Supporting Information for: Energy Landscapes and Hybridization Pathways for DNA Hexamer Duplexes

Shiyan Xiao,[†] Daniel J. Sharpe,[†] Debayan Chakraborty,[‡] and David J. Wales[*,†]

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom., and Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, USA

E-mail: dw34@cam.ac.uk

[*]To whom correspondence should be addressed
[†]University of Cambridge
[‡]The University of Texas at Austin

SUPPORTING SECTION 1: METHODS

Exploration of the energy landscapes.

Discrete path sampling[1,2] was used to construct and analyze the kinetic transition networks[3–5] for hybridization of DNA duplexes. A discrete path is an ordered sequence of pairs of minima, and the transition states connecting them, between two states of interest. A kinetic transition network is a coarse-grained representation of an energy landscape that is built up using geometry optimization methods to locate the local minima and intervening transition states of a potential energy surface. First, low energy minima, including the global minimum, are determined by basin-hopping global optimization[6–9] using the GMIN program.[10] Initial discrete paths between selected pairs of endpoints are determined using the doubly-nudged[11] elastic band[12,13] (DNEB) method to find transition state candidates, and then hybrid eigenvector-following[12,14,15] to converge high-energy images of the band to transition states. The two minima connected by each transition state are determined by a modified version of the limited-memory BFGS local minimization algorithm.[16] These calculations were carried out using the OPTIM program[17] interfaced with AMBER9.[18] The root mean square gradient convergence criterion for stationary point optimization was set to 10^{-6} kcal mol^{-1}Å^{-1}.

The initial discrete path between a pair of distant endpoints is unlikely to be kinetically relevant, and frequently has high energy barriers, and unnecessary detours. Hence, initial discrete paths between endpoints of interest are used as the starting point to grow the stationary point database and determine kinetically relevant pathways. Here, the SHORTCUT[19,20] and UNTRAP[20] schemes within the PATHSAMPLE driver program[21] for OPTIM were used to further sample the potential energy landscapes. The SHORTCUT scheme selects pairs of minima along the current fastest path, and aims to determine refined discrete paths with fewer steps and lower energy barriers. The UNTRAP scheme is used to eliminate artificial kinetic traps in the kinetic transition networks, which arise due to incomplete

sampling. Connections are attempted between pairs of low energy minima for which the energy difference is small, but which are separated by high energy barriers. Sampling to locate alternative pathways and explore new regions of the energy landscape is continued until observable quantities, such as phenomenological rate constants, have converged.

The starting coordinates of the selected sequences were for the B-form and were generated using the program *nab* from AMBER.[18] The DNA duplexes were modelled using the all-atom AMBER99/bsc0 force field,[22] employing the χOL4 torsional corrections.[23] Water was modelled by a generalized Born implicit solvent model[24,25] and the salt concentration was maintained at 1.0 M using the Debye-Hückel approximation.[26]

Our focus in the present work is to discover details of mechanisms that may correspond to long timescales in the range of tens of microseconds to seconds and even longer.[27] We have therefore chosen to use an implicit solvent model to facilitate sampling these rare events. Although such continuum models introduce a further approximation, there is evidence that the energetics can be reproduced with sufficient accuracy to produce useful predictions.[27,28] For example, Tsui *et al.* employed the generalized Born (GB) model to simulate the A- and B-forms of a duplex DNA d(CCAACGTTGG)$_2$ and the corresponding duplex RNA r(CCAACGUUGG)$_2$, demonstrating good agreement with simulations using explicit water in terms of both structure and energetics.[29] In particular, the energy differences between the A and B form duplexes derived from GB trajectories for both DNA and RNA closely matched those obtained using explicit water simulations.[29] Recent systematic evaluations of DNA force fields with implicit solvation have also produced reasonable agreement with explicit solvation and experimental data for both structural and dynamical properties.[30,31]

Within the present framework, an implicit solvent model is particularly useful because the additional degrees of freedom associated with explicit solvent molecules and counterions would complicate the landscape, making it harder to identify the key transition pathways.[32] Comparison with explicit solvent results in previous work has demonstrated that implicit solvent landscape can capture the mechanism and kinetics underlying conformational changes

of biomolecules.[33,34] Using the all-atom AMBER99/bsc0[22] + χOL4 torsional corrections[23] with a GB implicit solvent model, Chakraborty *et al.* characterised the folding mechanisms and kinetics of RNA tetraloop hairpins, and the calculated rate constants for folding of both UUCG and GCAA hairpins were in excellent agreement with the values reported for hairpins of similar size in temperature jump experiments.[35] In a recent work, we have demonstrated that the latest AMBER force fields in conjunction with the GB model can also reproduce the experimental free energy difference, as well interconversion rates for a RNA conformational switch, where the major and minor conformations only differ in the stacking orientation of the central adenines.[32] Similarly, we also investigated the transitions between Watson-Crick (WC) and Hoogsteen (HG) base pairing in DNA using an implicit solvent model, revealing multiple pathways.[36] Our analysis indicates that the transition state ensemble corresponds to base flipped-out states, in agreement with the experimental work of Nikolova *et al.*[37] Based on all this evidence, we believe that implicit solvent landscapes can capture the essential features of the underlying landscape, and provide key mechanic insight into the kinetics of conformational transitions.

Based on our previous experience, we anticipate that the topography of the free energy landscape and its emergent properties (transition mechanism, thermodynamics, and kinetics) would remain qualitatively similar if explicit ions or solvent molecules are included, although some reorganization of minima within the major free energy basins is expected. The kinetic transition networks described here can provide starting points for such studies. We note that unbiased all-atom simulations of DNA hybridization using explicit solvent (even for the short oligonucleotides considered here) would be extremely challenging, if not computationally intractable, due to the long time scales involved. Enhanced sampling of the kind employed by Zacharias and coworkers,[38] or similar, could be useful in addressing this sampling bottleneck, but these approaches also have their own limitations, be it the loss of kinetic information due to biasing along reaction coordinates or exchanges in temperature/Hamiltonian space, or difficulties/inaccuracies associated with partitioning of the phase space into a few metastable

states (Markov-State Model based frameworks[39]). These limitations could potentially lead to a misleading picture of thermodynamics and kinetics even when a more detailed description of the DNA and the surrounding solvent is invoked. Our results, which do not require such approximations, provide a powerful alternative approach, with a different set of assumptions. The ultimate validation must be comparison with experimental observables.

Analysis and visualization of the energy landscapes.

Given a database of stationary points on the potential energy landscape, a self-consistent regrouping scheme[40,41] was used to recursively group together minima, separated by free energy barriers below a chosen threshold, at a given temperature. Within the harmonic superposition approximation (HSA),[42] the partition function is a sum of individual contributions from local minima, each of which is represented by a locally harmonic vibrational density of states. The minimum-to-minimum rate constants are likewise assumed to be given by harmonic transition state theory.[43–45] The free energies of groups of minima and transition states are then estimated by appropriate sums over partition functions of the individual stationary points.

Assigning appropriate edge weights to a kinetic transition network,[1,46] the 'fastest' discrete path, *i.e.* that which makes the largest contribution to the steady-state rate constant, can be determined by Dijkstra's algorithm.[47] A path deviation algorithm[48–50] was used to determine the complete set of distinct paths in the network, where each discrete path is distinguished on the basis of having a unique rate-limiting step, and to characterise the transition state ensemble. The representative pathways are chosen by manually examining the pathway between the conformational ensembles corresponding to the dissociated and fully-hybridized states. A longer path with a distinct potential energy profile is recognized as a representative pathway. The new graph transformation (NGT) method[51] allows calculation of phenomenological rate constants.

Disconnectivity graphs[52–56] are used to visualize potential and free energy landscapes, in

a manner that preserves their full dimensionality. In a disconnectivity graph, the vertical axis corresponds to energy. Local minima are represented by the leaves of the graph, which terminate at their energies. The connected minima are classified into disjoint sets, or 'superbasins',[52] at regular threshold increments, by inducing a cut in the graph if groups of minima are separated by a barrier that exceeds the threshold. The disconnectivity graphs are coloured according to the total number of hydrogen-bonds between complementary DNA strands. The VMD program[57] was used to visualize the structures for representative stationary points.

Estimating free energies and interconversion rates.

The vibrational partition functions associated with the minima and the transition states in the network were estimated using a harmonic approximation.[58,59]

$$Z_i(T) = \frac{n_i e^{-V_i/kT}}{(h\overline{\nu}_i/kT)^{\kappa}}. \tag{1}$$

The local free energy, $F_i(T)$ of each minimum can be written as:

$$F_i(T) = -kT \ln Z_i(T), \tag{2}$$

The equilibrium occupation probability of each minimum $p_i^{eq}(T)$ is:

$$p_i^{eq}(T) = \frac{Z_i(T)}{Z(T)}. \tag{3}$$

$Z(T)$ is the full canonical partition function, which can be written as a sum of all the contributions from the different catchment basins.

$$Z(T) = \sum_i^N Z_i(T). \tag{4}$$

In Eq. (1) , V_i denotes the potential energy of minimum i, n_i is the number of distinct permutational isomers of i, $\overline{\nu}_i$ denotes the geometric mean normal mode frequency associated with minimum i, and $\kappa = 3N - 6$ is the number of vibrational degrees of freedom, where N is the number of atoms.

The partition functions and the free energies for the transition states are defined in an analogous way, except that the normal mode frequency corresponding to the unique negative Hessian eigenvalue (imaginary normal mode frequency) is excluded.

The minimum-to-minimum rate constants are estimated using harmonic transition state theory (TST).

$$k_i^{\dagger}(T) = \frac{kT}{h}\frac{Z^{\dagger}(T)}{Z_i(T)}e^{-\beta\Delta V}. \tag{5}$$

In Eq. (5), $Z^{\dagger}(T)$ denotes the partition function of the transition state; $Z_i(T)$ is the partition function of minimum i; ΔV is the potential energy difference between the transition state and minimum i. The total rate constant $k_{ji}(T)$ for an elementary transition from minimum i to minimum j is obtained by summing the $k_i^{\dagger}(T)$ values for all transition states that connect the two minima.

