Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins

Michael T. Woodside*[†][§], William M. Behnke-Parks[†], Kevan Larizadeh[†], Kevin Travers[¶], Daniel Herschlag[¶], and Steven M. Block[†][‖]

*National Institute for Nanotechnology, National Research Council of Canada, Edmonton, AB, Canada T6G 2V4; and Departments of [†]Biological Sciences, [¶]Biochemistry, and [‖]Applied Physics, Stanford University, Stanford, CA 94305

Nucleic acid hairpins provide a powerful model system for probing the formation of secondary structure. We report a systematic study of the kinetics and thermodynamics of the folding transition for individual DNA hairpins of varying stem length, loop length, and stem GC content. Folding was induced mechanically in a high-resolution optical trap using a unique force clamp arrangement with fast response times. We measured 20 different hairpin sequences with quasi-random stem sequences that were 6–30 bp long, polythymidine loops that were 3–30 nt long, and stem GC content that ranged from 0% to 100%. For all hairpins studied, folding and unfolding were characterized by a single transition. From the force dependence of these rates, we determined the position and height of the energy barrier, finding that the transition state for duplex formation involves the formation of 1–2 bp next to the loop. By measuring unfolding energies spanning one order of magnitude, transition rates covering six orders of magnitude, and hairpin opening distances with subnanometer precision, our results define the essential features of the energy landscape for folding. We find quantitative agreement over the entire range of measurements with a hybrid landscape model that combines thermodynamic nearest-neighbor free energies and nanomechanical DNA stretching energies.

DNA hairpin | energy landscape | force clamp | optical tweezers | single molecule

Hairpins formed from self-complementary sequences supply a model system for studying folding and duplex formation, the most fundamental processes for generating structure in nucleic acids. Using hairpins, repeated measurements can be made on the same molecule, facilitating single-molecule studies. Furthermore, by simply altering the nucleotide sequence, physical properties such as folding energies, kinetic rates, and distances to transition states all can be changed systematically. Hairpins also play essential roles *in vivo*. DNA hairpins bind proteins to regulate transcription (1), and hairpin intermediates are involved in both replication and recombination (2, 3). RNA hairpins form tertiary contacts (4), bind to proteins (2), regulate transcription (5), and mediate RNA interference (6). Understanding the factors that influence hairpin folding should therefore not only elucidate principles of structure formation in nucleic acids but may also shed light on the biological roles played by these structures.

Extensive calorimetric and melting studies have been carried out to generate predictive rules for the thermodynamic stability of arbitrary nucleic acid duplexes (7). The kinetic properties of duplex formation, however, remain less well understood, particularly those related to the nature of the transition state. Temperature-jump studies of annealing in short duplexes have been interpreted in terms of the nucleation of a transition state consisting of ≈1–3 bp, followed by zippering of the remaining stem (8). This interpretation, however, rests on the assumption that the enthalpy of activation arises solely from the formation of contiguous base pairs in the transition state, for which there is a lack of direct evidence. Studies of thermally induced folding in short hairpins (8–19) investigated the effects of loop structure mediated through intraloop and loop–stem interactions but have not provided any more direct means of probing the transition state. Furthermore, all of these kinetic measurements involved short, unstable duplexes because of limits on the stability of molecules that could be explored without resorting to extreme conditions.

Single-molecule studies of folding induced by mechanical means offer an alternative approach with distinct and complementary advantages. The application of force constrains a potentially complex admixture of folding pathways to a single process characterized by a dominant reaction coordinate: the molecular extension. Nanomechanical experiments follow individual folding trajectories by measuring the end-to-end extension of single molecules with extraordinary precision (20). Furthermore, this extension can be directly related to the number of base pairs formed, simplifying the interpretation of data. Especially useful is the fact that the location of the transition state along the reaction coordinate can be derived directly from the force dependence of the kinetic rates (21). The application of force also facilitates the study of molecules with a wide range of energetic stabilities under similar environmental conditions, as demonstrated by the force-induced unfolding of short duplexes (22) and hairpins (20) and very long duplexes (23–25). Nanomechanical measurements have been used to follow folding transitions in single and multiple hairpins of DNA (26) and RNA (27–30), but no attempt has yet been made to investigate systematically the effects of sequence on the folding process.

We report here optical trapping measurements comparing the folding of a variety of hairpin constructs under identical buffer and temperature conditions. Single DNA hairpins were unzipped and refolded at constant force by using an ultrastable, high-precision (≈0.1 nm/$\sqrt{\text{Hz}}$ resolution) optical force clamp capable of recording transitions with rates ranging from ≈0.001 to 2,000 s^{-1} (20). The two ends of each hairpin were attached to dsDNA "handles" bound to polystyrene beads held in two independently controlled laser traps (Fig. 1*A*). By monitoring the extension of the molecule as the hairpin folded and unfolded under constant force, we measured the opening distance, the unfolding force and free energy, the force-dependent transition rates, and the location of the transition state. We investigated how each of these quantities varied for 20 different hairpin structures with quasi-random stem sequences as we altered the stem length, the loop length, and the GC content of the stem.
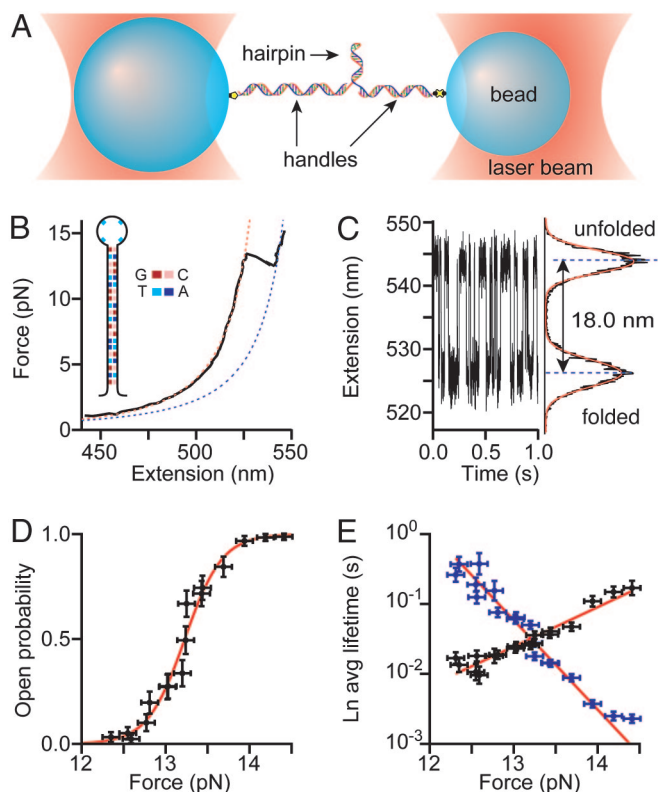

