Force-Induced Denaturation of RNA

Ulrich Gerland, Ralf Bundschuh, and Terence Hwa
Department of Physics, University of California at San Diego, La Jolla, California 92093-0319 USA

ABSTRACT We quantitatively describe an RNA molecule under the influence of an external force exerted at its two ends as in a typical single-molecule experiment. Our calculation incorporates the interactions between nucleotides by using the experimentally determined free energy rules for RNA secondary structure and models the polymeric properties of the exterior single-stranded regions explicitly as elastic freely jointed chains. We find that despite complicated secondary structures, force-extension curves are typically smooth in quasi-equilibrium. We identify and characterize two sequence/structure-dependent mechanisms that, in addition to the sequence-independent entropic elasticity of the exterior single-stranded regions, are responsible for the smoothness. These involve compensation between different structural elements on which the external force acts simultaneously and contribution of suboptimal structures, respectively. We estimate how many features a force-extension curve recorded in nonequilibrium, where the pulling proceeds faster than rearrangements in the secondary structure of the molecule, could show in principle. Our software is available to the public through an "RNA-pulling server."

INTRODUCTION

In recent years, single-molecule experiments using optical tweezers, atomic force microscopy, and other techniques have successfully probed basic physical properties of biomolecules through the application of forces in the pN range (see, e.g., Bockelmann et al., 1997; Essevaz-Roulet et al., 1997; Mehta et al., 1999, and references therein; Rief et al., 1997, 1999; Smith et al., 1996; Yang et al., 2000). Both simple elastic properties of the polymers (such as persistence length and longitudinal elasticity) and structural transitions (e.g., unfolding of protein domains) were characterized by recording and analyzing force-extension curves (FECs). For nucleic acids, a prominent experiment of the latter type is the "unzipping" of double-stranded DNA (Bockelmann et al., 1997; Essevaz-Roulet et al., 1997). The resulting FECs display clear sequence-specific features (e.g., local maxima), which may be attributed to small regions of the sequence that are more strongly bound than their neighbors (Essevaz-Roulet et al., 1997; Lubensky and Nelson, 2000; Thompson and Siggia, 1995). In contrast, long single-stranded DNA, which, like RNA, may fold into complicated branched structures by forming intra-strand basepairs, showed extremely smooth FECs in a very recent experiment by Maier et al. (2000). Thus, depending on its structure, DNA may show a broad range of FECs from very rugged to completely featureless. However, it is unclear *how* quantitatively the structure determines the outcome of the FEC measurement.

Here, we address this question theoretically, focusing on the case of RNA and restricting ourselves to secondary structure (i.e., basepairing patterns only instead of full, tertiary structure). In this context, RNA seems to be a more interesting object than DNA because RNA naturally occurs in many different and functionally important structures, while DNA is primarily found as a double strand. One may hope that pulling experiments generate new insights into the RNA folding problem (Tinoco and Bustamante, 1999, and references therein), including the folding pathways (Chen and Dill, 2000; Isambert and Siggia, 2000; Thirumalai and Woodson, 2000, and references therein). Also, force-induced denaturation of RNA is currently studied experimentally (C. Bustamante and I. Tinoco, private communication). The limitation to secondary structure allows us to draw upon the experimentally determined "free energy rules" for RNA secondary structure (Freier et al., 1986; Mathews et al., 1999; Walter et al., 1994), which yield minimum free energy structures that agree reasonably well with experimentally and phylogenetically determined ones (Mathews et al., 1999). Furthermore, it permits us to use and extend the efficient dynamic programming algorithms (Hofacker et al., 1994; McCaskill, 1990; Zuker and Stiegler, 1981) that can compute the exact partition function (including all possible secondary structures) and reconstruct the minimal free energy structures in polynomial time. Experimentally, the secondary structures may be probed in specific ionic conditions (e.g., those with only monovalent ions) such that the tertiary contacts are strongly disfavored (due to electrostatic repulsion of the sugar-phosphate backbone) (Tinoco and Bustamante, 1999, and references therein).

The type of experiment that we consider is sketched in Fig. 1. The distance R between the two ends of an RNA molecule is held fixed, e.g., by attaching them to two beads whose positions are controlled by optical tweezers, and the force f acting on the beads is recorded as a function of R. As long as the external change in force/extension is applied at a much slower time scale than that of structural transitions of the molecule, the equilibrium FEC is measured. In the main part of the present article we assume that this is always the case. Experimentally, this condition is usually checked

Received for publication 23 January 2001 and in final form 17 May 2001.
Address reprint requests to Dr. Ulrich Gerland, Physics Dept. 319, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0319. Tel.: 858-822-3487; Fax: 858-534-7697; E-mail: gerland@matisse.ucsd.edu.

0006-3495/01/09/1324/09 $2.00

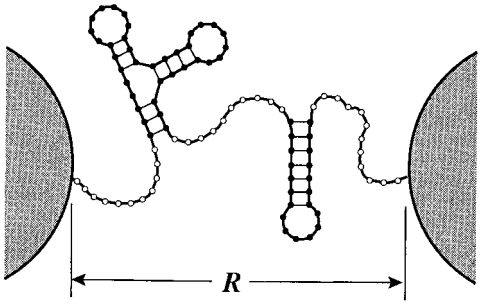

FIGURE 1 Sketch of the pulling experiment considered in the text: the two ends of an RNA molecule are attached to beads (*shaded gray*) and held fixed at distance R, while the force f acting on the beads is measured. The open circles represent the open bases of the exterior single strands, modeled here as elastic freely jointed chains.

by retracing the FEC (e.g., a hysteresis effect is a clear sign of a nonequilibrium situation).

Besides the above-mentioned free energy parameters for RNA secondary structure, we need a polymer model for single-stranded RNA as input to make quantitative predictions of FECs. To that end, we use an elastic freely jointed chain model that has been used to fit experimental FECs of single-stranded DNA (Montanari and Mézard, 2001; Smith et al., 1996). This introduces two polymer parameters, the Kuhn length characterizing the lateral rigidity, and the longitudinal elasticity, which is determined by the forces needed to stretch the chemical structure of the backbone. We estimate both from the experiments on DNA, so that we are left with no free parameters.

