SISSA

STOCHASTIC THERMODYNAMICS

Efficient methods for determining folding free energies in single-molecule pulling experiments

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Efficient methods for determining folding free energies in single-molecule pulling experiments

A Severino[1]**, A M Monge**[1]**, P Rissone**[1]** and F Ritort**[1,2]

[1] Departamento de Física de la Matèria Condensada, Universitat de Barcelona, C/ Martí i Franquès 1, 08028 Barcelona, Spain
[2] CIBER-BBN de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, 28029 Madrid, Spain
E-mail: fritort@gmail.com

Abstract. The remarkable accuracy and versatility of single-molecule techniques make new measurements that are not feasible in bulk assays possible. Among these, the precise estimation of folding free energies using fluctuation theorems in nonequilibrium pulling experiments has become a benchmark in modern biophysics. In practice, the use of fluctuation relations to determine free energies requires a thorough evaluation of the usually large energetic contributions caused by the elastic deformation of the different elements of the experimental setup (such as the optical trap, the molecular linkers and the stretched-unfolded polymer). We review and describe how to optimally estimate such elastic energy contributions to extract folding free energies, using DNA and RNA hairpins as model systems pulled by laser optical tweezers. The methodology is generally applicable to other force-spectroscopy techniques and molecular systems.

J. Stat. Mech. (2019) 124001

Contents

1. Introduction

Predicting free-energy differences is a central problem in molecular biophysics. Protein folding [1], DNA hybridization [2], ligand binding and CRISPR–Cas9 RNA editing [3, 4] are molecular reactions whose fate is determined by the free-energy difference between reactants and products. Finding methods to extract free-energy, enthalpy and entropy differences is an essential task in biochemistry, where most of these quantities are measured by employing bulk techniques such as calorimetry, UV absorbance, fluorescence and surface plasmon resonance, among others [5]. Bulk methods yield results that are incoherent temporal averages over a large population of molecules that are in different states. The signal is masked by the dominant species and reactions, limiting the capability of detecting rare non-native states and reaction pathways. Moreover, bulk molecular transformations often exhibit strong hysteresis effects rendering equilibrium differences inaccessible.

By monitoring molecules one at a time, techniques such as single-molecule fluorescence [6], single-molecule translocation across nanopores [7] and single-molecule force spectroscopy [8] overcome the previous limitations and therefore have become key experimental tools in many laboratories worldwide. In particular, force-spectroscopy

techniques using atomic-force microscopy, magnetic tweezers, acoustic-force spectroscopy and laser optical tweezers (LOTs) have been extremely fruitful, revolutionizing biophysics over the last three decades[3].

The main advantage of force-measuring techniques (as compared to fluorescence and other non-invasive optical technologies) lies in the possibility of measuring force and displacement simultaneously, giving direct access to mechanical work measurements in single-molecule pulling experiments. Similarly to bulk assays, pulling experiments are often carried out under irreversible conditions, in principle providing bounds (rather than direct estimates) of equilibrium free-energy differences. The development of the non-equilibrium thermodynamics of small systems (also known as stochastic thermodynamics) [10–13] during the past three decades has provided the theoretical concepts and methods needed to extract free-energy differences from repeated irreversible work measurements. Exact results such as the Jarzynski equality [14] and the Crooks fluctuation theorem [15] are now commonly employed to extract free-energy differences from single-molecule pulling experiments [16–21]. A particularly useful application is the measurement of the folding free energy of nucleic acids and proteins (ΔG_0), which is equal to the free energy difference between the folded structure and the unstructured random coil in the solvent. This quantity can be obtained from pulling experiments by measuring the free energy difference (ΔG) between the folded- and unfolded-stretched states of the considered experimental system taken at two force values, and by deriving from it the value of ΔG_0. However, a general problem in the manipulation of small systems using single-molecule techniques is that we cannot abstract away certain components or parts—generally known as instrumental artifacts—of the full experimental system. Hence, the quantity that can be obtained from pulling experiments using non-equilibrium thermodynamics is not ΔG_0 directly. It is instead the free energy difference ΔG between the folded- and unfolded-stretched states of the *entire* considered experimental system taken at two force values. In order to retrieve the 'bare' molecular properties such as the value of ΔG_0 in a single molecule, we therefore need to infer from ΔG contributions due to the experimental setup (e.g. the optical trap in LOTs or cantilever in AFM and the linkers used to manipulate the molecule under study). These so-called stretching corrections play a crucial role because their contribution to the total free energy difference ΔG is much larger than the free energy one wants to extract ΔG_0, making the accurate estimation of the latter a difficult task. Although there are several studies on the influence of the instrumental artifacts on the folding kinetics in single-molecule experiments [22–29], their influence regarding the determination of the folding free energies at zero force has, to the best of our knowledge, never been addressed in detail.

In this work we will rigorously examine these experimental contributions in LOTs showing how to efficiently and reliably estimate the free energies of formation of DNA and RNA hairpins in unzipping assays. The same methodology is applicable to proteins and ligand binding interactions using LOTs or other force measuring techniques as well (AFM, magnetic tweezers and so on). The development of novel and refined statistical analysis methods to extract differences in thermodynamic potentials (free energy,

enthalpy, entropy, chemical potential, ...) will become crucial with the recent boost of high-throughput single-molecule techniques (magnetic tweezers, acoustic force spectroscopy) that will require fast and efficient algorithms.

The content of this paper is organized as follows. In sections 2 and 3 we describe the typical experimental setup of LOTs and then define and discuss the different contributions to the total free energy. The two following sections 4 and 5 feature how to estimate these contributions when analyzing DNA and RNA molecules. Section 4 first covers the situations in which it is possible to introduce the so-called effective stiffness approximation, which considerably simplifies the computation of the large stretching terms. When this approximation fails, a more elaborate approach requiring a careful evaluation of the elastic response of the linkers and of the force probe is needed, and this is the focus of section 5. Finally, in section 6 we present the conclusions.

2. Model of the experimental setup

We consider the case of a nucleic acid (DNA or RNA) hairpin pulled by LOTs. In an LOT, the total distance λ between the tip of the micropipette and the center of the optical trap is the control parameter of the experiments. As shown in figures 1(a) and (b), the distance λ can be decomposed as:

$$\lambda(f) = \begin{cases} x_{\mathrm{b}}(f) + x_{\mathrm{h}}(f) + x_{\mathrm{d}}(f) + \mathrm{const} & (\text{folded state}), \\ x_{\mathrm{b}}(f) + x_{\mathrm{h}}(f) + x_{\mathrm{ss}}(f) + \mathrm{const} & (\text{unfolded state}), \end{cases} \tag{1}$$

depending on whether the molecule is folded or unfolded. Here $x_{\mathrm{b}}(f)$ is the displacement of the bead from the center of the optical trap, and $x_{\mathrm{h}}(f) = x_{\mathrm{h_1}}(f) + x_{\mathrm{h_2}}(f)$ accounts for the sum of the elongations of the two double-stranded handles, $x_{\mathrm{ss}}(f)$ is the end-to-end extension of the single-stranded unfolded molecule, and $x_{\mathrm{d}}(f)$ is the average extension of the folded hairpin. This last term is defined as the distance between the attachment points of the handles to the 5' and 3' ends of the hairpin and is usually called the 'hairpin diameter' (whence the index d). All these extensions are evaluated against the x-(pulling) axis and at a given force f. The 'const' stands for an arbitrary shift in the total distance λ, which does not affect the analysis.

In general, a small perturbation $\delta\lambda$ generates a small change in the applied force δf. The extent of this variation is the *effective stiffness* of the system $k_{\mathrm{eff}} = \delta f/\delta\lambda$ and it equals the slope of the experimental force-distance curve (FDC). Therefore, according to the above definition and to the prescription given in (1), the inverse effective stiffness of the folded (F) and unfolded (U) branches are respectively given by:

$$\frac{1}{k_{\mathrm{eff}}^{\mathrm{F}}(f)} = \frac{1}{k_{\mathrm{b}}(f)} + \frac{1}{k_{\mathrm{h}}(f)} + \frac{1}{k_{\mathrm{d}}(f)}, \tag{2a}$$

