

Energy Landscapes and Hybridization Pathways for DNA Hexamer Duplexes

Shiyao 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

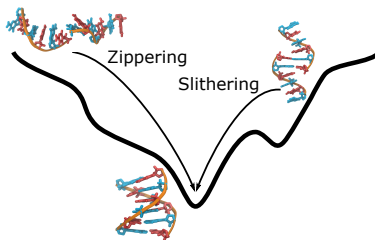
*To whom correspondence should be addressed

[†]University of Cambridge

[‡]The University of Texas at Austin

Abstract

Strand hybridization is not only a fundamental molecular mechanism underlying the biological functions of nucleic acids, but is also a key step in the design of efficient nanodevices. Despite recent efforts, the microscopic rules governing the hybridization mechanisms remain largely unknown. In this study, we exploit the energy landscape framework, to assess how sequence-specificity modulates the hybridization mechanisms in DNA. We find that GG-tracts hybridize much more rapidly compared to GC-tracts, via either zippering or slithering pathways. For the hybridization of GG-tracts, both zippering and slithering mechanisms appear to be kinetically relevant. In contrast, for the GC-tracts, the zippering mechanism is dominant. Our work reveals that even for the relatively small systems considered, the energy landscapes feature multiple metastable states and kinetic traps, which is at odds with the conventional ‘all-or-none’ model of DNA hybridization formulated on the basis of thermodynamic arguments alone. Interestingly, entropic effects are found to play an important role in determining the thermal stability of competing conformational ensembles, and in determining the preferred hybridization pathways.



Hybridization is a fundamental molecular mechanism that underlies both the natural biological functions of DNA^{1,2} and the role of DNA in nanotechnology.^{3,4} Precise control of the kinetics of DNA hybridization at the molecular level is critical to the processing of genetic information, since this is a key step in gene replication and regulation,^{5,6} and in the design of DNA-based molecular machines for varied applications, such as DNA bioassays,^{7,8} DNA computation,^{9,10} and the programmable self-assembly of DNA nanostructures.^{3,11} Although the well-established nearest-neighbour model of base stacking gives some insights into the thermodynamics of DNA hybridization,^{12–15}

the fundamental mechanism and detailed conformational changes driving the process remain poorly understood.¹⁶

The hybridization of DNA oligonucleotides has often been described as an ‘all-or-nothing’ mechanism,¹⁷⁻¹⁹ *i.e.* a first order-like transition. This assumption of two-state dissociated/fully-hybridized behaviour simplifies the analysis of experimental data.²⁰⁻²² However, the two-state model has limitations concerning the hybridization kinetics, and evident deviations from this idealized behaviour have been observed in recent experiments and simulation studies. Molecular dynamics (MD) simulations based on all-atom and coarse-grained potentials provide insight into the pathways of DNA hybridization, and suggest a sequential mechanism, initiated by critical nucleation of a few base pairs, followed by a rapid zippering or internal displacement of the remaining bases.²³⁻²⁵ Direct measurements of the folding of DNA hairpins have verified the existence of intermediate collapsed structures,²⁶⁻²⁸ and indicate that the locations and heights of energy barriers correlate strongly with the number and location of G:C base pairs.²⁹ These and other observations imply that the pathways and energy barriers for DNA hybridization are strongly sequence-dependent.³⁰⁻³⁵ Even for single base-pair formation, diverse folding pathways and multiple kinetic intermediates/traps have been recognized.^{36,37} While previous computational efforts,^{23-25,38,39} in conjunction with experiments, have greatly advanced our understanding of DNA hybridization, we stress that the multitude of conformational transitions, and the corresponding time scales must be naturally understood in the language of energy landscapes,⁴⁰⁻⁴⁵ which necessitates proper characterization of competing minima, and barrier heights. In the present work, we aim to provide these fundamental links.

Using a properly symmetrized AMBER99bsc0 force field⁴⁶ with χ OL4 corrections for DNA,⁴⁷ and generalized Born treatment of the surrounding solvent and counterion atmosphere,^{48,49} we map out the underlying hybridization landscapes for two DNA hexamer sequences, d(GGGGGG) and d(GCGCGC). Further details regarding the sampling strategy, and the theoretical framework are included in the supporting information. We show that even for these short duplexes, the landscapes can be quite

complex, with many metastable states, presenting a challenge for both theory and experiment. The different hybridization mechanisms identified in previous work are encoded in the sequence-dependent landscape topographies. The GG and GC tracts exhibit different propensities to hybridize via zippering or slithering, the two dominant mechanisms of DNA strand association. For GG-tracts, both zippering and slithering mechanisms are kinetically relevant, with comparable rate constants. In contrast, for GC-tracts, slithering mechanisms make a negligible contribution, since slipped structures (misaligned conformations with non-native GC contacts) are deep kinetic traps on the free energy landscape. The occurrence of well-defined metastable states along the hybridization pathways, and non-Arrhenius kinetics strongly hint at deviations from the conventional ‘all-or-none’ hypothesis.