We find that for different secondary structures with all other parameters (temperature, sequence length, etc.) fixed, the FECs of RNA vary over a broad range from very rugged to very smooth. Apart from the entropic elasticity of the exterior single strand, which smooths the features in the FEC independent of the secondary structure as already discussed by Thompson and Siggia (1995), there are two additional smoothing mechanisms. The first is a "compensation effect": the increase in the length of the exterior single strand upon opening of a structural element and the associated drop in the tension may be absorbed by rebinding of bases from the exterior single strand in other structural elements. The second is due to thermal fluctuations in the secondary structure, i.e., the contribution of suboptimal structures. We discuss both mechanisms and analyze the fluctuations in the FEC quantitatively. The equilibrium FECs of typical (natural or random) RNA sequences are smooth and display no distinguishable signatures of individual structural elements opening. This is consistent with the experimental result of Maier et al. (2000) for single-stranded DNA, but applies even for sequences with only a few hundred nucleotides, i.e., for much shorter sequences than used in their experiment.

For the purpose of obtaining information on the structure of RNA, the measurement of equilibrium FECs is therefore not very useful. More promising options include the measurement of the fluctuations about the equilibrium and nonequilibrium FECs, where the pulling proceeds faster than (some of) the rearrangements in the structure. Although the present approach is extended readily to include equilibrium fluctuations (Gerland, U., R. Bundschuh, and T. Hwa, in preparation), a quantitative treatment of the dynamics of force-induced denaturation of RNA presents a challenge to theoreticians.

The organization of the paper is as follows. In the next section, we explain the details of our model and the way we calculate the FECs. Readers interested in the results only should directly proceed to that section. The Discussion section explores the possibility of using experimental FECs of appropriately designed sequences as an alternative way to determine the RNA free energy parameters. In addition, we estimate to what extent features may be expected in nonequilibrium FECs.

MODEL AND METHODS

We assume that the force $f(R)$ acting on the beads (see Fig. 1) is measured as a function of the fixed distance $R = |\mathbf{R}|$, where $\mathbf{R}$ denotes the end-to-end vector of the RNA molecule, and that R is varied very slowly so that thermal equilibrium is always maintained. In practice, the force measurement requires a device acting as a spring, hence the distance cannot be kept exactly constant. However, we consider the situation where the stiffness of this spring is much higher than that of the single-stranded RNA, which has already been pulled out. This condition could only be violated in the very early part of the pulling experiment, which is not the focus of the present investigation. We may therefore neglect the presence of the spring altogether, which amounts to working in the "fixed-distance ensemble" (see Note 1 at end of text). Another difference between our model and actual experiments is that we neglect the presence of additional spacer sequences, which are used to connect the RNA molecule to the force-measuring device (e.g., the beads). Again, we assume that they are stiffer than the liberated single-stranded RNA because we are interested in the size of the features in the FEC, which are observable in an ideal measurement.

The partition function at fixed extension, $Z_N(R)$, for a given RNA sequence consisting of N nucleotides, may be written as a sum over the number m of exterior open bases (as represented by open circles in Fig. 1). For each m the secondary structure contributes a factor $\mathcal{Q}_N(m)$ to the partition function, according to the free energy rules for RNA/DNA secondary structure to be detailed shortly below. This contribution needs to be weighted by the probability $W(\mathbf{R}; m)$ that the chain of m exterior open bases has end-to-end vector $\mathbf{R}$, given by an appropriate polymer model for the single strand. Together, they yield

$$Z_N(R) = \sum_m \mathcal{Q}_N(m) W(\mathbf{R}; m). \quad (1)$$

The normalization $\int d^3R\, W(\mathbf{R}; m) = 1$ assures that the integral of $Z_N(R)$ over space yields the usual partition function Z_N for N nucleotides without any external constraints. Equation 1 clearly separates the contribution of the secondary structure, which is entirely contained in $\mathcal{Q}_N(m)$, from the contribution of the exterior single strand contained in $W(\mathbf{R}; m)$. Note that the polymer properties of the *interior* single strands (i.e., the single strands not subject to the external force) are contained in $\mathcal{Q}_N(m)$ through the loop-entropy parameters, which are part of the free energy rules derived from experiments (see Walter et al. (1994) and references therein).

Secondary structure

The number of possible secondary structures for a given sequence of length N grows exponentially with N. To each structure $\mathcal{S}$, a Boltzmann weight $\zeta(\mathcal{S})$ may be assigned with the help of the free energy rules (Walter et al., 1994) which contain a large number of experimentally determined energy and enthalpy parameters, e.g., those for the stacking of basepairs, formation of internal, hairpin, bulge or multi-loops, and dangling ends. Due to the large number of possible structures, the full partition function $Z_N \equiv \Sigma_{\mathcal{S}} \zeta(\mathcal{S})$ is impossible to evaluate by enumeration, except for very small N. However, one can make use of recursion relations that express the partition function for a subsequence with the help of the partition functions for even shorter subsequences (McCaskill, 1990; Zuker and Stiegler, 1981), and proceed to compute the full partition function exactly in $O(N^3)$ time. These recursion relations owe their existence to the fact that the class of secondary structures was defined to include only *nested* structures, e.g., two basepairs (i, j) and (k, l) with $i < k < j < l$ are not admitted (the occurrence of such pairings is called a pseudoknot and contributes relatively little to the free energy of natural RNAs (Tinoco and Bustamante, 1999). One implementation of this algorithm with very detailed free energy rules is the "Vienna package" (Hofacker et al., 1994, publically available at http://www.tbi.univie.ac.at/). In the following, we describe the modifications that we made to this package to obtain $\mathcal{Z}_N(m)$ and the corresponding minimum free energy structures.

The Vienna package calculates the auxiliary partition function $\Pi(i, j)$ for the substrand (i.e., a contiguous segment of the sequence) from base i to base j, under the condition that base i and base j are paired. These quantities can be used to calculate the partition function $Q(j; n)$ of the substrand from base 1 to base j, under the condition that the *exterior* part of the configurations is $0 \le n \le j$ bases long. The recursion formula for Q is

$$Q(j+1; n) = Q(j; n-1) + \sum_{i=n-\Delta+1}^{j} Q(i-1; n-\Delta)\Pi(i, j+1),$$

obtained by splitting the partition function $Q(j + 1; n)$ up according to all possible binding partners of base $j + 1$. [Here, the constant $\Delta = 3$ accounts for the fact that each stem branching from the exterior single strand contributes an additional segment, whose length is approximately equal to the length of three single-stranded bases.] This formula, together with the appropriate boundary conditions for $j = 0$ and $n = 0$, can be solved recursively by calculating $Q(j; n)$ first for all n at a given j, and then for increasing j. In the end, we have $\mathcal{Z}_N(m) = Q(N; m)$ for the m exterior bases in $O(N^3)$ time.