$$\frac{1}{k_{\mathrm{eff}}^{\mathrm{U}}(f)} = \frac{1}{k_{\mathrm{b}}(f)} + \frac{1}{k_{\mathrm{h}}(f)} + \frac{1}{k_{\mathrm{ss}}(f)} \tag{2b}$$
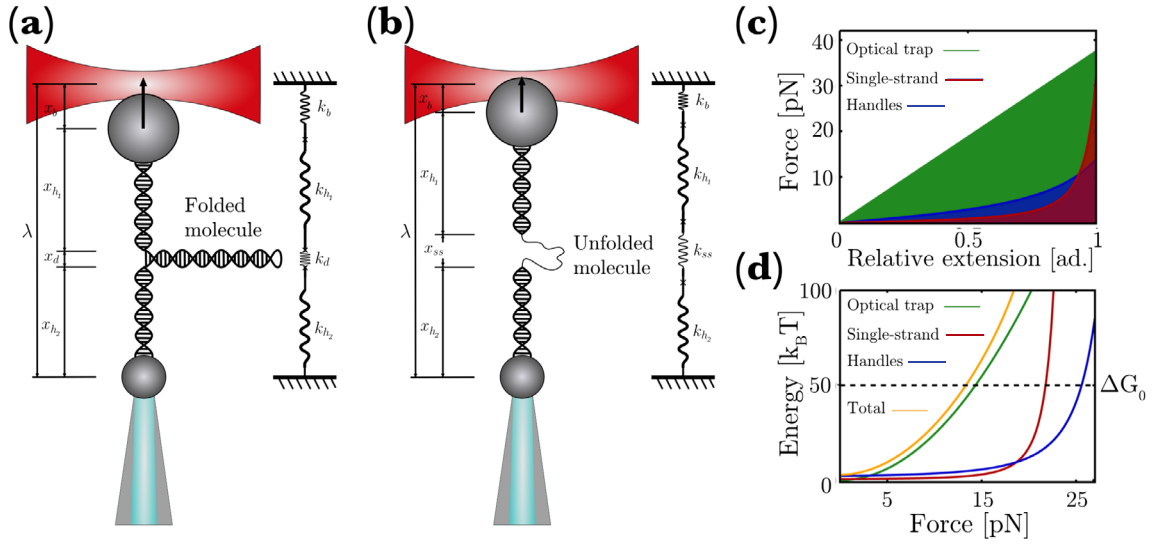
Figure 1. (a), (b). Laser optical tweezers (LOTs) experimental setup. The molecule is tethered between two polystyrene beads using two dsDNA (or dsRNA or even dsDNA/DNA hybrids) handles. The arrow towards the center of the optical trap indicates the direction of the force. $\lambda = x_{\rm b} + x_{\rm h} + x_{\rm m}$ (with $x_{\rm h} = x_{\rm h_1} + x_{\rm h_2}$) is the relative distance between the center of the optical trap and the tip of the micropipette. $x_{\rm m}$ equals $x_{\rm d}$ when the molecule is folded (a) or $x_{\rm ss}$ when the molecule is unfolded (b). (c) Sketch of the force versus relative extension (extension divided by contour length) for each elastic element showing their respective energy contributions (shaded areas). (d) Elastic energy contribution of each element versus force and comparison with the typical energy of formation (dashed line, ΔG_0) for a 20 bp DNA or RNA hairpin.

where $k_{\rm b}(f)$ corresponds to the stiffness of the bead in the optical trap, $k_{\rm h}(f)$ is the sum of the two handles' stiffness and $k_{\rm d}(f)$ and $k_{\rm ss}(f)$ stand for the molecular stiffness of the folded and unfolded molecule, respectively.

In particular, $k_{\rm d}(f)$ is modeled as the stiffness to orient a dipole of diameter d (typically $d = 2\,\text{nm}$ for DNA and RNA hairpins [30]) along the force axis [31]. Recalling that in general $k^{-1} = \delta x/\delta f$, the dipole stiffness can be derived from the well-known relation between a dipole average extension (which is here equal to the average extension of the folded hairpin) and the force f to which it is subjected:

$$x_{\rm d}(f) = d\left[\coth\left(\frac{fd}{k_{\rm B}T}\right) - \frac{k_{\rm B}T}{fd}\right] \tag{3}$$

where T is the temperature of the heat bath around the dipole and $k_{\rm B}$ is the Boltzmann constant.

An analytic expression for $k_{\rm ss}$ and $k_{\rm h}$ can be obtained by describing the elastic response of nucleic acids in their single-stranded and double-stranded form with the worm-like chain (WLC) polymer model and its interpolation formula [32],

$$f(x) = \frac{k_{\rm B}T}{4P}\left[\left(1 - \frac{x}{L_c}\right)^{-2} - 1 + 4\frac{x}{L_c}\right] \tag{4}$$

where x is the average extension of the molecule ($x = x_{ss}$ for the unfolded hairpin, $x = x_h$ for the double-stranded handles) and P is the persistence length, i.e. the typical distance along the polymer backbone over which there is an appreciable bending due to thermal fluctuations. L_c is the contour length, i.e. the end-to-end distance of the fully straightened polymer, which can also be written as $L_c = nd_b$ with n being the total number of monomers in the polymer and d_b the length per monomer. In general, inverting (4) to get $x(f)$ is not an easy task (the full computation is reported in the appendix) and the solution depends on the system parameters.

Finally, the stiffness of the polymer can be obtained by differentiation of (4):

$$k(x) \equiv \frac{\partial f(x)}{\partial x} = \frac{k_B T}{2L_c P} \left[\left(1 - \frac{x}{L_c} \right)^{-3} + 2 \right] . \tag{5}$$

Given (4), it is also possible to further take into account the elastic deformation of the stretched polymer by performing the substitution $L_c \to L_c(1 + f/Y)$, with Y the Young modulus of the stretchable polymer [33, 34], i.e. the resistance to deformation of the system to an applied uniaxial stress. In this case, the contour length becomes force-dependent and the corresponding model is called the *extensible* WLC. By contrast, equation (4) where L_c is constant is known as the *inextensible* WLC. The latter has been shown to describe the elastic properties of single-stranded nucleic acids (ssDNA and ssRNA) with good accuracy [35] while the former has for long been the standard to model the elastic properties of double-stranded nucleic acids in the entropic regime.

The persistence length P is a measure of the mechanical stiffness of the polymer being strongly sensitive to environmental conditions (e.g. ionic strength, temperature, solvation, etc). Polymers with $P \gg L_c$ effectively behave as rigid rods, whereas if $P \leqslant L_c$ polymers are bent at the scale of the contour length by thermal forces. It is important to mention that P does not only depend on the ionic concentration and temperature [36] (as predicted by polyelectrolyte theories) but also on experimental parameters such as contour length [35]. For example, at 1 M NaCl, recent single-molecule studies have shown that, for short (a few tens bases) ssDNA molecules, $P = 1.35$ nm [37] whereas for long ssDNAs ~ 13 kb $P = 0.76$ nm [38]. On the other hand, for short ssRNA molecules $P = 0.75$ nm [39] and for long ~ 1 kb ssRNAs $P = 0.83$ nm [35]. These values are significantly lower than for double-stranded nucleic acids (dsDNA and dsRNA) where $P = 50$ nm for dsDNA [40] and $P \simeq 60$ nm for dsRNA molecules [41].

3. Stretching contributions and free-energy recovery

Let us suppose that initially at $t = 0$ we have a molecule in thermal equilibrium at the folded (or native, N) state at a given value λ_0 of the control parameter. Then, we perturb the system by applying a predetermined time-dependent forward (F) protocol, $\lambda_F(t)$, which, starting at λ_0 at $t = 0$, ends at an arbitrary λ_1 at a time t_1. The mechanical work W done along this process is equal to:

$$W = \int_{\lambda_0}^{\lambda_1} f \mathrm{d}\lambda\,. \tag{6}$$

The Crooks fluctuation theorem (CFT) [15] relates to the mechanical work done on a system in a set of arbitrary irreversible measurements with the equilibrium free-energy difference of this system between λ_0 and λ_1, $\Delta G = G(\lambda_1) - G(\lambda_0)$. It reads:

$$\frac{P_{\mathrm{F}}(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_{\mathrm{B}}T}\right), \tag{7}$$

where $P_{\mathrm{F}}(W)$ is the probability distribution of the work done in the F process and $P_R(-W)$ is the probability distribution of the work measured in the time-reversed (R) process (i.e. starting in thermal equilibrium in λ_1 and performing the time-mirrored protocol so that $\lambda_R(t) = \lambda_{\mathrm{F}}(t_1 - t)$). The derivation of the CFT has become a milestone for single-molecule experimentalists, allowing the measurement of free-energy differences in conditions where traditional bulk experiments are unfeasible. By pulling single molecules using LOTs or magnetic tweezers, it is possible to recover molecular free-energy differences from irreversible work measurements [17, 42]. The CFT (equation (7)) implies the well-known Jarzynski equality [14]:

$$\left\langle \exp\left(-\frac{W}{k_{\mathrm{B}}T}\right)\right\rangle_{\mathrm{F}} = \exp\left(-\frac{\Delta G}{k_{\mathrm{B}}T}\right). \tag{8}$$

Note that the average $\langle\cdots\rangle_{\mathrm{F}}$ is evaluated over $P_{\mathrm{F}}(W)$ (an analogous equality holds for the reverse process). It is important to bear in mind that the free energy ΔG obtained using the CFT (7) (or the Jarzynski equality (8)) contains several contributions due to the stretching of the different parts forming the experimental setup. These are the molecules under study, the molecular handles and the optically trapped bead (figures 1(a) and (b)):

$$\Delta G = \Delta G_0 + \Delta W_{\mathrm{m}} + \Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}}, \tag{9}$$

where ΔG_0 is the free energy of formation of the molecule at zero force, which is equal to the free energy difference between the folded and unfolded hairpin conformations in solution (i.e. without optical trap and handles and without any applied force). The quantities ΔW_i ($i = m, b, h$) are the reversible work differences between the state of the ith setup element (optical trap, handles or molecule) at λ_0 (where the hairpin is folded and subjected to a minimum force $f_{\min}$) and λ_1 (where the hairpin is unfolded and subjected to a maximum force $f_{\max}$). Mathematical definitions of these quantities for the LOT setup are given in the subsections below.

As depicted in figures 1(c) and (d), for typical unfolding forces in DNA and RNA hairpins (15–25 pN), (9) is dominated by the trap contribution, while the other terms have the same order of magnitude. Therefore, an accurate measurement of ΔG_0 requires precise knowledge of all the different energetic contributions involved in the mechanical unfolding of the molecule.

