BIOPHYSICS LETTER

Probing DNA base pairing energy profiles using a nanopore

Virgile Viasnoff · Nicolas Chiaruttini ·
Ulrich Bockelmann

Abstract We experimentally show that the voltage driven unzipping of long DNA duplexes by an α-hemolysin pore is sensitive to the shape of the base pairing energy landscape. Two sequences of equal global stability were investigated. The sequence with an homogeneous base pairing profile translocates faster than the one with alternative weak and strong regions. We could qualitatively account for theses observations by theoretically describing the voltage driven translocation as a biased random walk of the unzipping fork in the sequence dependent energy landscape.

Keywords Nanopore · α-Hemolysin · Unzipping · DNA · Sequence energy landscape

Introduction

Determining the structures and the exchanges between structures of molecules such as nucleic acids and proteins is one of the keys to the understanding of their biological functions. Single molecule force measurements using either position or force clamp (Greenleaf et al. 2007) have been used to reveal the local heterogeneity of base pairing energies (Bockelmann et al. 2002; Collin et al. 2005), the interchange between alternative folded structures (Harlepp et al. 2003; Li et al. 2007; Greenleaf et al. 2008;

V. Viasnoff (✉) · N. Chiaruttini · U. Bockelmann
Laboratoire de Nanobiophysique, CNRS, ESPCI ParisTech,
10 rue Vauquelin, 75005 Paris, France
e-mail: virgile.viasnoff@espci.fr

Garcia-Manyes et al. 2007) as well as the mechanisms of enzymatically mediated mechanical opening of dsDNA (Larson et al. 2008). Theses studies were carried out using mostly optical traps or AFM. They rely on the direct measurement of the instantaneous force involved in the various above mentioned mechanisms. These techniques, although extremely powerful, demand a large amount of work and know-how in the preparation of the tethered probed sample as well as a long time for data collection. In addition they cannot be easily parallelized. In the past decade, the electrophoretic threading of biomolecules through a single nanopore (Kasianowicz et al. 1996) has emerged as an alternative time-efficient approach. The first studies focused on characterizing and understanding the passage of ssDNA and ssRNA through the α-hemolysin (Akeson et al. 1999; Meller et al. 2000). Theses molecules are small enough (1.3 nm in diameter) to translocate through the pore. Transient ionic current blockades are recorded. dsDNA with a helix diameter (2.2 nm) larger than the pore size (1.5 nm) can also translocate, presumably due to the mechanical action of the pore that either opens (Sauer-Budge et al. 2003; Mathe et al. 2004; Tropini and Marziali 2007) or stretch the paired structure (Zhao et al. 2008). Hence, it has been suggested that nanopores could be used to study the stem–loop structures of ssRNA by their sequential opening during the translocation (Bundschuh and Gerland 2005).

To date, however, this technique has been limited by its ability to measure only quantities that are averaged at best over a single translocation event but more generally over the distributions of all recorded events. Although a real time analysis of the blocked current noise has been shown to discriminate between different types of inserted molecules (Vercoutere et al. 2003), it is generally impossible to get information on the actual position of the molecule in the pore