To produce the minimum free energy structures at fixed m, we use an equivalent recursive scheme, but replacing the summations by maxima to obtain first the minimum free energy (Zuker and Stiegler, 1981). Then, we determine the corresponding structure by going through the scheme in reverse and reconstructing at each step which of the terms was maximal.

Polymer model

The simplest polymer model for the exterior single strand (the open circles in Fig. 1) is the Gaussian chain (de Gennes, 1979). However, as shown below, the force-induced denaturation of RNA occurs at forces of order 10 pN, where the exterior single strand is strongly stretched and the Gaussian model breaks down. In this regime, an elastic freely jointed chain (EFJC) model [Self-avoidance in the exterior single strand may be neglected, again because of its highly stretched state] yields a good fit to experimental FECs (Montanari and Mézard, 2001; Smith et al., 1996).

The distance along the backbone between two adjacent nucleotides is the segment length of the chain. We denote it by l and assign an elastic energy $V(\mathbf{r}) = \kappa(r - l)^2/2$ per segment, where $\mathbf{r}$ represents the end-to-end vector of the segment. Instead of attempting (the very cumbersome) exact computation of the end-to-end vector distribution $W(\mathbf{R}; m)$ of the chain, we use an asymptotic expression that becomes exact in the limit of large m and is sufficiently accurate for our purposes even for small m. It can be derived along the line of a similar calculation for the case of the regular (i.e., nonelastic) freely jointed chain given in Flory (1967). The result is conveniently expressed in terms of the quantity

$$q(h) = \int \mathrm{d}^3 r\, e^{-\mathbf{h}\cdot\mathbf{r} - \mathbf{V}(\mathbf{r})/k_B T} \Big/ \int \mathrm{d}^3 r\, e^{-\mathrm{V}(\mathbf{r})/k_B T},$$

where k_B denotes the Boltzmann constant and $\mathbf{h}$ is a vector of length h with fixed (but arbitrary) orientation in space. The asymptotic expression is then

$$W(\mathbf{R}; m) \approx C \frac{h}{2\pi R} [q(h)]^m e^{-hR}, \qquad (2)$$

where C is a normalization constant and h is determined from $R = m\,(\partial/\partial h) \log q(h)$. We incorporate the effect of a Kuhn length $b > l$ by rescaling the end-to-end vector distribution through $l \to b$ and $m \to ml/b$.

Observables

Apart from the force at fixed extension, which is calculated from Eq. 1 by

$$f(R) = -k_B T \frac{\partial}{\partial R} \log Z_N(R) \qquad (3)$$

(see, e.g., Flory (1967)), we also calculate the mean number of stems, n_{stem}, along the exterior chain (for the structure depicted in Fig. 1 this would be $n_{stem} = 2$). This may be determined by introducing an extra free energy penalty, ϵ_{stem}, for each external stem into the calculation of $\mathcal{Z}_N(m)$ and then differentiating numerically with respect to ϵ_{stem}, i.e.,

$$n_{stem}(R) = -k_B T \frac{\partial}{\partial \epsilon_{stem}} \log Z_N(R) \Big|_{\epsilon_{stem}=0}.$$

Choice of parameters

We work at room temperature, $T = 20°C$, and use the DNA polymer parameters obtained by Montanari and Mézard (2001) by fitting to the experiment of Maier et al. (2000) also for RNA, because we are not aware of the corresponding experimental data. (We do not expect a large difference in the single-strand properties between DNA and RNA because of the high similarity between their chemical structures.) The values are $l = 0.7$ nm, $b = 1.9$ nm, and $(\kappa/k_B T)^{-1/2} = 0.1$ nm. We take the free energy parameters for RNA secondary structure as supplied with the Vienna package. The salt concentrations at which these free energy parameters were measured are $[Na^+] = 1$ M and $[Mg^{2+}] = 0$ M.

RESULTS

Fig. 2, *a* and *b* show the FECs (*solid lines*) for two RNA sequences with practically the same total length and composition, both computed as described in the last section using the same set of parameters. Strikingly, the first curve is almost completely smooth, with no significant features, while the second is extremely jagged, with large "jumps" in the force. This dissemblance is entirely due to the difference between the secondary structures into which the two se-

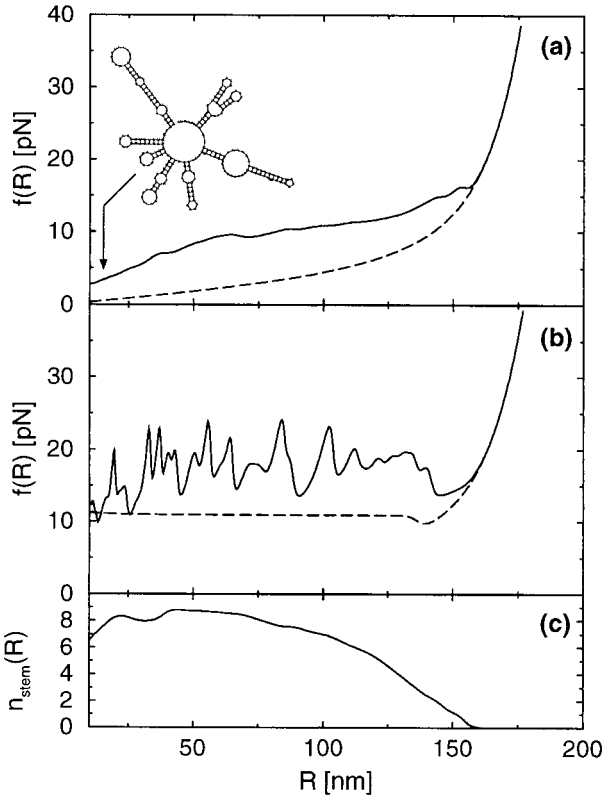

FIGURE 2 (*a*) Force-extension curve (FEC) for a group I intron (*solid line*, see text for details) and a homopolymeric RNA of the same length, $N = 251$ (*dashed line*). The depicted secondary structure is the minimum free energy structure at $R = 10$ nm. (*b*) FEC for a hairpin composed of randomly chosen basepairs (*solid line*) and a homogeneous hairpin of AU-basepairs (*dashed line*). In both cases the total sequence length is $N = 252$. (*c*) Mean number of exterior stems, $n_{stem}(R)$, for the group I intron.

quences fold. The sequence in Fig. 2 *a* originates from the group I intron of the methionine tRNA of *Scytonema hofmanii* with a sequence length of $N = 251$ (GenBank U10481). Its dominant secondary structure (according to our algorithm (see Note 2 at end of text)) at an extension of $R = 10$ nm is also depicted in Fig. 2 *a*. The sequence in Fig. 2 *b* was artificially generated by concatenating a randomly chosen sequence with its reverse complement, so that it folds into a single hairpin composed of random basepairs. Its FEC is very similar to the experimental force curve obtained upon unzipping double-stranded DNA by Essevaz-Roulet et al. (1997); the sawtooth-like oscillations correspond to a "molecular stick-slip process" (Bockelmann et al., 1997).