Within the steady-state approximation for intervening minima, the rate constants, k_{AB}^{SS} and k_{BA}^{SS} between reactant (A) and product (B) states, can be expressed as weighted sums over all discrete paths in the network, assuming that the dynamics between adjacent minima or lumped states[40] is Markovian:[1]

$$k_{AB}^{SS} = \frac{1}{p_B^{eq}}\sum_{a\leftarrow b}\frac{k_{ai_1}k_{i_1i_2}k_{i_2i_3}...k_{i_nb}p_b^{eq}}{\sum_{j_1}k_{j_1i_1}\sum_{j_2}k_{j_2i_2}\sum_{j_3}k_{j_3i_3}...\sum_{j_n}k_{j_ni_n}}. \tag{6}$$

In terms of transition probabilities, $P_{\gamma\alpha}$ between directly connected minima γ and α, Eq. (6) can be rewritten as:[1]

$$k_{AB}^{SS} = \frac{1}{p_B^{eq}}\sum_{a\leftarrow b}P_{ai_1}P_{i_1i_2}P_{i_2i_3}...P_{i_nb}p_b^{eq}\tau_b^{-1}. \tag{7}$$

The individual sums in the denominators of Eq. (6) consist of the unimolecular rate con-

stants for all direct transitions from minimum j_k to i_k. The discrete path that makes the largest contribution to the steady-state rate constant is termed the 'fastest path', and can be extracted from the network using Dijkstra's shortest path algorithm with appropriate edge weights[60] $-\ln P_{\alpha\beta}$, providing access to the product of transition probabilities in Eq. (7). To characterize additional paths in order of their contribution to k^{SS} we employ the recursive enumeration algorithm[61] within the same framework,[41] and these pathways were examined to deduce mechanistic details of the adenine conformational switch. The steady-state approximation for the intervening minima can be relaxed to yield rate constants k_{AB} and k_{BA} that correspond to the mean first passage times between reactants and products.[51] To extract these values we employ a graph transformation technique (NGT)[51,62] where minima in the intervening region are progressively removed, and the transition probabilities as well as the waiting times are renormalised to conserve the average mean first passage time (MFPT).[51,63]

We compute the rate constants in conjunction with a recursive free energy regrouping scheme,[40] which lumps together structures separated by free energy barriers below a certain threshold into a single macrostate. This approach is similar in spirit to the kinetic clustering schemes employed in methods based upon explicit dynamics.[64,65] In the regrouping, the original reactant and product states are expanded into ensembles of conformations assumed to be in local equilibrium, and hence the global dynamics can be directly compared to the observation time scale of experiments.[66] After regrouping, the equilibrium occupation probability and the free energy associated with group J are

$$p_J^{eq}(T) = \sum_{j \in J} p_j^{eq}(T), \tag{8}$$

$$F_J = -kT \ln \sum_{j \in J} Z_j(T). \tag{9}$$

where minimum j is a member of group J. The free energy of the group of transition states

linking J and K is:[40]

$$F^{\dagger}_{KJ} = -kT \ln \sum_{k \leftarrow j} Z^{\dagger}_{kj}(T) \equiv -kT \ln Z^{\dagger}_{KJ}(T), \tag{10}$$

To analyze global dynamics corresponding to regrouped databases, the rate constants corresponding to transitions between different free energy groups are required, which can then be used in the appropriate expressions for rate constants and committor probabilities.[51] The intergroup rate constant from J to K is:[40]

$$\begin{aligned} k_{KJ} &= \sum_{k \leftarrow j} \frac{p^{eq}_{j}(T)}{p^{eq}_{J}(T)} k_{kj}(T) = \sum_{k \leftarrow j} \frac{Z_j(T)}{Z_J(T)} \frac{kT}{h} \frac{Z^{\dagger}_{kj}(T)}{Z_j(T)}, \\ &= \frac{kT}{h} \frac{Z^{\dagger}_{KJ}(T)}{Z_J(T)} = \frac{kT}{h} e^{-\left[F^{\dagger}_{KJ}(T) - F_J(T)\right]/kT}. \end{aligned} \tag{11}$$

SUPPORTING SECTION 2: FIGURES

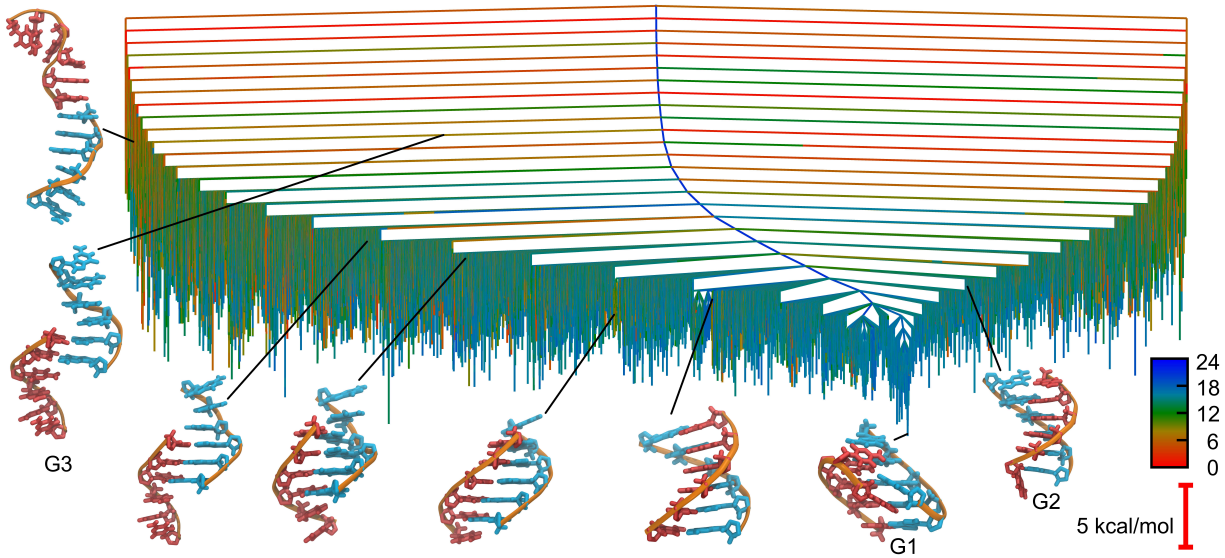

Figure S1: Potential energy disconnectivity graph for d(GGGGGG). The branches are coloured according to the total number of hydrogen-bonds between the two DNA strands. Some representative local minima from the different conformational ensembles are shown.

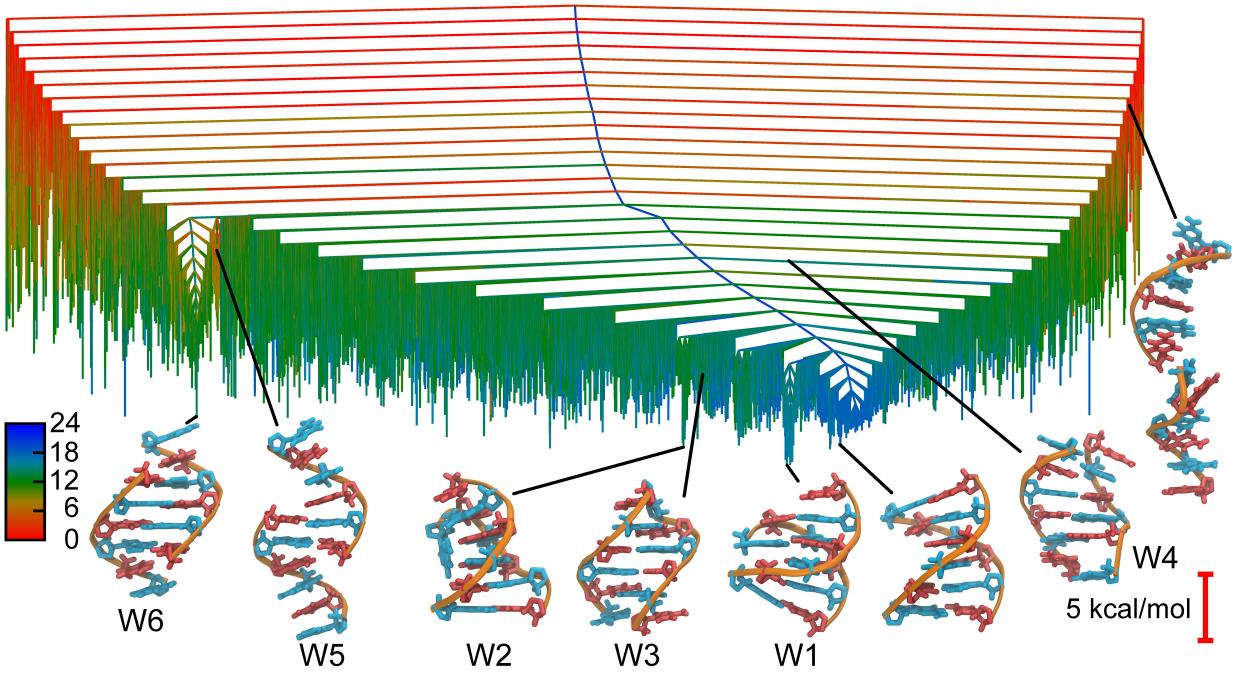

Figure S2: Potential energy disconnectivity graph for d(GCGCGC). The branches are coloured according to the total number of hydrogen-bonds between the two DNA strands. Some representative local minima from the different conformational ensembles are shown.

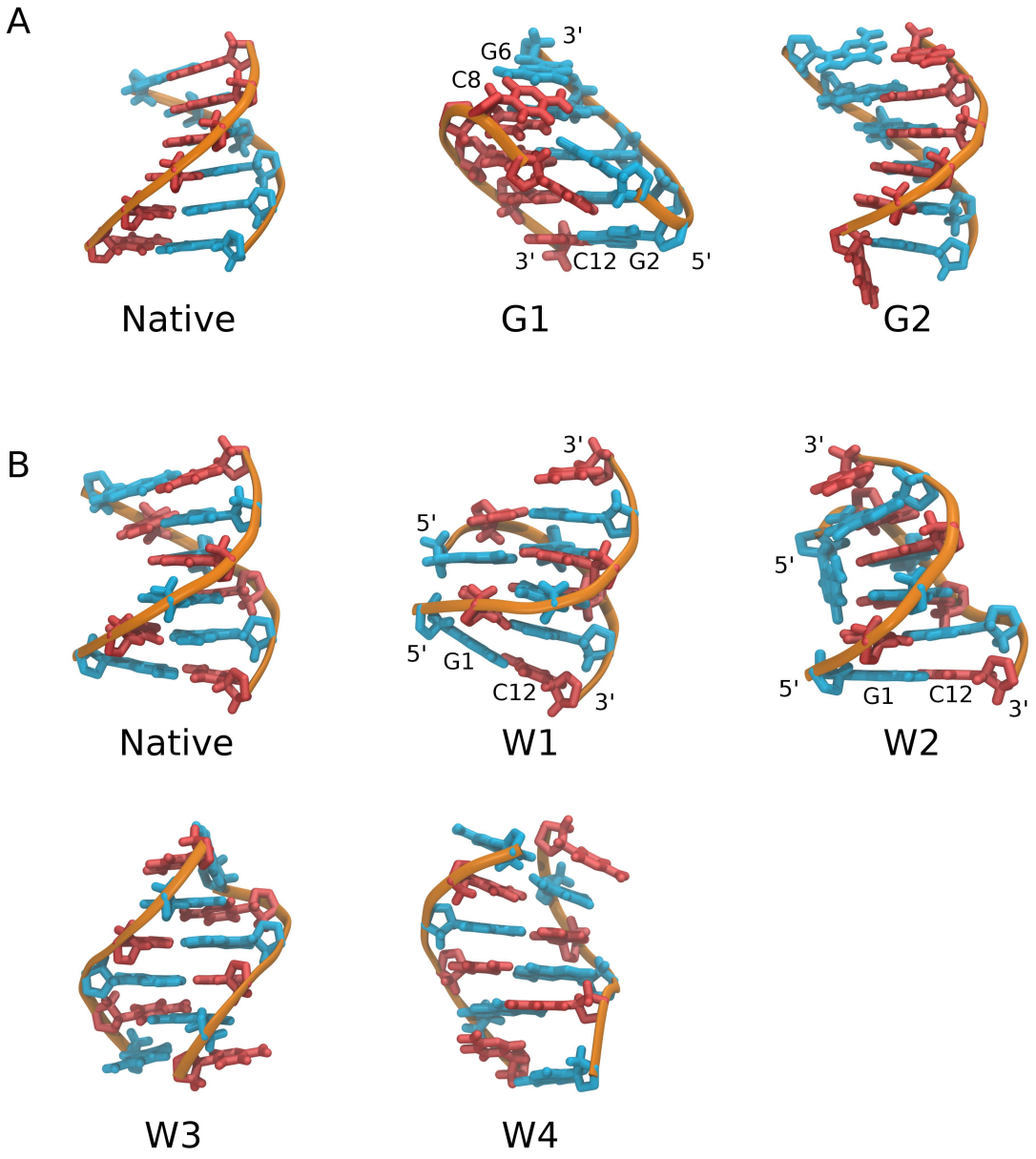

Figure S3: Some representative structures located in the low free energy region of the landscapes for (A) d(GGGGG) and (B) d(GCGCGC). Note that all the structures are the same as those in Figs 1 and 2 of the main text.