during its translocation. There is one reported exception though, for ssRNA sequences consisting in blocks of poly dA and poly dC (Akeson et al. 1999), where a current change indicates the passage of the different molecule parts. The lack of instantaneous information complicates the elucidation of the unzipping process of dsDNA by a nanopore. Indeed, while the translocation mechanism of unstructured ssDNA or ssRNA is rather well characterized and understood, the process by which a double stranded part can pass through the nanopore remains to be studied. It has been shown that the average translocation time of unstructured homogeneous single stranded DNA or RNA scales linearly with the sequence length, slightly depends on the sequence composition and seem to vary quadratically with the applied voltage (Meller et al. 2000, 2001; Meller and Branton 2002). The translocation times of double stranded DNA are much longer and were interpreted in terms of an unzipping event followed by the translocation of the opened structure. This mechanism was described as an activated process over the effective energy barrier that corresponds to a partial opening of the sequence. In the case of short blunt end duplexes (Vercoutere et al. 2001) (<8 mers) the energy barrier was related to the bulk free enthalpy ΔG_o of the entire hairpin. The helix thermally opens and is then threaded into the pore lumen. The applied voltage hardly biases the unzipping process. In the case of longer duplexes [10 mers as in Mathe et al. (2004) or 50 mers as in Sauer-Budge et al. (2003)] with larger ΔG_o, the spontaneous opening of all base pairs by thermal activation is not probable enough to allow the translocation. If a single stranded overhang is added at one end, a mechanical force can be applied on the duplex. The unzipping process is then biased by the applied voltage. The effective energy barrier extracted from the experiments corresponds to the opening of 3–5 bp but has not been related to any specific part of the sequence. The barrier rather accounts for a Kramers like behavior of the unzipping time with the applied voltage. A multi step mechanism where bases or groups of bases would sequentially open was also suggested (Dudko et al. 2007). In order to fit the experimental results, a two-dimensional (2D) multi barrier stochastic model (Lakatos et al. 2005) and a microscopic approach based on extended Kramers theory with one supplementary free energy parameter (Dudko et al. 2007) have been proposed. To account theoretically for the nanopore unzipping of long sequences (>30 mers), we recently suggested a model (Bockelmann and Viasnoff 2008) where the actual sequence is explicitly taken into account. The unzipping process is described as a discrete motion of the DNA into the pore where each step is an activated process over the energy barrier needed to open/close the first base outside the pore and thread it in/out of the pore. The step length is equal to the base size.

In order to further unravel the unzipping mechanism we designed two sequences of $N_b = 45$ bp with the same total base pairing enthalpies and entropies but with a different energy landscape. We show that the existence of local energy barriers located in the middle of the sequence landscape slows down the translocation process. The translocation time of long structures is found to depend on the local base-pair arrangements and not only on the global stability or the duplex length, which indicates that the unzipping mechanism of DNA in nanopores is sequential. Our model is able to predict the difference in translocation time between both sequences.

Experimental results

Description of the sequences

We designed two different sequences of 45 bp coined λ-RLS for Rough Landscape Sequence and λ-SLS for Smooth Landscape Sequence. The details of the sequences are given in the Supplementary Materials. λ-RLS is the sequence of the lambda phage from base 738 to 783. Its total GC content is 64.4%. It presents an AT rich region around base 18, and a GC rich region around base 28. We designed λ-SLS with a G–C content of 66.7% (1 additional GC pair) evenly spread along the sequence. We used Mfold (Zuker 2003) to estimate the base pairing enthalpies and entropies of our sequences. We optimized λ-SLS such that both its enthalpy and entropy match those of λ-RLS to 0.3%. The results are given in Table 1. Both sequences have (1) the same eight first bases in order to avoid any difference in the unzipping time due to the nucleation of the opening fork. (2) A 3′ end with a single stranded overhang of 30 adenines that serves as a handle to initiate the threading.

Melting curves

We tested the global stability of the sequences by measuring their bulk melting profiles (Mergny and Lacroix 2003). Absorbance at 260 nm was measured as a function of temperature in the following buffer: LiCaco 20 mM, KCl

Table 1 Mfold predicted values of the thermodynamical parameters for both sequences at 25°C and 1 M monovalent salt

Seq	ΔH_o (kcal/mol)	ΔS_o (cal/mol per K)	ΔG_o (kcal/mol, $T = 25$°C)
λ-RLS	−378.4	−995.5	−81.6
λ-SLS	−379.3	−998.5	−81.6