Fig. 1. Measurement of hairpin folding/unfolding. (*A*) Illustration of the experimental geometry: A hairpin is attached by means of dsDNA handles to beads held in two traps (not to scale). (*B*) FEC for hairpin 20R55/4T, displaying WLC behavior of handles at low *F* (dotted red line) followed by hairpin unfolding at ≈13 pN. WLC fit to the contour length increase after hairpin unfolding (dotted blue line) gives 17.5 ± 1 nm. (*C*) DNA extension vs. time at constant *F* reveals two-state behavior. Fit of extension histogram (black) to two Gaussian curves (red) gives an opening distance of 18.0 ± 0.5 nm. (*D*) Unfolded state probability (black) varies with *F* according to two-state Boltzmann relation (red). (*E*) Lifetimes of folded (blue) and unfolded (black) states vary exponentially with force.

Our results were then compared quantitatively with the predictions of a simple model of the folding energy landscape combining the nearest-neighbor free energies of duplex DNA with the elastic energy of unzipped DNA and the mechanical work performed by the optical trap.

Results and Discussion

Experimental Procedure. For each hairpin construct, we measured a force–extension curve (FEC), recording the force on the beads as they were separated at 10–15 nm/s. Fig. 1*B* shows a typical FEC for a hairpin consisting of a 20-bp stem with quasi-random (unpatterned) sequence, 55% GC content, and a thymidine tetraloop (denoted as hairpin 20R55/T4, where 20 indicates the stem length in base pairs, R represents the random sequence, 55 is the percentile stem GC content, and T4 indicates the loop sequence). At low forces, the FEC is well fit by a worm-like chain (WLC) model, as expected for a dsDNA molecule (31, 32). At $F \approx 13$ pN, this hairpin unfolds abruptly, producing a characteristic sawtooth pattern as the extension increases and the force drops (27). We fit the additional contour length released during unfolding to derive the opening distance, $\Delta x = 17.5 \pm 1$ nm. This distance supplies a rough estimate of 0.44 ± 0.02 nm per nt, taking into account the 2.0-nm width of duplex DNA, in good agreement with previous results (33).

Repeated folding transitions were measured at constant force

by using the passive force clamp to observe from dozens to hundreds of individual transitions. The hairpin remained in thermal equilibrium throughout these measurements. A representative segment of a record for hairpin 20R55/T4 under a 13.2-pN load displays the two-state behavior noted previously (19, 20, 27), with the hairpin switching rapidly between folded and unfolded states (Fig. 1*C*). At this force, the hairpin spent roughly equal time in each state: a histogram of the extension displays two peaks well fit by Gaussian curves separated by $\Delta x = 18.0 \pm 0.5$ nm, in agreement with the opening distance obtained from the FEC. To measure directly the folded and unfolded lifetimes, we partitioned each trace into two states, using a threshold set at the midpoint between the histogram peaks. Lifetimes were exponentially distributed, as expected for a two-state system (see Fig. 4, which is published as supporting information on the PNAS web site).

The presence of two mechanical states implies that the energy landscape of the hairpin is dominated by two potential wells separated by a single barrier. The application of force tilts the energy landscape (see Fig. 5, which is published as supporting information on the PNAS web site), changing the free-energy difference between folded and unfolded states (21). By measuring the force dependence of the populations of the two states, we determined both the unfolding force, $F_{1/2}$ (defined as the force at which the hairpin has a 50% probability of being unfolded), and the free energy released, ΔG. The probability of the unfolded state as a function of force (Fig. 1*D*) fits well to the Boltzmann relation $P_u(F) = \{1 + \exp[(F_{1/2} - F) \cdot \Delta x]\}^{-1}$ (27). We found $F_{1/2} = 13.2 \pm 0.7$ pN and $\Delta x = 18 \pm 2$ nm, in good agreement with the value for Δx obtained above. The product of the unfolding force and distance gave $\Delta G = 143 \pm 12$ kJ/mol. As discussed in ref. 21, ΔG equals the hairpin unfolding free energy at zero force plus the free energy for stretching the unfolded ssDNA from $F = 0$ to $F = F_{1/2}$.