Why does the group I intron not display an abundance of features in the FEC like the hairpin does? Its secondary structure consists of many structural elements (e.g., stem-loop structures), the opening of which one might expect to produce clear signatures in the FEC. Indeed, in their theoretical study of force-induced denaturation of DNA/RNA, Thompson and Siggia (1995) concluded that the opening of individual basepairs in double-stranded DNA cannot readily be observed, but the opening of stem-loop structures in RNA should be.

One fairly obvious effect that could cause the smooth FEC is thermal superposition of alternative secondary structures. Because one may expect that typical RNA structures (such as the one depicted in Fig. 2 *a*) are less well-designed than a perfect hairpin, force-induced denaturation should make more alternative structures accessible in the former case than in the latter. In our analysis below we find that this effect is indeed non-negligible, but the largest loss of features originates from another, more subtle mechanism, which we call the "compensation effect," and which persists even when no alternative secondary structures are allowed. The compensation effect depends on the fact that when several structural elements are pulled at *in parallel*, the optimization process that determines the minimum free energy structure with a given number m of external open bases may *reclose* stretches of basepairs that had already been opened at a lower value $m' < m$.

In our approach (see Model and Methods above), the information on the secondary structure energetics for a given sequence is entirely contained in the function $\mathcal{Q}(m)$. With the help of the polymer model (contained in $W(\mathbf{R}; m)$) this information is translated into an FEC via Eq. 1. Our investigation therefore comprises two steps. First, we seek to understand what property of $\mathcal{Q}(m)$ determines the size of the fluctuations in the FEC, and second, how this property depends on the secondary structure.

The first question is addressed most readily for the special case of the random hairpin of Fig. 2 *b*. It is known that in the fixed-force ensemble, unzipping of a random hairpin may be mapped onto the problem of a particle in a tilted one-dimensional random potential (de Gennes, 1975; Lubensky and Nelson, 2000). The random potential is correlated and has the statistical properties of a one-dimensional random walk. In the fixed-distance ensemble, we may perform a very similar mapping (see Fig. 3). [For these mappings, alternative structures of the hairpin sequence are neglected, which is a good approximation due to the perfect design of the hairpin. Also, the nearest-neighbor correlations in the random potential caused by the stacking energies are not taken into account because they would not change the qualitative predictions of the model.] Here, the bias for the direction of movement of the particle is not caused by a tilt of the potential, but instead by a spring that is attached to the particle. The position of the other end of the spring is externally controlled, i.e., it is determined by R, the given end-to-end distance of the RNA molecule.

In the following, we review the relation between the parameters of the particle-in-a-random-potential problem, i.e., the spring constant γ and the variance of the random potential, and the parameters of the unzipping problem. This will also serve us to introduce our notation for the subsequent discussion. We may write the free energy $G(m) =$

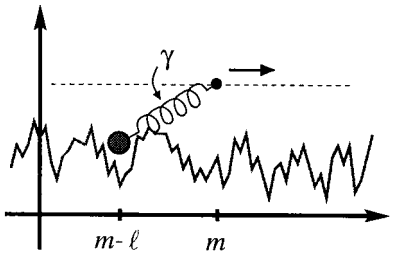

FIGURE 3 The problem of RNA pulling (in the fixed-distance ensemble) may be mapped onto the statistical mechanics problem of a particle with a spring attached to it moving in a one-dimensional disordered potential. The other end of the spring is externally controlled and slowly advanced into one direction.

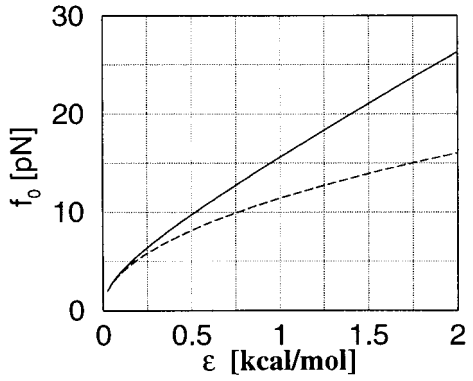

FIGURE 4 Threshold force f_0 for unzipping of a homogeneous hairpin as a function of the binding energy per base ϵ (*solid line*). The dashed line indicates the Gaussian approximation $f_0 = (6k_BT\epsilon/lb)^{1/2}$, which is obtained by using the end-to-end distance distribution $W(R; m)$ of a Gaussian chain. Note that the Gaussian approximation breaks down already at low forces, and the more detailed treatment according to Eq. 2 is necessary. The stacking energy for AU-pairs at $T = 20°C$ is $2\epsilon \approx 1.21$ kcal/mol corresponding to a threshold force $f_0 \approx 11$ pN, which agrees with the value observed in Fig. 2 *b* (*dashed line*).