3.1. Molecular stretching contribution

The molecular contribution ΔW_{m} in (9) accounts for the reversible work needed to stretch the molecule under study and it can be written as:

$$\Delta W_{\mathrm{m}} = \int_0^{x_{\mathrm{ss}}(f_{\max})} f(x_{\mathrm{ss}})\,\mathrm{d}x_{\mathrm{ss}} - \int_0^{x_{\mathrm{d}}(f_{\min})} f(x_{\mathrm{d}})\,\mathrm{d}x_{\mathrm{d}}\,, \tag{10}$$

where $f(x_{\mathrm{ss}})$ and $f(x_{\mathrm{d}})$ are the equilibrium force-extension curves of the unfolded and folded molecule, respectively (albeit different mathematical functions, the same letter f will be used to lighten the notation). The first term in the right-hand side of (10) corresponds to the reversible work needed to stretch the unfolded molecule from its single-stranded random coil conformation at $f = 0$ up to $f_{\max}$ and it can be computed from the WLC model, equation (4). The second term in the right-hand side of (10) is the reversible work required to orientate the molecular diameter along the force axis. It can be written as:

$$\int_0^{x_{\mathrm{d}}(f_{\min})} f(x_{\mathrm{d}})\mathrm{d}x_{\mathrm{d}} = f_{\min} \cdot x_{\mathrm{d}}(f_{\min}) - \int_0^{f_{\min}} x_{\mathrm{d}}(f)\mathrm{d}f \tag{11}$$

where $x_{\mathrm{d}}(f)$ is given by (3).

3.2. Bead and handles stretching contributions

The term $\Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}}$, which corresponds to the sum of the reversible work required to displace the bead from the center of the optical trap (ΔW_{b}) plus the reversible work needed to stretch the handles (ΔW_{h}), can be generally written as:

$$\begin{aligned}
\Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}} &= \int_{x_{\mathrm{b}}(f_{\min})}^{x_{\mathrm{b}}(f_{\max})} f(x_{\mathrm{b}})\,\mathrm{d}x_{\mathrm{b}} + \int_{x_{\mathrm{h}}(f_{\min})}^{x_{\mathrm{h}}(f_{\max})} f(x_{\mathrm{h}})\,\mathrm{d}x_{\mathrm{h}} \\
&= \int_{f_{\min}}^{f_{\max}} f\left(\frac{\partial f}{\partial x_{\mathrm{b}}}\right)^{-1}\mathrm{d}f + \int_{f_{\min}}^{f_{\max}} f\left(\frac{\partial f}{\partial x_{\mathrm{h}}}\right)^{-1}\mathrm{d}f \\
&= \int_{f_{\min}}^{f_{\max}} \frac{f}{k_{\mathrm{b}}(f)}\mathrm{d}f + \int_{f_{\min}}^{f_{\max}} \frac{f}{k_{\mathrm{h}}(f)}\mathrm{d}f\,.
\end{aligned} \tag{12}$$

Note that each element in the setup is substantially different. In particular, the bead in the optical trap can be well approximated by a Hookean spring, whereas the elastic response of the handles and the single-stranded molecule (plus the diameter) is strongly nonlinear (see below). The contribution of these two terms in equation (9) is often large. In particular, the energy required to displace the bead from the center of the optical trap is considerably higher as compared to the other terms. A schematic depiction of this fact can be seen in figure 1(c), where the shaded areas below the curves represent the work W obtained according to (6) using realistic elastic parameters for DNA and RNA hairpins.

3.3. Effective stiffness approximation

A further important simplification can be carried out when the FDC along the folded branch is approximately linear over the integration range of forces. Such a situation corresponds by definition to a scenario where the slope (or stiffness) is constant, i.e. $k_{\mathrm{eff}}^{\mathrm{F}} \neq k_{\mathrm{eff}}^{\mathrm{F}}(f)$. It allows one to readily perform the integration in equation (12), which is now reduced to the simple task of integrating an affine function:

$$\Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}} = \int_{f_{\min}}^{f_{\max}} f \left(\frac{1}{k_{\mathrm{b}}} + \frac{1}{k_{\mathrm{h}}} \right) \mathrm{d}f \cong \int_{f_{\min}}^{f_{\max}} \frac{f}{k_{\mathrm{eff}}^{\mathrm{F}}} \mathrm{d}f = \frac{f_{\max}^2 - f_{\min}^2}{2k_{\mathrm{eff}}^{\mathrm{F}}} , \quad (13)$$

where we used the fact that the stiffness of the dipole modeling the folded hairpin is considerably larger than the other terms in (2a), so that $k_{\mathrm{eff}}^{\mathrm{F}} = (k_{\mathrm{h}}^{-1} + k_{\mathrm{b}}^{-1} + k_{\mathrm{d}}^{-1})^{-1} \cong (k_{\mathrm{h}}^{-1} + k_{\mathrm{b}}^{-1})^{-1}$, and where the constant stiffness assumption is used in the last equality of the right-hand side of (13). Linearity of the FDC is a good approximation if the integration range is not too large (for example, when $f_{\max} - f_{\min} \approx 5$ pN, the case of the DNA and RNA hairpins considered in the next section.). Above all, linearity of the FDC certainly requires a linear optical trap of constant stiffness [31]. We will refer to this approximation as the *effective stiffness method*.

4. The effective stiffness method

The effective stiffness approximation discussed in section 3.3 provides an easy method to treat the elastic contributions of the experimental setup. Here we provide two typical scenarios where (13) provides a reliable estimation of the free energy of formation ΔG_0 of DNA and RNA hairpins. In section 4.1, the case of the CD4 DNA hairpin with short handles is reported. Then in section 4.2 we discuss the case of the CD4 RNA hairpin with long handles.

4.1. Short handles: the CD4 DNA hairpin

The use of short dsDNA handles ($\sim$29 bp each) in single-molecule experiments has been shown to increase the precision of kinetic measurements due to their enhanced signal-to-noise ratio as compared to long handles [31]. Short handles also make the evaluation of the stretching contributions easier. In fact, the large stiffness of short handles as compared to the trap stiffness, $k_{\mathrm{h}} \gg k_{\mathrm{b}}$, implies that $k_{\mathrm{eff}} \simeq k_{\mathrm{b}}$ to first order. As the trap stiffness itself can be considered nearly force independent $k_{\mathrm{eff}}^{\mathrm{F}}$ is, therefore, constant along the folded branch, and the effective stiffness approximation (13) becomes applicable.

We tested this approach using a 20 bp DNA hairpin ending in a tetraloop (sequence in top panel of figure 2(a)) flanked by two dsDNA handles, each one of 29 bp. The assembled molecular construct (DNA hairpin + handles) shown in figures 1(a) and (b), is repeatedly pulled between λ_0 and λ_1. In the forward (reverse) process the system starts in thermal equilibrium at λ_0 (λ_1) and it is driven out of equilibrium following a predetermined protocol $\lambda_{\mathrm{F}}(t)$ ($\lambda_{\mathrm{R}}(t)$) until λ_1 (λ_0) is reached. For each experimental
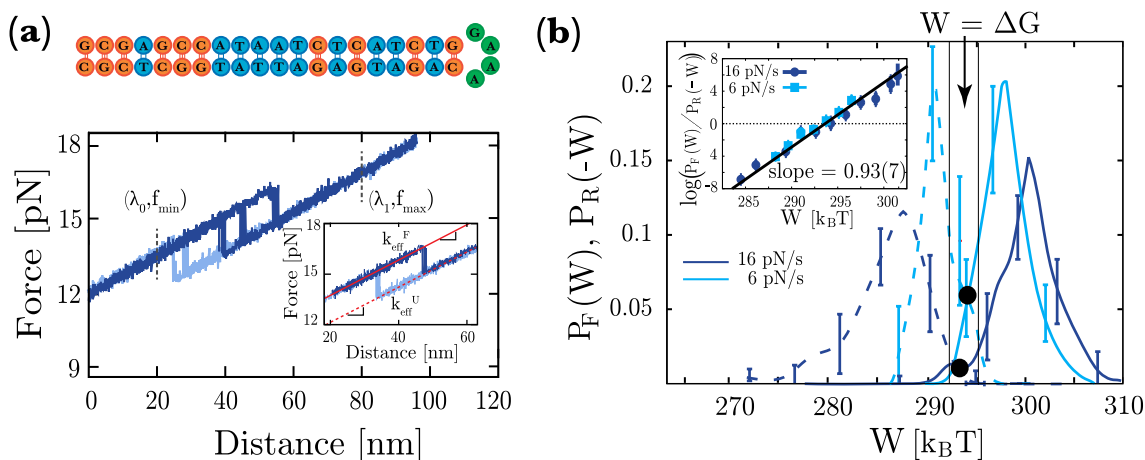
Figure 2. Free-energy recovery of CD4 DNA with short handles (a). Sequence of CD4 DNA (top panel), FDCs and integration range for the work W (bottom panel). Demonstration of the linearity of the FDCs in the integration range (inset) plus linear fits to the folded (solid line in the inset) and the unfolded branches (dashed line in the inset). (b) Forward (solid lines) and reverse (dashed lines) work distributions for two different pulling speeds calculated in the integration range indicated in panel (a). Crossing points between work distributions are tagged as solid points. The CFT verification is shown inset. Error bars have been obtained using the Bootstrap method.

realization the work W is calculated according to (6). Note that, in the force range at which the molecule typically unfolds and refolds (12–17 pN in figure 2(a)), the FDCs are linear in force (inset of figure 2(a)). Therefore, the conditions required to use the effective stiffness method are fulfilled (13).