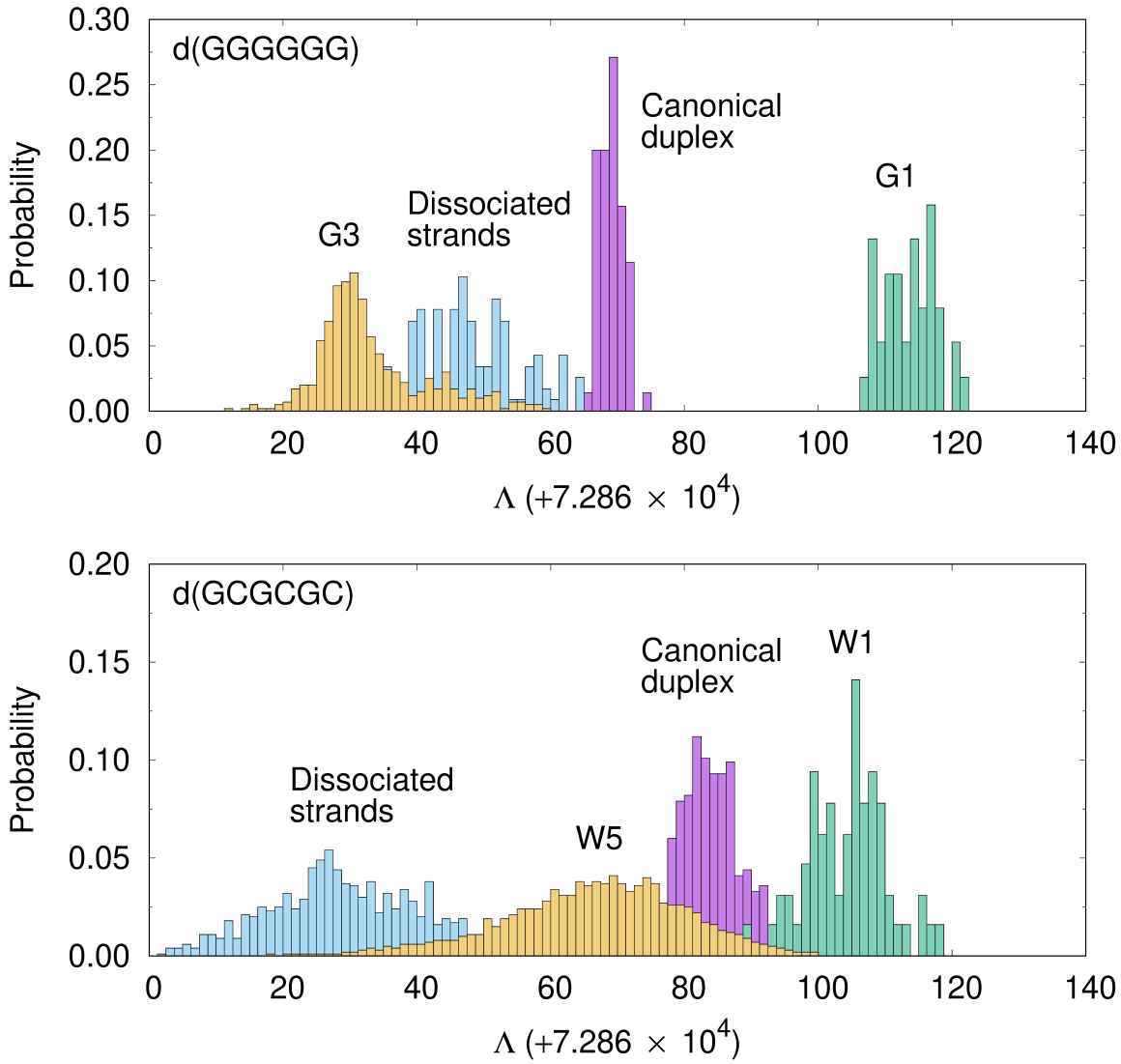

Figure S4: Top: Distribution of the normal mode frequencies (Λ) for minima constituting the free energy groups corresponding to the canonical duplex, dissociated strands, misaligned structure G3, and compact structure G1 for d(GGGGGG). The regrouping calculations were performed using the recursive regrouping scheme,[40,41] with a barrier threshold of 3.5 kcal mol^{-1}. Bottom: Distribution of Λ values for minima constituting the free energy groups corresponding to the canonical duplex, dissociated strands, misaligned structure W5, and compact structure W1 for d(GCGCGC). The regrouping calculations were performed using the recursive regrouping scheme,[40,41] with a barrier threshold of 4.5 kcal mol^{-1}. Within the harmonic approximation, vibrational entropy is $-k_B\Lambda$[67–69] and more flexible structures are therefore expected to have lower Λ values. Note that the Λ values for slipped conformations G3 are lower than the those of the dissociated structures for d(GGGGGG). This is because the nucleobases of the two single-stranded oligonucleotides for the dissociated strands of d(GGGGGG) stack with each other, and even form hydrogen-bonds with their neighbours, producing compact-like structures rather than fully extended single-stranded oligonucleotides. The 'dissociated' structures do not refer to separated strands here. All the conformations are the same as those in Figures 1 and 2 of the main text.

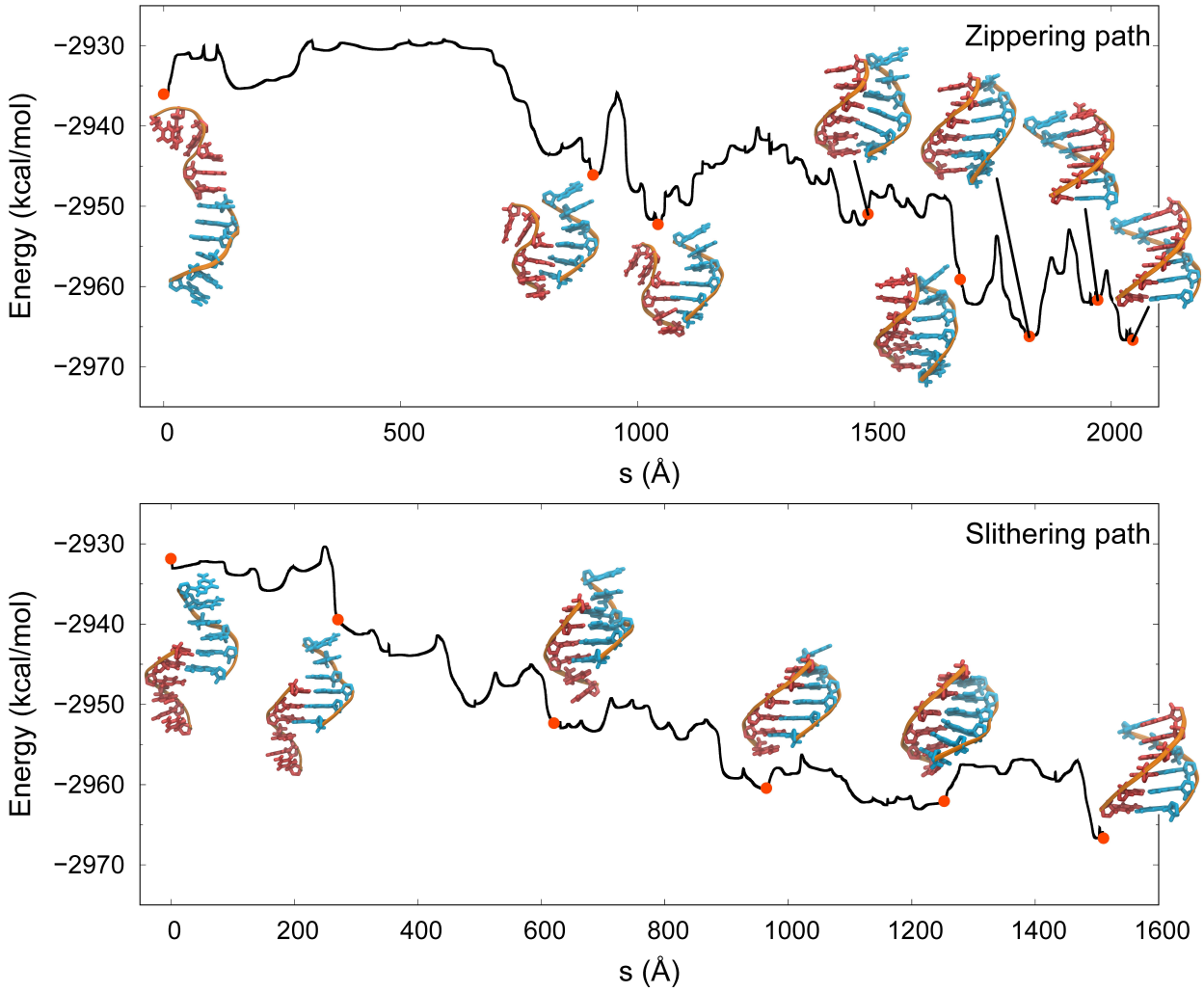

Figure S5: Top: The potential energy as a function of the integrated path length (**s**) for the fastest zippering pathway between the conformational ensembles corresponding to dissociated and fully-hybridized states for d(GGGGGG). Bottom: The fastest slithering pathway between the ensemble of structures slipped by four bases and the fully-hybridized state for d(GGGGGG). Each step along the pathway is one stationary point. Selected structures are illustrated along the path, as indicated by the red dots.