$-k_BT \log \mathcal{Q}(m)$ of the random hairpin as $G(m) = -\sum_{i=1}^{N-m} \eta(i)$, where the $\eta(i)$ are random with mean $\langle\eta\rangle = \epsilon$ and variance $\langle\eta(i)\eta(j)\rangle - \langle\eta\rangle^2 = \delta_{ij}(\Delta\epsilon)^2$. Here, ϵ represents the mean binding energy per base, which depends on the GC-content of the hairpin, the temperature, and the salt concentrations; and $\Delta\epsilon$ measures the fluctuations of ϵ, both along a given hairpin and between different realizations of the random sequence. The difference between two free energies that are ℓ units apart, $\Delta G(\ell) = G(m) - G(m - \ell)$, then has the variance

$$\mathrm{var}(\Delta G(\ell)) = \ell(\Delta\epsilon)^2. \tag{4}$$

In the particle picture (see Fig. 3), $m - \ell$ corresponds to the position of the particle, and m to the position of the other end of the spring. For fixed m, the particle therefore sees the effective potential

$$\Delta G(\ell) + \frac{\gamma}{2}\ell^2, \tag{5}$$

i.e., Eq. 4 determines the variance of the random potential. The spring constant γ is determined by ϵ as follows. If $\Delta\epsilon$ were zero, the unzipping force would take a constant value f_0 (cf. the *dashed line* in Fig. 2 *b*, which shows the FEC of a homogeneous AU-hairpin). The dependence of f_0 on ϵ can be calculated analytically by evaluating the sum in Eq. 1 by the saddle point method (see also D. K. Lubensky and D. R. Nelson, manuscript in preparation). The result is shown in Fig. 4 (*solid line*). Now $\gamma = l^2\Gamma$, where Γ is the local spring constant of a nonbinding RNA of m bases at force $f_0(\epsilon)$. Because the spring constant of a homopolymer scales with the inverse of the number of segments, we write $\Gamma = \Gamma_0/m$, where Γ_0 depends only on f_0, but not on m. Graphically, $\Gamma_0(f_0)$ is the slope at $f = f_0$ of the dashed line in Fig. 2 *a* (FEC of a homopolymeric RNA), multiplied by 251 (the number of bases in that example). In this way $\Gamma_0(f_0)$ may also be determined from an experimental FEC.

When the fluctuations in the random potential are not too weak, the particle follows the other end of the spring in *discrete jumps*. The typical size of a jump, $\Delta\ell_{\mathrm{jump}}$, is given by the value of ℓ for which the two terms in Eq. 5 are of equal size, $\Delta\ell_{\mathrm{jump}} \simeq (2m\Delta\epsilon/l^2\Gamma_0)^{2/3}$. A typical jump then leads to a drop in the force by $\delta f \simeq \Gamma l \Delta\ell_{\mathrm{jump}}$, i.e.,

$$\delta f \simeq (4\Gamma_0\Delta\epsilon^2/R)^{1/3}. \tag{6}$$

This is valid as long as the thermal broadening of the particle position, $\Delta\ell_T \simeq (2m/l^2\Gamma_0\beta)^{1/2}$, is less than the typical jump size $\Delta\ell_{\mathrm{jump}}$. In the opposite case, the particle is *sliding* more or less smoothly, and $\delta f \propto \Delta\epsilon$.

Equation 6 furnishes an estimate for the size of the fluctuations in the FEC for the case of a random hairpin. However, because we used an arbitrary function $G(m)$ as input, the above argument may be made in general for any structure, as long as Eq. 4 holds sufficiently well. Alternatively, if for a particular structure the dependence of $\mathrm{var}(\Delta G(\ell))$ on ℓ is determined numerically, this could be used to replace Eq. 4, and Eq. 6 would have to be modified accordingly.

We now address the question of how the fluctuations in $G(m)$ depend on the secondary structure. An essential difference between unzipping of a hairpin and force-induced denaturation of a typical RNA structure is that in the latter case, several stems are being pulled on simultaneously (see Note 3 at end of text) for most of the extension interval (see Fig. 2 *c*, which shows the number of stems as a function of the extension for the group I intron studied above). To analyze the effect of multiple stems, we constructed artificial sequences that form a given number n of random hairpins in a row (i.e., the sequences are a concatenation of n random hairpin sequences, each of which is constructed as explained above). For each n in the range $1 \le n < 10$, we computed $G(m)$ and the FECs for 1000 different sequence

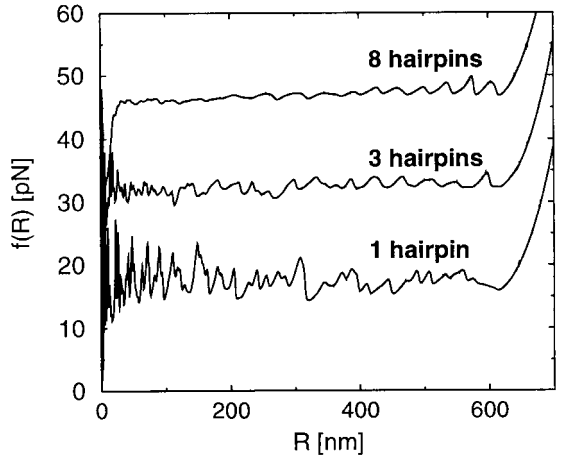

FIGURE 5 Force-extension curves for 1, 3, and 8 hairpins with random basepair composition in a row (sequence length $N = 1000$; the middle and upper curves are vertically shifted by 15 and 30 pN, respectively). Clearly, the fluctuations in the force curve decrease with increasing number of hairpins, except for the last third of the extension interval, where some of the hairpins of the 8-hairpin curve have already completely disappeared. In our analysis described in the main text only the first two-thirds of all FECs were used. The decrease of the force fluctuations with increasing extension is due to the entropic elasticity of the exterior single strand as described by the R-dependence in Eq. 6.

realizations, all with an approximate total length of $N = 1000$. As an example, Fig. 5 shows the FECs for three sequences, which fold into $n = 1$, 3, and 8 hairpins, respectively. Clearly, the fluctuations in the force curve decrease with increasing n. We obtained $\mathrm{var}(\Delta G(\ell))$ as an average over the 1000 realizations and a small interval of m. Some of the resulting curves are shown in Fig. 6. Although the dependence of $\mathrm{var}(\Delta G(\ell))$ on ℓ is not completely linear, the deviation from linearity over the small range of ℓ-values relevant here (typically, $0 \le \ell \le 12$) is not very large. For the sake of simplicity, we chose to interpret the data with the theory for a linear $\mathrm{var}(\Delta G(\ell))$ developed above. To this end,

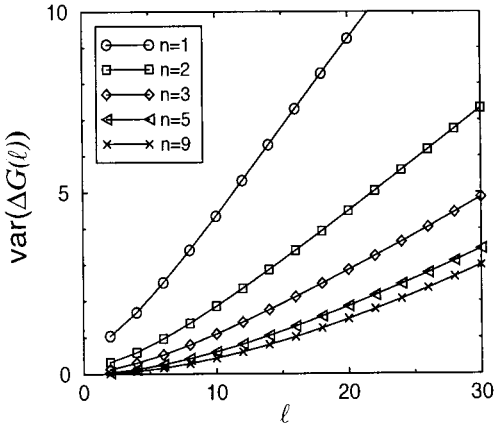

FIGURE 6 The variance of $\Delta G(\ell)$ for different numbers of hairpins.