In figure 2(b) we show the F and R work distributions calculated for two pulling speeds (6 and 16 pN s^{-1}) in the same integration range. According to the CFT (7), the work value at which both distributions cross (black solid points) are equal to ΔG. Note that, since the integration range is the same, ΔG does not change with pulling speed, as it is required for an equilibrium quantity. We emphasize the validity of the CFT by plotting the function $\log P_F(W)/P_R(-W)$ as a function of W in $k_B T$ units. According to (7), this function is linear in W with slope 1 and with a y-intercept equal to ΔG (both in $k_B T$ units). As expected, the experimental data (solid points) satisfy the previous relation (see inset of figure 2(b), where the solid line is a linear fit to the experimental data).

Once we have measured ΔG using the CFT, we subtract the stretching contributions to recover ΔG_0. According to (9), we have:

$$\Delta G_0 = \Delta G - \Delta W_m - \Delta W_b - \Delta W_h. \tag{14}$$

The term ΔW_m is calculated using (10). In order to model the ssDNA elastic response (i.e. $f(x_{ss})$ in (10)), we use the WLC model (4) with a persistence length P equal to 1.35 nm and an interphosphate distance d_b equal to 0.59 nm/base [37], so that $L_c = (2n_{bases} + 4) \times 0.59$ nm/base ≈ 26 nm. On the other hand, the term $\Delta W_h + \Delta W_b$ is calculated using the effective stiffness method (13) with $k_{eff}^F = 0.065 \pm 0.002$ pN nm^{-1} (obtained by a linear fit of the FDCs, see the inset in 2(a)).

Table 1. Fluctuation theorem and stretching contributions for the CD4 DNA hairpin (short handles) and CD4 RNA hairpin (long handles). (DNA short, first row) Reported energies for the integration range $[\lambda_0, \lambda_1] = [20, 80]$ nm corresponding to a force range $(f_{\min}, f_{\max}) = (13, 17)$ pN. (RNA long, second row) Reported energies for the integration range $[\lambda_0, \lambda_1] = [30, 85]$ nm corresponding to a force range $(f_{\min}, f_{\max}) = (18, 22)$ pN. Error bars obtained after averaging the results over four (DNA short) and five (RNA long) molecules at two pulling speeds, respectively.

	ΔG [$k_{\mathrm{B}}T$]	ΔW_{m} [$k_{\mathrm{B}}T$]	$\Delta W_{\mathrm{h}} + \Delta W_{\mathrm{b}}$ [$k_{\mathrm{B}}T$]	ΔG_0 [$k_{\mathrm{B}}T$]
DNA short	295 ± 1	17.4 ± 0.5	225 ± 4	52 ± 4
RNA long	342 ± 2	28 ± 1	244 ± 5	70 ± 5

In table 1 we report the values we obtained for ΔG, ΔG_0, as well as the aforementioned stretching contributions.

Results for ΔG_0 are in very good agreement with the theoretical ones obtained using the nearest-neighbour model for DNA either using Mfold parameters ($\Delta G_0 = 51 k_{\mathrm{B}}T$) [43] or the ones derived from unzipping experiments ($\Delta G_0 = 48 k_{\mathrm{B}}T$) [44].

4.2. Long handles: the CD4 RNA hairpin

In what follows, we first discuss the characteristics of long handles in section 4.2.1, explaining why sometimes the effective stiffness method can be applied, while other times it cannot. To illustrate the two distinct situations, we first present in section 4.2.2 a scenario based on the CD4 RNA hairpin, where the effective stiffness method is applicable with long handles, just as with short handles (section 4.1). Secondly, the development of a general approach for long handles, beyond the effective stiffness approximation, is covered in section 5 and exemplified with the CD4L12 RNA hairpin.

4.2.1. Characteristics of long handles. Long handles, $\sim$500 bp each, typically represent a bigger challenge than their short counterpart because they are significantly softer. This implies that long handle stiffness features a noticeable force dependence $k_{\mathrm{h}} = k_{\mathrm{h}}(f)$, especially in the lower range of forces experimentally accessible with LOTs. Moreover, the magnitude of k_{h} is now lower and typically comparable to the trap stiffness, $k_{\mathrm{h}} \sim k_{\mathrm{b}}$. Thus, since $k_{\mathrm{eff}}^{\mathrm{F}} \simeq (k_{\mathrm{h}}^{-1} + k_{\mathrm{b}}^{-1})^{-1}$, the term k_{h} significantly contributes to $k_{\mathrm{eff}}^{\mathrm{F}}$. This, together with the clear force dependence of k_{h}, implies in turn that the effective stiffness is not constant but depends on force: $k_{\mathrm{eff}}^{\mathrm{F}} = k_{\mathrm{eff}}^{\mathrm{F}}(f)$. Consequently, upon calculating stretching contributions, the terms ΔW_{b} and ΔW_{h} need to be evaluated more carefully. At closer inspection, however, the use of long handles does not invalidate *per se* the effective stiffness approximation (13). The validity of (13) relies on the assumption that $k_{\mathrm{eff}}^{\mathrm{F}}$ is constant over the integration range $[\lambda_0, \lambda_1]$. Indeed, in many situations, such as with CD4 RNA hairpin, the actual integration range occurs at forces high enough so that $k_{\mathrm{h}} \gg k_{\mathrm{b}}$ and $k_{\mathrm{eff}}^{\mathrm{F}}$ can be taken as constant. Whenever this assumption does not hold, another approach must be used. There are two typical scenarios. On the one hand, if the integration range is large (e.g. for molecules featuring a pronounced hysteresis), the force-dependence of the stiffness $k_{\mathrm{eff}}^{\mathrm{F}} = k_{\mathrm{eff}}^{\mathrm{F}}(f)$ cannot be neglected (note that even if $k_{\mathrm{eff}}^{\mathrm{F}}$ changes marginally from pN to pN, the overall change

on the whole integration range can be significant). On the other hand, if we reach low enough forces (e.g. by using a molecule that refolds at very low forces), the effective stiffness also exhibits force dependence. Indeed at low forces $k_\mathrm{h} \ll k_\mathrm{b}$, hence $k_\mathrm{eff}^\mathrm{F} \sim k_\mathrm{h}$, and as $k_\mathrm{h} = k_\mathrm{h}(f)$ is steep at low f, so will $k_\mathrm{eff}^\mathrm{F}$ be.

Provided that the handle stiffness $k_\mathrm{h}(f)$ and the force stiffness of the trap $k_\mathrm{b}(f)$ are known with a good precision, the integrals in (12) can in principle be carried out easily, irrelevant of $k_\mathrm{eff}^\mathrm{F}$ being non-constant. This corresponds however to an idealized scenario that rarely occurs in practice. To begin with, the elastic properties of the handles are typically characterized for a few ionic and temperature ranges, and often, parameterizations are lacking for generic experimental situations (non-standard buffer, unusual temperatures). Furthermore, for certain types of handles, such as DNA-RNA hybrids, there is a missing gap regarding the elastic description in the literature. But the largest problem comes by far from the usually imprecise knowledge of the optical trap stiffness, since the slightest deviations in the expected value of k_b can have a very significant energetic impact. For instance, a modest deviation of 5% from $k_\mathrm{b} = 0.08$ pN nm^{-1} to $k_\mathrm{b} = 0.075$ pN nm^{-1}, results in a change of a dozen $k_\mathrm{B}T$ in ΔW_b when integrated between 2 an 12 pN. Changes in the value of k_b and even non-linear force corrections in $k_\mathrm{b}(f)$ do inevitably occur in LOT, not only on a day-to-day basis (depending on the laser focusing, alignment, power, intensity or temperature) but also within the same day on a molecule-to-molecule basis, since the beads used for performing experiments can usually slightly vary in size, and the trap stiffness directly depends on this. A slight force dependence in $k_b(f)$ also occurs if the optical plane of the bead shifts with force. Hence, we see that $k_\mathrm{h}(f)$ and $k_\mathrm{b}(f)$ are usually not characterized precisely enough for the integrals in (12) to be computed reliably.

To address the aforementioned issues, we will introduce in section 5 a novel methodology to retrieve the optimal stiffness profile $k_\mathrm{b}(f)$ and $k_\mathrm{h}(f)$ directly from the FDCs obtained in pulling experiments with LOTs. Before doing so, let us however show an example where long handles and the effective stiffness approximation go in pairs: the CD4 RNA hairpin.

4.2.2. Stretching contributions and folding free energy of CD4 RNA. The effective stiffness method can be applied to the CD4 RNA hairpin, which is a molecule showing nearly reversible folding-unfolding kinetics at the accessible pulling speeds [17, 39]. The molecule has the same sequence as hairpin CD4-DNA presented in section 4.1 but replacing thymines by uracils (top panel of figure 3(a)). In the present case, the RNA hairpin is inserted between two ∼500 base-long hybrid RNA/DNA handles [45]. Thus, the molecular construct is formed by the RNA hairpin plus the two long hybrid handles. Pulling experiments were performed analogously as described in section 4.1.

