Coupled Dynamics of RNA Folding and Nanopore Translocation

Ralf Bundschuh[1] and Ulrich Gerland[2]

[1]*Department of Physics, The Ohio State University, Columbus, Ohio 43210-1117, USA*
[2]*Arnold-Sommerfeld Center for Theoretical Physics and Center for Nanoscience (CeNS), LMU München,*
Theresienstrasse 37, 80333 München, Germany
(Received 27 July 2005; published 10 November 2005)

The translocation of structured RNA or DNA molecules through narrow pores necessitates the opening of all base pairs. Here, we study the interplay between the dynamics of translocation and base pairing theoretically, using kinetic Monte Carlo simulations and analytical methods. We find that the transient formation of base pairs that do not occur in the ground state can significantly speed up translocation.

The dynamics of base pairing in DNA and RNA molecules plays an important role in many biological processes ranging from DNA replication to RNA folding [1]. Often, the dynamics is coupled to other kinetic processes. For instance, RNA folds as it is synthesized, while it unfolds and refolds as it passes through the ribosome during translation. A series of recent single-molecule experiments [2] used electric fields to drive RNA and DNA through tiny pores, which let single but not double strands pass. The case of unstructured, e.g., homopolymeric molecules, is particularly well characterized, with many theoretical studies complementing the experiments [3–5]. In contrast, the effect of base pairing on translocation [see Fig. 1] is only beginning to be explored, with a few existing experiments [6–8] and first steps towards a theoretical description [7,9]. In this Letter, we specifically examine the interplay between base-pairing and translocation dynamics.

Already homopolymer translocation displays rich dynamics. Experimentally, one can obtain the full distribution $p(\tau)$ of the times τ for the translocation of a molecule from the *cis* to the *trans* side of the pore. Hence, the goal of theoretical descriptions is to determine $p(\tau)$ or its moments as a function of system parameters such as the chain length N, the voltage drop across the pore, and temperature. Most treatments reduce the dynamics to a one-dimensional reaction coordinate, m, which measures how many bases have reached the *trans* side [3,4]. This approach is justified only when the polymer degrees of freedom equilibrate rapidly compared to the time scale of translocation. It fails for long chains, where the internal polymer dynamics and friction with the solvent limit the translocation speed [5]. However, most experiments so far were performed with short chains where the friction between the polymer and the pore dominates the translocation dynamics [2,4]. In a minimal model for these experiments the index m is increased with a forward rate k_+ and decreased with a backward rate k_-. When expressed in terms of a continuous reaction coordinate $0 < x < N$, the drift-diffusion equation

$$\partial_t P(x, t) = D\partial_x^2 P(x, t) - v\partial_x P(x, t) \qquad (1)$$

describes the time evolution of the probability distribution

for x, where D is an effective diffusion constant and the drift velocity v is induced by the applied voltage. The ratio $\ell = D/v$ defines a diffusive length scale, below which the drift v is negligible compared to the Brownian motion [4]. From (1), one obtains $p(\tau)$ as the probability current into the absorbing boundary at $x = N$.

Base-pairing interactions affect the translocation dynamics significantly by introducing sequence- and structure-dependent kinetic barriers [6–8]. Previous theoretical work [9] argues that these barriers can be exploited to measure the secondary structure of an RNA molecule, if its sequence is already known. However, the interplay of the translocation and base-pairing dynamics also poses new physical questions, which will be important for the design of applications. For instance, one expects two limiting cases for the dynamics: *slow translocation*, during which the base-pairing pattern remains equilibrated at almost all times and on both sides of the pore, and *fast translocation*, where an essentially frozen base-pairing pattern on the *cis* side is unzipped as the molecule passes through the pore. Reference [9] considers only the latter limit. Here, we are interested in the crossover from the slow to the fast translocation limit and address the following questions: (i) Can the physics of the crossover be described in simple terms? (ii) What is the nature and
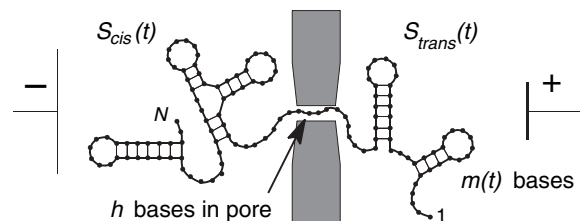