The lifetimes of the folded and unfolded states also depend on the applied force, because force changes the height of the energy barrier (Fig. 5). Assuming that the position of this barrier is force-independent, the lifetime of the folded state, τ_f, is given by $\tau_f(F) = \tau_{f,0} \cdot \exp(F \Delta x_f^{\ddagger}/k_B T)$, where $\Delta x_f^{\ddagger}$ is the distance from the folded to the transition state and $\tau_{f,0}$ is the lifetime of the folded state at $F = 0$ (21). An analogous expression holds for the lifetime of the unfolded state. From the slopes of the logarithmic average lifetimes of the folded and unfolded states as functions of force (Fig. 1*E*), we found that the transition barrier was located 12 ± 2 nm from the folded state and 5.3 ± 0.6 nm from the unfolded state. From the intercept of the fit to folded-state lifetime, we derived the unfolding rate at $F = 0$ for this hairpin, $k_{u,0} = 10^{-15 \pm 2}$ s^{-1}, which, along with the unfolding free energy, supplies a measure of hairpin stability.

Energy Landscape Model. For each hairpin construct, we made a quantitative model of the folding landscape, adapting elements from previous models for dsDNA stretching (33) and unzipping (34) and RNA unfolding (27, 21, 35, 36). This model incorporates five distinct components: the energetic contributions arising from (*i*) base stacking and hydrogen bonding within the folded helix, (*ii*) stretching of the ssDNA liberated by the hairpin unfolding, and (*iii*) molecular motions in the optical trap; and the effects of elastic compliance associated with (*iv*) the ssDNA of the unfolded state and (*v*) the dsDNA handles attached to the hairpin.

The free energy required to break the base pairs in the hairpin stem sequentially was estimated from nearest-neighbor energy parameters (37) for our experimental conditions (200 mM monovalent salt and 23°C) by using MFOLD 3.1 (38). This calculation produced an energy landscape for $F = 0$ (Fig. 2*A*), to which we added the free energy for stretching the ssDNA generated as stem base pairs broke under load, estimated by using a WLC
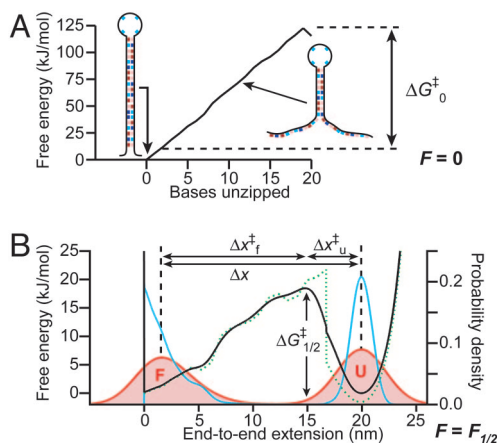
Fig. 2. Hairpin energy landscape model. (*A*) Computed free energy for sequential unzipping of base pairs in the stem for hairpin 20R55/T4 at $F = 0$ based on MFOLD. (*B*) Energy landscape at $F_{1/2}$ before (dotted green line) and after (solid black line) smoothing due to ssDNA elasticity. Barrier is close to the unfolded state; its height ($\Delta G^{\ddagger}_{1/2}$) is dominated by loop length. Hairpin extension probability density (blue line) is further smoothed by dsDNA handle elasticity to create bead position probability density (red line).

approximation (27, 32). A repulsive Morse potential (39) was added to model fluctuation-driven compression of the stem helix (i.e., negative extension), and the energy associated with stretching the fully unfolded state was included, using a WLC approximation. We then subtracted the mechanical work performed by the trap on the hairpin during unfolding, calculating the extension of the partially unzipped hairpin from previous FEC measurements of ssDNA (33) but taking into account the finite width of the stem helix as measured from NMR structures (40). Finally, we incorporated the effect of thermal fluctuations in the extension of the partially unzipped hairpin by smoothing the energy landscape by an amount proportional to the ssDNA stiffness, estimated by a WLC approximation. This model uses several simplifying assumptions: for example, we treat ssDNA as a noninteracting polymer, we neglect its enthalpic elasticity, and we assume that the applied force does not affect base-pair stacking free energies. Despite such simplifications, the model provides a practical, quantitative estimate of the folding energy landscape. See *Supporting Text*, which is published as supporting information on the PNAS web site, for more details.

An energy landscape computed for hairpin 20R55/4T is illustrated in Fig. 2*B*. The two deep wells correspond to the folded and unfolded states. Because of the unpatterned stem sequence, the steady rise in energy between these two states displays only relatively weak, sequence-specific fluctuations, followed by a steep fall upon unzipping of the last base pair and consequent release of the loop. The abrupt drop in energy near the unfolded state, attributable primarily to the release of entropy stored in the folded loop, is spread out over ≈ 2 nm because of the effect of ssDNA elasticity. Well depths associated with sequence-dependent fluctuations in the energy landscape are too small to trap the hairpin in intermediate minima, resulting in two-state behavior.

The behavior of hairpins was predicted quantitatively from this landscape model. First, $F_{1/2}$ was calculated as the force yielding equal probabilities for folded and unfolded states. Equal probabilities, rather than equal energy minima, is the appropriate criterion, because the potential wells associated with the two states have unequal widths (Fig. 2*B*). The compliance of the dsDNA handles attached to the hairpin elastically filters any hairpin motions transmitted to the trapped beads. This effect was incorporated by smoothing the probability density for the hairpin

extension, generating a modeled extension histogram (Fig. 2*B*) similar to the experimental result (Fig. 1*C*). From the energy landscape and extension histograms, we obtained predictions for Δx, ΔG, the barrier heights at F_0 and $F_{1/2}$, and the barrier location along the reaction coordinate. For all hairpins, model predictions were averaged over the same range of WLC parameters consistent with previous measurements: ssDNA persistence length of 1.0–1.5 nm (33, 41–44) and contour length per nucleotide of 0.58–0.67 nm (45).