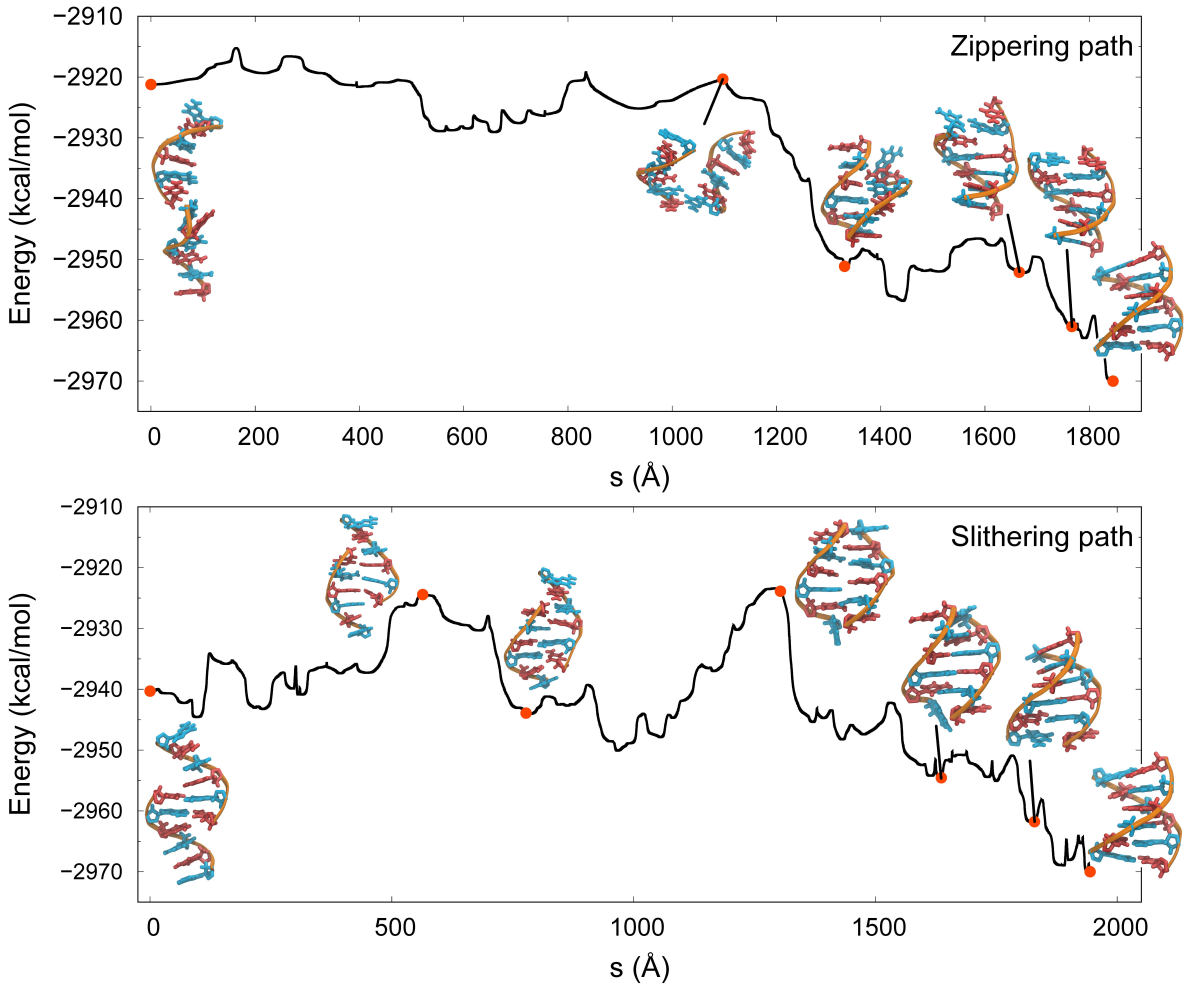

Figure S6: Top: The potential energy as a function of the integrated path length (**s**) for the fastest zippering pathway between the conformational ensembles corresponding to melted and fully-hybridized states for d(GCGCGC). Bottom: The fastest slithering pathway between the ensemble of structures slipped by four bases and the fully-hybridized state for d(GCGCGC). Each step along the pathway represents one stationary point. Selected structures are illustrated along the path, as indicated by the red dots.

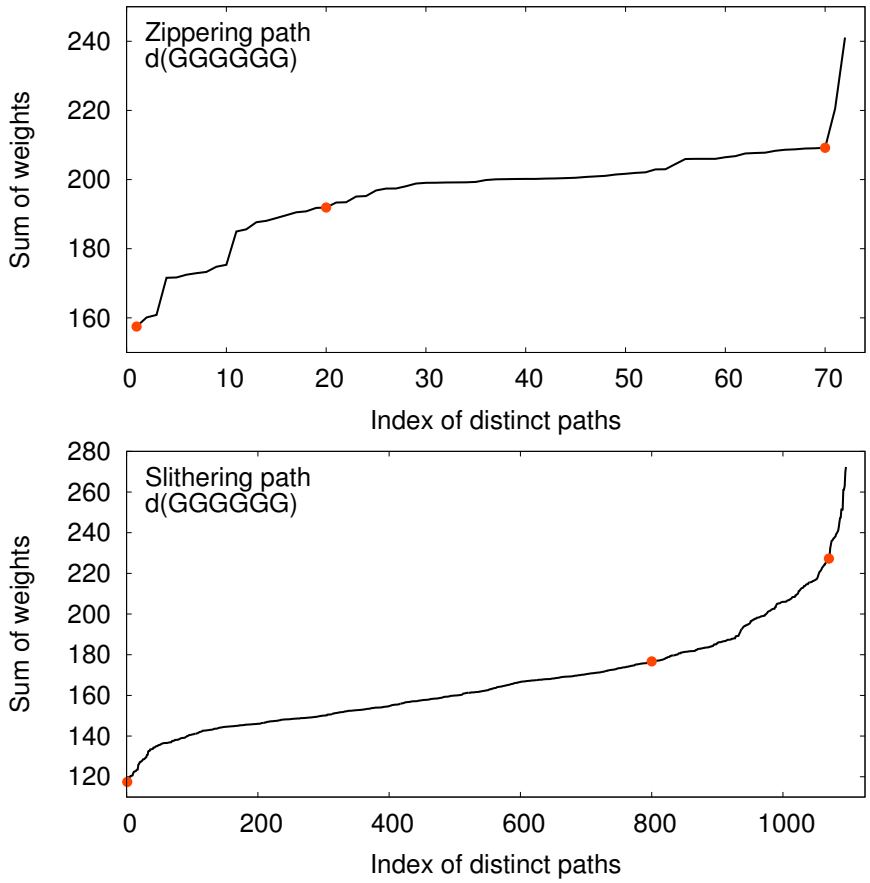

Figure S7: The sum of edge weights as a function of distinct paths on the network for the hybridization of d(GGGGGG), determined by the path deviation algorithm.[48,49] Each distinct path is distinguished on the basis of having a unique rate-limiting step. The distinct paths are indexed from one, starting from the fastest. For both zippering (top) and slithering (bottom) mechanisms, three representative pathways belonging to families with distinct potential energy profiles shown in Figures S8 and S9 are indicated using filled circles in red on the profiles.

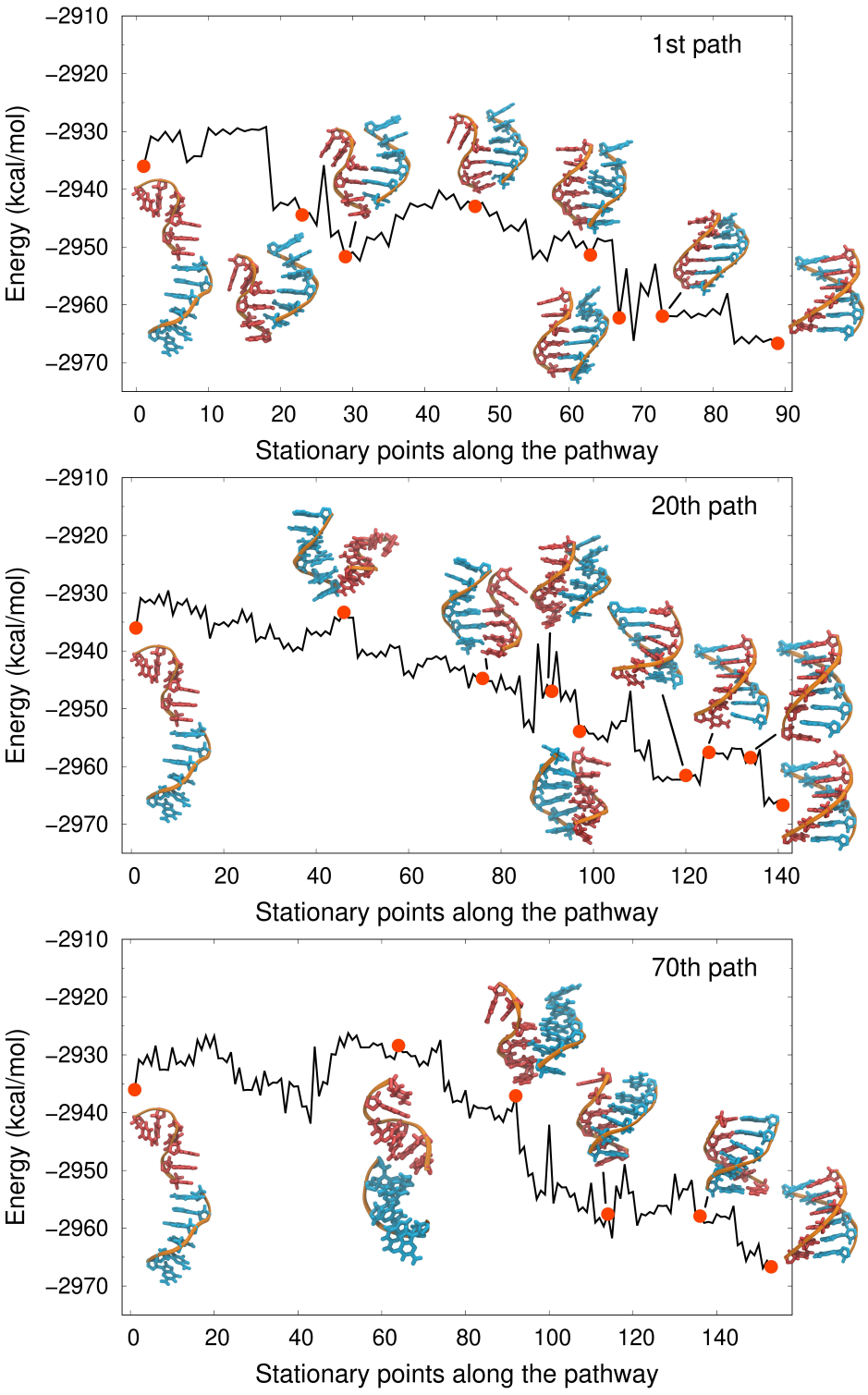

Figure S8: Potential energy profiles of three representative pathways for the hybridization of d(GGGGGG) by a zippering mechanism, determined by the path deviation algorithm[48,49] to find kinetically distinguishable pathway ensembles. The first path (top) shown is the fastest, and the second (middle) and third (bottom) paths are the 20th and 70th fastest distinct paths, respectively, which are highlighted in Figure S7. Snapshots of some representative local minima that are encountered at different stages of the hybridization process are shown.

