.8309$, $b(R_0, D_0) = 2.6614$ and $c(R_0, D_0) = 2.323$. In general these quantities depend on R_0 and D_0, so different simulations and fits need to be determined for each choice of these model parameters.

To compare the average FRET efficiency $\langle I \rangle$ of the model with experiment where the linker length, L, is given in terms of the total number of residues, N, it is necessary to relate L and N. This was done by first computing the mean square center-to-center distance between the macro particles, $\langle R^2 \rangle$, as a function of L, to find $\langle R^2 \rangle \approx d(R_0, D_0)L^2 + e(R_0, D_0)$. Here the quantities $d(R_0, D_0) = 243.06$ and $e(R_0, D_0) = 285.30 \text{ Å}^2$ are determined again by fitting the $\langle R^2 \rangle$ obtained from our Monte Carlo calculations, as detailed in section 1 of the supplementary information. For an experimental system, if the linker is sufficiently flexible, the center-to-center distance can be approximated as a Gaussian random walk for which we find, $\langle R^2 \rangle - D_0^2 = C_\infty N b_0^2$. Equating these two expressions for the mean square distance we obtain

$$N = \frac{dL^2 + e - D_0^2}{C_\infty b_0^2} \tag{4}$$

This equation can be combined with Eq.3 to predict the dependence of $\langle I \rangle$ on the number of residues, N.

By combining Eq.3 and Eq.4 we can estimate C_∞ for a flexible peptide of N residues directly from the experimentally measured $\langle I \rangle$ according to the following result:

$$C_\infty = \frac{[d(b(\langle I \rangle^{-1} - 1)^{1/c} + a)^2] + e - D_0^2}{N b_0^2} \tag{5}$$

The experimental points presented in Fig.1 (symbols) were used in the above result to obtain values of C_∞ for various systems. The averages of the experimentally determined C_∞ values for each system are given in Table 1. Using these results in our analytical expression for the dependence of $\langle I \rangle$ on N gives the green dashed lines displayed in Fig.1 which very accurately reproduce the available experimental results. Additional studies show that $\langle I \rangle$ is slightly sensitive to the choice of R_0, and practically invariant to small changes in D_0 as shown in Fig.S2 suggesting that application of Eq.5 is robust to variations in R_0 and D_0.

Deviation between our calculated model results and experiments observed for short length flexible peptides are due to the inadequacy of our continuum model as explained by Evers *et al.*[28]

Table 1 Table showing different flexible linker systems used in experiments and their corresponding C_∞ predicted by our model.

Linker	Residues	Chromphore Pair	R_0	Predicted C_∞
$(SAGG)_{13-61}$ [5]	52-244	ECFP-YPet	49	1.3
$(GGSGGS)_{1-8}$ [28]	23-71	ECFP-EYFP	49	2.1
$(GGGGS)_5$ [31]	25	ECFP-EYFP	49	2.6
$(GGGGS)_{3-4}$ [11]	19,25	EBFP-EGFP	48	3.0

We have shown that the trend in FRET efficiency with linker length predicted by the simple effective potential model is very similar to that observed experimentally. To further explore to validity of our simple model we have conducted a series of more
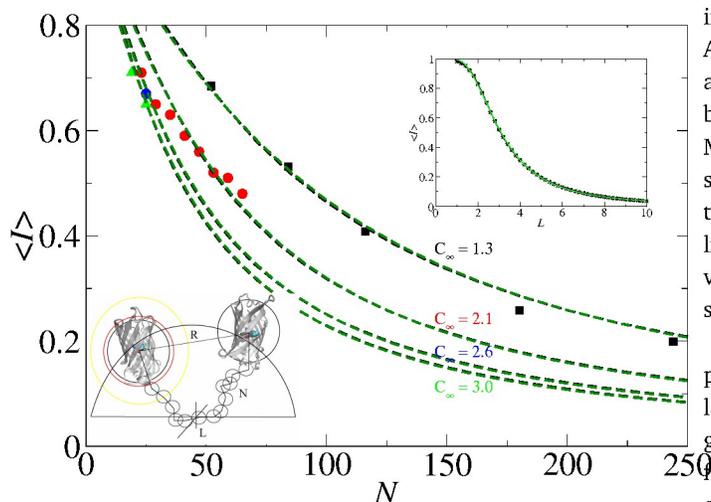
Fig. 1 FRET efficiency $\langle I \rangle$ as a function of number of linker residues N, presented for experimental measurements in the basal state using linkers made of (1) $(SAGG)_{13-61}$ [5] (filled black squares), (2) $(GGSGGS)_{1-8}$ [28] (filled red circles) (3) $(GGGGS)_5$ [31] (filled blue circles), and (4) $(GGGGS)_{3-4}$ [11] (filled green triangles). Superimposed on these points are using our theoretical predictions for flexible linker model computed by mapping L of to N for various values of C_∞ (black dashed lines), and using the algebraic expression (green dashed lines). Inset (Top Right) showing FRET efficiency $\langle I \rangle$ as a function of length L (measured in units of σ) computed through Monte Carlo simulation. Inset (Bottom Left) showing the schematic representation of the effective linker model.