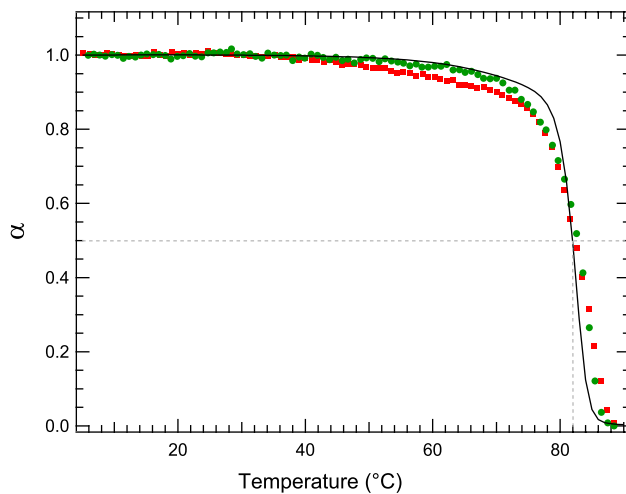**Fig. 1** The hybridized fraction $\alpha = (dsDNA)/[(ssDNA) + (dsDNA)]$ of duplexes versus temperature [λ-RLS (*square*) and λ-SLS (*circle*)]. Buffer: LiCaco 20 mM, KCl 100 mM, pH 7.2, 0.2 °C/min. The measured $T_m = 82.6 \pm 0.3$°C. The *continuous line* represents the melting profile generated by the Mfold server ([DNA] = 1 μm, [Na$^+$] = 120 mM). The predicted melting temperature is $T_m = 82.2$°C

100 mM, pH 7.2. Figure 1 shows that the melting curves cannot be distinguished within experimental uncertainty. The melting temperature is 82.6 ± 0.3°C. The clear overlay of the profiles indicates that both sequences have very similar global stability. In addition the agreement of the Mfold estimated melting profile with our experimental data is good. Hence, we conclude that the ΔH_o and ΔS_o of our sequences match their predicted values.

Nanopore unzipping

Then, we studied the distributions of translocation time τ of both duplexes across a single α-hemolysin pore embedded in a suspended bilayer. The bilayer was formed on a 30 μm hole made in a Teflon cell mounted on an inverted microscope. Figure 2 depicts schematically our experimental set-up. The blockade events are detected with a current amplifier (Axopatch 200B). The blockade signals are filtered at 10 kHz and digitized at 20 kS/s. We worked at $T = 24$°C, pH 7.4, TBE 1×, 1 M KCl. Figure 3a shows the distributions of blocked current levels versus blockade durations for λ-SLS at 250 mV. We distinguish fast translocating events corresponding to non-hybridized ssDNA from long events corresponding to the unzipping of dsDNA. The distribution of fast events matches the distribution of ssDNA only (data not shown). Theses events were subsequently withdrawn from the analysis. The distributions of current levels are identical for both duplexes. We call $P(\tau) = \int_0^\tau p(t)dt$ the integrated distribution of translocation times τ, where $p(\tau)$ is the distribution of blockade duration. Figure 3b shows $P(\tau)$ for both sequences at 250 mV. We observe that λ-RLS
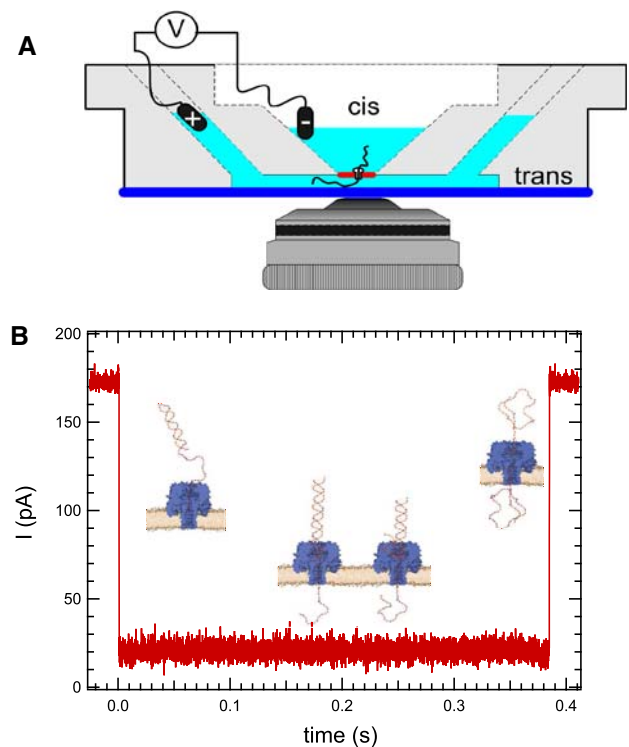