Due to the narrowness of the region in the FDCs (figure 3(a), bottom panel) where folding-unfolding events of CD4 RNA take place, the effective stiffness $k_\mathrm{eff}^\mathrm{F}$ remains fairly constant over the force range experimentally probed. This linearity of the FDCs is evidenced in the inset of figure 3(a) and justifies the use of the effective stiffness approximation. By fitting the FDCs slopes in the highlighted region, we obtain a value for $k_\mathrm{eff}^\mathrm{F}$ equal to 0.067 ± 0.001.

Next, we integrate all FDCs in the range, $[\lambda_0, \lambda_1] = [30, 85]$ nm, which corresponds to the force interval $(f_\mathrm{min}, f_\mathrm{max}) = (18, 22)$ pN. As we did in section 4.1, the F and R
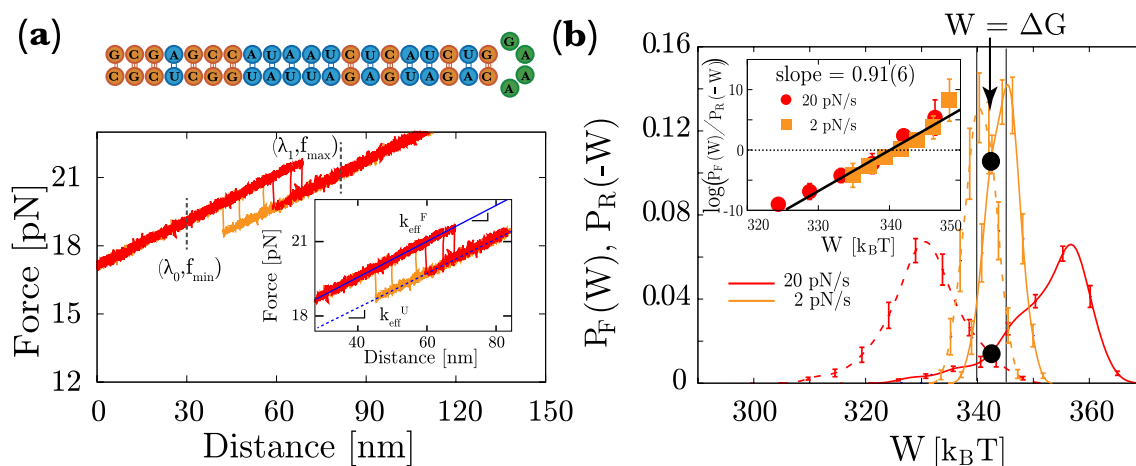
Figure 3. Free-energy recovery CD4 RNA hairpin with long handles (a). Sequence of CD4 RNA (top panel). FDCs and integration range for the work W (bottom panel). Visual evidence of the linearity of the FDCs in the integration range (inset) plus linear fits to the folded (solid line in the inset) and the unfolded branches (dashed line in the inset). (b) Forward (solid lines) and reverse (dashed lines) work distributions for two different pulling speeds calculated in the integration range are indicated in (a) panel. Crossing points between work distributions are tagged as solid points. The CFT verification is shown as an inset. Error bars have been obtained using the Bootstrap method.

work distributions are calculated for two pulling speeds (2 and 20 pN s^{-1}) and are shown in figure 3(b). Note that the crossing point between both distributions corresponds to the work value equal to ΔG. The CFT (7) is satisfied for CD4 with long handles, as can be seen in the inset of figure 3(b). We can thus subtract from the obtained ΔG the stretching contributions $\Delta W_{\rm h} + \Delta W_{\rm b}$ using the effective stiffness method, along the exact same lines as in 4.1. As a last step, the term $\Delta W_{\rm m}$ in (9) is calculated using the WLC model (4) with $P = 0.75\,{\rm nm}$ and an interphosphate distance $d_{\rm b}$ equal to $0.665\,{\rm nm/base}$, so that $L_c \approx 29$ nm, higher than for the CD4 DNA molecule.

We report in table 1 the values we obtained for ΔG, ΔG_0, as well as for the stretching contributions. The measured value for ΔG_0 ($70 \pm 5 k_B T$) is compatible with the previous single-molecule measurements obtained in LOT assays at 100mM Tris HCl pH 8 and 1M NaCl ($\Delta G_0 \approx 65$ $k_B T$) [39] and with the Mfold prediction ($\Delta G_0 = 68$ $k_B T$) [43]. We conclude that the effective stiffness approximation is valid for determining folding free energies from irreversible work measurements if the integration range is narrow enough so that FDCs along the folded branch have a constant slope in such a range (i.e. the effective stiffness $k_{\rm eff}^{\rm F}$ can be taken as constant).

5. Beyond the effective stiffness method

In the previous sections we introduced the effective stiffness method, testing its reliability in addressing the analysis of both short and long handles. We also gave evidence that its validity is limited to the case of a linear elastic response and that when this

condition is not fulfilled a more general methodology becomes necessary. This is the subject covered by section 5.1 where we present a novel technique going beyond the effective stiffness approximation. Then, in section 5.2 we present an application of this method to the case of CD4L12, a dodecaloop RNA hairpin exhibiting large hysteresis in pulling experiments.

5.1. Estimating the stretching contributions in the general case

As can be seen in (1) the force-extension profile $\lambda(f)$ depends on x_b and x_h, and these are, by definition, related to the stiffness through:

$$x_i(f) = \int_0^f k_i^{-1}(f')\mathrm{d}f', \qquad \frac{\mathrm{d}x_i}{\mathrm{d}f} = \frac{1}{k_\mathrm{i}(f)} \qquad \text{for } i = b, h\,. \tag{15}$$

This hints at the fact that FDCs (i.e. the $\lambda(f)$ profile) might allow us to retrieve the stiffness profiles needed to estimate the elastic energy contributions from bead and handles in (12). To realize this in practice, we must assume the elastic response of the trap and the one of the handles can be parametrized by some reasonable physical model. Starting with the handles, we will assume that the extensible WLC model (ext-WLC) is a good description.

$$k_\mathrm{h}(f) = k_\mathrm{h}^{\mathrm{ext-WLC}}(f; \{P, d_\mathrm{b}, Y\})\,, \qquad x_\mathrm{h}(f) = x_\mathrm{h}^{\mathrm{ext-WLC}}(f; \{P, d_\mathrm{b}, Y\})\,, \tag{16}$$

where we introduced the usual WLC elastic parameters (i.e. persistence length P, Young modulus, Y, and monomer length d_b). Then, we can either model the trap stiffness as constant, or as a linear function of force:

$$k_\mathrm{b}(f) = k_{\mathrm{b},0} + \alpha f\,, \qquad x_\mathrm{b}(f) = \frac{1}{\alpha}\log\left(1 + \frac{\alpha}{k_{b,0}}f\right)\,, \tag{17}$$

where α quantifies the linear dependence and $k_{\mathrm{b},0}$ is the stiffness at zero force ($x_\mathrm{b}(f)$ is obtained by integrating as in (15)).

Note that we can rewrite (1) as:

$$\lambda(f) = x_\mathrm{h}(f) + x_\mathrm{b}(f) + x_\mathrm{d}(f)\delta_N + x_\mathrm{ss}(f)\delta_U + \lambda_0\,, \tag{18}$$

where we used a delta-Kronecker-like notation ($\delta_{N(U)} = 1$ if the molecule is in the Native (Unfolded) state and zero otherwise) and explicitly introduced the offset λ_0, which accounts for the fact that the molecular extension is always measured with respect to the micropipette. If we now rewrite the explicit dependence with respect to our model parameters, (18) becomes:

$$\begin{aligned}\lambda(f) &= x_\mathrm{h}(f; \{P, d_\mathrm{b}, Y\}) + x_\mathrm{b}(f; \{\alpha, k_{\mathrm{b},0}\}) + x_\mathrm{d}(f)\delta_N + x_\mathrm{ss}(f)\delta_U + \lambda_0\\ &\equiv \mathcal{M}(f; \{P, d_\mathrm{b}, Y, \alpha, k_{\mathrm{b},0}, \lambda_0\})\,,\end{aligned} \tag{19}$$

where we have denoted $\mathcal{M}$ as the overall model underpinning the $\lambda(f)$ response. As (1) illustrates, the knowledge of a handful of physical parameters fully determines the FDC for the N and U branches. The key idea behind our methodology is that the inverse implication is also true: knowing $\lambda(f)$ and given $\mathcal{M}$ we can extract $P, d_\mathrm{b}, Y, \alpha, k_{\mathrm{b},0}$ through (19), using a least squares method, a Bayesian approach, or generally any regression method. Once these parameters are fitted, we can recompute $k_\mathrm{h}(f)$ and

$k_{\mathrm{b}}(f)$ for all f using the models in (16) and (17), eventually obtaining the stretching contributions through the numerical integration of (12). Crucially, this can be done without any *a priori* knowledge of the parameters of the experimental setup.