This sequence-specific molecular mechanisms described here should be helpful in understanding the colocalization of short complementary regions of DNA, which is an elementary step in gene replication.^{5,6} Furthermore, they will provide the necessary foundations for understanding various aspects of DNA based computation nanotechnology,^{3,9,10,50} such as the functioning of toehold mediated strand displacement based circuits.⁵⁰

The free energy landscape associated with the hybridization of d(GGGGGG) is depicted in the form of a disconnectivity graph^{51,52} in Figure 1. The corresponding potential energy disconnectivity graph is included in the Supporting Information (Figure S1). The kinetic transition network contains a fully connected component of 146,167 potential energy minima and 184,849 transition states.

Unlike fast folding biomolecules, which exhibit a structure-seeking landscape with low downhill barriers, the free energy landscape for the DNA duplex is somewhat frustrated, with many competing low energy structures, separated by moderately high energy barriers. Nonetheless, there is a weak overall funnelling towards the native state, and there are no large free energy barriers that segregate the landscape, corresponding overall to a downhill free energy gradient and relatively low barriers. Surprisingly, the global free energy minimum is not the canonical WC duplex, but instead a compact

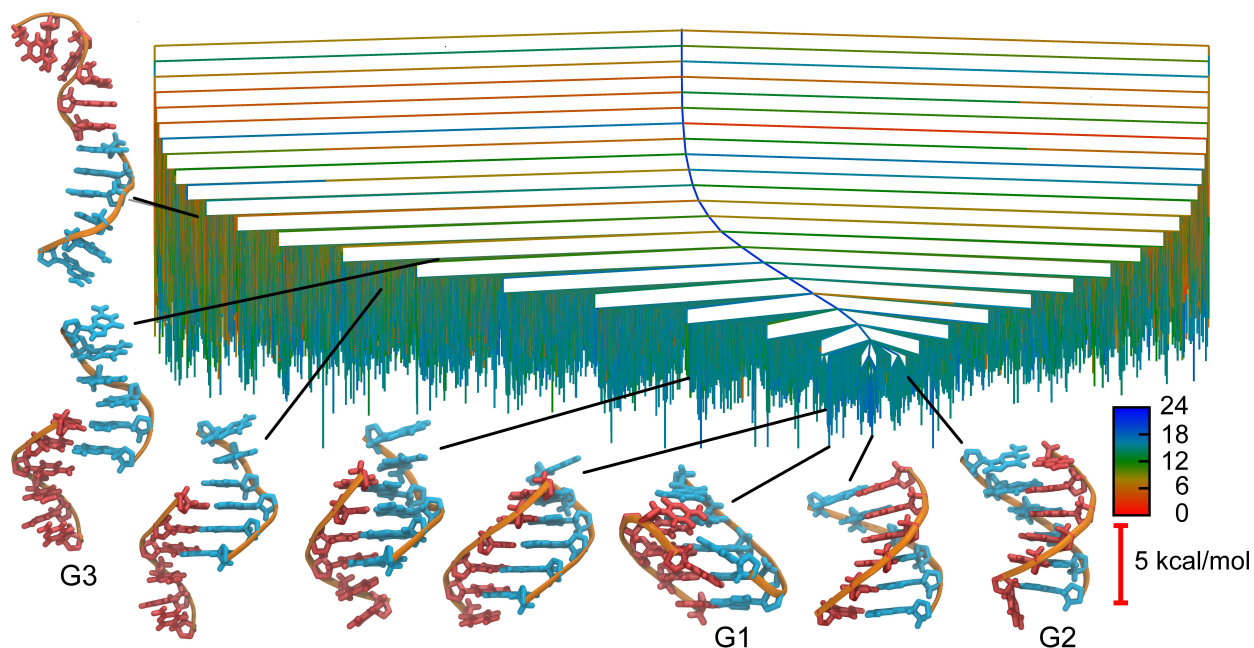


Figure 1: Free energy disconnectivity graph for the hybridization of d(GGGGGG), obtained at a temperature of 298 K and a regrouping threshold^{53,54} of $-10 \text{ kcal mol}^{-1}$. The branches are coloured according to the total number of hydrogen-bonds between the two DNA strands. Some representative minima from the different conformational ensembles are also shown.

structure (G1). The complementary strands of G1 are misaligned by one base compared to the canonical hybridized duplex, and the cytosine and guanine bases of the 5' terminals of the two strands are folded back into the major groove. The base stacking interaction thus formed probably makes a significant contribution to the observed thermal stability of this compact structure. A similar bonding mode has been reported by Ng and Dickerson,⁵⁵ where the terminal base pair is packed in the minor groove through hydrogen-bonding. Analysis of the structural parameters for the central fragment of five base pairs using 3DNA⁵⁶ demonstrates that the G1 conformation is similar to the B-DNA form of the native state, but is characterised by significantly different values of certain helical parameters, such as roll, twist, tilt and slide (Table S1).

