A fit-less approach to the elasticity of the handles in optical tweezers experiments

Alessandro Mossa\textsuperscript{1,2} · Ciro Cecconi\textsuperscript{3,4}

Received: 22 February 2022 / Revised: 1 May 2022 / Accepted: 2 May 2022 / Published online: 23 May 2022 © European Biophysical Societies’ Association 2022

Abstract
The elastic properties of the double-stranded DNA handles used in optical tweezers experiments on biomolecules are customarily modeled by an extensible worm-like chain model. Fitting such a model to experimental data, however, is no trivial task, as the function depends on four parameters in a highly non-linear fashion. We hereby propose a method to bypass the fitting procedure and obtain an empirical force vs. extension curve that accurately reproduces the elasticity of the handles.

Keywords Optical tweezers · Data analysis

Introduction
Optical tweezers are nowadays one of the main experimental techniques used in single-molecule biophysics (Zaltron et al. 2020; Lehmann et al. 2020). Proteins or nucleic acids can be manipulated by force in the piconewton scale by exploiting the ability of focused laser light to trap micrometer-sized polystyrene beads (Wang et al. 2021). To avoid non-specific interactions, the molecule of interest is tethered to the beads by means of a pair of double-stranded DNA handles (see Fig. 1), whose length is usually chosen in the range 20–800 base pairs (Synakewicz et al. 2019).

Among the various equilibrium and non-equilibrium experiments that can be performed within this setup, here we are concerned with the force-ramp protocol, where the distance between the trap and a reference point (which is usually either the tip of a pipette, or another optical trap) is increased (stretching phase) or decreased (relaxation phase) at a constant speed (Caldarini et al. 2014). Throughout the process, both the force that the trap exerts on the trapped bead and the distance between the beads are measured. The force vs. extension curve (shortened to FEC henceforth) obtained in this way is dominated by the elastic behaviour of the handles (Smith et al. 1996), which one needs to estimate and subtract during data analysis to get useful information about the molecule under study.

The most common approach is to model the elasticity of the handles with an extensible worm-like chain (Wang et al. 1997). This requires a non-linear fit with four parameters, which can (actually, very often does) give poor agreement with the experimental data, or assign very unlikely values to the parameters. In the following, we will illustrate the shortcomings of this standard recipe, and propose a novel method to deal with handles elasticity which does not depend on any fitting procedure. The first step is to build an empirical FEC by combining data belonging to the low-force regime of a stretching trace with data from the high-force regime of a relaxation trace. From such baseline a reference grid of FECs can be constructed and employed to ease the data analysis process.

To make the presentation more definite, we will refer to a specific example, taken from a recent experiment\textsuperscript{1}. In the concrete application we describe, the system under scrutiny is a construct comprised of four repeats of maltose-binding

---

\textsuperscript{1} Choudhary D, Mediani L, Avellaneda MJ, Alberti S, Boczek E, Mossa A, Tans S, Carra S, Cecconi C, Single-molecule investigation of HSPB8 holdase activity (to be submitted).
Elastic models

When considering the response to tension of the molecular construct depicted in Fig. 1, it is useful to divide the distance \( x \) between the beads into three contributions:

\[
x = x_h + x_f + x_u, \tag{1}
\]

where \( x_h \), \( x_f \), \( x_u \) are the extensions of the handles, of the folded portion of the protein, and of the unfolded peptide chain, respectively.

The equilibrium elastic behaviour of the folded protein \( x_f(F) \) at force \( F \) may be modeled by a freely-jointed chain of a polymer with identical Kuhn and contour lengths (Alemany and Ritort 2014):

\[
x_f(F) = d \left[ \coth \left( \frac{Fd}{k_BT} \right) - \frac{k_BT}{Fd} \right], \tag{2}
\]

where \( k_b \) is Boltzmann’s constant, \( T \) is the absolute temperature and \( d \) is the contour length of the folded protein, which can be estimated from the structures deposited in the Protein Data Bank\(^3\) as \( d = (3.99 \pm 0.01) \text{ nm} \) (reduced to \( d = (1.756 \pm 0.003) \text{ nm} \) for the core structure alone).