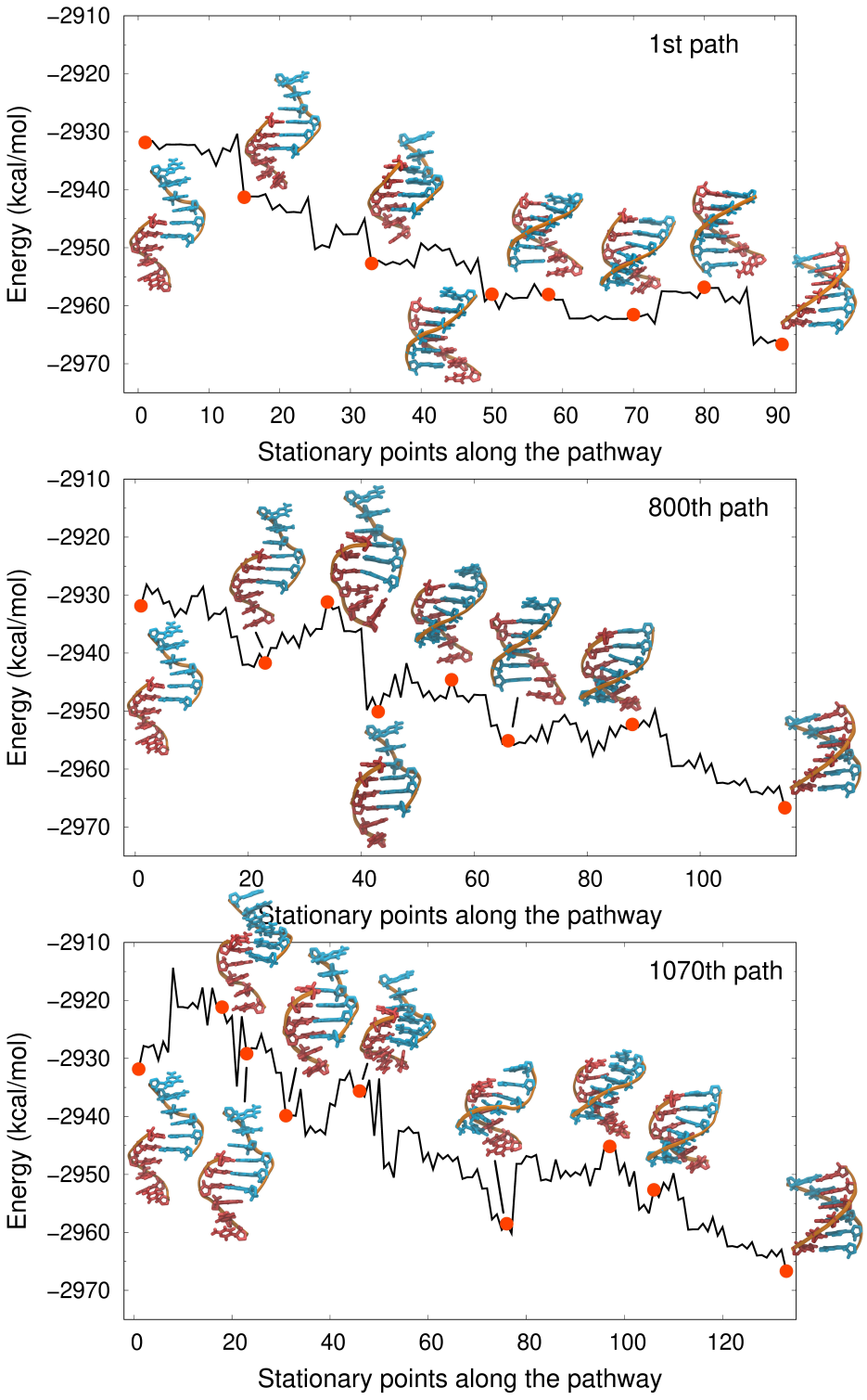

Figure S9: Potential energy profiles of three representative pathways for the hybridization of d(GGGGGG) by a slithering mechanism, determined by the path deviation algorithm to find kinetically distinguishable pathway ensembles. The first path (top) shown is the fastest, and the second (middle) and third (bottom) paths are the 800th and 1070th fastest distinct paths, respectively, which are highlighted in Figure S7. Snapshots of some representative local minima that are encountered at different stages of the hybridization process are shown.

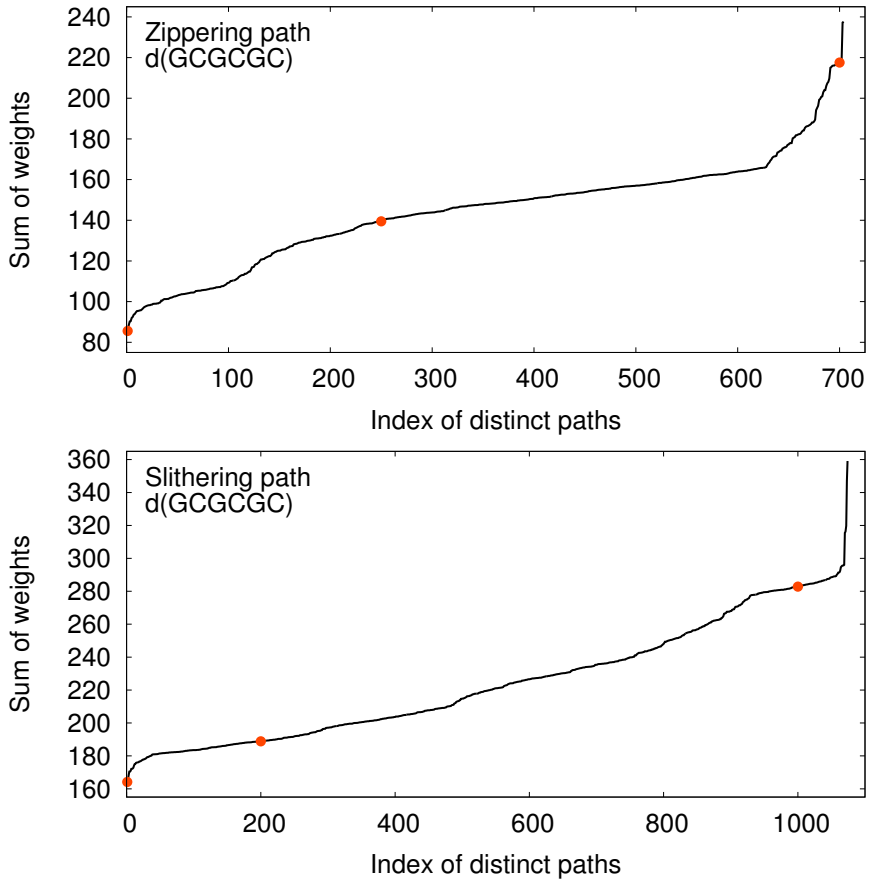

Figure S10: The sum of edge weights as a function of distinct paths on the network for the hybridization of d(GCGCGC), determined by the path deviation algorithm.[48,49] Each distinct path is distinguished on the basis of having a unique rate-limiting step. The distinct paths are indexed from one, starting from the fastest. For both zippering (top) and slithering (bottom) mechanisms, three representative pathways belonging to families with distinct potential energy profiles shown in Figures S11 and S12 are highlighted using filled circles in red on the profiles.

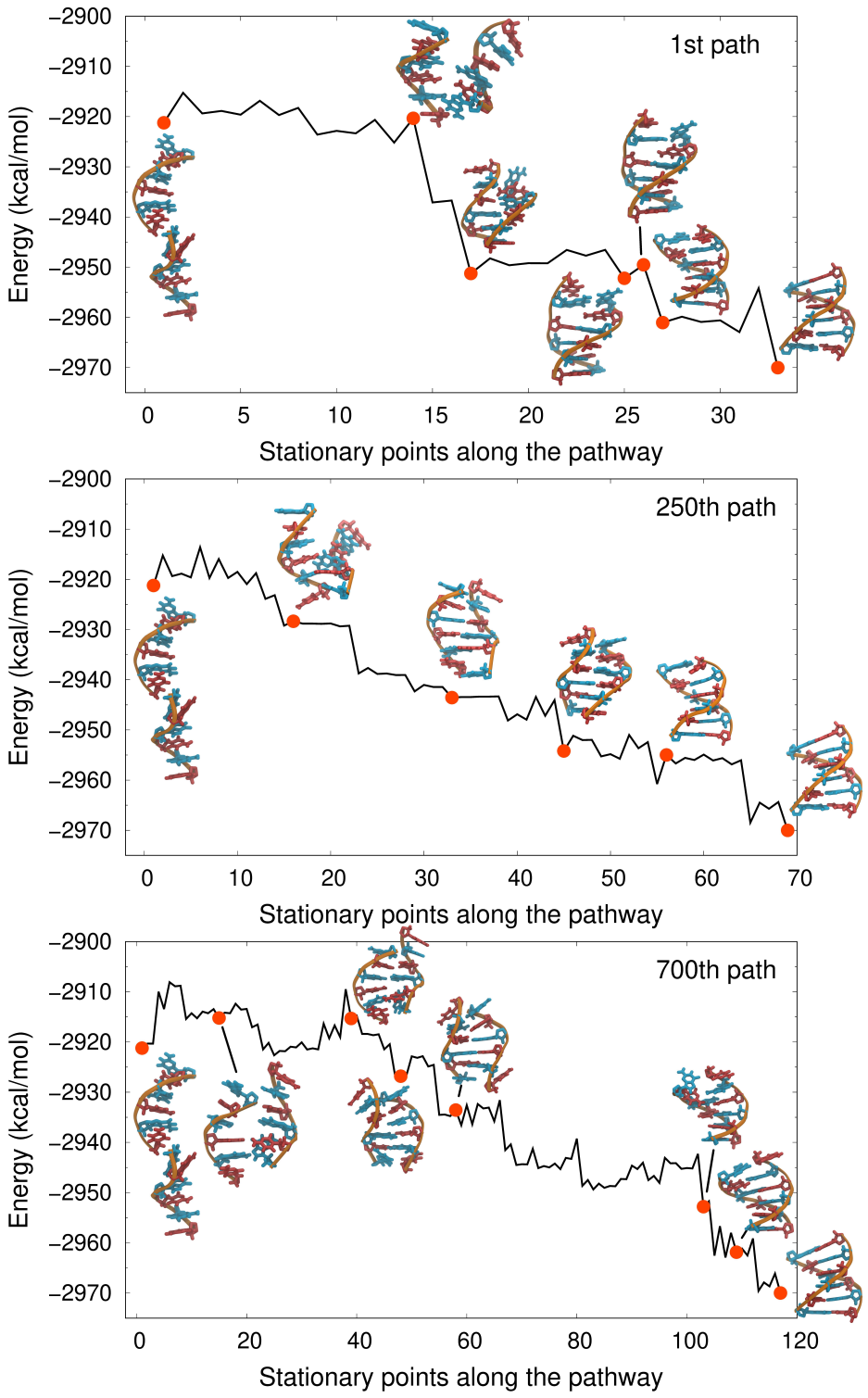

Figure S11: Potential energy profiles of three representative pathways for the hybridization of d(GCGCGC) by a zippering mechanism, determined by the path deviation algorithm[48,49] to find kinetically distinguishable pathway ensembles. The first path (top) shown is the fastest, and the second (middle) and third (bottom) paths are the 250th and 700th fastest distinct paths, respectively, which are highlighted in Figure S10. Snapshots of some representative local minima that are encountered at different stages of the hybridization process are shown.