The G1 structure is only marginally stable with respect to the native WC duplex, with a free energy difference close to 1 kcal mol^{-1} . In contrast, G1 is around 15 kcal mol^{-1} lower in potential energy than the canonical duplex (Figure S1). The geometric mean of the vibrational normal mode frequencies for the canonical duplex,

calculated within the harmonic superposition approximation (HSA),^{57,58} is much lower than for G1 (Figure S4), which suggests that the WC structure is stabilized by entropy as compared to misaligned and distorted structures. Previous work has shown that the WC duplex tends to exhibit greater conformational plasticity compared to other polymorphs,^{41,59} and suggests a possible role for intrinsic flexibility in modulating DNA-protein binding during gene replication. As shown in Figure 1, there is a diverse range of thermally accessible conformations in the low-energy region of the free energy landscape, including flexible structures with frayed terminal bases (G2), which are relatively unstable in terms potential energy alone (see supplementary information). These observations demonstrate that entropic effects play an important role in modulating the equilibrium behaviour of DNA duplexes at physiological temperatures. Although the high stabilities of some of the compact structures could be an artefact of the force field, this prevalence could be a genuine manifestation in terms of the tendency for guanine bases to form strong stacking interactions, owing to their high aromaticity. The native WC duplex is kinetically accessible from many of the low-free energy conformational ensembles, suggesting that compact, misaligned, and frayed structures can all play a role in the dynamics of hybridization.

The stationary point database for the hybridization of d(GCGCGC) consists of a single connected component of 137,847 potential energy minima and 170,245 transition states. The free energy disconnectivity graph is shown in Figure 2, and the potential energy disconnectivity graph is shown in Figure S2 of the Supporting Information. The landscapes differ in topography from those of d(GGGGGG). In particular, the free energy landscape is more strongly funnelled towards the global minimum, which in this case is the canonical WC duplex. As for d(GGGGGG), there are many thermally accessible conformational ensembles on the free energy landscape. However, in the case of the GC-tract sequence, both the free energy differences between the various low-lying conformations, and the free energy barriers separating them, are generally larger than those for the GG-tract sequence. Strikingly, structures with initial misaligned two- (W5) and four-base pairings (W6) are relatively deep kinetic traps on the landscape,

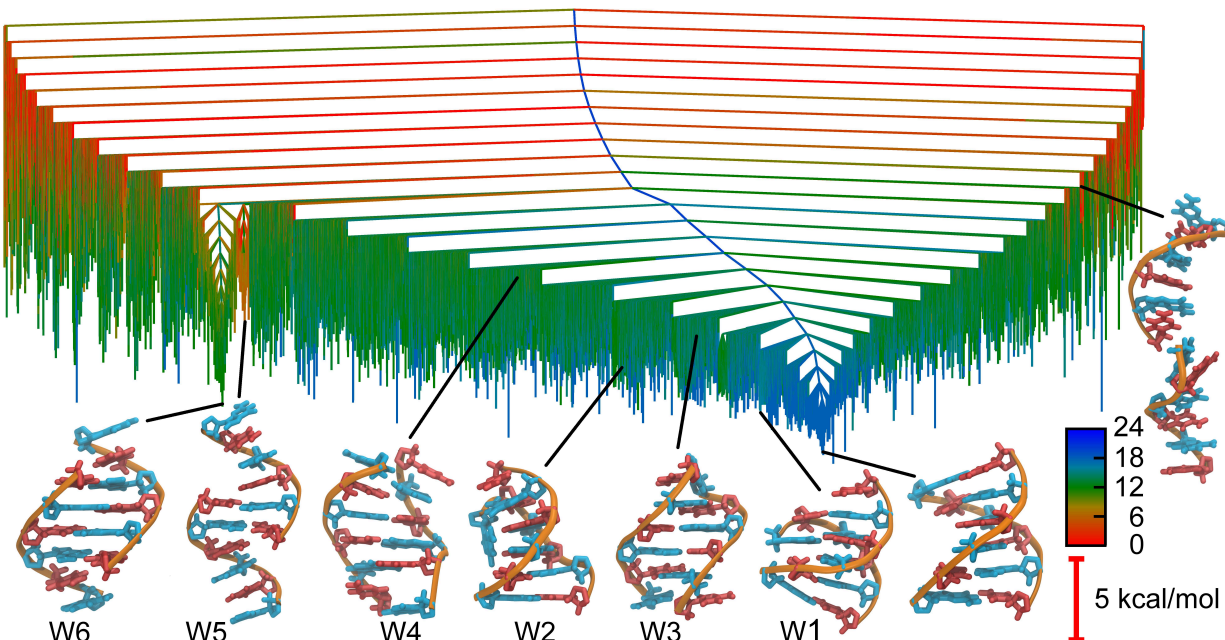


Figure 2: Free energy disconnectivity graph for the hybridization of d(GCGCGC), obtained at a temperature of 298 K and a regrouping threshold^{53,54} of $-10 \text{ kcal mol}^{-1}$. The branches are coloured according to the total number of hydrogen-bonds between the two DNA strands. Some representative local minima from the different conformational ensembles are shown.

separated from the native funnel by free energy barriers of around 10 kcal mol^{-1} . In contrast, for d(GCGCGC), misaligned structures are relatively unstable. This crucial difference between the two sequences arises because, for the alternating GC-tracts, the complementary strands of misaligned structures must slip by two bases, and not one, to retain G:C WC pairings. For both sequences, dissociated states, with no hydrogen-bonds between base pairs, are associated with relatively low equilibrium probabilities.