Comparison of Measurements and Model. We measured the folding of 20 different hairpins with a wide variety of structures and stabilities, observing the effects of systematic changes in stem length, loop length, and stem GC content. Results are graphed in Fig. 3. Numerical values are listed in Table 1, which is published as supporting information on the PNAS web site, and representative records are shown in Fig. 6, which is published as supporting information on the PNAS web site. Fig. 3 *A*–*F* shows results for hairpins with identical tetraloops and stems ranging in length from 6 to 30 bp, sharing the same sequence at each end of the stem (hairpins 6R50/4T, 8R50/4T, 10R50/4T, 15R53/4T, 20R50/4T, 25R52/4T, and 30R50/4T; sequences are listed in Table 2, which is published as supporting information on the PNAS web site). Fig. 3 *G*–*L* shows results for hairpins with identical 15-bp stems but loop sizes ranging from 3 to 30 nt (hairpins 15R60/3T, 15R60/4T, 15R60/6T, 15R60/8T, 15R60/12T, 15R60/15T, 15R60/20T, and 15R60/30T). Polythymidine loops were chosen to minimize any loop structure arising from intraloop stacking interactions (15, 46). Fig. 3 *M*–*R* shows results for hairpins with thymidine tetraloops and 20-bp unpatterned stem sequences as the GC content was varied from 0 to 100% (hairpins 20R0/4T, 20R25/4T, 20R50/4T, 20R55/4T, 20R75/4T, and 20R100/4T).

The opening distance measured directly from extension histograms (Fig. 3 *A*, *G*, and *M*) increased roughly linearly with stem length and loop length but varied little with GC content, as expected. The small rise with increasing GC content is caused by the increase in $F_{1/2}$ for more stable hairpins, which produces a larger extension per nucleotide (33). The measured Δx values agree well with predictions of the landscape model, confirming previous measurements of the mechanical properties of ssDNA. The only discrepancies were observed for the hairpins with the longest loops, which we attributed to measurement drift associated with the long lifetimes of these hairpins. We note that the model predicts that the nominally "folded" state actually consists of an admixture of fully folded and partially unfolded hairpins (by 1–3 bp; Fig. 2*B*). This fraying, analogous to thermal breathing of the duplex (46), is consistent with statistical mechanical calculations of RNA hairpin unfolding under load (21). It represents a general, but previously neglected, feature of mechanically induced folding that must be taken into account when interpreting such measurements. Based on our findings, failure to account for fraying results in predicted values for the opening distance that are consistently at least 1–2 nm too large.

The unfolding force rose nonlinearly with stem length, decreased with increasing loop length, and increased linearly with stem GC content, in excellent agreement with model predictions (Fig. 3 *B*, *H*, and *N*). The increase in $F_{1/2}$ with stem length and GC content follows from the increasing energetic stability of the hairpin. The decrease with increasing loop length follows from the increasing energy stored in longer loops, which reduces the work performed by the optical trap to equalize the populations of folded and unfolded states. These data cover the complete range of forces that can be expected for unfolding DNA. Previous work on phage λ DNA found that kilobase-long duplexes with $\approx 50\%$ GC content unzip at an average force of ≈ 15 pN (23–25). Our results indicate that this asymptotic value is reached at a duplex length of only ≈ 25 bp. Extrapolating the
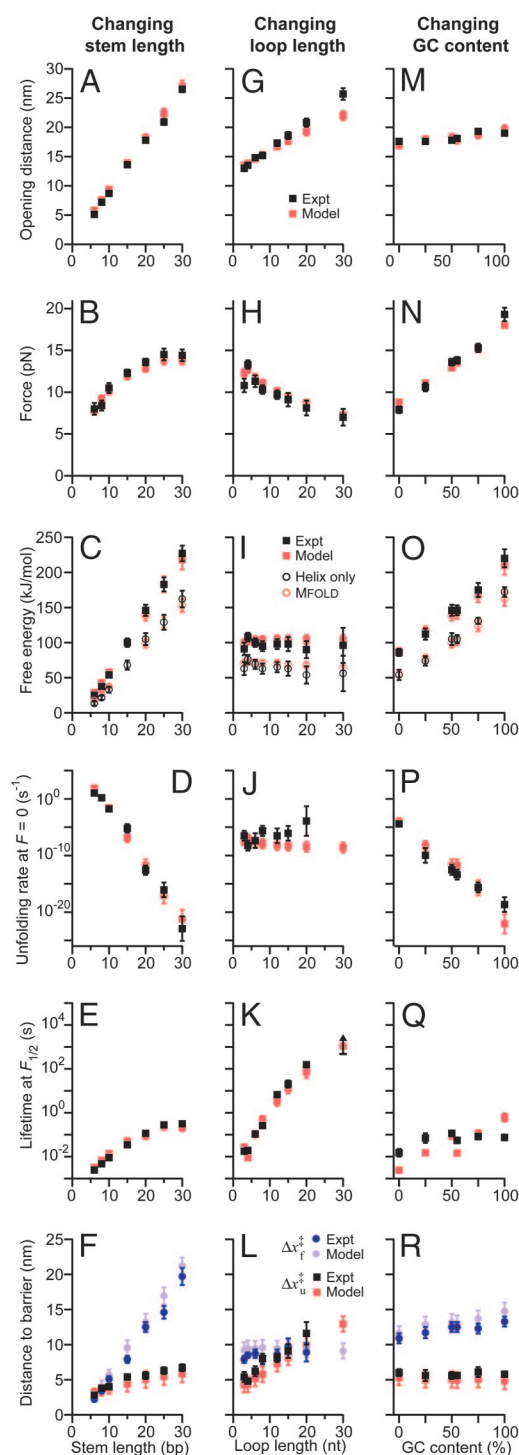
Fig. 3. Summary of results for hairpins with varied stem length (*A–F*), loop length (*G–L*), and stem GC content (*M–R*). Error bars represent sum of standard and systematic errors (experiment) or standard deviation of predictions over the full parameter set (model). (*A*, *G*, and *M*) Hairpin opening distance rises linearly with stem and loop length but is little changed with GC content (experiment, black; model, red). (*B*, *H*, and *N*) Unfolding force rises nonlinearly with stem length to a plateau at ≈25 bp, falls with loop length, and rises linearly with GC content. (*C*, *I*, and *O*) Unfolding free energy rises linearly with stem length and GC content but is little changed with loop length. Open circles show free energy of helix stacking only. (*D*, *J*, and *P*) Unfolding rate extrapolated to $F = 0$ decreases exponentially with increasing stem length and GC content but is little changed with loop length. (*E*, *K*, and *Q*) Folding lifetime at $F_{1/2}$ rises exponentially with loop length but depends much less strongly on stem length and GC content. (*F*, *L*, and *R*) Distance to transition state from