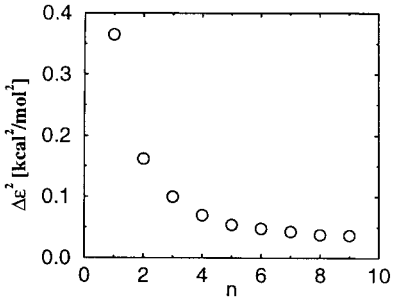

FIGURE 7 Dependence of $\Delta\epsilon^2$ on the number of hairpins (*circles*).

we define an effective $\Delta\epsilon$ for each n from the slope of $\mathrm{var}(\Delta G(\ell))$ at $\ell = 4$.

Fig. 7 shows that $\Delta\epsilon^2$ decreases monotonically with the number of stems that are being pulled on simultaneously. This decrease is almost entirely due to the compensation effect, which we may intuitively understand as follows. When a single hairpin is being unzipped, the stick-slip process described in Essevaz-Roulet et al. (1997) is topologically inevitable because the basepairs have to be opened in the order in which they occur. A strongly bound region that is followed by a weakly bound one always leads to a rise and subsequent drop of the FEC. However, with several hairpins, only the total number of exterior open bases is externally constrained, while the individual hairpins may freely open and reclose basepairs (for equilibrium FECs there is no kinetic constraint). Therefore, if in a particular hairpin a strongly bound region is followed by a weakly bound one, both regions can open together and another hairpin can reclose a few basepairs to compensate for the released single-strand. Obviously, with a growing number of hairpins, this mechanism will be increasingly effective. Clearly, in the fixed-force ensemble the compensation effect is equivalent to an average over the FECs of the individual hairpins. Moreover, with a large number of hairpins, the fixed-force and the fixed-distance ensembles become equivalent (D. K. Lubensky and D. R. Nelson, manuscript in preparation).

To quantitatively analyze the force fluctuations, we calculated the FECs for all of the 1000 sequence realizations of the n parallel hairpins, and defined $\Delta f(R)$ as the standard deviation of the force at extension R (the so-defined Δf is smaller than the typical size of a force jump, δf, but should have the same scaling behavior). Fig. 8 shows a plot of the force fluctuations against the free energy fluctuations, where the horizontal axis, $\Delta\epsilon(2\beta^3 R/l^3\Gamma_0)^{1/4} = (\Delta\ell_{\mathrm{jump}}/\Delta\ell_{\mathrm{T}})^{3/2}$, is scaled such that it separates the jumping regime from the sliding regime at a crossover value of one. The vertical axis is scaled such that the data should collapse onto a straight line in the jumping regime according to Eq. 6. To guide the eye, Fig. 8 also displays artificial data (*crosses*) for which $G(m)$ was generated by drawing random numbers

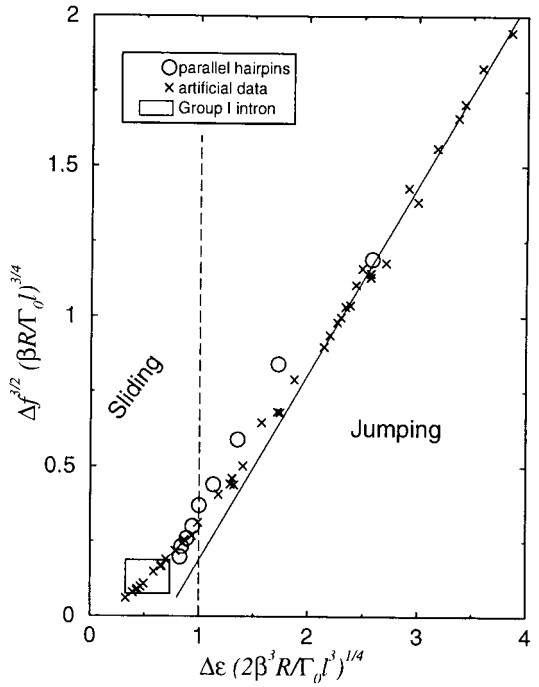

FIGURE 8 Scaling plot of the force fluctuations against the free energy fluctuations. The dashed vertical line marks the crossover region between the jumping regime and the sliding regime. The solid line is a linear fit to the data with abscissae larger than two. It confirms the scaling behavior expected for the jumping regime. See text for details.

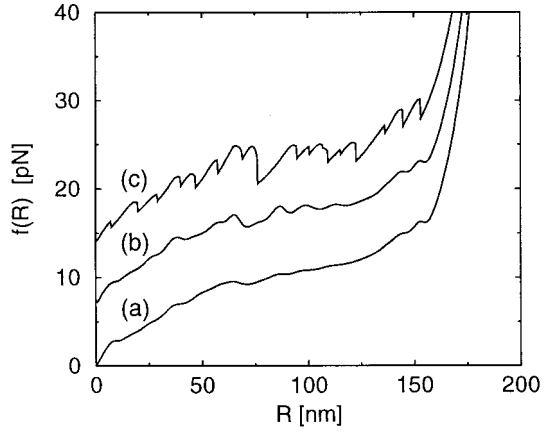

FIGURE 9 Force-extension curves for the group I intron under different conditions. Curve (*a*) is a copy of the full thermodynamic curve of Fig. 2 *a*. Curve (*b*) (vertically shifted by 7pN) was calculated by taking only the minimum free energy structures along the unfolding pathway into account, i.e., the thermal smoothening due to suboptimal structures is suppressed. For curve (*c*) (vertically shifted by 14 pN), the rebinding of basepairs that had already opened at smaller extension has also been suppressed to simulate nonequilibrium pulling.

$\eta(i)$ and taking $G(m) = -\sum_{i=1}^{N-m} \eta(i)$ (the different points are for different values for the mean and variance of $\eta(i)$). The circles mark the data points for the parallel hairpins, and the rectangular symbol in the lower left indicates in what region the group I intron is situated. [The rectangular area marks the range of points that we obtained by determining δf, $\Delta\epsilon$, and Γ_0 by averaging over different extension intervals, all within the range 50–110 nm, which is a region where the mean force is relatively constant (this is required to separate fluctuations in the force from a gradual change in the mean value).]