In the potential energy disconnectivity graph for the hybridization of GC-tracts (Figure S2), two stable compact structures (W1 and W2) can be distinguished. Both of these structures are associated with small funnels on the potential energy landscape, and are lower in potential energy than the canonical duplex. In W1 the guanine base at the 5' terminal is folded back into the major groove, and forms a G*GC triplet with the G5:C8 base pair, resulting in disrupted base stacking. The enhanced twist and roll angles at the triplet site, compared with the ideal B-DNA structure, indicate an interruption of base-pair steps (Table S1 of the Supporting Information). In the W2

structure, the guanine base at the 5' terminus of one strand is directed into the major groove, and forms a stable contact *via* hydrophobic interactions. Similar to the compact G1 structure observed for d(GGGGGG), the W1 and W2 structures have relatively high vibrational frequencies compared to the WC duplex (Figure S4). Hence, again, entropic effects appear to play an important role in stabilizing the canonical WC duplex, as opposed to defective or higher-order structures, at physiological temperatures.^{60,61}

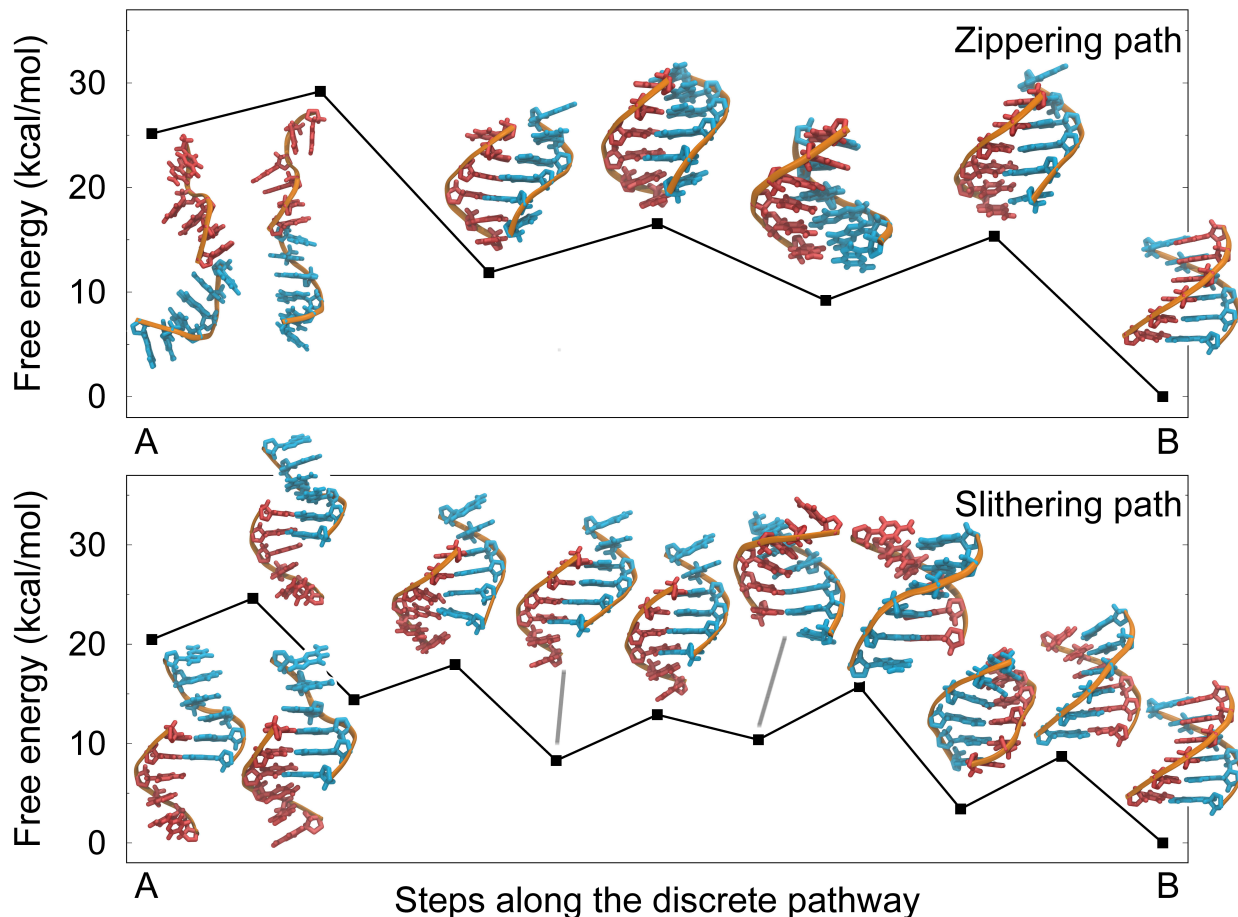


Figure 3: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GGGGGG). The points correspond to groups of minima and transition states, obtained by the recursive regrouping scheme^{53,54} using a barrier threshold of $5.5 \text{ kcal mol}^{-1}$. 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 $3.5 \text{ kcal mol}^{-1}$. These results are for a temperature of 298 K.