folded state (experiment, blue; model, purple) rises linearly with stem length, whereas distance from unfolded state (experiment, black; model, red) rises linearly with loop length, consistent with a transition state located 1–2 bp from loop.

results of Fig. 3*N*, we expect $F_{1/2}$ for long duplexes to vary linearly with GC content from ≈9 pN (at 0% GC content) to ≈21 pN (at 100%), consistent with previous work on poly(AT) and poly(GC) duplexes (26).

The unfolding free energy, ΔG, depended only on the properties of the stem, rising linearly with both stem length and GC content (as expected) but varying little with loop length (Fig. 3 *C*, *I*, and *O*). We found excellent agreement with the free energy predicted by our model over one order of magnitude, confirming that single-molecule unfolding measurements offer a robust, accurate means of determining equilibrium free energies. Subtracting the entropic contribution to ΔG from ssDNA stretching (representing approximately a quarter of the total free energy), we obtained the duplex stacking free energies (open circles in Fig. 3*C*, *I*, and *O*). These values agree well with the predictions of MFOLD after taking into account fraying of the stem as discussed above. Our data therefore provide independent confirmation of the validity of the thermodynamic nearest-neighbor energies for DNA, showing clearly that values measured by single-molecule force spectroscopy over a wide range of sequences agree very well with those derived from bulk assays.

The unloaded unfolding rate, $k_{u,0}$, decreased exponentially with increasing stem length and GC content but was insensitive to loop length (Fig. 3 *D*, *J*, and *P*). These trends can be understood in terms of the energy landscape for $F = 0$ (Fig. 2*A*). For all hairpins we studied, the model predicts that $\Delta G_0^{\ddagger}$, the barrier height at $F = 0$, corresponds to the energy for melting all but the last base pair in the stem (minus the energy of any frayed base pairs in the folded state under load). Hence, linear increases in stem free energy arising from changes in stem length or GC content produce exponential decreases in $k_{u,0}$. To compare model and experiment, we expressed $k_{u,0}$ as $k_{u,0} = \kappa_u \cdot \exp(-\Delta G_0^{\ddagger}/k_B T)$, where the prefactor, κ_u, depends on the viscous damping because of drag on the hairpin, the handles, and the beads, as well as on the shape of the potential landscape (47). Assuming that drag on the beads and handles predominates, κ_u should be similar for all constructs. We therefore fit the data to the modeled barrier heights by using a single parameter, $\kappa_u = 3 \times 10^6$ s^{-1}. The predictions and measurements agree remarkably well over ≈25 orders of magnitude in rate for all hairpins studied. The rates at $F = 0$ are also consistent with rates from fluorescence and temperature-jump studies of short DNA and RNA hairpins with 2- to 9-bp stems (14–19, 48, 49), from optical trapping studies of longer RNA hairpins with ≈20-bp stems (27, 28), and from temperature-jump studies of the helix-coil transition in duplexes (8). The present measurements, however, greatly expand the range of stability that has been explored and connect these diverse findings (see Fig. 7, which is published as supporting information on the PNAS web site).

The unfolding lifetime at $F_{1/2}$, $\tau_{1/2}$, varied roughly exponentially with both stem length and loop length but was ≈100 times more sensitive to loop length, rising approximately five orders of magnitude (Fig. 3 *E*, *K*, and *Q*; only a lower bound could be placed on $\tau_{1/2}$ for the most stable hairpin, 15R60/T30). In contrast, $\tau_{1/2}$ varied little with GC content. This behavior is precisely that predicted by the energy landscape model (Fig. 2*B*). For hairpins with random stem sequences, the barrier height is set mostly by the energy for unfolding or reforming the loop under load. Expressing the lifetime as $\tau_{1/2} = \tau_{1/2}^0 \cdot \exp(\Delta G_{1/2}^{\ddagger}/k_B T)$, where $\Delta G_{1/2}^{\ddagger}$ is the barrier height at $F_{1/2}$, we fit the data by using a single parameter, $\tau_{1/2}^0 = 1 \times 10^{-5}$ s (allowing different values for prefactors $\tau_{1/2}^0$ and κ_u because of qualitatively different

landscapes at $F_{1/2}$ and $F = 0$). The model agrees very well over six orders of magnitude in lifetime with the data for varying stem and loop lengths. It also captures the qualitative effects of changing stem GC content but predicts slightly larger changes in $\tau_{1/2}$ than were observed. This discrepancy is likely caused by simplifying assumptions in the WLC model used to calculate the loop energies to which $\Delta G_{1/2}^{\ddagger}$ is sensitive. Effects such as the enthalpic component of DNA elasticity, ignored here, would tend to reduce the predicted rise in $\Delta G_{1/2}^{\ddagger}$ with GC content.