In practice, however, the fitting procedure requires a FDC featuring enough curvature to be able to constrain the model, and even so, the number of parameters to fit is too large for a two-dimensional curve, so that some additional considerations must be taken into account. Firstly, reasonable bounds/priors on the allowed values for the parameters must be set. Secondly, it is convenient to assume that certain parameters play a minor role in the overall FDC shape (such as Y) or are characterized well enough (e.g. the monomer length for dsDNA) to be fixed at some nominal value and not fitted. Thirdly, computing the handles extension $x_{\mathrm{h}} = x_{\mathrm{h}}(f)$ using the extensible WLC can be slow and numerically inaccurate as it normally requires the performance of a numerical inversion of $f = f(x_{\mathrm{h}})$. To address this, we introduce in the appendix a formula to explicitly invert the WLC, which can then be used in (19). Fourthly, to get as many points as possible to constrain the fit, we have aligned all the FDCs in the starting point so they share an identical λ_0 offset (i.e. 'const' in (1)(*a*) and (*b*)). After all these steps, fitting $\lambda(f) = \mathcal{M}(f; \{P, d_{\mathrm{b}}, Y, \alpha, k_{b,0}, \lambda_0\})$ is affordable.

In the following section we will show concrete examples of the FDCs fitting procedure and its application to extract the stretching contributions.

5.2. Application to the specific example of the CD4L12 RNA hairpin

The effective stiffness method may work well when the range of force integration is not too large. This condition is met in molecules exhibiting mild hysteresis. For molecules showing large hysteresis in pulling cycles, the limits of integration $f_{\min}$ and $f_{\max}$ are far away and the effective stiffness $k_{\mathrm{eff}}^{\mathrm{F}}$ cannot be considered constant anymore. Here we present results for an RNA molecule (CD4L12) falling in this category and present a general procedure to extract the free energy of formation. CD4L12 shares the same stem than the previously discussed CD4 RNA in section 4.2.2, but with the original tetraloop replaced by a dodecaloop (i.e. 12-loop bases); see the sequence in figure 4(a). A large loop yields a larger entropic barrier for refolding and large hysteresis in the FDC. Pulling experiments were performed as described in section 4.1, with a pulling speed of $100\,\mathrm{nm\,s}^{-1}$ and $300\,\mathrm{nm\,s}^{-1}$ and a buffer containing $4\,\mathrm{mM}$ MgCl, $50\,\mathrm{mM}$ NaCl, and $10\,\mathrm{mM}$ Tris. The values of $P = 0.75\,\mathrm{nm}$ and $d_{\mathrm{b}} = 0.665$ nm were used to describe the elastic properties of the ssRNA for this buffer [39].

As can be seen in figure 4(b), CD4L12 behaves as a two-state system being either folded or unfolded along the FDCs. As expected, pulling cycles feature large hysteresis, with a maximal difference of nearly 20 pN between the lowest folding and largest unfolding force rips. In order to compute the work needed for the CFT (7), we must integrate the area under the FDC within a large force range with a very low $f_{\min}$. It is clear that in this case the constant stiffness approximation described in section 4.2.1 does not apply, as shown in the inset of figure 4(b) where $k_{\mathrm{eff}}^{\mathrm{F}}$ markedly changes with force. To estimate the stretching contributions we follow the previous section 5.1 and (19) to obtain ΔW_{b}, ΔW_{h}, and, from (14), the value of ΔG_0.

In order to carry out the fit prescribed by (19), we need to introduce some further assumptions to simplify the problem. Regarding the hybrid DNA-RNA handles, we
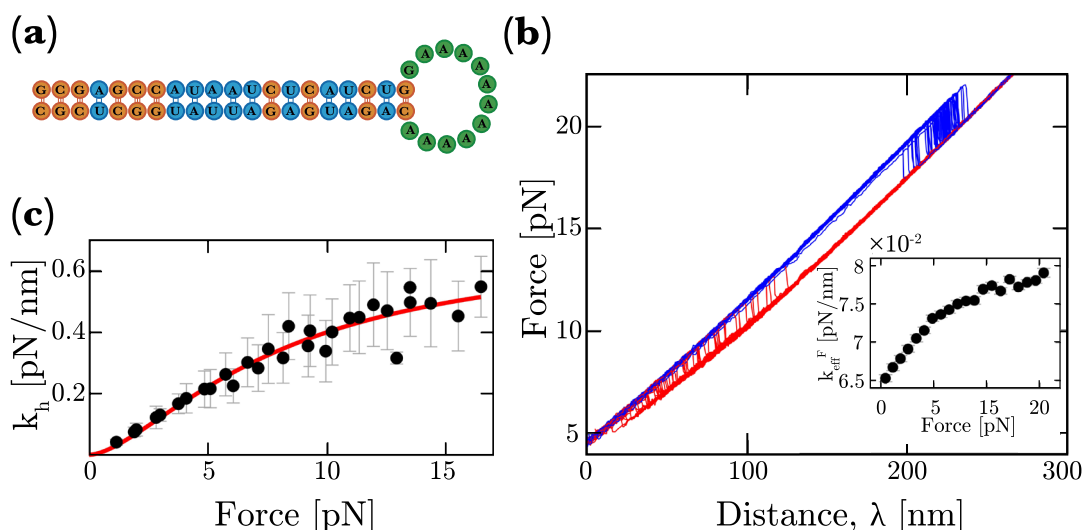
Figure 4. Free-energy recovery CD4L12 RNA hairpin. (a) Sequence and secondary structure of CD4L12 RNA. (b) Aligned FDCs folding (red) and unfolding (blue) for a given molecule pulled at $100\,\mathrm{nm\ s^{-1}}$. *Inset*: effective stiffness profile measured along the folded branch. (c) Stiffness profile of the hybrid DNA-RNA handles that form the molecular construct used with CD4L12 and CD4 RNA [39]. Data points have been obtained using the high frequency power-spectrum method described in [31]. The red line is a fit of the extensible WLC model, yielding $P = 20 \pm 4$ nm and $Y = 200 \pm 14$ pN (d_{b} was not fitted but fixed to the interphosphate distance of A-form RNA, $d_{\mathrm{b}} = 0.27$ nm [45]).

use the value of the interphosphate distance $d_{\mathrm{b}} = 0.27$ nm of A-form RNA and Young modulus $Y = 200$ pN obtained by fitting the stiffness of the handle profile (figure 4(c)). While changes in d_{b} only moderately affect the overall curvature of k_{h} (but they impact the overall contour length, an effect already captured by fitting λ_0), changes in Y do not. Hence fixing these two values gives a better constrained model. For the persistence length of the handles P it is convenient to fit the deviation ΔP (in %) with respect to a plausible expected nominal value P_0, i.e. $P_{\mathrm{eff}} = P_0 (1 + \Delta P)$, which we take from the fit in figure 4 as $P_0 = 20\,\mathrm{nm}$. Lastly, we also include the number of nucleotides n released in the transition between the folded and the unfolded branches as an extra free parameter of the model. We are thus eventually left with five free parameters which we fit (18) and (19) using a standard non-linear least square regression (Levenberg–Marquardt):

$$\lambda(f) = \mathcal{M}(f) = \mathcal{M}(f; \{k_{\mathrm{b},0}, \alpha, \Delta P, \lambda_0, n)\}) \ . \tag{20}$$

An example of such a fitting procedure is shown in figure 5(a). As can be seen, the agreement between the experimental points and the reconstructed curve is remarkable. Furthermore, all the values obtained from the fit dovetail with prior expectations. Firstly, the value of n matches with the expected number of released nucleotides (i.e. 52). Secondly, the zero-force trap stiffness k_{b} falls in an expected range [31]. Thirdly, the force-dependence parameter α of the trap stiffness is of the same order of magnitude than values already reported in the literature for similar LOT settings [31]. Fourthly, ΔP is small so P is reasonably close to the assigned nominal value P_0. Another good generic indicator is the very low error on the fitted parameters, hinting at
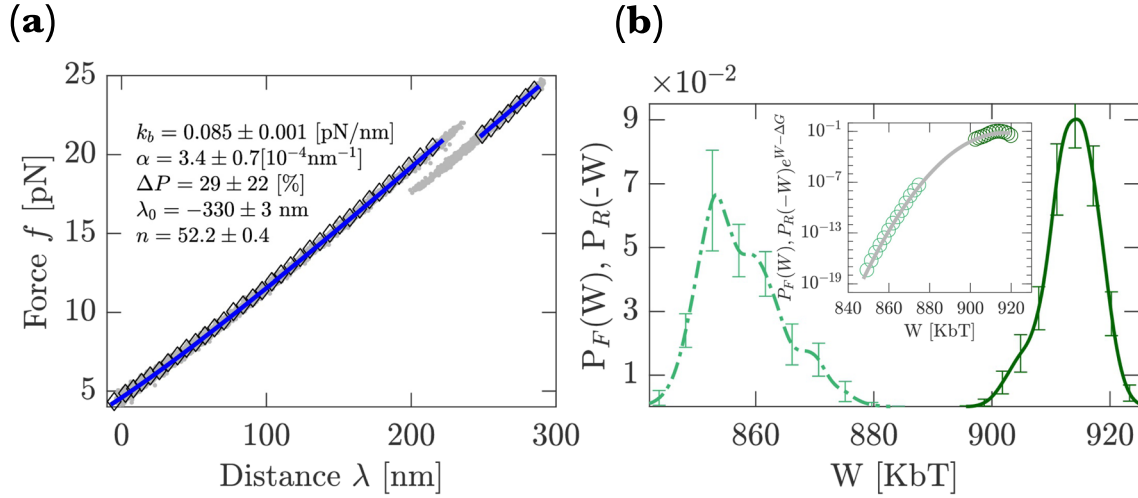
(a)

(b)

Figure 5. Fitting the folded and unfolded branches of CD4L12 RNA hairpin (a). The solid blue line is an example of curve fitting based on (20). Data points used for the fit are the black diamonds. They are obtained by smoothing and filtering the gray dots, which are themselves obtained by aggregating the unfolding FDCs of different pulling cycles from figure 4(b). (b) An example of forward and reverse (solid and dashed lines) work distributions for the same molecule pulled at $100\,\mathrm{nm\ s^{-1}}$. Due to the large hysteresis, work distributions do not overlap. *Inset:* Illustration of the *matching* method to retrieve ΔG by imposing continuity between $P_F(W)$ (light green) and $P_R(-W)\mathrm{e}^{(W-\Delta G)/k_\mathrm{B}T}$ (dark green) in log-normal scale. The solid grey line is the fitted Gaussian, see [46] for details.

a well-constrained model; a fact that is further confirmed by the observation that in the correlation matrix of the fit, most off-diagonal entries are near-zero (details not shown). We must finally stress that the choice of free parameters in (20) is convenient for the considered situation but is by no means customary. In a context where the trap would be well characterized and the handles would not, we may have, for instance, fixed k_b but fitted d_b. Equation (19) can be adapted at will, depending on the requirement.