To draw mechanistic insights into the renaturation of DNA duplexes from melted

and misaligned states, we apply shortest path algorithms⁶²⁻⁶⁴ with appropriate edge weights,^{65,66} to determine the pathways that make the largest contributions to the steady-state rate constants. We distinguish between two principal mechanisms of DNA hybridization, namely zippering and slithering. The zippering pathway starts from disordered states with no hydrogen-bonds, and forms the duplex through helix propagation after the initial native base-base contacts. The slithering path starts from a misaligned structure with two initial non-native base pairings. The chains then slide along one another along a Watson-Crick interface until the formation of a full helix with all native base-base contacts. The single fastest pathways corresponding to the zippering and slithering mechanisms are shown in Figure 3 and Figure 4 for the GG- and GC-tracts, respectively. The complete sets of fastest pathways corresponding to zippering and slithering mechanisms for the two systems comprise many similar alternatives in the path ensembles, and some representative trajectories are shown in Figure S5 and Figure S6 of the Supporting Information. The profiles of the fastest pathways for all mechanisms and sequences considered show that the hybridization process is essentially downhill in energy. We also find that the steps in each particular hybridization pathways are largely dictated by the nature of the initial contact between the complementary strands.

For GG-tracts, the fastest zippering pathway (Figure 3) proceeds *via* initial fast association of the three guanine bases at the 3' terminal of one strand to the central three cytosine bases of the complementary strand. The free energy barrier for this step is small (less than 5 kcal mol⁻¹) and the resulting misaligned structure is significantly thermodynamically and kinetically stable with respect to the dissociated structure. These observations are attributable to the fact that, in this associated state, the ends of both strands are dangling, and hence the entropy remains relatively high. Thus, although statistically, collision occurs most frequently at the termini of both strands, there is a lower free energy barrier to nucleation if the initial contact involves the centre of one strand. This effect was also noted in a previous study by Louis and coworkers.²³ After the rapid nucleation step, both DNA strands undergo a succession of ordered

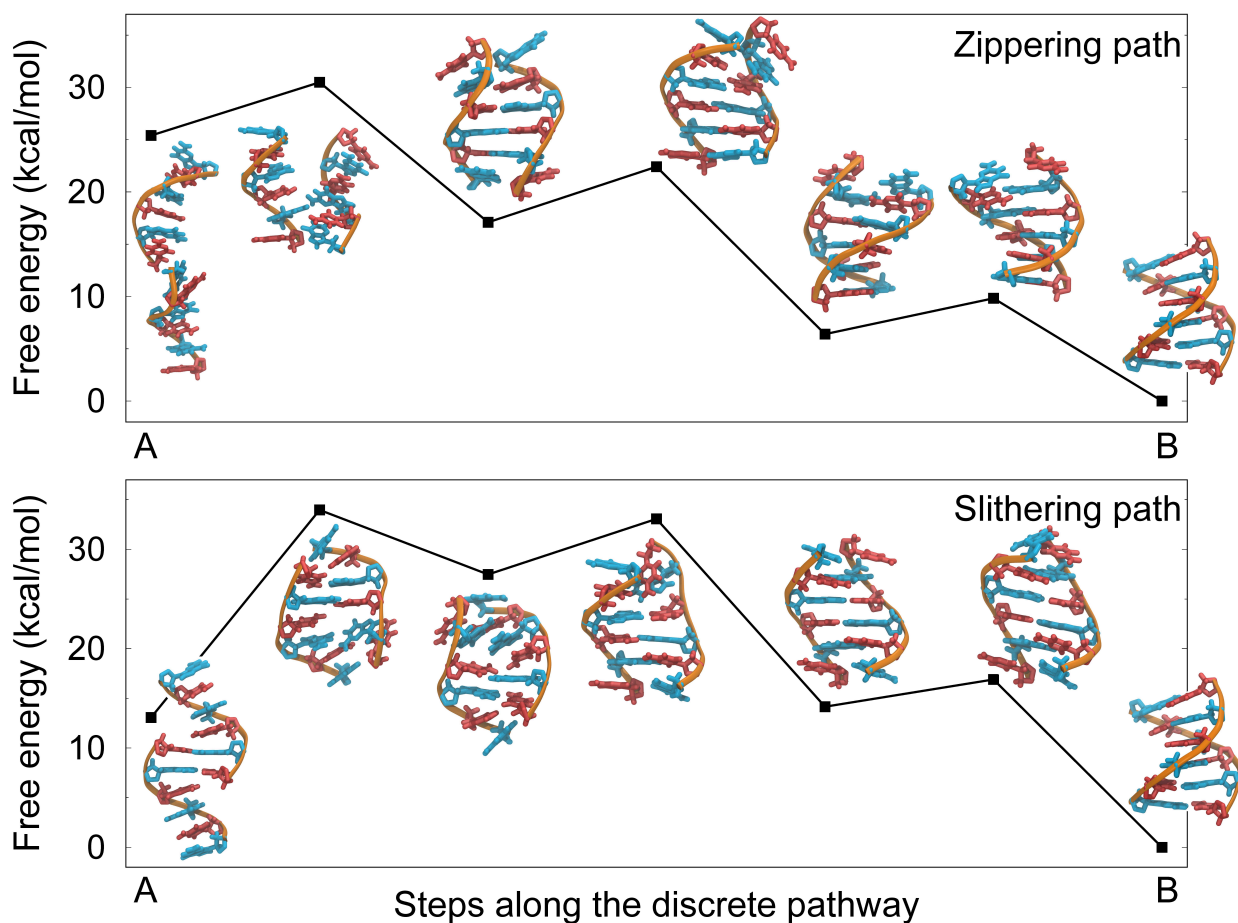


Figure 4: Top: The fastest free energy zippering pathway between the conformational ensembles corresponding to dissociated (A) and fully-hybridized (B) states for d(GCGCGC). The points correspond to free energy groups of minima and transition states, obtained by the recursive regrouping scheme^{53,54} 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. These results are for a temperature of 298 K.