Changing the closing base pair (i.e., the base pair nearest the loop) from AT to GC, which increased the stem GC content by 5% and only minimally raised hairpin stability as measured by ΔG or $k_{u,0}$, nevertheless had a pronounced effect on $\tau_{1/2}$, reducing it by a factor of approximately three (compare hairpins 20R50/T4 and 20R55/T4). The lifetime is reduced because the transition state is very close to the loop (see below and Fig. 2B); hence, stabilizing the closing base pair reduces the barrier between the unfolded and transition states. To achieve a comparable variation in $\tau_{1/2}$ by changing base pairs distributed randomly in the stem, one-quarter or more of the stem sequence had to be changed, emphasizing the dominant contribution of the loop to the height of the energy barrier. Single base-pair changes thus exert significant effects on the folding of hairpins with random stem sequences only when located adjacent to the loop.

Turning finally to the position of the energy barrier, we found that it was located roughly midway between folded and unfolded states for short hairpins but remained close to the unfolded state for long hairpins, never moving $> \approx 6$ nm away (Fig. 3F). As the loop length increased, the barrier maintained a constant distance from the folded state (Fig. 3L). As the GC content was varied, the barrier distance was constant from both the folded and unfolded states (Fig. 3R). These observations are fully consistent with the model, which predicts a barrier located in the stem a few base pairs from the loop. Its position changes little with GC content but moves away from the unfolded state when the loop is lengthened and away from the folded state when the stem is lengthened.

Taken at face value, the measured opening distances imply that the transition state involves the formation of $\approx 1.5–5$ bp adjacent to the loop, depending on stem length. In comparison, the "raw" energy landscapes (without elastic smoothing effects, illustrated by the dotted line in Fig. 2B) suggest that approximately 1–2 bp would be formed in the transition state, placing it closer to the unfolded state than actually measured. This apparent discrepancy arises from the neglect of the elastic compliance of the ssDNA generated during the unfolding. The ssDNA elasticity smoothes the otherwise abrupt drop in energy upon release of the loop, in effect pushing the energy barrier further from the unfolded state. When elasticity is taken properly into account, our experimental results imply a transition state consisting of 1–2 bp adjacent to the loop. Previous measurements of the activation enthalpies for short duplexes and hairpins were consistent with the formation of 1–3 contiguous, canonical bp, but such studies could not distinguish this simple model of the transition state from alternatives that would return identical enthalpy values through combinations of hydrogen bonding, base stacking, and solvation. Our measurements agree with the 1–3 bp previously postulated (8, 15, 22, 44, 50–52) and provide the most direct evidence thus far for the formation of contiguous base pairs in the transition state for duplex association and dissolution. We note that the absence of tertiary structure has been proposed to generate a "compliant" transition, where the transition state is located midway between folded and unfolded states (27, 53). Our results show clearly that this situation does not obtain in general. These earlier conclusions regarding the location of the transition state therefore appear to relate only to the specific sequences studied, which happened to generate barriers closer to the middle of the stem (35).

By measuring the mechanically induced folding of individual DNA hairpins with a variety of sequences at high resolution and under constant environmental conditions, we have systematically studied the sequence-dependent folding landscape for nucleic acid duplexes. The power of these techniques, the simplicity of the folding landscape under force, and the wealth of available biochemical and physical information have allowed us to describe quantitatively the folding of hairpins by a straightforward model for the energy landscape that incorporates key features of the thermodynamic and mechanical properties of DNA. Our results provide strong evidence that the transition state for duplex formation involves ≈ 2 consecutive base pairs. The agreement of the model with the data is especially remarkable in light of the physical details neglected, such as solvent effects and viscous drag on the hairpin (52), base stacking and other self-interactions in ssDNA, and enthalpic elasticity. Further precision measurements of hairpin folding should enable more stringent tests of the energy landscape formalism through experiments designed to probe the limits of applicability of the physical assumptions and studies of more irregular landscapes formed by hairpins containing patterned stem sequences and structured loops.

Methods

DNA Hairpin Constructs. Each construct consisted of a single short DNA hairpin attached at its 5′ and 3′ ends to long dsDNA handles used for pulling. Constructs were based on PAGE-purified DNA oligomers (Integrated DNA Technologies, Coralville, IA) containing the desired hairpin sequence (Table 2). Abasic sites (deoxyribose spacers) were incorporated in the oligomers at each end of the hairpin sequence to disrupt potential base-stacking interactions between the hairpin stem and the DNA handles. A 23-nt PCR primer was placed at the 3′ end of the oligomers, and a nonpalindromic ligation sequence was placed at the 5′ end. Autosticky PCR (54) using the primer region of the oligomer generated a 621-bp segment of dsDNA with the hairpin as a 5′ overhang at one end and a digoxigenin label at the other end, adapting a previous approach (22). The PCR product was ligated at 24°C with T4 DNA ligase (New England Biolabs) to a biotin-labeled, 1,036-bp fragment of dsDNA with a 5′ overhang complementary to the ligation sequence at the 5′ end of the hairpin oligomer. The final ligation product was gel-purified.