Following a well established practice, the equilibrium elastic properties of the unfolded peptide chain are modeled by a Marko–Siggia (Bustamante et al. 1994) worm-like chain (WLC for short) with the correction terms computed by Bouchiat et al. (1999)

\[
F(x_u) = \frac{k_BT}{L_p} \left[ \frac{1}{4(1 - \ell^2)} - \frac{1}{4} + \ell + \sum_{i=2}^{7} a_i \ell^i \right], \tag{3}
\]

with \( \ell = x/L_p \) and

\[
\begin{align*}
    a_2 &= -0.5164228 \quad a_3 = -2.737418 \\
    a_4 &= 16.07497 \quad a_5 = -38.87607 \\
    a_6 &= 39.49944 \quad a_7 = -14.17718.
\end{align*}
\]

The choice of parameters is the same that has been successfully adopted in other past experiments (Alemany et al. 2016; Sonar et al. 2020), that is persistence length \( L_p = 0.65 \text{ nm} \) and contour length \( L_c = n \ell_0 \), where the inter-residue distance \( \ell_0 = 0.34 \text{ nm} \) is multiplied by the number \( n \) of unfolded amino acids\(^4\).

The two models in Eqs. (2) and (3) only consider entropic effects, which is adequate for typical optical tweezers experiments. The DNA handles, however, get usually stretched to the point that it is no longer possible to disregard enthalpic effects.

\(^3\) PDB ID: 2MV0, doi:10.2210/pdb2MV0/pdb.

\(^4\) Actually, in the cited references we used \( L_p = 0.60 \text{ nm} \); the larger value has been chosen here to optimize Fig. 4, as explained towards the end of Sec. 3.
It is customary to deal with them by modifying the WLC into an extensible version of it (eWLC in the following) (Wang et al. 1997) which is still written as Eq. (3) with the understanding that

$$\ell = \frac{x}{L_c} - \frac{F}{Y},$$

(4)

where $Y$ is the Young modulus. Modeling the handles elasticity with an eWLC, therefore, requires a highly non-linear fitting procedure over four parameters: $L_p$, $L_c$, $Y$ and $x_0$ because, as we have already remarked, the measured extension is shifted with respect to the physical extension by an unknown additive term.

It is easy to realize what can go wrong with such a fit. With reference to Fig. 2, the range of data where only the elastic effect of the handles is relevant is restricted to the region below 14 pN, which is precisely the region where the effect of the enthalpic correction is negligible, so we can expect to have huge uncertainty in the determination of $Y$. However, the Young modulus so poorly determined is essential when we extrapolate the theoretical curve for handles elasticity to the high-force region, where it is actually more useful. Fortunately, as we will show in the next section, there is a way to bypass the fit altogether.

**Building a baseline, then a grid**

While fitting the eWLC to handles elasticity is tricky, the model in Eq. 3 does a very good job of describing the elastic behaviour of the unfolded peptide chain. So good that we can actually use it to subtract its contribution from the relaxation trace in Fig. 2, thus obtaining an empirical curve that, in the region $F > 10$ pN, before compaction begins, depends only on the behaviour of the handles.

Let $x_h(F)$ denote the experimental relaxation trace. We know that at very high forces the tetramer is completely denatured, so it behaves like a WLC with $L_p = 0.65$ nm and $L_c = 512.38$ nm (the total number of residues is 1507, including 27 amino acids employed as linkers between MBP repeats). Let $x_u(F)$ be the function defined by numerical inversion of Eq. (3) (see Appendix A); we can compute a shifted relaxation curve

$$\tilde{x}_h(F) \equiv x_h(F) - x_u(F),$$

(5)

and compare it to the stretching trace $x_s(F)$, as done in Fig. 3. The reasons why we are disregarding in this analysis the elastic properties of the folded MBP repeats given by Eq. (2) are discussed towards the end of this section.