FIG. 1. Translocation of a structured RNA molecule through a narrow pore, which allows single but not double strands to pass. The driving force is an applied voltage, which acts on the negatively charged backbone of the RNA. An appropriate reaction coordinate is the number of bases $m(t)$ that have reached the *trans* side. The dynamics of $m(t)$ is coupled to the dynamics of the base-pairing patterns $S_{cis}(t)$ and $S_{trans}(t)$ on the *cis* and *trans* sides.

size of the effective free energy barriers for translocation? (iii) How is the distribution $p(\tau)$ affected by the presence of secondary structure?

Whether an experiment is closer to the slow or the fast translocation limit can depend on the sequence or structure of the molecule under study, the bias voltage, and the friction coefficient between the polymer and the pore. The latter determines the rate k_0 for translocation by a single (unpaired) base at zero bias, which we use as a control parameter for the crossover. Microscopically, k_0 is determined by the diameter of the pore and its surface chemistry, both of which are in principle tunable, either by mutation in the case of protein pores or through the fabrication process in the case of solid-state pores [10,11]. We will see below that the pore depth is also a relevant parameter for the questions raised above.

Model.—We focus on the case of RNA, even though our results apply equally to single-stranded DNA. Since kinetic barriers slow the translocation of structured RNA, we expect the description by the reaction coordinate m to be valid over an even wider regime of lengths N than for unstructured RNA. However, for structured RNA, the dynamics of $m(t)$ is coupled to the dynamics of the RNA base-pairing patterns $S_{cis}(t)$, $S_{trans}(t)$ on both sides of the pore. In our model, illustrated in Fig. 1, only unpaired bases can enter the pore. We describe the dynamics of $S_{cis}(t)$ and $S_{trans}(t)$ using Monte Carlo kinetics with three elementary moves: opening of a pair, closing, or a shift in the binding partner of a base [12]. We use a single rate k_{bp} for all moves that are energetically favorable, and the rate $k_{bp}e^{-\Delta G}$, if the move increases the Gibbs free energy by ΔG (all energies are in units of the thermal energy $k_B T$). While this assumption certainly does not hold on the microscopic level, we expect that it will not affect the qualitative features of the dynamics on long time scales. Indeed, a recent model of force-induced RNA unfolding [13] has shown a remarkably good agreement with experiment using similar assumptions.

To calculate the Gibbs free energies, we use a simplified energy model, which allows for CG, AU, and GU base pairs with a binding energy of 2, 1, and 0, respectively. The free energy cost for any loop (interior, bulge, multiloop, or hairpin) of size ℓ is $3\nu \ln(\ell)$, where $\nu \approx 0.6$ is the Flory exponent. We exclude hairpin loops of size $\ell < 3$, which cannot occur in RNA structures due to steric constraints. The applied voltage drops primarily directly across the pore, leading to an energy gain ΔU when a monomer traverses the pore. Thus, the ratio of the forward and backward rates is given by the Boltzmann factor $e^{\Delta U}$. We assume the symmetric choice $k_{\pm} = k_0 e^{\pm \Delta U/2}$ corresponding to a centered transition state along the microscopic reaction pathway for translocation by one monomer. We denote by h the number of bases that fit inside the pore and assume these cannot base pair at all. While translocation consists of an entrance stage followed by a passage stage, we focus only on the passage dynamics. Hence, we choose

$m(0) = 0$ as the initial condition and stop the simulation when $m(\tau) = N$, whereas we do not allow the RNA to exit on the *cis* side.