Fig. 2 a Experimental set-up. The bilayer is formed on a 30 μm hole in a Teflon septum 100 μm above a glass slide. A single α-hemolysin pore is incorporated in the bilayer. We used 1 M KCl TBE 1× pH 7.4. **b** Typical translocation curve at 150 mV. We superimpose the putative position of the DNA in the pore during the translocation process. The pore and the DNA are drawn to scale

translocates more slowly than λ-SLS. This observation remains true over our accessible voltage range 180–350 mV. The typical ratio of the passage times is around 3. Comparison of the distributions at six different voltages are given in the Supplementary Materials. Figure 4a shows how $P(\tau)$ of λ-SLS depends on voltage. We observe that the distributions steepen and that the unzipping times decrease as the voltage U is increased. Our distributions do not fit to the functional form $1 - \exp(-\tau/\tau_K)$ expected from a Kramers approach and observed for shorter hairpins (Mathe et al. 2004). We thus characterize the distributions by their median value $\tau_{1/2}$, defined as $P(\tau_{1/2}) = 1/2$. As demonstrated by Fig. 4c, $\tau_{1/2}$ diverges at low voltages and does not have the exponential dependance on the applied voltage U expected from a Kramers approach. At low voltage, both sequences tend to have the same translocation time whereas at higher voltages, λ-RLS is always slower than λ-SLS by a factor 2–4, even at our highest achievable voltage of 350 mV.

Discussion

If we interpret our data as a single activated event, an increase of the unzipping time by a factor of 2–4 corresponds
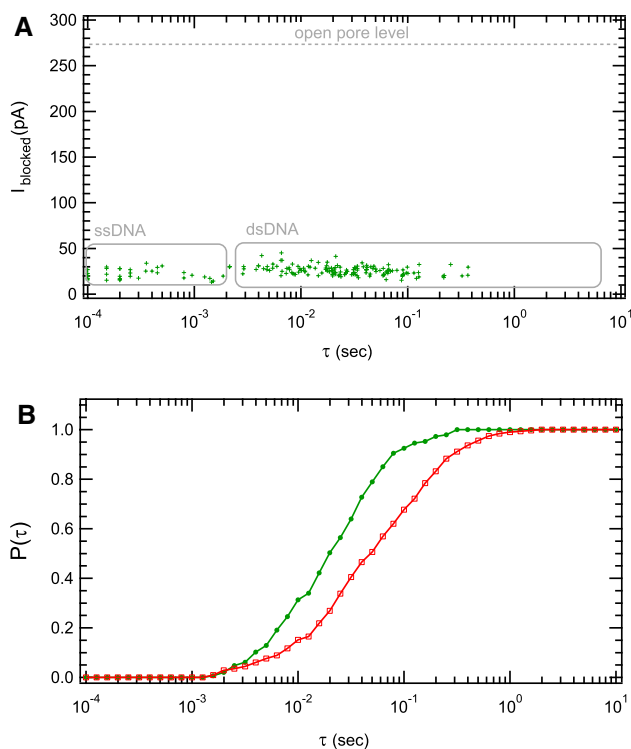
Fig. 3 a Event diagram showing the blocked current levels versus the translocation durations for λ-SLS at 250 mV. Two populations clearly appear. The fast translocating events correspond to non-hybridized molecules and are subsequently withdrawn from analysis. **b** Integrated probability $P(\tau)$ of the translocation times for λ-RLS (*open square*) and λ-SLS (*filled circle*) at 250 mV. Notice that λ-RLS translocates more slowly than λ-SLS