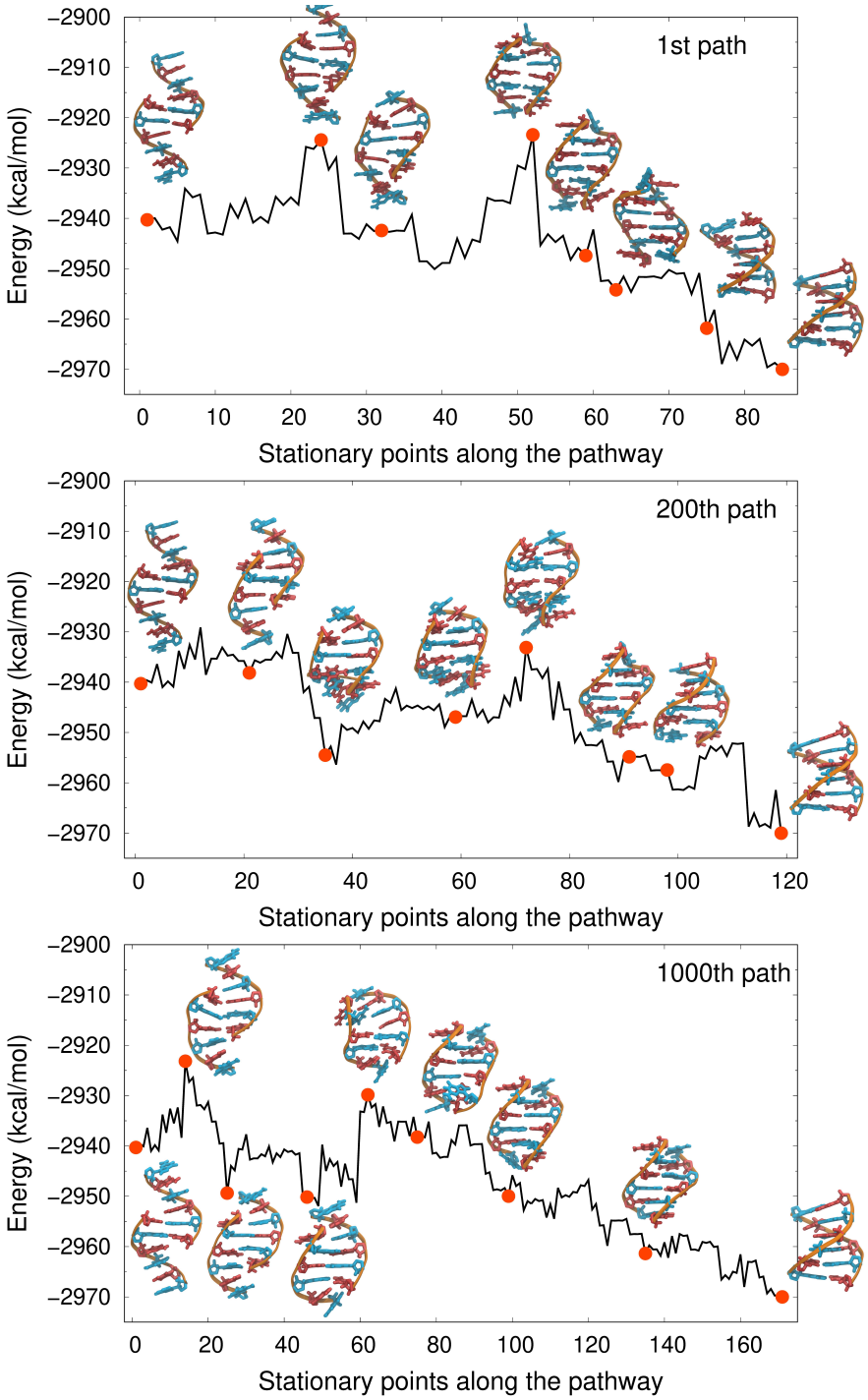

Figure S12: Potential energy profiles of three representative pathways for the hybridization of d(GCGCGC) by a slithering mechanism, determined by the path deviation algorithm[48,49] to find kinetically distinguishable pathway ensembles. The first path (top) shown is the fastest, and the second (middle) and third (bottom) paths are the 200th and 1000th fastest distinct paths, respectively, which are highlighted in Figure S10. Snapshots of some representative local minima that are encountered at different stages of the hybridization process are shown.

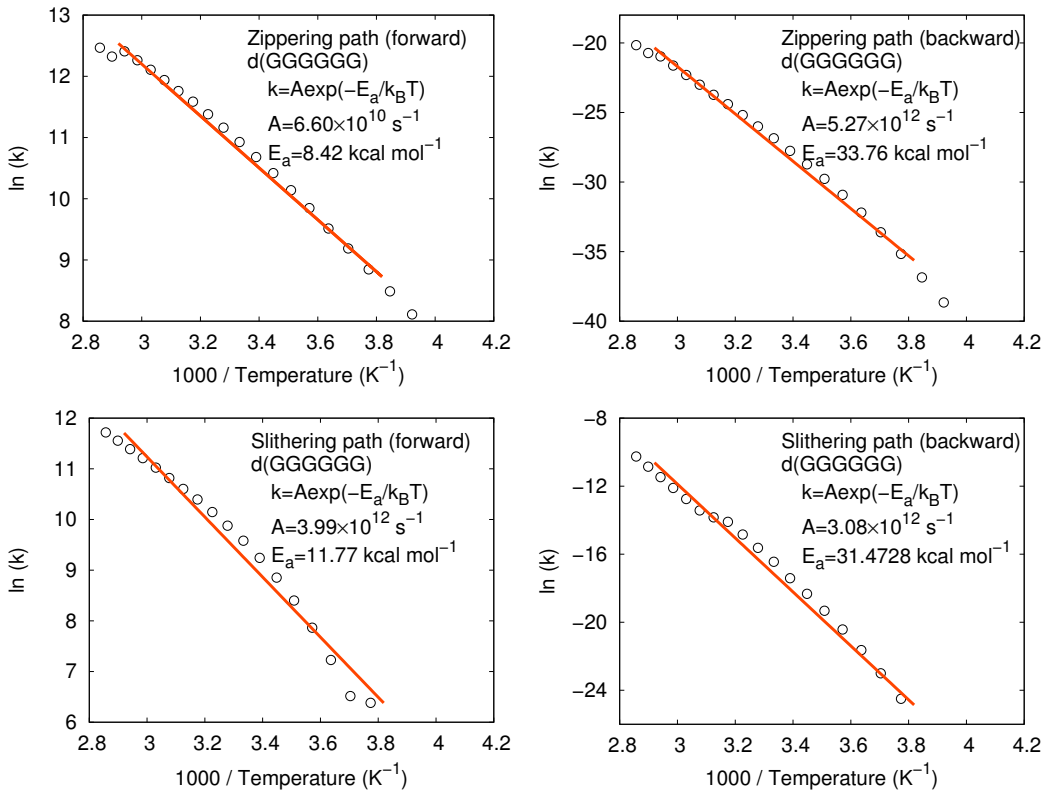

Figure S13: Variations of the rate constants (k, s^{-1}) corresponding to the zippering (forward and backward transition, top) and the slithering (forward and backward transition, bottom) pathways with reciprocal temperature ranging from 250 K to 350 K for d(GGGGGG), depicted using open circles. The best fits to the Arrhenius equation, $k(\mathrm{T}) = \mathrm{Aexp}(-\mathrm{E_a}/\mathrm{k_B T})$ are depicted as solid lines in red. Significant deviations from Arrhenius behaviour are evident in each case.

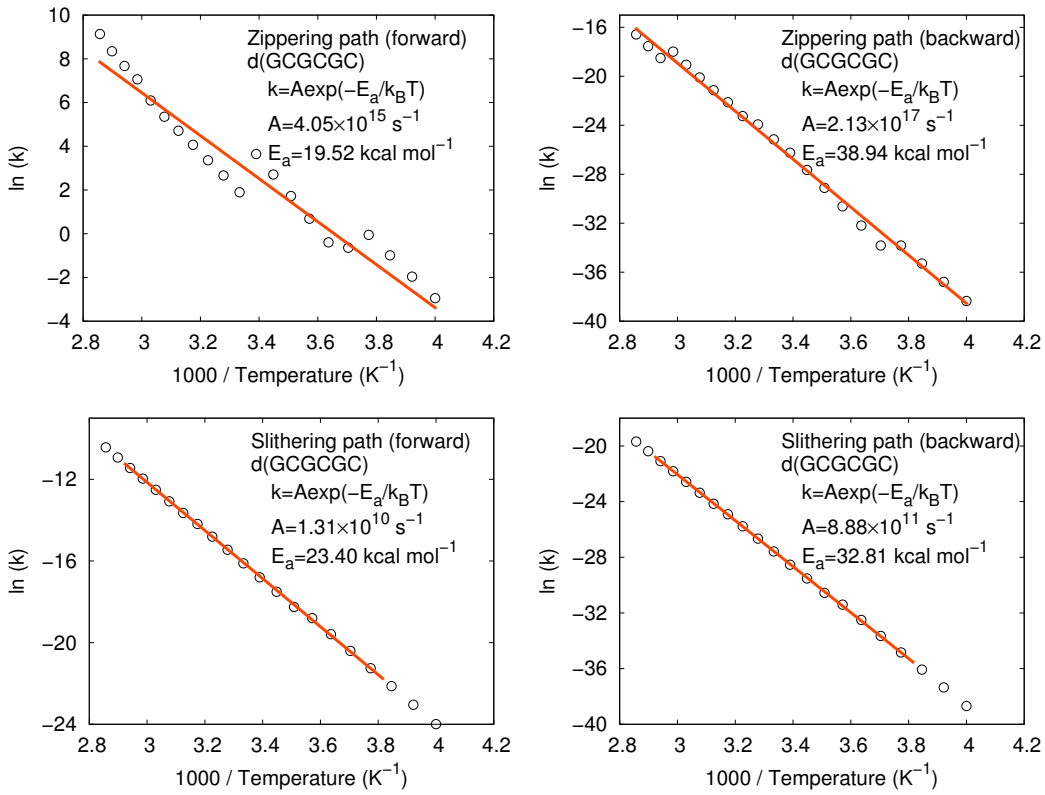

Figure S14: Variations of the rate constants (k, s^{-1}) corresponding to the zippering (forward and backward transition, top) and the slithering (forward and backward transition, bottom) pathways with reciprocal temperature ranging from 250 K to 350 K for d(GCGCGC), depicted using open circles. The best fits to the Arrhenius equation, $k(\mathrm{T}) = \mathrm{Aexp}(-\mathrm{E_a}/\mathrm{k_B T})$ are depicted as solid lines in red. Significant deviations from Arrhenius behaviour are evident in the zippering pathway.

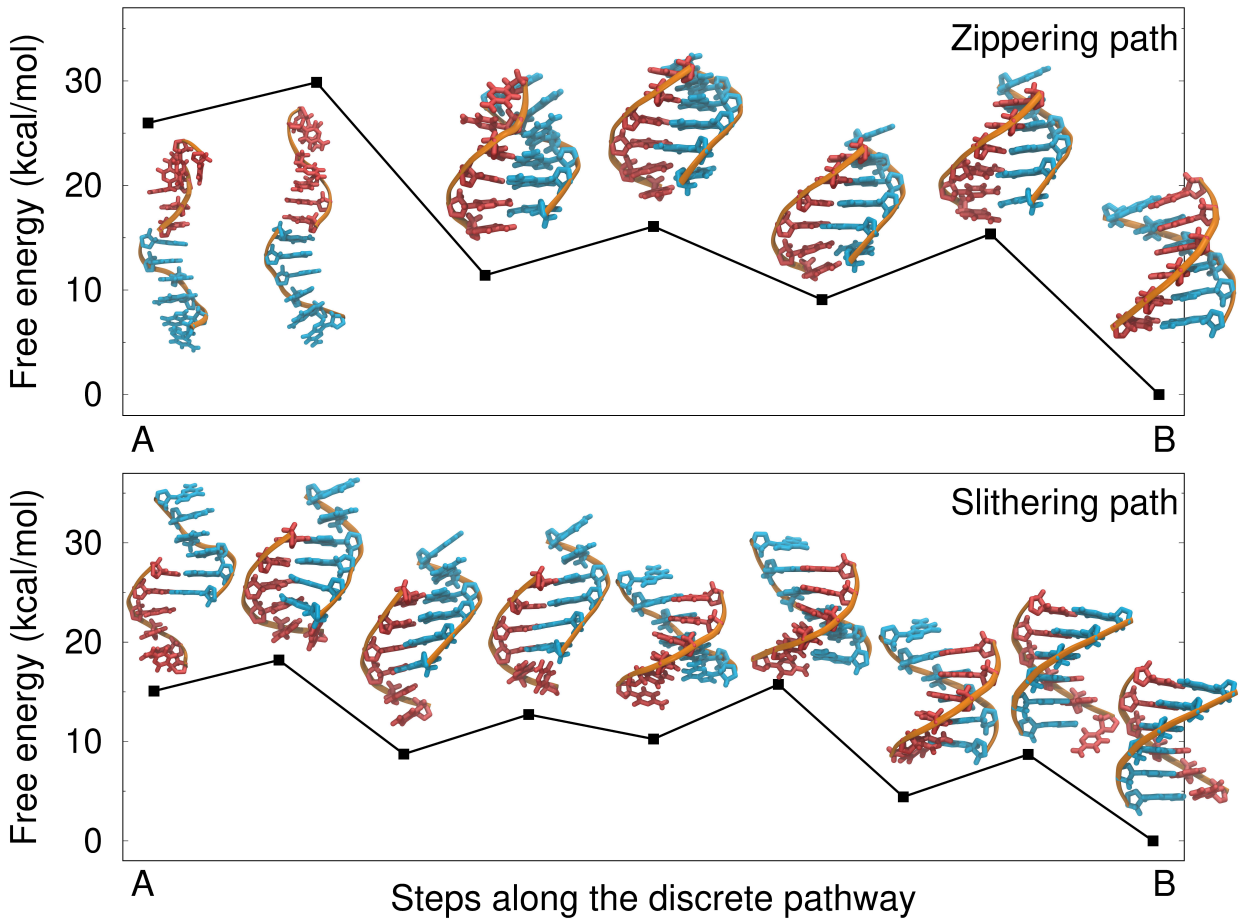

Figure S15: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GGGGGG). The points correspond to groups of minima and transition states, obtained by the recursive regrouping scheme[40,41] using a barrier threshold of 5.5 kcal mol^{-1}. Bottom: The fastest free energy slithering pathway between the ensemble of structures slipped by four bases (A) and the fully-hybridized state (B) for d(GGGGGG), using a regrouping threshold of 3.5 kcal mol^{-1}. In the calculations, the temperature was set to 270 K.