Hairpin constructs were incubated for 1 h at ≈ 22°C with 600-nm diameter polystyrene beads (Bangs Laboratories, Carmel, IN) coated with avidin-DN (Vector Laboratories) and 730-nm diameter beads coated with polyclonal anti-digoxigenin (Roche Diagnostics). Incubations were performed with all ingredients at ≈ 50 pM in 6 mM Tris/40 mM phosphate buffer, pH 7.5/0.05% Tween 20, producing "dumbbells" consisting of the DNA construct bound to one bead of each type at either end (55). Dumbbells were diluted 100-fold in assay buffer consisting of 50 mM Mops, pH 7.5/200 mM KCl/2% oxygen scavenger solution (250 mg/ml glucose/37 mg/ml glucose oxidase/1.7 mg/ml catalase).

Optical Trap. Measurements were performed at 23 ± 0.5°C with an optical trapping apparatus as described (20). Two traps produced by orthogonally polarized infrared laser beams were used to exert force on beads at either end of a dumbbell (Fig. 1A). The positions and intensities of the traps were controlled by acousto-optic deflectors. Bead displacements in each trap were measured independently with two orthogonally polarized detector beams from a HeNe laser. Light scattered by the beads was separated by polarization and collected by position-sensitive detectors (Pacific Silicon Sensor, Westlake Village, CA), whose

signals were amplified by home-built normalizing differential amplifiers, filtered, and digitized. Trap stiffness was determined by standard techniques, and bead positions in each trap were calibrated for each dumbbell measured (56). For steady-state measurements, hairpin extension was recorded for 5–4,500 s (depending on the folding rate), sampled at 0.2–20 kHz, filtered online with an eight-pole Bessel filter at 0.1–10 kHz, and filtered offline by a median filter with a 1- to 10,000-ms window.

Repeated folding transitions at constant force were measured by using a passive force clamp: one of the traps was made weaker than the other, and the bead held in it was pulled to the edge of the trap, where the local stiffness became zero, and hence the force was nearly constant for small displacements (20). In this arrangement, changes in the end-to-end extension of the hairpin construct were measured in the weak trap, without any need for elastic compliance corrections. Force was measured in the strong trap, where the bead remains stationary, based on a calibrated stiffness of 0.3–0.4 pN/nm. Force was set by modulating the light intensity in the weak trap. The passive force clamp used here aids the study of hairpin folding by removing artifactual force changes caused by finite feedback response times and by allowing the measurement of rapid transitions with lifetimes down to 0.5 ms.

Hairpins with very short unfolding distances or low unzipping forces (hairpins 6R50/4T, 8R50/4T, 10R50/4T, 15R50/4T, 15R60/3T, 15R60/4T, and 20R0/4T) were also measured in an unclamped configuration, where both beads were held in the harmonic regions of their respective traps (stiffness of each trap was 0.3–0.4 pN). This method allowed folding transitions to be better resolved by increasing the total system stiffness and hence reducing the Brownian noise. Comparisons of force-clamped and unclamped measurements on the same hairpins showed no appreciable difference in behavior (data not shown).

Dumbbells with more than one DNA molecule were detected by measuring the FEC at forces below the unzipping force and determining the persistence length, L_p, from a fit to the WLC model. Multiple tethers, characterized by low L_p values, were discarded.