slithering rearrangements, associated with moderate energy barriers, to achieve the native base pairings. The last step in the formation of the WC duplex is the zippering of the termini. Alternative zippering pathways, where the formation of base pairs proceeds in a strict order from one terminus to the other (Figure S8), make a much smaller contribution to the steady-state rate constant. This result is in agreement with recent infrared spectroscopy experiments on the reverse process, which show that the dissociation of the duplex begins with the fraying of the terminal base pairs.³¹

For GC-tracts, the fastest zippering pathway (Figure 4) follows a broadly similar mechanism, with fast initial nucleation of a few key contacts, followed by slow zippering of the remaining base pairs. Whereas for the homo-duplex, the formation of initial base pairings along the fastest path results in a misaligned structure, for the heteroduplex, it is strongly favourable for the initial contacts to be native base pairings. This is because, for GC-tracts, the slipping of misaligned structures by a single base leads to mismatched G:G and C:C pairings, which are highly unstable. Hence strand slippage by two bases is required to reach a metastable intermediate state, and consequently there are high energy barriers associated with slithering (Figure S6). Indeed, we see that misaligned structures constitute prominent kinetic traps on the free energy landscape for d(GCGCGC) (Figure 2). Thus, the alternating sequence of the GC-tracts behaves similarly to a random sequence, for which productive hybridization pathways are restricted to those where the nucleation event involves key native base pairings.³⁰ In the fastest zippering pathway, the complementary strands of the GC-tracts first form native base pairs at one terminus, followed by the propagation of hydrogen-bonding to the opposite terminus of the duplex, in strict order (Figure S6). To remove the local kinetic traps in stationary databases along the hybridization pathway, we grouped the stationary and transition points using the recursive regrouping scheme^{53,54} with a barrier threshold of 4.5 kcal mol⁻¹ (Figure 4). The fastest zippering pathway for the grouped d(GCGCGC) database is slightly different from that without grouping, involving initial formation of native base pairs at the strand centre, and the remaining base pairs are formed by simultaneous propagation of hydrogen-bonding outwards to

both termini. We find that the normal mode frequencies are typically higher, producing lower entropy for structures with a three native base pair contact involving the centres of the two strands, than for structures with three native base pairs involving the 5' ends of both strands. Although initial contacts form preferentially at the end of the strands,²³ owing to their higher flexibility and entropy, intermediate states with frayed terminal bases are entropically disfavoured to avoid local kinetic traps, and can readily rearrange to the native WC duplex, thereby facilitating hybridization. The differences between the fastest pathways with and without regrouping indicate the important role of entropic effects in determining the mechanism of DNA hybridization.

We also consider slithering pathways for the two systems, starting from conformations with two non-native G:C pairings, involving the 5' ends of both strands. In the first step to renaturation, the end base pairings are broken, and new hydrogen-bonds form with the nearest free complementary base, resulting in a bulge loop of one (for GG-tracts) or two (for GC-tracts) bases. This strand slippage transition is repeated until the canonical WC duplex is formed. The slithering mechanism we have characterised is different from the 'defect diffusion' proposed in previous work^{24,67} and from the 'internal displacement' and 'inchworm' mechanisms²³ observed in coarse-grained simulations with longer sequences of more than ten base pairs. The introduction of defects in short oligonucleotides is probably less favourable.³⁹ Additionally, the coarse-grained model has a single, relatively strong hydrogen-bonding interaction between base pairs, which may suppress slithering. Hydrogen-bonds in the atomistic model are individually weaker and probably have a lower orientational preference, which facilitates sliding strands past each other in a process 'devoid of significant energy barriers'.²³ Interestingly, using a high resolution CG model that incorporates multiple sites to describe base-pairing, Scheraga and coworkers²⁵ did not see any evidence of 'internal displacement' or 'inchworm' mechanisms in the hybridization pathways of a 12mer DNA duplex, and reported that strand association primarily occurs via zippering or slithering. Therefore, it seems likely that the interplay of various factors, such as duplex length, resolution of the model, and sequence-specificity could dictate the

occurrence of alternate hybridization mechanisms than those reported here.

The free energy barrier for a single slippage step in GG-tracts is relatively small, around 5 kcal mol^{-1} , and hence the overall slithering mechanism is competitive with the zippering mechanism. In contrast, the free energy barrier for a single strand slippage step in GC-tracts is much larger, around 20 kcal mol^{-1} . This barrier significantly exceeds the thermal energy available at physiological temperature. The highest barrier in the slithering pathway of GC-tracts is the rearrangement of the slipped strand to the canonical base pairing, in which the complementary strands of misaligned structures slip by two bases to retain G:C WC pairings. Hence, for heteroduplexes, the overall slithering mechanism is not kinetically relevant, and the zippering mechanism is predicted to dominate. The prediction of a single dominant mechanism for heteroduplexes agrees with recent MD simulations.⁶⁸