For the artificial data (*crosses*) the above scaling arguments should rigorously apply. Indeed, the artificial data fall onto a straight line in the jumping regime (the solid line represents a linear fit to the points with abscissae larger than two), and in the sliding regime Δf is proportional to $\Delta\epsilon$ (not shown). For the real data, Fig. 8 shows that passing from a single hairpin through structures with several parallel perfect hairpins to a typical natural RNA may be viewed as passing from the jumping regime to the sliding regime for a particle in a (correlated) random potential. At the same time, the FECs change from jagged to smooth.

As mentioned above, thermal superposition of alternative secondary structures also contributes to the smoothing of the FECs: as the structural elements in each suboptimal structure open at different values of m, the thermal average over all these structures smoothes $G(m)$. To assess the importance of this effect, we suppressed it by taking only the minimum free energy secondary structures into account instead of calculating the full partition function $\mathbb{Z}(m)$. For the group I intron, the FEC without the contribution of suboptimal structures is shown in Fig. 9 *b*. Compared to the full thermodynamic curve (shown in Fig. 9 *a*) some structure is gained, but not nearly as much as in the FEC for the random hairpin of the same length, Fig. 2 *b*. This indicates that the compensation effect is the dominant source for the smoothing of the FEC.

DISCUSSION

In the preceding section we found that the equilibrium FECs for typical RNA molecules (such as the group I intron that served us as an example) are quite smooth and do not reveal any features that can be associated with the opening of structural elements. The compensation effect is the primary cause for this result, and we expect it to be responsible, in part, also for the experimental observation of extremely smooth FECs for single-stranded DNA by Maier et al. (2000). Nevertheless, the measurement of equilibrium FECs for RNA or single-stranded DNA might still be useful, e.g., for an experimental determination of the RNA/DNA free energy parameters. Usually, these are extracted from melting curves of oligomers (Freier et al., 1986), which requires variation of the temperature away from the temperature of interest up to the melting point of the oligomers, where the free energy and its temperature derivative are determined. The free energy parameters at the temperature of interest are

then obtained by extrapolation, which introduces an error inherent to the method. For pulling experiments, the temperature can be kept constant at the value of interest, which is an obvious advantage. Here, the limiting factor is only the precision of the force measurement. The quantitative relationship between stacking energy and threshold force expressed by Fig. 4 furnishes the necessary link between force and energy. Measuring FECs for periodic hairpins composed of different building blocks would lead to curves like the dashed line in Fig. 2 *b* with different values for the threshold force. From these values the stacking energies could then be determined, which might lead to more accurate parameters at the desired temperature and salt concentrations.

There are (at least) two options to obtain FECs with more features, which in turn might allow one to obtain information on RNA secondary structure from pulling experiments. One could either record *nonequilibrium* FECs or analyze the *fluctuations* around the equilibrium curve. For our theoretical investigation, the latter option is not available as long as we work in the fixed-distance ensemble because the force fluctuations around the thermodynamic average diverge in that ensemble. We will pursue this option in a separate publication by working in a mixed ensemble (U. Gerland, R. Bundschuh, and T. Hwa, manuscript in preparation). Here, we briefly consider nonequilibrium FECs, where the rate of external increase in the force/extension is higher than (some of) the rates associated with internal rearrangements in the secondary structure. In the case of long *proteins*, either naturally occurring as an array of globular domains (Rief et al., 1997) or synthesized protein arrays (Yang et al., 2000), mechanical stretching experiments resolved the unfolding of up to 20 individual domains. These experiments were performed under nonequilibrium conditions (Rief et al., 1998) with typical pulling speeds of 1 μm/s.

To estimate whether nonequilibrium conditions are attainable for RNA with reasonable pulling speeds, we need a rough idea of the timescales involved in secondary structure rearrangements of RNA. For this, we again assume that RNA and single-stranded DNA behave similarly, so that we may draw on an experiment by Bonnet, Krichevsky, and Libchaber (Bonnet et al., 1998) measuring the opening and closing rates of DNA stem-loops using fluorescence correlation spectroscopy. From their results, we extract 10 μs as an estimate for the closing time (at $T = 20$°C) of a stem-loop structure with three basepairs and a loop of four nucleotides, which may be considered as a minimal secondary structure element. We expect that the formation of the stem-loop takes place in a single step whose reaction pathway goes through a transition state where the basepairs of the stem have not yet formed, but the corresponding bases are already closely together (see Fig. 10). In the presence of an external force, the closing time must then be multiplied with an Arrhenius factor $e^{\Delta W/k_B T}$, where ΔW is the work that has to be exerted against the force to pull in the amount

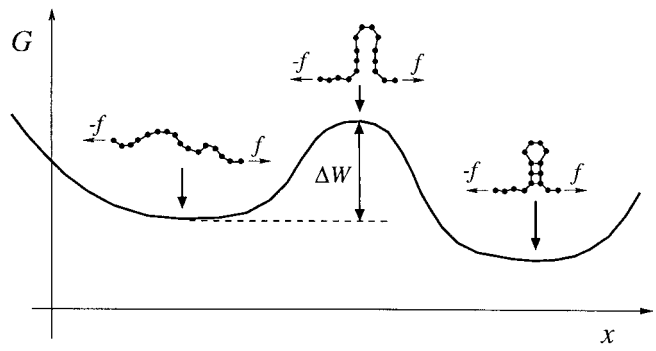

FIGURE 10 Sketch of the assumed pathway for the formation of a stem-loop structure in the presence of a stretching force f. A generalized reaction coordinate x is plotted along the horizontal axis and the free energy G along the vertical axis. The work that has to be exerted against the force to pull in the single strand needed for the formation of the stem-loop structure is denoted by ΔW. In principle, the entropy difference between the random coil state on the left and the transition state also contributes to the barrier height; however, we assume that at typical stretching forces it is negligible compared to ΔW.

of single strand needed for the formation of the stem-loop (Rief et al., 1998). With a typical force of 6 pN we obtain $\Delta W \approx 4$ kcal/mol, which results in a closing time on the order of 10 ms. This timescale has to be compared to the time it takes to stretch out the stem-loop. At a pulling speed on the order of 1 μm/s the two timescales are comparable, and hence both the formation of new secondary structure elements and the restoration of already opened ones are likely to be suppressed. [This estimate *does not apply* for the *rezipping* of partially opened, perfectly complementary long hairpins, which is faster than closing of a stem-loop. However, in real RNA structures, long stems are usually interrupted by internal or bulge loops, which we expect to reclose on similar timescales as the stem-loops.] Although it is beyond the scope of this paper, we want to note that in the presence of pseudoknots and/or tertiary interactions, the formation or re-formation of structural elements is expected to be slowed down even further, due to long search times for the interaction partners.