to a rather small change in the effective activation energy of about 1 kcal/mol. Translocation of our long DNA duplexes are however very unlikely to happen in a single microscopic activated event. It would imply an activation barrier of the order of $\Delta G_0 = 135kT$. In a simple Kramers approach, τ_K would read $\tau_K = \frac{1}{v_0}\exp[(\Delta G_0 - Q_{\text{eff}}U)/k_bT]$. Using the literature values for $Q_{\text{eff}} = 0.1\text{–}0.2\text{e}^-/bp$, v_0 should be comprised between 10^{20} and 10^{40} Hz in order to account for the observed values of $\tau_{1/2}$. Hence the activated state should correspond to a partial opening of the DNA. Given that (1) for dsDNA, the measurable unzipping times are between 10 ms and 10 s, (2) the applied voltage is of the order of 100 mV, (3) the average free enthalpy per base pair is around $3k_bT$ and that typical τ_0 are around 10^{-6} s, the number of base pairs involved in the transition state is always between 4 and 8 when derived from the measurements of τ. In addition the value of the effective barrier depends on on the actual shape of the energy profile and may not be readily translated into a number of involved base pairs. Hence the Kramers approach is not sufficient to understand the microscopic mechanism of the unzipping and translocation of long DNA duplexes. It also fails to describe the dependence of $\tau_{1/2}$ with U.

For long duplexes we rather think that the DNA first nucleate an unzipping fork that progressively slide along the sequence. Notice that due to the pore geometry, the helix is shear open rather than pull open (as it happens with optical tweezers). The number of base pairs sustaining the mechanical force in a sheared DNA was theoretically estimated (deGennes 2001) to be 6. Since the first eight base pairs are the same for both sequences, we believe that the nucleation process is not responsible for the slower translocation of λ-RLS. We propose to describe the translocation process of our sequences as the progressive threading of the bases into the pore as they individually unzip. We hypothesized that the observed discrepancy is ascribable to the local probing of the sequence energy profile by the pore. The energy profile differences between both sequences is due to (1) the sequence dependence of base pairing energies as measured in the first neighbor model. (2) The difference of interactions of each base with the pore. This last contribution is believed to account for the sequence dependance of the translocation times for single stranded DNA (Meller et al. 2000; Luo et al. 2007; Gauthier and Slater 2008). However, these interactions involve changes in the translocation time of around 10–100 µs. They thus seem too weak to explain the differences ($\sim$1 s) observed in this study. Furthermore the base content of both sequences being almost similar, we expect their effects to be comparable in both sequences.

To explore the influence of the base pairing energy landscape on the translocation process, we recently developed a model (Bockelmann and Viasnoff 2008) where the unzipping process is described as a biased random walk of the opening fork in a 1D energy landscape calculated using the nearest neighbors model (SantaLucia 1998). The energy required to open the first j ($1 < j < N_b = 45$) bases pairs reads:

$$E_j^{\text{bulk}} = \sum_{i=1}^{j} \Delta H_i - T\Delta S_i \tag{1}$$

where ΔH_i and ΔS_i are the bulk literature values for the ith base pair. We call W the difference in electrostatic potential energy between two bases paired on the *cis* side of the membrane and two bases, unpaired, one on each side of the membrane. We estimate the energy of the DNA strand with j open and translocated bases as

$$E_j = E_j^{\text{bulk}} - jW. \tag{2}$$

W has to be above $\Delta G_0/N_b = 3k_bT$ to favor the unzipping process. We modeled the DNA translocation as the random walk of the last open base in the sequence energy landscape using a Monte Carlo simulation previously described (Bockelmann and Viasnoff 2008) and identified the translocation time to the first passage time at base 45. The