Mean translocation times.—To explore the effect of base pairing on the translocation dynamics, we use three different sequences with identical length $N = 50$; one well-defined structure (a hairpin with 23 random base pairs and a loop of 4 bases), one random sequence, and an unstructured homopolymer for comparison. Figure 2 shows their mean translocation times $\langle \tau \rangle$ for different combinations of the parameters k_0, k_{bp}, and ΔU (here we have used an idealized pore with $h = 1$). The average is taken over 1000 independent simulations, and $\langle \tau \rangle$ is plotted in units of the "hopping time" k_0^{-1} for an unpaired base across the pore at zero bias.

Figure 2(a) displays the transition from the slow to the fast translocation limit by varying the ratio of the two characteristic time scales in our model, k_0/k_{bp}, at constant bias $\Delta U = 0.5$. For the homopolymer, the base-pairing time scale is irrelevant and hence $k_0\langle \tau \rangle$ is constant. For the structured RNAs, the curves also approach a constant limit at small k_0/k_{bp}, indicating that translocation is sufficiently slow to allow equilibration of the base-pairing pattern whenever the RNA translocates by a base. Note that this limit value of $k_0\langle \tau \rangle$ depends strongly on the sequence or structure of the RNA. In the other extreme, $k_0 \gg k_{bp}$, the base-pair opening in front of the pore is the rate limiting process. The crossover between these limits occurs at $k_0 \sim k_{bp}$. We will now seek a quantitative understanding of these simulation results.

Slow translocation limit.—In the limit $k_0 \ll k_{bp}$, the translocation dynamics $m(t)$ becomes independent of the base-pairing kinetics. The effect of base pairing then
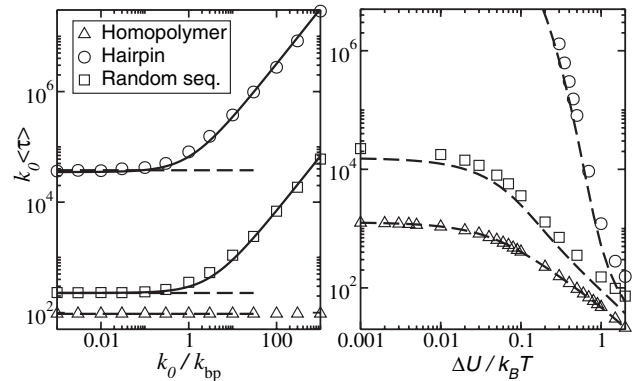

FIG. 2. Mean translocation time $\langle \tau \rangle$ for RNAs with the same length $N = 50$, but different structures [we plot $\langle \tau \rangle$ in units of k_0^{-1}, which is $O(\mu s)$ in typical experiments [2]]. Symbols represent simulation data, while dashed lines show the adiabatic limit (2) and solid lines the expression (3) for the two random walker model. (a) Crossover from slow to fast translocation at a fixed bias $\Delta U = 0.5 k_B T$. Typical base-pairing rates are $k_{bp} \sim 1\ \mu s^{-1}$ [17], so that typical experiments would fall into the crossover region. (b) Dependence on the bias at a fixed ratio $k_0/k_{bp} = 1$.

amounts to reducing the translocation rates k_+, k_- by factors equal to the probability that the base entering the pore is unpaired. Such equilibrium properties of the base-pairing interaction can be expressed through the partition function $Z_{i,j}$, the total statistical weight of all secondary structures for the substrand from base i to base j of the RNA sequence. For instance, the probability for the jth base on the *trans* side to be unbound is $Z_{1,j-1}/Z_{1,j}$. Hence, the translocation dynamics $m(t)$ becomes a 1D random walk with site-dependent hopping rates $k_-(m) = k_- Z_{1,m-1}/Z_{1,m}$ and $k_+(m) = k_+ Z_{m+h+2,N}/Z_{m+h+1,N}$. Using the mean first passage time formalism [14], we then find the mean translocation time

$$\langle \tau \rangle = \frac{1}{k_+} \sum_{j=m(0)}^{N-1} \sum_{k=0}^{j} e^{\mathcal{F}_{h+1}(j) - \mathcal{F}_h(k) - (j-k)\Delta U}, \qquad (2)$$

where $\mathcal{F}_h(m) = -\ln(Z_{1,m} Z_{m+h+1,N}/Z_{1,N})$ is the free energy cost of placing a free RNA from solution into position m of a pore with depth h. We obtain the partition functions $Z_{i,j}$ by generalizing the well-known recursion relations of [15] to a logarithmic multiloop cost [16], in order to derive the exact adiabatic limit of our model.