The striking feature of Fig. 3 is the overlapping in the range of forces from 6 pN to 14 pN between the shifted relaxation curve (in colour blue) of Eq. (5) and the stretching trace (in colour red). This suggests a simple recipe to deal with handles elasticity: if we join the low-force portion of the stretching trace, which corresponds to a phase where the protein is fully folded and no structure has been denatured yet, with the high-force portion of the shifted relaxation trace, which corresponds to a phase where the protein is fully unfolded and no structure as been formed yet, we obtain what we call the baseline, an empirical reference curve that describes the elastic properties of the handles alone. In practice, we choose some force threshold $F^*$ in the region where the shifted relaxation curve and the stretching trace overlap (for instance, in the case of Fig. 3 it is $F^* = 13.22$ pN), and define the baseline $x_b(F)$ as

$$x_b(F) = \begin{cases} x_s(F) & \text{for } F < F^* \\ x_u(F) & \text{for } F > F^*. \end{cases}$$

(6)

Once the baseline is built, we can smooth it by applying a suitable filter, or we can fit an eWLC model to it: the already mentioned drawbacks of the fitting procedure are mitigated by the fact that now we are working with data that reach into the high-force regime. Either way, using the function $x_u(F)$ we can add back to the baseline the elastic behaviour of an $n$-residue peptide chain, with $n$ going from 0 to the total size of the protein (1507 in our specific example), thus obtaining a reference grid of curves that can be used to reveal the presence of intermediate structures, and estimate their size.

The use of a tetramer construct allows to perform a neat consistency check of the whole approach: Fig. 4 shows the stretching trace of Fig. 2 plotted against the grid lines corresponding to (from top to bottom) the wholly denatured protein ($n = 1507$), one to four core structures still folded ($n = 1228, 949, 670, 391^5$), and the completely folded native structure ($n = 0$). The excellent agreement between the computed lines and the experimental trace must be interpreted as

---

5 The alpha-helices account for 91 residues for each MBP repeat, plus there are 27 amino acids employed as linkers, for a total of 391 aa involved in the transition between the two lower lines of the grid.
a statement about the goodness of the choice made for the parameters in Eq. (3).

The same grid can be used to analyze all the experimental traces obtained from the same individual molecule; see for instance Fig. 5 where the grid, built for this example so far, allows to recognize and size up, in another stretching trace belonging to the same dataset, the presence of a core that breaks around 14 pN, a smaller structure that breaks at about 40 pN, and an aggregate that remains unbroken at very high forces.

One could argue that instead of Eq. (5) the proper subtraction should be

$$\tilde{x}_r(F) \equiv x_r(F) - x_u(F) + x_f(F),$$

(7)

taking into account also the contribution of the folded structures when they are present. After all, four MBP proteins in the native state at 12 pN tension are barely shorter than 16 nm, a shift that would be clearly visible, e.g. in Fig. 4. The point though is that, while in a cycle that starts from the native state it is easy to know with a certain confidence what is folded and what is not at various stages of the experiment, this is impossible for a generic trace like the one in Fig. 5. In a pragmatic perspective, it is more useful to adjust the parameters for the WLC model of the peptide chain ignoring the folded structures’ contribution: the small error is absorbed into the choice of the elastic model for the unfolded protein and we can use the grid to study any trace. This also explains why the value for the persistence length of the denatured protein that we used here ($L_p = 0.65$ nm) is slightly larger than the value $L_p = 0.60$ nm we used in Alemany et al. (2016) and Sonar et al. (2020): it must compensate for our neglect of the size of folded structures.

**Conclusion**

In accord with the principle of making science as open and reproducible as possible, the data and the program used to produce the figures for this paper are publicly available on a GitHub repository6. The interested readers can easily adapt the Python code to apply the baseline and grid method to their own data.