To obtain an impression of how many features a nonequilibrium FEC might show for the group I intron we change our equilibrium algorithm, such that the rebinding of bases is disabled once they have been unbound, and include only the contribution of the minimum free energy structures instead of all possible secondary structures. This is clearly a very crude approximation. In a proper treatment, only those kinetic processes whose energy barrier is higher than a certain threshold as determined by the pulling speed should be suppressed. Also, we did not account for the fact that the opening of basepairs occurs at higher forces in nonequilibrium as a consequence of Kramers theory (Evans and Ritchie, 1997). Nevertheless, the FEC shown in Fig. 9 *c* gives an idea of the large number of structural transitions that take place during force-induced denaturation (for com-

parison, the equilibrium FEC is shown again in Fig. 9 *a*). We therefore believe that nonequilibrium stretching experiments of RNA could lead to interesting and useful results.

We made most of the software tools developed for the present work available to the public by creating an "RNA pulling server" at http://bioinfo.ucsd.edu/RNA.

NOTES

1. In the "fixed-distance ensemble" only the average force is well-defined, whereas the fluctuations about the average diverge. This reflects the fact that it takes increasingly higher forces to compensate thermal fluctuations on shorter and shorter timescales to keep the extension *exactly* fixed. Therefore, if one is interested in the fluctuations (of either the force or the extension), the external spring should not be neglected, which would amount to working in a mixed ensemble between "fixed-distance" and "fixed-force."

2. The known native secondary structure of this sequence contains two helical regions forming a pseudoknot. Because pseudoknots are excluded from our approach (as explained above), we removed it from the structure computationally by replacing six basepairs in the less stable of the two helical regions (positions no. 79–84 and 157–162) by artificial bases that are excluded from basepairing. With this modification, the minimum free energy structure at zero force (as determined by the Vienna package) is almost identical with the secondary structure known from comparative sequence analysis (Gutell et al., 2001, manuscript in preparation, available at http://www.rna.icmb.utexas.edu/) outside of the pseudoknot region. Beyond the distance at which the pseudoknot is pulled apart, our modification of the sequence should not significantly affect the FEC. This expectation is supported by our numerical observation that the FECs for the unmodified sequence (ignoring the pseudoknot) and for our modified sequence become close to identical beyond a distance of $R \approx 70$ nm.

3. In principle, a situation where several stems are pulled on in parallel can also arise in the process of unzipping a single long hairpin, due to accidental palindromic regions in the single strand that has already been pulled out. However, these non-native interactions have to overcome the energetic advantage of the native single-hairpin interactions for the effect to become relevant. Hence, the palindrome needs to be extremely GC-rich. For a single hairpin consisting of random basepairs, we estimated that a non-negligible palindrome would typically occur only in sequences of at least several thousand bases in length, which is beyond the length of the sequences studied here.

We thank D. Bensimon, C. Bustamante, and J. D. Moroz for stimulating discussions.

U.G. is supported by the Hochschulsonderprogramm III of the DAAD. R.B. and T.H. acknowledge support by the National Science Foundation through Grant DMR-9971456, DBI-9970199, and the Beckmann foundation.

REFERENCES

Bockelmann, U., B. Essevaz-Roulet, and F. Heslot. 1997. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Phys. Rev. Lett.* 79:4489–4492; DNA strand separation studied by single molecule force measurements. *Phys. Rev. E.* 58:2386–2394.

Bonnet, G., O. Krichevsky, and A. Libchaber. 1998. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci. USA.* 95:8602–8606.

Chen, S.-J., and K. A. Dill. 2000. RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA.* 97:646–651.

de Gennes, P.-G. 1975. Brownian motion of a classical particle through potential barriers. Application to the helix-coil transitions of heteropolymers. *J. Stat. Phys.* 12:463–481.

de Gennes, P.-G. 1979. Scaling concepts in polymer physics. Cornell University Press, Ithaca, NY.

Essevaz-Roulet, B., U. Bockelmann, and F. Heslot. 1997. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA.* 94:11935–11940.

Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.

Flory, P. J. 1967. Statistical mechanics of chain molecules. Interscience Publishers, New York.

Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA.* 83: 9373–9377.

Hofacker, I. L., W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte f. Chemie* 125:167–188.

Isambert, H., and E. D. Siggia. 2000. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. USA.* 97:6515–6520.

Lubensky, D. K., and D. R. Nelson. 2000. Pulling pinned polymers and unzipping DNA. *Phys. Rev. Lett.* 85:1572–1575.

Maier, B., D. Bensimon, and V. Croquette. 2000. Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl. Acad. Sci. USA.* 97:12002–12007.

Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288:911–940.

McCaskill, J. S. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers.* 29: 1105–1119.

Mehta, A. D., M. Rief, J. A. Spudich, D. A. Smith, and R. M. Simmons. 1999. Single-molecule biomechanics with optical methods. *Science.* 283:1689–1695.

Montanari, A., and M. Mézard. 2001. Hairpin formation and elongation of biomolecules. *Phys. Rev. Lett.* 86:2178–2181.

Rief, M., H. Clausen-Schaumann, and H. E. Gaub. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 6:346–349.

Rief, M., J. M. Fernandez, and H. E. Gaub. 1998. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* 81:4764–4767.

Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science.* 276:1109–1112.

Smith, S. B., Y. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science.* 271:795–799.

Thirumalai, D., and S. A. Woodson. 2000. Maximizing RNA folding rates: a balancing act. *RNA.* 6:790–794.

Thompson, R. E., and E. D. Siggia. 1995. Physical limits on the mechanical measurement of the secondary structure of biomolecules. *Europhys. Lett.* 31:335–340.

Tinoco, I., Jr., and C. Bustamante. 1999. How RNA folds. *J. Mol. Biol.* 293:271–281.

Walter, A. E., D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker. 1994. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA.* 91:9218–9222.

Yang, G., C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante. 2000. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. USA.* 97:139–144.

Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acid. Res.* 9:133–148.