In Fig. 2(a), the dashed lines show that our kinetic simulations indeed converge to the adiabatic limit (2) for small k_0/k_{bp}. However, besides providing a check on our simulations, Eq. (2) also helps one to understand the nature of the relevant free energy barriers. For example, the hairpin structure has a total binding free energy of $\approx 33 k_B T$. Yet, its translocation time compared to that of the homopolymer is not more than 1000 times longer (at $\Delta U = 0.5 k_B T$ and $k_0 < k_{bp}$). The applied voltage lowers the free energy barrier only about $10 k_B T$ and does not suffice to explain the comparatively small effect of the stable structure on the translocation time. Inspection of the free energy landscape $\mathcal{F}_1(m)$ of the hairpin, shown in the inset of Fig. 3(a), reveals that the formation of non-native base pairs on both sides of the pore lowers $\mathcal{F}_1(m)$ considerably in the central part and explains the effect (for comparison, the landscape for only the native hairpin structure is shown with a dashed line). Hence, the major barrier for the translocation of structured molecules is not unfolding of the complete structure, but partial unfolding of the structure until the formation of non-native base pairs speeds up translocation.

Crossover to fast translocation.—In Fig. 2(a), the crossover between slow and fast transition occurs in the range $0.1 < k_0/k_{bp} < 10$ for both structured RNA sequences. Incidentally, typical experimental values for the base-pairing and translocation rates are comparable with $k_{bp} \sim k_0 \sim 1 \; \mu s^{-1}$ [2,17]. Hence it is important to characterize the translocation behavior in the transition region.

Note that while the higher stability of the hairpin yields a longer translocation time, the shape of the prominent bend in the curves of Fig. 2(a) is independent of the sequence as long as it is structured. This observation encourages us to
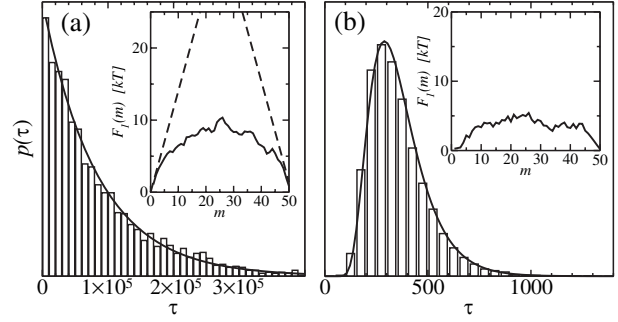

FIG. 3. Translocation time distributions for (a) the RNA hairpin and (b) the random sequence at $\Delta U = 0.5$ and $k_0 = k_{bp}$. The histograms show simulation data, whereas the solid line in (a) is an exponential fit and in (b) is a fit to the drift-diffusion form [4]. The insets show the free energy landscapes $\mathcal{F}_1(m)$ of the adiabatic limit (2); see main text for details.

develop a simplistic model: The dominant effect giving rise to the prominent bend is a kinetic competition between the opening and closing dynamics of the base pair next to the *cis* side of the pore, and the position of the pore with respect to the RNA, which is driven into the structure by the external bias. This process resembles two impenetrable random walkers biased to run into each other. Walker P representing the pore has a hopping rate k_+ towards walker R representing the RNA and a rate k_- away from R. Walker R, in turn, moves towards P at the rate k_{bp} for base-pair closing, and away at an unbinding rate k_u that is reduced by the stability of the base pair. If none of these rates depends on the positions of R and P, one finds that the center of mass position of the walkers drifts with a velocity $v = (k_u k_+ - k_{bp} k_-)/(k_{bp} + k_+)$. Assuming this drift dominates over diffusion, $\langle \tau \rangle$ takes the simple form