We have seen that the elastic behaviour of the handles can be represented by an empiric baseline obtained from experimental data without the need to fit the eWLC model. In addition, a grid of reference curves can be built on top of the baseline and used to identify and size up intermediate states along the stretching or relaxation curves. It remains to be discussed how general is this method, and what are its limitations. First of all, there is nothing strictly peculiar to the tetramer example that has been used so far; Fig. 6 demonstrates the effectiveness of the baseline and grid construction in the case of data obtained with a single MBP, while the Jupyter Notebook available as an online resource for this paper contains other

---

6 https://github.com/AlessandroMossa/fit-less-handles
examples about the Acyl-CoA-Binding Protein (Sonar et al. 2020) and the Neuronal Calcium Sensor-1 (Naqvi et al. 2015).

Of course, the more details we know about the protein that is being pulled, the more possibilities we have of setting non-trivial consistency checks like the one in Fig. 4. While the general idea is still valid, it is more difficult to be completely confident in the baseline built for a molecule about which one knows nothing except the structure. Another limitation that comes to mind is that for some systems it might be impossible to achieve the overlap between the stretching trace and the shifted relaxation trace shown in Figs. 3 and 6: this is the case if unfolding in the stretching curve starts at a force for which refolding is already underway in the relaxation curve.

Apart from such peculiar cases, however, we think that the approach we have introduced is a definite improvement over the standard practices in the field of optical tweezers force-ramp experiments. It can be used either to provide a better sample for an eWLC fit, or to avoid the fit entirely. In any way, it provides a reliable method to factor out the elasticity of the handles, allowing to focus the attention on the properties of the system we are actually interested in.

Appendix A An inversion formula for the WLC

The original Marko–Siggia formula for the WLC model

\[ f(\ell) = \frac{1}{4(1-\ell)^2} - \frac{1}{4} + \ell, \]  

with \( \ell = x/L_c \) and \( f = FL_p/(k_B T) \) can actually be inverted explicitly by solving a cubic equation (Severino et al. 2019). The solution is given by Cardano’s formula

\[ \ell = \frac{4f^9 + 9}{12} + \sqrt{-\frac{\delta(f) + 108}{1728} + \sqrt{\frac{\delta(f) + 108}{48\sqrt{3}} + \sqrt{-\frac{\delta(f) + 108}{1728} - \frac{\delta(f)}{48\sqrt{3}}}}} \]

(9)

if \( f < f_0 \) and by Viète’s formula

\[ \ell = \frac{4f^9 + 9}{12} - \frac{4f - 3}{6} \cos \left[ \frac{\pi}{6} - \frac{1}{3} \arcsin \frac{\delta(f) + 108}{(4f - 3)^{3/2}} \right] \]

(10)

if \( f > f_0 \), with \( f_0 = 3(1 + \sqrt{4})/4 \) and

\[ \delta(f) = -64f^3 + 144f^2 + 108f + 135. \]

(11)

This may be used to provide a very good initial estimate to a numerical routine for inverting Eq. (3), as is done in the online resource that complement this paper, or to help setting up a fit routine. In this regard, note that the extensible version of Eq. (8), which is obtained by replacing \( \ell \) with \( \tilde{\ell} \) defined as

\[ \tilde{\ell} = \ell - kf, \quad k \equiv \frac{k_B T}{L_p Y}. \]

(12)

can also be inverted in the same way, yielding

\[ \tilde{\ell} = \ell + kf \]

(13)

where \( \ell \) is given by Eqs. (9) and (10).

Author contributions AM devised the method here described while analyzing data from CC’s experiments. Both authors wrote the paper.

Funding CC gratefully acknowledges the University of Modena and Reggio Emilia for financial support through the grant 020145_17_FDA_CARRAFAR2016INTER.