Fig. 4 **a** Integrated distribution of the translocation times $P(\tau)$ for λ-SLS at various voltages: 150, 180, 210, 250, 300, 350 mV increasing from *right* to *left*. The duration of events lasting longer than 60 s was not measured, but they were nonetheless taken into account in the normalization of $P(\tau)$. About 400 events were typically measured per distribution with a minimum of 200 for the smallest voltage. **b** $P(\tau)$ for λ-SLS calculated from the model. We used a typical microscopic attempt frequency of $\nu_0 = 10^6$ Hz to set the time scale of our Monte Carlo simulation. **c** Characteristic rise time $\tau_{1/2}$ versus voltage for λ-RLS (*open square*) and λ-SLS (*filled circle*). The *dotted lines* are guides to the eye. Notice the strong increase of $\tau_{1/2}$ at low voltages. Above 200 mV λ-RLS translocates between two and four times slower than λ-SLS. **d** $\tau_{1/2}$ Versus W as calculated from the model at $T = 24°C$
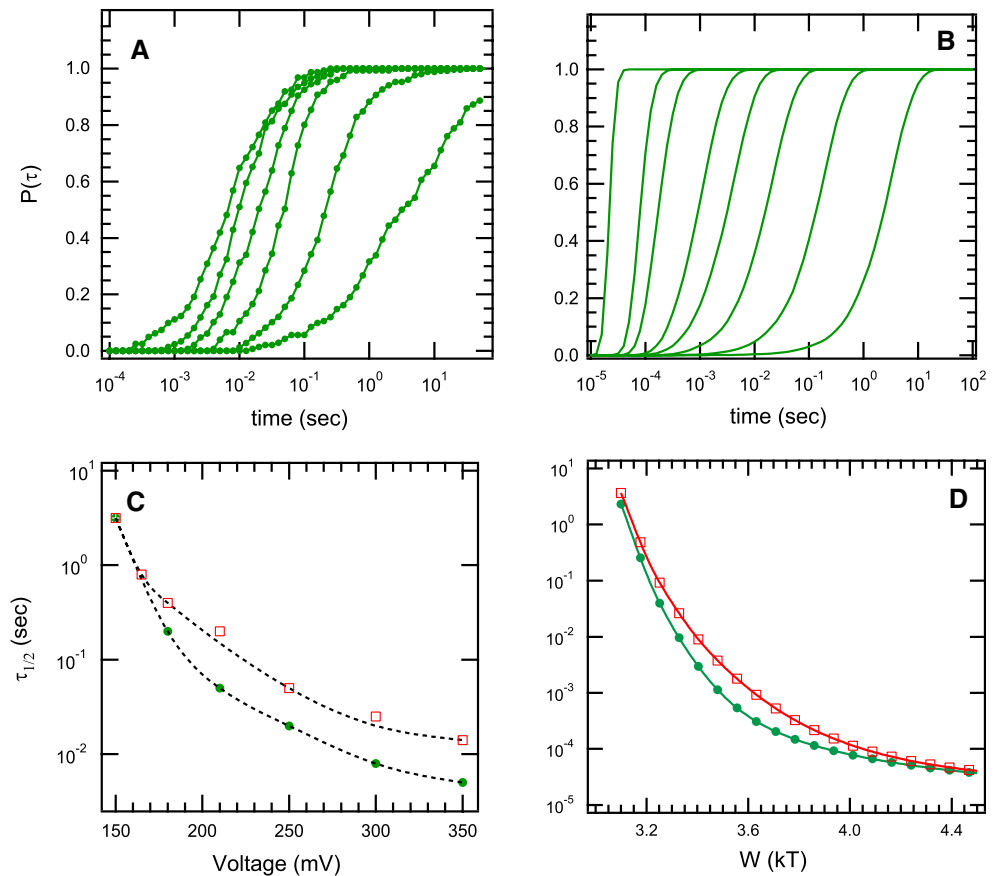
energy profiles for both sequences at $W = 3.5$ are exemplified on Fig. 5. Figure 4b, d shows the predicted distributions and $\tau_{1/2}$ for both sequences. The experimentally observed trends are well accounted by the model: (1) λ-RLS translocates about three times more slowly than λ-SLS due to longer trapping of the unzipping fork in the energy well around base pair 18. (2) The translocating time diverges faster than exponentially at low W. The divergence comes from the apparition of barriers increasing either in number or height as the energy landscape is less tilted. The ratio of the unzipping times of both sequences converges to unity at low W since the sequences have the same ΔG_0. (3) The distributions steepen with increasing voltage due to the vanishing number of local energy minima resulting in a purely diffusive motion of the fork in a 'flat' landscape. (4) The order of magnitude of the predicted values for the translocation times is correct given a typical attempt rate $\nu_0 = 10^6$ Hz.