$$\langle \tau \rangle = N/v = \frac{N}{k_0}\left(a + b\frac{k_0}{k_{bp}}\right) \qquad (3)$$

with dimensionless constants a and b. The solid lines in Fig. 2(a) show Eq. (3), where we determined a from the slow translocation limit (2) and used b as the fitting parameter. The agreement with the simulation data suggests that the shape of the curves is indeed largely determined by a kinetic competition between translocation and the local base-pairing dynamics in the vicinity of the pore. The effects of large-scale rearrangements in the base-pairing pattern during translocation are too subtle to be discernible in Fig. 2(a). However, reformation of base pairs on the *trans* side, which we have ignored to obtain Eq. (3), strongly influences the parameter b.

Voltage dependence.—In Fig. 2(b) the mean translocation time of our three test sequences is plotted against the bias ΔU, where symbols represent simulation data at $k_{bp} = k_0$ and dashed lines the adiabatic limit (2). For the homopolymer, the adiabatic limit is exact and displays the well-known crossover from a diffusion-dominated $\langle \tau \rangle \sim N^2/k_0$ at $\Delta U \ll 1/N$ to a drift-dominated $\langle \tau \rangle \sim N/k_+$ at $\Delta U \gg$

$1/N$. For the structured RNAs, the mean translocation time also levels off at $\Delta U \sim 1/N$, since the molecule "feels" the bias only if its total free energy differs by at least one $k_B T$ between the two sides of the pore. In this case, however, the simulation data lie somewhat above the adiabatic limit (2), consistent with the fact that $k_{bp} = k_0$ is within the crossover regime of Fig. 2(a). Clearly, as we leave the adiabatic limit, the effective kinetic barriers for translocation increase and (2) gives only a lower bound on the mean translocation time. As ΔU is increased to $\sim k_B T$ (right edge of the plot), the system is more strongly perturbed and the deviation from the adiabatic limit becomes more significant.

Translocation time distributions.—In Fig. 2(b), the translocation dynamics of the homopolymer is drift dominated for $\Delta U > 1/N$. Is this equally true for the translocation dynamics of the structured RNAs? To address this question, we examine the full distribution $p(\tau)$, which displays more detailed signatures of the translocation dynamics than the average $\langle \tau \rangle$ and is routinely measured experimentally. The histograms in Figs. 3(a) and 3(b) show $p(\tau)$ for (a) the hairpin and (b) the random sequence, from simulations at $k_0 = k_{bp}$ and $\Delta U = 0.5 \gg 1/N$. We observe that the distributions are markedly different: For the hairpin, $p(\tau)$ follows an exponential law (solid line) characteristic for thermally activated crossing of a single free energy barrier. In contrast, $p(\tau)$ for the random sequence has the shape characteristic for the drift-diffusion process (1), which is appropriate for *unstructured* polymers [the solid line in Fig. 3(b) is a fit with the associated mean first passage time distribution [4]].

Indeed, the free energy landscape $\mathcal{F}_1(m)$ of the random sequence is remarkably flat [see inset of Fig. 3(b)]. This is due to the fact that the random sequence can accommodate many dissimilar structures with similar free energies: when the pore breaks up part of the structure of the random sequence, a (non-native) structure can reform on the *trans* side, which is almost equally stable. Hence, the translocation dynamics of the random sequence is qualitatively similar to that of the homopolymer, only with a reduced diffusion coefficient D and drift velocity v due to the roughness of the free energy landscape.