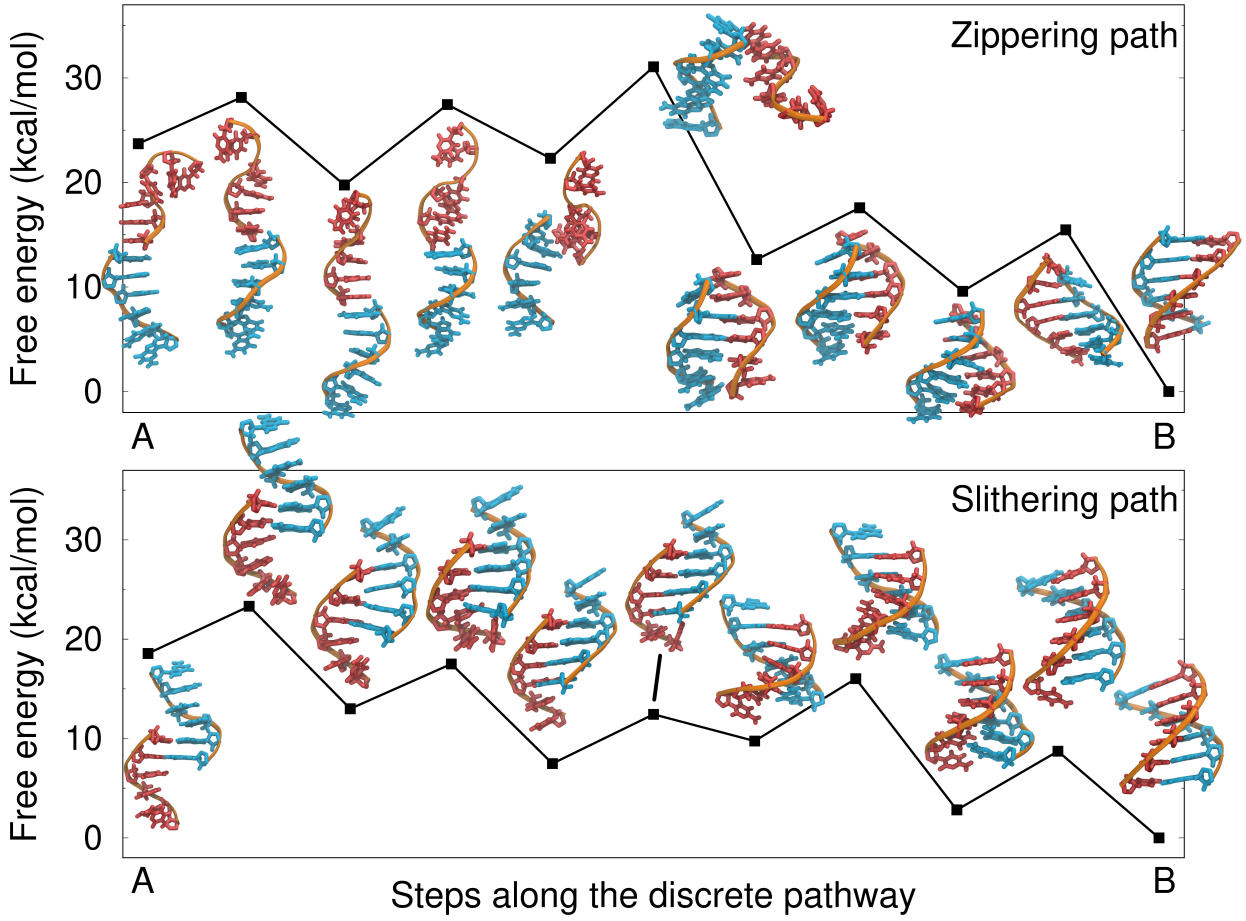

Figure S16: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GGGGGG). The points correspond to groups of minima and transition states, obtained by the recursive regrouping scheme[40,41] using a barrier threshold of 5.5 kcal mol^{-1}. Bottom: The fastest free energy slithering pathway between the ensemble of structures slipped by four bases (A) and the fully-hybridized state (B) for d(GGGGGG), using a regrouping threshold of 3.5 kcal mol^{-1}. In the calculations, the temperature was set to 350 K.

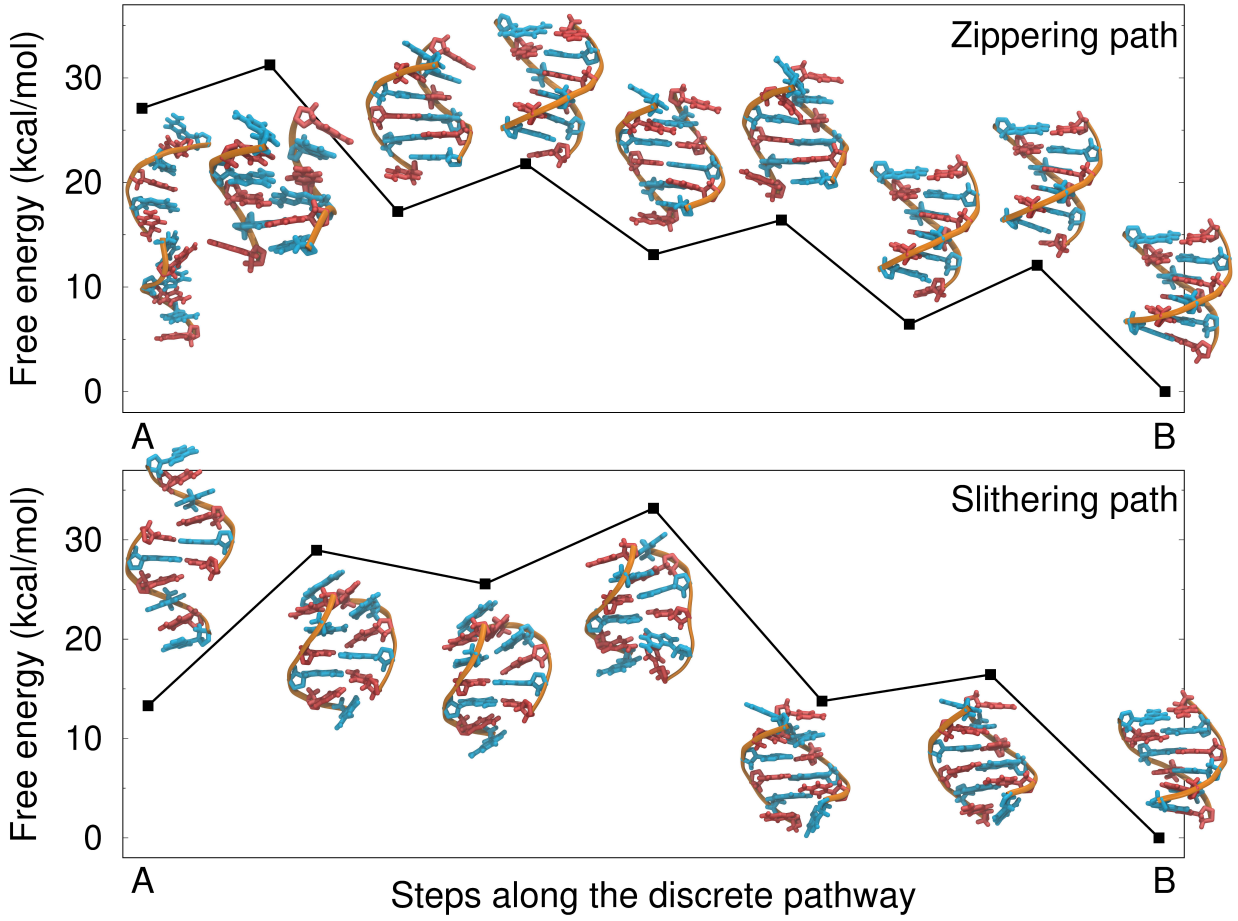

Figure S17: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GCGCGC). The points correspond to free energy groups of minima and transition states, obtained by the recursive regrouping scheme[40,41] using a barrier threshold of 4.5 kcal mol^{-1}. Bottom: The fastest free energy slithering pathway between the ensemble of structures slipped by four bases (A) and the fully-hybridized state (B) for d(GCGCGC), using a regrouping threshold was 6.0 kcal mol^{-1}. In the calculations, the temperature was set to 270 K.

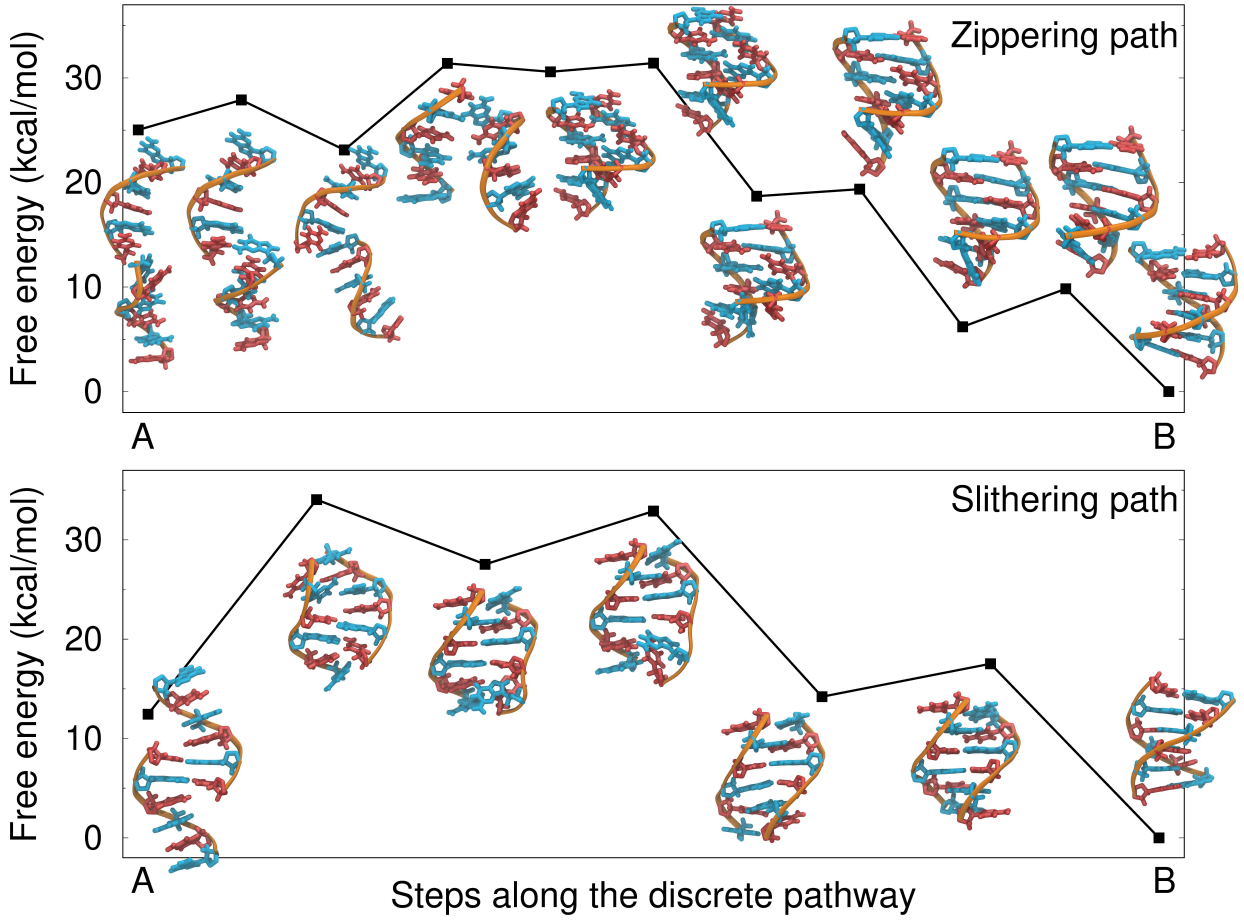

Figure S18: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GCGCGC). The points correspond to free energy groups of minima and transition states, obtained by the recursive regrouping scheme[40,41] using a barrier threshold of 4.5 kcal mol^{-1}. Bottom: The fastest free energy slithering pathway between the ensemble of structures slipped by four bases (A) and the fully-hybridized state (B) for d(GCGCGC), using a regrouping threshold was 6.0 kcal mol^{-1}. In the calculations, the temperature was set to 350 K.