Interestingly, however, we find two main discrepancies that prevent the fitting of the experiments with the theory (1) the dependance of W versus U is far from linear. Hence W cannot be written $Q_{eff}U$. The predicted $\tau_{1/2}$ varies by five decades for a 25% increase in W whereas doubling U only leads to a three decades decrease of the observed $\tau_{1/2}$. The predicted divergence is too strong. One explanation may be

that the base pair opening involves some collective unzipping. Taking theses effects into account fuels some ongoing work. (2) The predicted shapes of the distributions are not quite correct. At low W, the theoretical distributions become exponential since the unzipping process is dominated by the activated jump over the highest barrier of the landscape. In the low voltage regime the measured distributions are larger than exponential. It can be that the unzipping/translocation process is not a simple 1D process. The position and bent of the helix and the open strand in the pore can potentially induce several opening modes involving different local energy barriers to unzip a few bases. At high W the translocation boils down to a pure diffusion process where the energy landscape plays a minor role and the theoretical distributions become more gaussian like. This high voltage region may not be experimentally accessible. Despite theses discrepancies, our model is able to qualitatively predict, without any ad hoc parameters, the translocation behavior of the tested sequences.

Conclusion

In conclusion, we experimentally showed that two sequences with similar global stability can unzip and
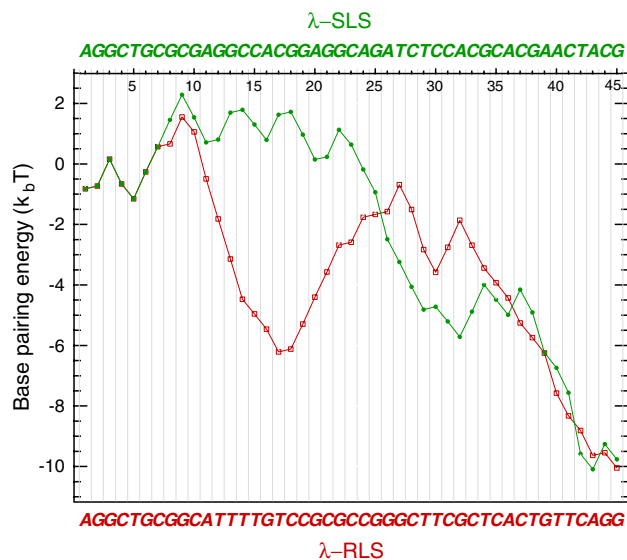
λ–SLS

AGGCTGCGCGAGGCCACGGAGGCAGATCTCCACGCACGAACTACG

AGGCTGCGGCATTTTGTCCGCGCCGGGCTTCGCTCACTGTTCAGG

λ–RLS

Fig. 5 The biased energy landscape $E_j = \sum_{i=1}^{j} \Delta H_i - T\Delta S_i - jW$ for λ-RLS (*open square*) and λ-SLS (*filled circle*) for $W = 3.5$. Notice the existence of a pronounced energy well around base 18 for λ-RLS. The pinning of the unzipping process around this region slows down the translocation of λ-RLS compared to λ-SLS

translocate through an α-hemolysin pore at different speeds. This difference indicates that the translocation of dsDNA involves a sequential unzipping process and that the pore probes the local energy profile of the sequence. The divergence of the translocation times at low voltages is faster than exponential and their distributions cannot be accounted by a jump over an effective energy barrier. We showed that describing the unzipping process as 1D random walk in the sequence dependant energy landscape can qualitatively account for our observations.