With the fitted values of α, $k_{\mathrm{b},0}$ and ΔP in hand and our assumptions for Y and d_b (legitimated retrospectively by the agreement of the fit in figure 5), we are now in a position to precisely establish the profiles of $k_\mathrm{h}(f)$ and $k_\mathrm{b}(f)$ through the use of equations (15)–(17). We can now quantify the terms ΔW_b and ΔW_h using (12) and ΔG using the FT. These numbers together with (14) allow us to extract ΔG_0.

Figure 5(b) shows the work distributions $P_\mathrm{F}(W)$ and $P_R(-W)$ obtained from the FDC (figure 4(b)). The very pronounced hysteresis and the large value of the average dissipated work in a pulling cycle (about 60 $k_\mathrm{B}T$) is such that F and R work distributions lie far apart without overlapping. Previous methods based on the overlapping of F and R work distributions are not applicable and an alternative approach must be used, such as the Bennett acceptance ratio [47] and the 'matching method'. This last method consists of finding the optimal ΔG value so that $P_\mathrm{F}(W)$ is the analytical continuation of $P_R(-W)\mathrm{e}^{(W-\Delta G)/k_\mathrm{B}T}$. This procedure is graphically illustrated in the inset of figure 5(b) and further explained in [46]. Results obtained for different molecules are shown in table 2. We note that the values of ΔG obtained with the two methods yield compatible results (*matching* being systematically 3–5 $k_\mathrm{B}T$ lower than *Bennett*). Our

Table 2. Fluctuation theorem and stretching contributions for CD4L12 RNA hairpin with long handles. Overview of the values of ΔG, the stretching corrections, and the final ΔG_0 estimate for six different molecules. All values are given in $k_{\mathrm{B}}T$. $\Delta G_{\mathrm{Bennet}}$ and $\Delta G_{\mathrm{Matching}}$ provide two ways to extract ΔG using the CFT. The value of ΔG_0 is obtained through (9) using the value of the Bennett estimate. The last line corresponds with the only experimental setting in which the pulling speed is $300\,\mathrm{nm\ s^{-1}}$, and all the other results were obtained at $100\,\mathrm{nm\ s^{-1}}$.

$\Delta G_{\mathrm{Bennet}}$ $[k_{\mathrm{B}}T]$	$\Delta G_{\mathrm{Matching}}$ $[k_{\mathrm{B}}T]$	ΔW_{m} $[k_{\mathrm{B}}T]$	$\Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}}$ $[k_{\mathrm{B}}T]$	ΔG_0 $[k_{\mathrm{B}}T]$
1045 ± 3	1040	35	944 ± 3	66
950 ± 2	947	35	846 ± 1	68
863 ± 3	859	35	758 ± 2	70
888 ± 2	886	35	790 ± 2	63
938 ± 4	935	35	838 ± 1	66
1107 ± 2	1105	38	1003 ± 2	68
			Mean:	$67 \pm 2\ \mathbf{k_B T}$

estimated value $\Delta G_0 = 67 \pm 2k_{\mathrm{B}}T$ is not far from the Mfold prediction ($\Delta G_0 = 63\ k_{\mathrm{B}}T$) showing the reliability of the approach.

We want to stress the sensitivity of the value of ΔG_0 on the accurate estimation of the stretching contributions which, being one order of magnitude larger, can lead to inconsistent results. Had we used a methodology assuming 'average' or 'standard' stretching contributions, we would have obtained erroneous numbers. Consider, for instance, subtracting the average value $\langle \Delta W_{\mathrm{b}} + \Delta W_{\mathrm{h}} + \Delta W_{\mathrm{m}} \rangle = 898k_{\mathrm{B}}T$ derived from table 2 to the highest and the lowest estimates of ΔG shown in the same table: it results in two widely off values $\Delta G_0 = 1107 - 898 = 209k_{\mathrm{B}}T$ and $\Delta G_0 = 863 - 898 = -35k_{\mathrm{B}}T$. Therefore a tailored molecule-to-molecule estimation of the stretching contribution is absolutely essential for molecules like CD4L12 where the effective stiffness approximation cannot be used.

6. Conclusions

We have presented a brief tutorial on the approaches commonly used to extract folding free energies of single molecules pulled with optical tweezers in unzipping assays. A recurrent issue in these calculations is the large magnitude of the stretching contributions to the full free-energy difference measured in a pulling experiment using the CFT. Such contributions arise from the experimental setup and include the optical trap, the elastic stretching of the handles used in the molecular construct and the extension release of the unfolded polymer. A great simplification in the analysis of these correction terms can be performed when the effective stiffness of the experimental system can be approximated as constant, as we saw in section 4. In this so-called effective stiffness approximation a single parameter $k_{\mathrm{eff}}^{\mathrm{F}}$ suffices to quantify the stretching contributions of handles and trap. We exemplified this case in the study of a DNA hairpin in section 4.1. For long handles the stiffness of the handles turns out to be comparable to that of the trap and a force-dependent $k_{\mathrm{eff}}^{\mathrm{F}}$ is apparent. In this case, as

we showed in section 4.2.2, one can still use the effective stiffness approximation if the range of integration to evaluate the work is narrow enough. This is possible if the pulling curves are not too irreversible and forward and reverse work distributions overlap. In contrast, for strong irreversible pulling experiments one needs to accurately characterize all elastic contributions from the experimental setup. Here we have introduced a novel method (section 5) based on least-squares fitting of the elastic response of the folded and unfolded branches. It relies on adapting the elastic parameters extracted from the literature (inter-monomer distance, persistence length, Young modulus) to the experimental data as well as accurately retrieving the stiffness of the optical trap using the same data.

One problem that remains open is the magnitude of the statistical error committed in the estimation of ΔG_0. In fact, ΔG_0 is the difference of two large numbers (ΔG and the stretching contributions) each with a large error and extracted from the same experimental FDC data. How to combine the errors from these two large quantities remains largely unclear as they are not really uncorrelated. A rule of thumb in single-molecule experiments is that the largest errors come from molecule to molecule experimental variability. It is then recommended to first extract ΔG_0 values for different molecules by subtracting elastic contributions from ΔG on a single-molecule basis, and then derive the mean value of ΔG_0 and the corresponding statistical error.

The large contribution of the stretching term (14) to the full free energy ΔG makes the prediction of the (comparably small) value of ΔG_0 a difficult task. This situation is reminiscent of the enthalpy-entropy compensation problem in biochemistry [48, 49]. In this case, free-energy differences of intra- and intermolecular weak interactions (e.g. folding, binding, allostery, enzymatic reactions and so on) are typically one order of magnitude smaller than entropies and enthalpies, i.e. $\Delta G = \Delta H - T\Delta S$ with $\Delta G \ll \Delta H, T\Delta S$. In this regard, enthalpy-entropy compensation in biochemistry appears to be similar to the ΔG-stretching compensation in force spectroscopy. The analogy is not pure coincidence as the stretching contributions are essentially also of an entropic nature and much larger than the bare free-energy difference ΔG_0.

The methodology we have described should be generally useful and applicable to force spectroscopy studies of single-molecule constructs whenever elastic contributions are present. Applications go beyond the case of measuring folding free energies such as extracting molecular free-energy landscapes [30], measuring ligand binding energies [50], protein-protein and RNA-protein interactions, and characterizing heterogeneous molecular ensembles [51].