SUPPORTING SECTION 3: TABLE

Table 1: Average values of base pair step parameters for standard B-DNA,[70] for the G1 structure of d(GGGGGG), and for the W1 and W2 structures of d(GCGCGC). For the G1, W1 and W2 structures, only the base paired segment was used for the calculations of local base pair step parameters. The local base pair parameters were calculated using 3DNA.[71]

Local base-pair step parameter	B-DNA	G1	W1	W2
Tilt (deg.)	-0.10 ± 2.5	0.99 ± 2.04	-0.68 ± 5.58	-0.74 ± 5.29
Roll (deg.)	0.60 ± 5.2	9.26 ± 2.84	5.84 ± 5.42	6.90 ± 7.20
Twist (deg.)	36.00 ± 6.8	31.13 ± 2.18	36.14 ± 1.92	38.03 ± 12.99
Shift (Å)	-0.02 ± 0.45	0.00 ± 0.49	-0.22 ± 0.79	0.11 ± 1.18
Slide (Å)	0.23 ± 0.87	-1.29 ± 0.12	-0.17 ± 0.14	-0.04 ± 0.71
Rise (Å)	3.32 ± 0.18	3.41 ± 0.04	3.28 ± 0.17	3.18 ± 0.20
Form	B-DNA	B-DNA	B-DNA	B-DNA

References

(1) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.

(2) Wales, D. J. Some Further Applications of Discrete Path Sampling to Cluster Isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(3) Noé, F.; Fischer, S. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.

(4) Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.

(5) Röder, K.; Joseph, J. A.; Husic, B. E.; Wales, D. J. Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems. *Adv. Theory Simul.* **2019**, *2*, 1800175.

(6) Li, Z.; Scheraga, H. A. Monte Carlo-Minimization Approach to the Multiple-minima Problem in Protein Folding. *Proc. Nat.l Acad. Sci. U S A* **1987**, *84*, 6611–6615.

(7) Li, Z.; Scheraga, H. A. Structure and Free Energy of Complex Thermodynamic Systems. *J. Mol. Struct.* **1988**, *179*, 333 – 352.

(8) Oakley, M. T.; Johnston, R. L.; Wales, D. J. Symmetrisation Schemes for Global Optimisation of Atomic Clusters. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3965–3976.

(9) Strodel, B.; Lee, J. W. L.; Whittleston, C. S.; Wales, D. J. Transmembrane Structures for Alzheimer's Aβ_{1-42} Oligomers. *J. Am. Chem. Soc.* **2010**, *132*, 13300–13312.

(10) Wales, D. J. GMIN: A Program for Finding Global Minima and Calculating Thermodynamic Properties. `http://www-wales.ch.cam.ac.uk/GMIN`, 2018.

(11) Trygubenko, S. A.; Wales, D. J. A Doubly Nudged Elastic Band Method for Finding Transition States. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(12) Henkelman, G.; Jónsson, H. A Dimer Method for Finding Saddle Points on High Dimensional Potential Surfaces Using Only First Derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(13) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(14) Munro, L. J.; Wales, D. J. Defect migration in crystalline silicon. *Phys. Rev. B* **1999**, *59*, 3969–3980.

(15) Zeng, Y.; Xiao, P.; Henkelman, G. Unification of algorithms for minimum mode optimization. *J. Chem. Phys.* **2014**, *140*, 044115.

(16) Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program.* **1989**, *45*, 503–528.

(17) Wales, D. J. OPTIM: A Program for Optimizing Geometries and Calculating Pathways. `http://www-wales.ch.cam.ac.uk/OPTIM`, 2018.

(18) Case, D. A.; Darden, T. A.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M. et al. AMBER 9. `http://ambermd.org/`, 2006.

(19) Strodel, B.; Whittleston, C. S.; Wales, D. J. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(20) Carr, J. M.; Wales, D. J. Global Optimization and Folding Pathways of Selected Alpha-helical Proteins. *J. Chem. Phys.* **2005**, *123*, 234901.

(21) Wales, D. J. PATHSAMPLE: A Driver for OPTIM to Create Stationary Point Databases Using Discrete Path Sampling and Perform Kinetic Analysis. `http://www-wales.ch.cam.ac.uk/PATHSAMPLE`, 2018.

(22) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophy. J.* **2007**, *92*, 3817–3829.

(23) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(24) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-scale Conformational Changes With a Modified Generalized Born Model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383–394.

(25) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(26) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of A Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(27) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.

(28) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 1153–1164.

(29) Tsui, V.; Case, D. A. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.

(30) Gaillard, T.; Case, D. A. Evaluation of DNA Force Fields in Implicit Solvation. *J. Chem. Theory Comput.* **2011**, *7*, 3181–3198.

(31) Chocholoušová, J.; Feig, M. Implicit Solvent Simulations of DNA and DNA-Protein Complexes: Agreement with Explicit Solvent vs Experiment. *J. Phys. Chem. B* **2006**, *110*, 17240–17251.

(32) Chakraborty, D.; Wales, D. J. Dynamics of an Adenine-adenine RNA Conformational Switch From Discrete Path Sampling. *J. Chem. Phys.* **2019**, *150*, 125101.

(33) Wales, D. J. Exploring Energy Landscapes. *Annu. Rev. Phys. Chem.* **2018**, *69*, 401–425.

(34) Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring Biomolecular Energy Landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.

(35) Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J. Energy Landscapes, Folding Mechanisms, and Kinetics of RNA Tetraloop Hairpins. *J. Am. Chem. Soc.* **2014**, *136*, 18052–18061.

(36) Chakraborty, D.; Wales, D. J. Energy Landscape and Pathways for Transitions between Watson-Crick and Hoogsteen Base Pairing in DNA. *J. Phys. Chem. Lett.* **2018**, *9*, 229–241.

(37) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* **2011**, *470*, 498–502.

(38) Kannan, S.; Zacharias, M. Simulation of DNA Double-strand Dissociation and Formation During Replica-exchange Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10589–10595.

(39) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.

(40) Carr, J. M.; Wales, D. J. Folding Pathways and Rates for the Three-Stranded β-Sheet Peptide Beta3s using Discrete Path Sampling. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.

(41) Carr, J. M.; Wales, D. J. Refined Kinetic Transition Networks for the GB1 Hairpin Peptide. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3341–3354.

(42) Strodel, B.; Wales, D. J. Free Energy Surfaces From an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide. *Chem. Phys. Lett.* **2008**, *466*, 105–115.

(43) Forst, W. *Theory of Unimolecular Reactions*; Academic Press: New York, 1973.

(44) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.

(45) Miller, W. H. Unified Statistical Model for "complex" and "direct" Reaction Mechanisms. *J. Chem. Phys.* **1976**, *65*, 2216–2223.

(46) Evans, D. A.; Wales, D. J. Folding of the GB1 Hairpin Peptide From Discrete Path Sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090.

(47) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269–271.

(48) Frigioni, D.; Marchetti-Spaccamela, A.; Nanni, U. Fully Dynamic Algorithms for Maintaining Shortest Paths Trees. *J. Algorithms* **2000**, *34*, 251–281.

(49) Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *J. Chem. Theory Comput.* **2006**, *2*, 840–857.

(50) Sharpe, D. J.; Wales, D. J. Identifying Mechanistically Distinct Pathways in Kinetic Transition Networks. *J. Chem. Phys.* **2019**, *151*, 124101.

(51) Wales, D. J. Calculating Rate Constants and Committor Probabilities for Transition Networks by Graph Transformation. *J. Chem. Phys.* **2009**, *130*, 204111.

(52) Becker, O. M.; Karplus, M. The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517.

(53) Krivov, S. V.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 14766–14770.

(54) Krivov, S. V.; Karplus, M. Free Energy Disconnectivity Graphs: Application to Peptide Models. *J. Chem. Phys.* **2002**, *117*, 10894–10903.

(55) Evans, D. A.; Wales, D. J. Free Energy Landscapes of Model Peptides and Proteins. *J. Chem. Phys.* **2003**, *118*, 3891–3897.

(56) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal Energy Landscapes. *Nature* **1998**, *394*, 758–760.

(57) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(58) Hoare, M. R.; McInnes, J. Statistical Mechanics and Morphology of Very Small Atomic Clusters. *Faraday Discuss. Chem. Soc.* **1976**, *61*, 12–24.

(59) Hoare, M. *Advances in Chemical Physics*; John Wiley & Sons: USA, 1979; Vol. 40; pp 49 – 129.

(60) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. Finding Pathways Between Distant Local Minima. *J. Chem. Phys.* **2005**, *122*, 234903.

(61) Jiménez, V. M.; Marzal, A. Computing the K Shortest Paths: A New Algorithm and an Experimental Comparison. Algorithm Engineering. Berlin, Heidelberg, 1999; pp 15–29.

(62) Stevenson, J. D.; Wales, D. J. Communication: Analysing Kinetic Transition Networks for Rare Events. *J. Chem. Phys.* **2014**, *141*, 041104.

(63) MacKay, R. S.; Robinson, J. D. Aggregation of Markov Flows I: Theory. *Philos. Trans. A Math. Phys. Eng. Sci.* **2018**, *376*, 20170232.

(64) Wales, D. J. Decoding the Energy Landscape: Extracting Structure, Dynamics and Thermodynamics. *Philos. Trans. A Math. Phys. Eng. Sci.* **2012**, *370*, 2877–2899.

(65) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know About Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52*, 99–105.

(66) Wales, D. J.; Salamon, P. Observation time scale, free-energy landscapes, and molecular symmetry. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 617–622.

(67) Wales, D. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge Molecular Science; Cambridge University Press, 2004.

(68) Wales, D. J. Coexistence in Small Inert Gas Clusters. *Mol. Phys.* **1993**, *78*, 151–171.

(69) Stillinger, F. H.; Weber, T. A. Packing Structures and Transitions in Liquids and Solids. *Science* **1984**, *225*, 983–989.

(70) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakked, Z. et al. A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry. *J. Mol. Biol.* **2001**, *313*, 229–237.

(71) Zheng, G.; Lu, X.-J.; Olson, W. K. Web 3DNA−A Web Server for the Analysis, Reconstruction, and Visualization of Three-dimensional Nucleic-acid Structures. *Nucleic Acids Res.* **2009**, *37*, W240–W246.