A path deviation algorithm^{69,70} was used to identify the complete set of kinetically distinct pathways without regrouping the database (Figure S7-S12). We find that there are distinguishable pathway ensembles for hybridization by zippering and slithering in the kinetic transition networks. In general, slower pathways are characterized by the formation of collapsed structures, stabilized by stacking interactions, which must undergo slow intrachain diffusion to achieve the native base pairings.²⁶ Thus our results demonstrate that, even for short sequences, DNA hybridization is a complex process exhibiting rich dynamics, with multiple transition pathways involving many metastable basins on the free energy landscape. These observations, consistent with recent experimental and simulation studies,^{23,24,26,31} suggest significant deviations from the ideal ‘all or nothing’ model of DNA hybridization,¹⁷⁻¹⁹ which assumes a simple two-state dissociated/fully-hybridized behaviour based on thermodynamic arguments. We note that the free energy barriers for escaping some of the intermediate states along the most kinetically relevant pathways are relatively small, less than or around 5 kcal mol^{-1} . This result explains why partially hybridized structures have rarely been detected in previous experiments.^{31,32}

The graph transformation method⁷¹ was used to compute phenomenological rate

constants for DNA hybridization at a temperature of 298 K, after using a self-consistent regrouping scheme⁵³ to lump minima separated by barriers less than a threshold of 5 kcal mol⁻¹. The rate constants for the zippering pathways are estimated to be 7.5×10^4 s⁻¹ and 2.3×10^2 s⁻¹ for the GG-tracts and GC-tracts, respectively. The rate constants for the slithering pathways are estimated as 10^5 s⁻¹ for GG-tracts, and 9.5×10^{-8} s⁻¹ for GC-tracts. Thus the timescales for the hybridization of GG-tracts, *via* either zippering or slithering mechanisms, are many orders of magnitude faster than for GC-tracts. This contrasting trend for the hybridization of homo and hetero-duplexes agrees with the experiments of Wyer *et al.*,³³ who demonstrated that the timescale for the hybridization of AT-tracts is significantly longer than for AA-tracts. The faster hybridization kinetics of GG-tracts compared to GC-tracts also agrees with predictions based on sequence information using a six-feature weighted neighbour voting (WNV) model.⁷² Furthermore, for GG-tracts, both zippering and slithering mechanisms are kinetically relevant, with comparable rate constants. In contrast, for GC-tracts, slithering mechanisms make a negligible contribution, since slipped structures are deep kinetic traps on the free energy landscape (Figure 2).

The variation of rate constants with temperature is depicted in the form of Arrhenius plots for both zippering and slithering mechanisms of each system (Figure S13-S14). For both GG-tracts and GC-tracts, the apparent activation energy (E_a) of the forward reaction from dissociated or slipped structures to duplex is much lower than for the backward reaction. The zippering and slithering mechanisms of GG-tracts have comparable values of E_a around 10 kcal mol⁻¹, which are significantly smaller than for GC-tracts (around 20 kcal mol⁻¹), verifying the faster hybridization kinetics of the homo-duplex. Although the kinetics of DNA hybridization can be quantified by fitting the phenomenological rate constant to an Arrhenius form, the temperature dependence exhibits significant deviations from Arrhenius behaviour for zippering and slithering in GG-tracts and for zippering in GC-tracts. These deviations suggest that DNA hybridization is associated with intermediates and kinetic traps, rather than a single well-defined transition state. The off-pathway traps of compact structures (such

as, G1, W1 and W2), and on-path intermediate states (such as misaligned and frayed structures) represent competing free energy minima in addition to the native WC duplex. Moreover, the distinct competing pathway ensembles of DNA hybridization increase the deviation from Arrhenius kinetics. Although a two-state interpretation of the thermodynamics of DNA hybridization has proved useful, a multistate kinetic model seems more appropriate, especially in view of the organisation of the free energy landscapes, and the non-Arrhenius behaviour.

Our results also feature populated configurations with fraying of the helical termini, which have been suggested by recent experiments using nonlinear infrared (IR) spectroscopy.^{31,32} For DNA helix, the dimer ensemble has competing configurations separated by barriers smaller than the enthalpy penalty required for dissociation of the central base pairings. The stable fraying of the helical termini, in addition to the pathways for dimer separation, corresponds to a multi-step process, and hence deviations from Arrhenius behavior in duplex dissociation.³²

To clarify the mechanistic changes at different temperatures, we calculated the fastest pathways corresponding to the zippering and slithering mechanisms at 270 K and 350 K for GG- (figures S15 and S16) and GC-tracts (figures S17 and S18). In both cases slithering follows essentially the same pathway. For zippering, the free energy is essentially downhill at low temperature, but we see barriers at the higher temperature for both sequences. The highest free energy barrier along the path from dissociated strands to duplex at 350 K corresponds to the initial association and the nucleation of a few base pairs between the strands. It is also interesting that, for GC-tracts, the zippering starts with the initial contact between the central three base pairs at low temperature, whereas the strand termini first forms base pairs, followed by propagation from one end to the other at the higher temperature. In the experiments of Chen *et al.*, the apparent hybridization energy changes from negative at high temperature to positive at low temperature.⁷³ Meanwhile, Pröschke and co-workers^{74,75} reported positive hybridization energies of around 10 kcal/mol for duplexes containing GC base pairs at temperatures around T_m . Our results are in overall agreement with these

experimental observations.