Acknowledgments

Appendix. WLC explicit inversion

The inextensible WLC model described in (4) gives a very direct way to compute $f = f(x)$, but it is not straightforward to use it to retrieve $x = x(f)$. Although numerical inversion using Mathematica and other software is possible (e.g. as in [52]), it is useful to have explicit inversion formulae. Hence let us now quickly show that (4) can be easily inverted to express $z := x/L_c$ as a function of f. We first define the normalized quantity $\tilde{f} = (4P/k_{\mathrm{b}}T) f$. We can then re-write (4) as $\tilde{f} = (1-z)^{-2} - 1 + 4z$. By multiplying both sides of the previous by $(1-z)^2$ and by moving all terms to the same side, we obtain:

$$0 = z^3 + a_2 z^2 + a_1 z + a_0 \qquad \text{with} \qquad a_2 = -\frac{9}{4} - \frac{\tilde{f}}{4}, \quad a_1 = \frac{3}{2} + \frac{\tilde{f}}{2}, \quad a_0 = -\frac{\tilde{f}}{4}. \tag{A.1}$$

Thus we directly see that obtaining z as a function of f simply maps to finding the roots of a cubic polynomial—a problem solved since the fifteenth century. The approach taken here is the canonical one [53, 54]. We start defining the following intermediate quantities:

$$R := \frac{9 a_1 a_2 - 27 a_0 - 2 a_2^3}{54} \qquad Q := \frac{3 a_1 - a_2^2}{9} \tag{A.2}$$

from which we obtain the standard determinant D for cubic equations:

$$D := Q^3 + R^2. \tag{A.3}$$

If $D > 0$, there is only one real solution to (A.1), and we have to define the following intermediate quantities to express the answer:

$$T := \sqrt[3]{R + \sqrt{D}} \qquad S := \sqrt[3]{R - \sqrt{D}} \tag{A.4}$$

(since $D > 0$, we also have that $\sqrt{D}$ is real, and thus there is indeed at least one real cubic root for T and S). The desired inverse value $z^* = z(f)$ is then finally obtained as:

$$z^* = -\frac{1}{3} a_2 + S + T. \tag{A.5}$$

If $D < 0$, there are three real roots to the cubic equation. These roots can be obtained by re-using the quantities S and T defined above, but doing so requires using complex number algebra—which may not be handy. Instead, we also can define the following intermediate quantity:

$$\theta := \arccos\left(\frac{R}{\sqrt{-Q^3}}\right). \tag{A.6}$$

From this, the three real roots z_1, z_2, z_3 can be obtained directly as:

$$z_i = 2\sqrt{-Q} \cos\left(\frac{\theta + \theta_i}{3}\right) - \frac{1}{3} a_2 \qquad \text{with} \qquad \theta_1 = 0, \quad \theta_2 = 2\pi, \quad \theta_3 = 4\pi. \tag{A.7}$$

The root of interest is the one lying in the interval $[0, 1]$, since we defined above $z = x/L_c$ and a property of the inextensible WLC is that the extension x is always

smaller than the contour length L_c. Using trigonometric standard formula and the fact that $2\sqrt{-Q} > 0$ it is quite easy to verify that $z_1 - z_2 > 0$ and $z_3 - z_2 \geqslant 0$ for the given range of θ (which must belong to $[0, \pi]$ by definition of the arccosine), which implies that z_2 is the smallest of all the roots. Moreover, we note that all the roots must be positive, since we see in (4) that $\forall z < 0$, $f(z) < 0$ and is strictly monotonically decreasing. As all the roots are positive and z_2 is the smallest of them, it therefore has to be the one we are looking for, in $[0, 1]$, and hence $z_2 = z^* = z(f)$ when $D < 0$. The previous result also covers the $D = 0$ situation, because we then have from (A.6), $\theta = 0$, and so we are in the limiting case $z_3 = z_2$.

Let us finally note that in the case of the *extensible* WLC, the key difference with the inextensible case is the replacement $L_c \to L_c(1 + f/Y)$ with Y the Young Modulus, i.e. the contour length is now force dependent. It can be shown that this implies the following relationship between the two models:

$$x^{\text{ext}}_{\text{WLC}}(f) = x^{\text{inext}}_{\text{WLC}}(f)(1 + f/Y) \tag{A.8}$$

and so we see that knowing the explicit inversion for the inextensible model directly yields an explicit formula for the extensible model too.

References

[1] Aabert B, Johnson A L J, Raff M, Roberts K and Walter P 2002 *Molecular Biology of the Cell* (New York: Garland Science)
[2] Felsenfeld G and Miles H T 1967 *Ann. Rev. Biochem.* **36** 407–48
[3] Makarova K S *et al* 2011 *Nat. Rev. Microbiol.* **9** 467
[4] Sander J D and Joung J K 2014 *Nat. Biotechnol.* **32** 347
[5] Cantor C R and Schimmel P R 1980 *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function* (London: Macmillan)
[6] Shashkova S and Leake M C 2017 *Biosci. Rep.* **37** BSR20170031
[7] Meller A, Nivon L and Branton D 2001 *Phys. Rev. Lett.* **86** 3435
[8] Neuman K C and Nagy A 2008 *Nat. Methods* **5** 491
[9] Ashkin A, Mourou G and Strickland D 2018 *Curr. Sci.* **115** 1844
[10] Bustamante C, Liphardt J and Ritort F 2005 *Phys. Today* **58** 43
[11] Seifert U 2008 *Eur. Phys. J.* B **64** 423–31
[12] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
[13] Ciliberto S 2017 *Physical Review* X **7** 021051
[14] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
[15] Crooks G E 2000 *Phys. Rev.* E **61** 2361
[16] Liphardt J, Dumont S, Smith S B, Tinoco I and Bustamante C 2002 *Science* **296** 1832–5
[17] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco I Jr and Bustamante C 2005 *Nature* **437** 231
[18] Ritort F 2008 *Adv. Chem. Phys.* **137** 31
[19] Hummer G and Szabo A 2010 *Proc. Natl Acad. Sci.* **107** 21441–6
[20] Gupta A N, Vincent A, Neupane K, Yu H, Wang F and Woodside M T 2011 *Nat. Phys.* **7** 631
[21] Ross D, Strychalski E A, Jarzynski C and Stavis S M 2018 *Nat. Phys.* **14** 842
[22] Dudko O K, Hummer G and Szabo A 2008 *Proc. Natl Acad. Sci.* **105** 15755–60
[23] Maitra A and Arya G 2010 *Phys. Rev. Lett.* **104** 108301
[24] Maitra A and Arya G 2011 *Phys. Chem. Chem. Phys.* **13** 1836–42
[25] Berkovich R, Hermans R I, Popa I, Stirnemann G, Garcia-Manyes S, Berne B J and Fernandez J M 2012 *Proc. Natl Acad. Sci.* **109** 14416–21
[26] Chang J C, de Messieres M and La Porta A 2013 *Phys. Rev.* E **87** 012721
[27] Hinczewski M, Gebhardt J C M, Rief M and Thirumalai D 2013 *Proc. Natl Acad. Sci.* **110** 4500–5
[28] Cossio P, Hummer G and Szabo A 2015 *Proc. Natl Acad. Sci.* **112** 14248–53
[29] Neupane K and Woodside M T 2016 *Biophys. J.* **111** 283–6

[30] Woodside M T, Anthony P C, Behnke-Parks W M, Larizadeh K, Herschlag D and Block S M 2006 *Science* **314** 1001–4

[31] Forns N, de Lorenzo S, Manosas M, Hayashi K, Huguet J M and Ritort F 2011 *Biophys. J.* **100** 1765–74

[32] Marko J F and Siggia E D 1995 *Macromolecules* **28** 8759–70

[33] Wang M D, Yin H, Landick R, Gelles J and Block S M 1997 *Biophys. J.* **72** 1335–46

[34] Petrosyan R 2017 *Rheol. Acta* **56** 21–6

[35] Camunas-Soler J, Ribezzi-Crivellari M and Ritort F 2016 *Ann. Rev. Biophys.* **45** 65–84

[36] de Lorenzo S, Ribezzi-Crivellari M, Arias-Gonzalez J R, Smith S B and Ritort F 2015 *Biophys. J.* **108** 2854–64

[37] Alemany A and Ritort F 2014 *Biopolymers* **101** 1193–9

[38] Bosco A, Camunas-Soler J and Ritort F 2013 *Nucl. Acids Res.* **42** 2064–74

[39] Bizarro C V, Alemany A and Ritort F 2012 *Nucl. Acids Res.* **40** 6922–35

[40] Bustamante C *et al* 1994 *Science* 1599

[41] Abels J, Moreno-Herrero F, Van der Heijden T, Dekker C and Dekker N H 2005 *Biophys. J.* **88** 2737–44

[42] Monge A M, Manosas M and Ritort F 2018 *Phys. Rev.* E **98** 032146

[43] Zuker M 2003 *Nucl. Acids Res.* **31** 3406–15

[44] Huguet J M, Bizarro C V, Forns N, Smith S B, Bustamante C and Ritort F 2010 *Proc. Natl Acad. Sci.* **107** 15431–6

[45] Wen J D, Manosas M, Li P T, Smith S B, Bustamante C, Ritort F and Tinoco I Jr 2007 *Biophys. J.* **92** 2996–3009

[46] Alemany A, Rey-Serra B, Frutos S, Cecconi C and Ritort F 2016 *Biophys. J.* **110** 63–74

[47] Bennett C H 1976 *J. Comput. Phys.* **22** 245–68

[48] Petruska J and Goodman M F 1995 *J. Biol. Chem.* **270** 746–50

[49] Sharp K 2001 *Protein Sci.* **10** 661–7

[50] Camunas-Soler J, Alemany A and Ritort F 2017 *Science* **355** 412–5

[51] Monge A M 2019 Free energy and information-content measurements in thermodynamic and molecular ensembles *PhD Thesis* Universitat de Barcelona

[52] Broekmans O D, King G A, Stephens G J and Wuite G J 2016 *Phys. Rev. Lett.* **116** 258102

[53] Abramowitz M and Stegun I A 1965 *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* vol **55** (Washington DC: Courier Corporation)

[54] Borwein P and Erdélyi T 1995 *Polynomials and Polynomial Inequalities* vol 161 (Berlin: Springer)