References

Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW (1999) Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. Biophys J 77(6):3227–3233

Bockelmann U, Viasnoff V (2008) Theoretical study of sequence-dependent nanopore unzipping of DNA. Biophys J 94(7):2716–2724

Bockelmann U, Thomen P, Essevaz-Roulet B, Viasnoff V, Heslot F (2002) Unzipping DNAn with optical tweezers: high sequence sensitivity and force flips. Biophys J 82(3):1537–1553

Bundschuh R, Gerland U (2005) Coupled dynamics of RNA folding and nanopore translocation. Phys Rev Lett 95(20):208104

Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I, Bustamante C (2005) Verification of the crooks fluctuation theorem and recovery of RNA folding free energies. Nature 437(7056):231–234

deGennes PG (2001) Maximum pull out force on DNA hybrids. C R Acad Sci Paris IV:1505–1508

Dudko OK, Mathe J, Szabo A, Meller A, Hummer G (2007) Extracting kinetics from single-molecule force spectroscopy: nanopore unzipping of DNA hairpins. Biophys J 92(12):4188–4195

Garcia-Manyes S, Brujic J, Badilla CL, Fernandez JM (2007) Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of i27 and ubiquitin. Biophys J 93(7):2436–2446

Gauthier MG, Slater GW (2008) A monte carlo algorithm to study polymer translocation through nanopores. i. theory and numerical approach. J Chem Phys 128(6):065–103

Greenleaf WJ, Woodside MT, Block SM (2007) High-resolution, single-molecule measurements of biomolecular motion. Annu Rev Biophys Biomol Struct 36:171–190

Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM (2008) Direct observation of hierarchical folding in single riboswitch aptamers. Science 319(5863):630–633

Harlepp S, Marchal T, Robert J, Leger JF, Xayaphoumine A, Isambert H, Chatenay D (2003) Probing complex RNA structures by mechanical force. Eur Phys J E 12(4):605–615

Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci USA 93(24):13770–13773

Lakatos G, Chou T, Bergersen B, Patey GN (2005) First passage times of driven DNA hairpin unzipping. Phys Biol 2(3):166–174

Larson MH, Greenleaf WJ, Landick R, Block SM (2008) Applied force reveals mechanistic and energetic details of transcription termination. Cell 132(6):971–982

Li PTX, Bustamante C, Tinoco I (2007) Real-time control of the energy landscape by force directs the folding of RNA molecules. Proc Natl Acad Sci USA 104(17):7039–7044

Luo KF, Ala-Nissila T, Ying SC, Bhattacharya A (2007) Heteropolymer translocation through nanopores. J Chem Phys 126(14):145101

Mathe J, Visram H, Viasnoff V, Rabin Y, Meller A (2004) Nanopore unzipping of individual DNA hairpin molecules. Biophys J 87(5):3205–3212

Meller A, Branton D (2002) Single molecule measurements of DNA transport through a nanopore. Electrophoresis 23(16):2583–2591

Meller A, Nivon L, Brandin E, Golovchenko J, Branton D (2000) Rapid nanopore discrimination between single polynucleotide molecules. Proc Natl Acad Sci USA 97(3):1079–1084

Meller A, Nivon L, Branton D (2001) Voltage-driven DNA translocations through a nanopore. Phys Rev Lett 86(15):3435–3438

Mergny JL, Lacroix L (2003) Analysis of thermal melting curves. Oligonucleotides 13(6):515–537

SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95(4):1460–1465

Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Branton D (2003) Unzipping kinetics of double-stranded DNA in a nanopore. Phys Rev Lett 90(23):238101

Tropini C, Marziali A (2007) Multi-nanopore force spectroscopy for DNA analysis. Biophys J 92(5):1632–1637

Vercoutere W, Winters-Hilt S, Olsen H, Deamer D, Haussler D, Akeson M (2001) Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. Nat Biotechnol 19(3):248–252

Vercoutere WA, Winters-Hilt S, DeGuzman VS, Deamer D, Ridino SE, Rodgers JT, Olsen HE, Marziali A, Akeson M (2003) Discrimination among individual watson–crick base pairs at the

termini of single DNA hairpin molecules. Nucleic Acids Res 31(4):1311–1318

Zhao Q, Comer J, Dimitrov V, Yemenicioglu S, Aksimentiev A, Timp G (2008) Stretching and unzipping nucleic acid hairpins using a synthetic nanopore. Nucleic Acids Res 36(5):1532–1541

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13): 3406–3415