and more realistic and detailed microscopic simulations. We thus compare various pair distribution functions obtained using the simple model with results from different detailed microscopic model calculations. This comparison has been conducted for five different probe systems. For the first two systems, the linker sequences $(SAGG)_{13}$, $(SAGG)_{21}$ were devised by Komatsu et.al.[5], the third and fourth $(GGSGGS)_4$ and $(GGSGGS)_8$ were examined by Merkx *et al.*[28], and the fifth $(GGGGS)_5$[31] considered by Li *et al.*[31] For each of these systems we computed center-to-center pair distribution function of the macroparticles, $P^{macro}(R)$, and the end-to-end distance distribution of the isolated linker (*i.e.* in the absence of the FP's), $P^{linker}(r)$. These results were obtained using our effective potential model as well as: (1) Gaussian chain model, (2) non-overlapping spheres at the ends of bead and spring polymer, and (3) atomistic simulation using a coarse-grained MARTINI force field[32] (details are given in section 2 of the Supplementary Information).

The pair distribution functions of the first system are plotted in Fig.2, and for the remaining systems in Fig.S3. Our results show that these distributions have similar shapes, with some notable differences. For instance, the $P^{macro}(R)$ computed with the simple model drops abruptly to zero for low and high values of R, unlike the more detailed GCM and bead and spring models that show long-range tails. In marked contrast to the smooth distributions of the simpler models, the most realistic MARTINI model has a

complex structure that arises from the realistic treatment of the anisotropy of the interactions between the FPs that are cylindrical in shape rather than spherical as assumed in our simpler models. Again, with our simple model, the $P^{linker}(r)$ distribution cuts off at large values of r but reproduces the general shape of the distributions predicted by the simpler models at small r. The detailed MARTINI model, on the other hand, includes a coarse grained description of the excluded solvation volume resulting from explicit treatment of the water solvent around the ends of the flexible linker. The simpler models that have implicit treatment of solvent cannot reproduce the MARTINI model $P^{linker}(r)$ behaviour at short distances.

While Fig.2 suggests that the detailed differences in the computed distribution functions for the different models can be rather large, when the different $P^{macro}(R)$ distributions are used, together with the function $I(R)$ from Eq. 1, to compute $\langle I \rangle$, we find, as summarized in Table S1, that the average FRET efficiencies for these markedly different models are actually very similar due to cancellation effects of different peaks in the distributions and the fact that $I(R)$ decays quickly beyond R_0, which is typically shorter than the linker length R. Since these differences in distributions have little effect on $\langle I \rangle$ they do not bear significantly on the estimates of the characteristic ratio C_∞ obtained from Eq. 5.

In conclusion, the current study proposes a novel approach for predicting the characteristic ratio, C_∞, of a given polymer linker system whose FRET efficiency $\langle I \rangle$ has been measured in the absence of binding interaction between the FPs. The present work enables the accurate measurement of flexibility of intrinsically disordered linkers *in vivo*, in particular in complex local sub-cellular environments using existing FRET microscopy methods.