1. Glucksmann-Kuis, M. A., Dai, X., Markiewicz, P. & Rothman-Denes, L. B. (1996) *Cell* **84,** 147–154.
2. Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 379–404.
3. Gellert, M. (2002) *Annu. Rev. Biochem.* **71,** 101–132.
4. Brion, P. & Westhof, E. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26,** 113–137.
5. Wang, D., Meier, T. I., Chan, C. L., Feng, G., Lee, D. L. & Landick, R. (1995) *Cell* **81,** 341–350.
6. Yu, Y. J., DeRuiter, S. L. & Turner, D. L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6047–6052.
7. Serra, M. J. & Turner, D. H. (1995) *Methods Enzymol.* **259,** 242–261.
8. Pörschke, D. (1977)) *Mol. Biol. Biochem. Biophys.* **24,** 191–218.
9. Senior, M. M., Jones, R. A. & Breslauer, K. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6242–6246.
10. Antao, V. P. & Tinoco, I., Jr. (1992) *Nucleic Acids Res.* **20,** 819–824.
11. Vallone, P. M., Paner, T. M., Hilario, J., Lane, M. J., Faldasz, B. D. & Benight, A. S. (1999) *Biopolymers* **50,** 425–442.
12. Proctor, D. J., Ma, H., Kierzek, E., Kierzek, R., Gruebele, M. & Bevilacqua, P. C. (2004) *Biochemistry* **43,** 14004–14014.
13. Baxter, S. M., Greizerstein, M. B., Kushlan, D. M. & Ashley, G. W. (1993) *Biochemistry* **32,** 8702–8711.
14. Hilbers, C. W., Haasnoot, C. A., de Bruin, S. H., Joordens, J. J., van der Marel G. A. & van Boom, J. H. (1985) *Biochimie* **67,** 685–695.
15. Ansari, A., Kuznetsov, S. & Shen, Y. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 7771–7776.
16. Bonnet, G., Krichevsky, O. & Libchaber, A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8602–8606.
17. Goddard, N., Bonnet, G., Krichevsky, O. & Libchaber, A. (2000) *Phys. Rev. Lett.* **85,** 2400–2403.
18. Wallace, M. I., Ying, L., Balasubramanian, S. & Klenerman, D. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5584–5589.
19. Grunwell, J. R., Glass, J. L., Lacoste, T. D., Deniz, A. A., Chemla, D. S. & Schultz, P. G. (2001) *J. Am. Chem. Soc.* **123,** 4295–4303.
20. Greenleaf, W. J., Woodside, M. T., Abbondanzieri, E. A. & Block, S. M. (2005) *Phys. Rev. Lett.* **95,** 208102.
21. Tinoco, I., Jr., & Bustamante, C. (2002) *Biophys. Chem.* **101–102,** 513–533.
22. Lang, M. J., Fordyce, P. M., Engh, A. M., Neuman, K. C. & Block, S. M. (2004) *Nat. Methods* **1,** 133–139.
23. Essevaz-Roulet, B., Bockelmann, U. & Heslot, F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 11935–11940.
24. Bockelmann, U., Thomen, P., Essevaz-Roulet, B., Viasnoff, V. & Heslot, F. (2002) *Biophys. J.* **82,** 1537–1553.
25. Danilowicz, C., Coljee, V. W., Bouzigues, C., Lubensky, D. K., Nelson, D. R. & Prentiss, M. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 1694–1699.
26. Rief, M., Clausen-Schaumann, H. & Gaub, H. E. (1999) *Nat. Struct. Biol.* **6,** 346–349.
27. Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I., Jr., & Bustamante, C. (2001) *Science* **292,** 733–737.
28. Onoa, B., Dumont, S., Liphardt, J., Smith, S. B., Tinoco, I., Jr., & Bustamante, C. (2003) *Science* **299,** 1892–1895.
29. Harlepp, S., Marchal, T., Robert, J., Léger, J.-F., Xayaphoummine, A., Isambert, H. & Chatenay, D. (2003) *Eur. Phys. J. E Soft Matter* **12,** 605–615.
30. Li, P. T. X., Collin, D., Smith, S. B., Bustamante, C. & Tinoco, I., Jr. (2006) *Biophys. J.* **90,** 250–260.
31. Smith, S. B., Cui, Y. & Bustamante, C. (1996) *Science* **271,** 795–799.
32. Marko, J. F. & Siggia, E. D. (1995) *Macromolecules* **28,** 8759–8770.
33. Dessinges, M.-N., Maier, B., Zhang, Y., Peliti, M., Bensimon, D. & Croquette, V. (2002) *Phys. Rev. Lett.* **84,** 248102.
34. Lubensky, D. K. & Nelson, D. R. (2002) *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **65,** 031917.
35. Cocco, S., Marko, J. F. & Monasson, R. (2003) *Eur. Phys. J. E Soft Matter* **10,** 153–161.
36. Manosas, M. & Ritort, F. (2005) *Biophys. J.* **88,** 3224–3242.
37. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1460–1465.
38. Zuker, M. (2003) *Nucleic Acids Res.* **31,** 3406–3415.
39. Peyrard, M. & Bishop, A. R. (1989) *Phys. Rev. Lett.* **62,** 2755–2758.
40. van Dongen, M. J., Mooren, M. M., Willems, E. F., van der Marel, G. D., van Boom, J. H., Wijmenga, S. S. & Hilbers, C. W. (1997) *Nucleic Acids Res.* **25,** 1537–1547.
41. Tinland, B., Pluen, A., Sturm, J. & Weill, G. (1997) *Macromolecules* **30,** 5763–5765.
42. Rivetti, C., Walker, C. & Bustamante, C. (1998) *J. Mol. Biol.* **280,** 41–59.
43. Wuite, G. J. L., Smith, S. B., Yong, M., Keller, D. & Bustamante, C. (2000) *Nature* **404,** 103–106.
44. Kuznetsov, S. V., Shen, Y., Benight, A. S. & Ansari, A. (2001) *Biophys. J.* **81,** 2864–2875.
45. Murphy, M. C., Rasnick, I., Cheng, W., Lohman, T. M. & Ha, T. (2004) *Biophys. J.* **86,** 2530–2537.
46. Saenger, W. (1984) *Principles of Nucleic Acid Structure* (Springer, New York).
47. Hänggi, P., Talkner, P. & Borkovec, M. (1990) *Rev. Mod. Phys.* **62,** 251–341.
48. Menger, M., Eckstein, F. & Pörschke, D. (2000) *Biochemistry* **39,** 4500–4507.
49. Jung, J. & van Orden, A. (2005) *J. Phys. Chem. B* **109,** 3648–3657.
50. Zhang, W. & Chen, S.-J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1931–1936.
51. Sorin, E. J., Rhee, Y. M., Nakatani, B. J. & Pande, V. S. (2003) *Biophys. J.* **85,** 790–803.
52. Ansari, A. & Kuznetsov, S. (2005) *J. Phys. Chem. B* **109,** 12982–12989.
53. Tinoco, I., Jr., Collin, D. & Li, P. T. X. (2004) *Biochem. Soc. Trans.* **32,** 757–760.
54. Gal, J., Schnell, R., Szekeres, S. & Kalman, M. (1999) *Mol. Gen. Genet.* **260,** 569–573.
55. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. (2003) *Nature* **426,** 684–687.
56. Neuman, K. C. & Block, S. M. (2004) *Rev. Sci. Instr.* **75,** 2787–2809.