Data availability The data and the program used to produce the figures for this paper are publicly available in the GitHub repository https://github.com/AlessandroMossa/fit-less-handles.

References

Alemany A, Ritort F (2014) Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins. Biopolymers 101(12):1193–1199. https://doi.org/10.1002/bip.22533

Alemany A, Rey-Serra B, Frutos S, Cecconi C, Ritort F (2016) Mechanical folding and unfolding of protein barnase at the single-molecule level. Biophys J 110(1):63–74. https://doi.org/10.1016/j.bpj.2015.11.015

Bechtluft P, van Leeuwen RGH, Tyreman M, Tomkiewicz D, Nouwen N, Tepper HL, Driessen AJM, Tans SJ (2007) Direct observation of chaperone-induced changes in a protein folding pathway. Science 318(5855):1458–1461. https://doi.org/10.1126/science.1144972

Bouchiat C, Wang MD, Allemand JF, Strick T, Block SM, Croquette V (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. Biophys J 76(1):409–413. https://doi.org/10.1016/S0006-3495(99)77207-3

Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of λ-phage DNA. Science 265(5178):1599–1600. https://doi.org/10.1126/science.8079175

Caldarini M, Sonar P, Valpaparam I, Tavella D, Volonté C, Pandini V, Vanoni MA, Aliverti A, Broglia RA, Tiana G, Cecconi C (2014) The complex folding behavior of HIV-1-protease monomer revealed by optical-tweezer single-molecule experiments and molecular dynamics simulations. Biophys Chem 195:32–42. https://doi.org/10.1016/j.bpc.2014.08.001

Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D (2012) Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proc Natl Acad Sci USA 109(27):10,873–10,878. https://doi.org/10.1073/pnas.1203013109

Lehmann K, Shayegan M, Blab GA, Forde NR (2020) Optical tweezers approaches for probing multiscale protein mechanics and assembly. Front Mol Biosci 7:577314. https://doi.org/10.3389/fmolb.2020.577314

Naqvi MM, Heidarsson PO, Otazo MR, Mossa A, Kragelund BB, Cecconi C (2015) Single-molecule folding mechanisms of the
apo- and Mg$^{2+}$-bound states of human neuronal calcium sensor-1. Biophys J 109(1):113–123. https://doi.org/10.1016/j.bpj.2015.05.028

Severino A, Monge AM, Rissone P, Ritort F (2019) Efficient methods for determining folding free energies in single-molecule pulling experiments. J Stat Mech 2019:124001. https://doi.org/10.1088/1742-5468/ab4e91

Smith SB, Cui Y, Bustamante C (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science 271(5250):795–799. https://doi.org/10.1126/science.271.5250.795

Sonar P, Bellucci L, Mossa A, Heidarsson PO, Kragelund BB, Cecconi C (2020) Effects of ligand binding on the energy landscape of Acyl-CoA-binding protein. Biophys J 119(9):1821–1832. https://doi.org/10.1016/j.bpj.2020.09.016

Synakewicz M, Bauer D, Rief M, Itzhaki LS (2019) Bioorthogonal protein-DNA conjugation methods for force spectroscopy. Sci Rep 9:13820. https://doi.org/10.1038/s41598-019-49843-1

Wang MD, Yin H, Landick R, Gelles J, Block SM (1997) Stretching DNA with optical tweezers. Biophys J 72(3):1335–1346. https://doi.org/10.1016/S0006-3495(97)78780-0

Wang Y, Kumar A, Jin H, Zhang Y (2021) Single-molecule manipulation of macromolecules on GUV or SUV membranes using optical tweezers. Biophys J 120(24):5454–5465. https://doi.org/10.1016/j.bpj.2020.11.2884

Zaltron A, Merano M, Mistura G, Sada C, Seno F (2020) Optical tweezers in single-molecule experiments. Eur Phys J Plus 135(11):896. https://doi.org/10.1140/epjp/s13360-020-00907-6

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.