References

1   S. D. Briggs and T. Smithgall, *J. Biol. Chem.*, 1999, **274**, 26579–26583.
2   R. Gokhale, S. Tsuji, D. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.
3   M. Ikebe, T. Kambara, W. Stafford, M.Sata, E. Katayama and R. Ikeb, *J. Biol. Chem.*, 1998, **273**, 17702–17707.
4   R. George and J. Heringa, *J. Protein. Eng.*, 2002, **15**, 871–879.
5   N. Komatsua, K. Aoki, M. Yamadac, H. Yukinagac, Y. Fujitac, Y. Kamioka, and M. Matsuda, *Mol. Biol. Cell.*, 2011, **22**, 4647–4656.
6   R. D. Fritz, M. Letzelter, A. Reimann, K. Martin, L. Fusco, L. Ritsma, B. Ponsioen, E. Fluri, S. Schulte-Merker, J. v. Rheenen and O. Pertz, *Science*, 2013, **6**, rs12.
7   J. S. Klein, S. Jiang, R. Galimidi, J. R. Keeffe and P. Bjorkman, *Protein Engineering, Design and Selection*, 2014, **27**, 325–330.
8   W. Wriggers, S. Chakravarty and P. Jennings, *Biopolymers*, 2005, **80**, 736–746.
9   Y. Maeda, H. Ueda, T. Hara, J. Kazami, G. Kawano, E. Suzuki, and T. Nagammune, *Biotechniques*, 1996, **20**, 116–121.
10 Y. Maeda, H. Ueda, J. Kazami, G. Kawano, E. Suzuki and T. Nagammune, *Anal Biochem*, 1997, **249**, 147–152.
11 R. Arai, H. Ueda and T. Nagamune, *J. Ferment. Bioeng.*, 1998, **86**, 440–445.
12 C. Suzuki, H. Ueda, K. Tsumoto, W. Mahoney, I. Kumagai and T. Naggamune, *J. Immunol. Methods*, 1999, **224**, 171–184.
13 L. Bulow, *Eur. J. Biochem*, 1987, **163**, 443–448.
14 J. Lichty, J. Malecki, H. Agnew, D. Michelson-Horowitz and S. Ta, *Protein Expr. Purif*, 2005, **41**, 98–105.
15 Y. Zou, W. I. Weis and B. Kobilka, *PLoS ONE*, 2012, **7**, e46039.
16 X. Chen, J. Zaro and W. Shen, *Adv. Drug Deliv. Rev.*, 2013, **65**, 1357–1369.
17 R. Arai, W. Wriggers, Y. Nishikawa, T. Nagamune and T. Fujisawa, *Proteins Struct Funct Bioinformatics*, 2001, **14**, 529–532.
18 P. R. Schimmel and P. J. Flory, *Proc. Natl. Acad. Sci. U. S. A.*, 1967, **58**, 52–59.
19 D. A. Brant and P. J. Flory, *Proc. Natl. Acad. Sci. U. S. A.*, 1965, **58**, 428–435.
20 L. Feigin and D. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, NY: Plenum Press, 1987.
21 D. Svergun, *J. Appl. Crystallogr.*, 1992, **25**, 495–503.
22 M. L. Finnegan and B. Bowler, *Biophys J.*, 2012, **102**, 1969–1978.
23 W. Miller, D. A. Brant and P. Flory, *J. Mol. Biol.*, 1967, **23**, 47–65.
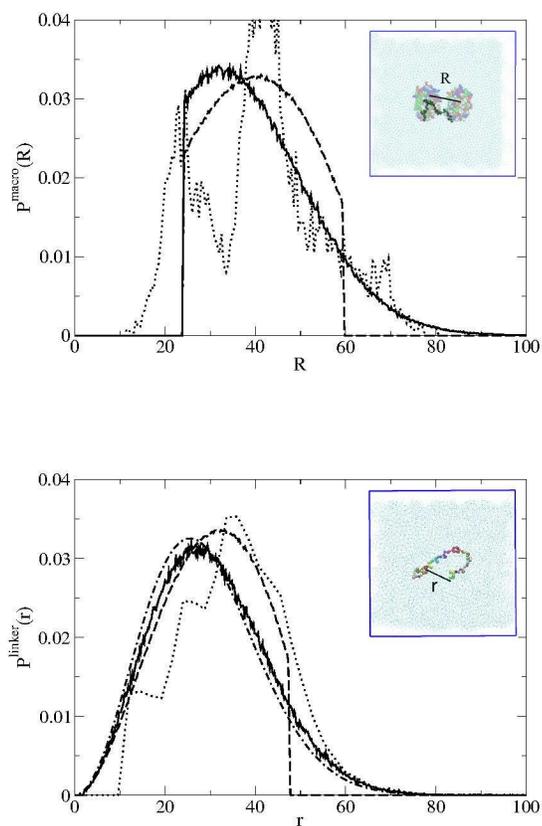
Fig. 2 Center-to-center distance distributions of the macroparticles $P^{macro}(R)$ (Top) and the end-to-end linker distance distribution $P^{linker}(r)$ (Bottom) for $(SAGG)_{13}{}^{5}$ (distances are measured in Angstroms), computed using: (1) Effective Potential (Dashed Line), (2) GCM (Dot and dashed line), (3) Bead and spring (Full line) and (4) MARTINI (Dotted line). (Insert top) The linker and FP system modelled using MARTINI. (Inset bottom) The linker modelled using MARTINI.

24 P. J. Flory, *Statistical Mechanics of Chain Molecules*, Interscience Publishers, 1969, vol. 38.
25 O. Kratky and G. Porod, *Recl. TraV. Chim. Pays-Bas.*, 1949, **68**, 1106–1122.
26 I. V. Gopich and A. Szabo, *J. Phys. Chem. B.*, 2003, **107**, 5058–5063.
27 J. A. Drake and B. M. Pettitt, *J. Comput. Chem.*, 2016, **36**, 1275–1285.
28 T. H. Evers, E. M. v. Dongen, A. C. Faesen, E. W. Meijer and M. Merkx, *Biochemistry*, 2006, **45**, 13183–13192.
29 H. Wallrabe and A. Periasamy, *Curr. Opin. Biotechnol.*, 2005, **16**, 19–27.
30 D. Frenkel and B. Smit., *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, Inc., 1996.
31 G. Li, Z. Huang, C. Zhang, B. DonG, R. Guo, H. Yue, L. Yan and X. Xing, *Appl. Microbiol. Biotechnol.*, 2015, **100**, 215–225.
32 L. Monticelli, S. Kandasamy, X. Periole, R. G. Larson, D. Tieleman and S. Marrink., *JCTC*, 2007, **4**, 819–834.