Conclusions.—Compared to the thermodynamics, the dynamics of base pairing is far less understood. Translocation experiments are emerging as a valuable tool to probe these dynamics on the single-molecule level. Here, we have studied the interplay between translocation and base pairing within a simplified but explicit theoretical model. In the slow translocation limit, the translocation dynamics of our model can be understood analytically, which is useful to interpret our simulations even outside the slow translocation regime. We find that the formation of non-native base pairs can significantly speed up translocation, rendering the effective kinetic barriers much smaller than might be naively expected. Furthermore, we

showed that the crossover between slow and fast translocation can be described by a simple phenomenological model. In the future, nanopores might be used to sort RNAs according to their structures.

[1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland, New York, 2002).
[2] J. Kasianowicz, E. Brandin, D. Branton, and D. Deamer, Proc. Natl. Acad. Sci. U.S.A. **93**, 13 770 (1996); A. Meller, J. Phys. Condens. Matter **15**, R581 (2003); A. J. Storm, J. H. Chen, H. W. Zandbergen, and C. Dekker, Phys. Rev. E **71**, 051903 (2005).
[3] W. Sung and P. J. Park, Phys. Rev. Lett. **77**, 783 (1996); M. Muthukumar, J. Chem. Phys. **111**, 10 371 (1999); E. Slonkina and A. Kolomeisky, J. Chem. Phys. **118**, 7112 (2003).
[4] D. Lubensky and D. Nelson, Biophys. J. **77**, 1824 (1999).
[5] J. Chuang, Y. Kantor, and M. Kardar, Phys. Rev. E **65**, 011802 (2002); A. Milchev, K. Binder, and A. Bhattacharya, J. Chem. Phys. **121**, 6042 (2004).
[6] W. Vercoutere *et al.*, Nat. Biotechnol. **19**, 248 (2001).
[7] A. Sauer-Budge, J. Nyamwanda, D. Lubensky, and D. Branton, Phys. Rev. Lett. **90**, 238101 (2003).
[8] J. Mathe *et al.*, Biophys. J. **87**, 3205 (2004).
[9] U. Gerland, R. Bundschuh, and T. Hwa, Phys. Med. Biol. **1**, 19 (2004).
[10] J. Li *et al.*, Nature (London) **412**, 166 (2001).
[11] A. J. Storm *et al.*, Nat. Mater. **2**, 537 (2003).
[12] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster, RNA **6**, 325 (2000).
[13] S. Harlepp *et al.*, Eur. Phys. J. E **12**, 605 (2003).
[14] C. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1983).
[15] J. McCaskill, Biopolymers **29**, 1105 (1990).
[16] Whereas [15] linearizes the free energy of multiloops, we retain the logarithmic form derived from polymer theory. The detailed calculation will be published elsewhere. Briefly, we split $Z_{i,j}$ into a contribution $B_{i,j}$ from structures including the base pair (i, j), and contributions $U_{i,j}^\ell$ without this base pair, $Z_{i,j} = B_{i,j} + \sum_{\ell=0}^{j-i+1} U_{i,j}^\ell$, where ℓ is the number of exterior unbound bases. We denote by $Q_{j,j-1}^{i,i+1}$ the Boltzmann factor for the stacking of the base pairs (i, j) and $(i + 1, j - 1)$. For $B_{i,j}$ we then obtain $B_{i,j} = Q_{j,j-1}^{i,i+1} B_{i+1,j-1} + \sum_\ell \ell^{-3\nu} U_{i+1,j-1}^\ell$, if the bases i and j match, and zero otherwise. Similarly, we find $U_{i,j}^\ell = U_{i+1,j}^{\ell-1} + \sum_k B_{i,k}(U_{k+1,j}^\ell + \delta_{\ell,0} B_{k,j}) + \delta_{\ell,1} B_{i+1,j}$. With these relations, and the initial conditions $B_{i,j} = 0$ and $U_{i,j}^\ell = \delta_{\ell,j-i+1}$ for $j - i < 4$, the partition functions are calculated in $O(N^4)$ time—slower than the $O(N^3)$ time of [15], but feasible for sequences up to 1000 bases.
[17] M. E. Craig, D. M. Crothers, and P. Doty, J. Mol. Biol. **62**, 383 (1971).