In summary, we have employed the computational energy landscape framework^{40,42–45} to study the dynamics and thermodynamics of DNA hybridization in hexamer duplexes at an atomistic level of detail. DNA renaturation begins with rapid initial nucleation of a few contacts, followed by complex structural rearrangements *via* zippering or slithering pathways to reach the native Watson-Crick duplex. Multiple metastable intermediates exist along the pathways, separated by relatively high barriers, indicating significant deviations from the ideal ‘all-or-nothing’ two-state model. The preferred hybridization mechanism is strongly sequence-dependent, with significantly different behaviour for homo- and hetero-duplexes. For the homogeneous GG-tracts, the most kinetically relevant pathway begins with the formation of base pairs involving the centre of one strand and the terminus of the complementary strand. This misaligned structure then rearranges to the WC duplex by coupled zippering and slithering motions. In contrast, the hybridization of GC-tracts occurs predominantly by a zippering pathway, in which the formation of native base pairs is propagated from one terminus to the other in strict order, and pathways involving slithering make a negligible contribution to the steady-state rate constant. The hybridization of GG-tracts is much faster than for GC-tracts, since misaligned states are deep kinetic traps on the free energy landscape of the heteroduplex. The sequence-dependent kinetics of DNA hybridization are critical in modulating the rates of DNA strand exchange reactions in DNA computation.^{3,9,10,50}

The free energy landscapes for hybridization are relatively frustrated, with many distinct conformational ensembles populated at equilibrium. Besides the canonical Watson-Crick duplex, off-pathway compact, on-pathway distorted structures, and flexible structures with frayed terminal base pairs exist in the low-lying region of the landscape. Frayed structures are kinetically accessible from the native state, and *vice versa*, and appear to play an important role in facilitating the renaturation and denaturation processes. Analysis of the landscapes and pathways demonstrates that entropic effects are important in determining both the equilibrium and dynamic features of the

hybridization process.

The hexamer duplex sequences considered in this study are an ideal length for efficient exploration of the conformational space using the energy landscape framework,^{40,42,43} revealing various facets of strand hybridization in atomistic level. Longer sequences may support additional pathways and structures, such as the ‘pseudoknot’,^{23,24} and we will investigate these possibilities in future work. Starting from shorter sequences will enable us to characterise the feasibility of different mechanisms as a function of system size. This insight will help to inform the rational design of DNA nanodevices and probes, and also lead to a better understanding of the genetic machinery that is sustained by strand hybridization.

Acknowledgement

SX, DC and DJW acknowledge funding from the Engineering and Physical Sciences Research Council UK. DJS gratefully acknowledges the Cambridge Commonwealth, European and International Trust for a PhD scholarship.

Supporting Information Available

Details regarding the exploration of the energy landscapes, analysis, and visualization, and the calculations of free energies and interconversion rates; potential energy disconnectivity graph; representative structures; distribution of the log product of normal mode frequencies; the potential energy as a function of the integrated path length (s) for hybridization paths; the identification of distinct pathways; variations of the rate constants as a function of reciprocal temperature; fastest pathway analysis at low and high temperatures; base pair step parameters. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Bell, S. P.; Dutta, A. DNA Replication in Eukaryotic Cells. *Ann. Rev. Biochem.* **2002**, *71*, 333–374.
- (2) Mandal, M.; Breaker, R. R. Gene Regulation by Riboswitches. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451–463.
- (3) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. Programmable Materials and the Nature of the DNA Bond. *Science* **2015**, *347*, 1260901.
- (4) Chen, Y.-J.; Groves, B.; Muscat, R. A.; Seelig, G. DNA Nanotechnology From the Test Tube to the Cell. *Nat. Nanotechnol.* **2015**, *10*, 748–760.
- (5) Yu, D.; Pendergraff, H.; Liu, J.; Kordasiewicz, H. B.; Cleveland, D. W.; Swayze, E. E.; Lima, W. F.; Crooke, S. T.; Prakash, T. P.; Corey, D. R. Single-Stranded RNAs Use RNAi to Potently and Allele-Selectively Inhibit Mutant Huntingtin Expression. *Cell* **2012**, *150*, 895–908.
- (6) Lima, W. F.; Prakash, T. P.; Murray, H. M.; Kinberger, G. A.; Li, W.; Chappell, A. E.; Li, C. S.; Murray, S. F.; Gaus, H.; Seth, P. P. et al. Single-Stranded siRNAs Activate RNAi in Animals. *Cell* **2012**, *150*, 883–894.
- (7) Khodakov, D.; Wang, C.; Zhang, D. Y. Diagnostics Based on Nucleic Acid Sequence Variant Profiling: PCR, Hybridization, and NGS Approaches. *Adv. Drug. Deliv. Rev.* **2016**, *105*, 3–19.
- (8) Zhao, Y.; Chen, F.; Li, Q.; Wang, L.; Fan, C. Isothermal Amplification of Nucleic Acids. *Chem. Rev.* **2015**, *115*, 12491–12545.
- (9) Sun, X.; Wei, B.; Guo, Y.; Xiao, S.; Li, X.; Yao, D.; Yin, X.; Liu, S.; Liang, H. A Scalable “Junction Substrate” to Engineer Robust DNA Circuits. *J. Am. Chem. Soc.* **2018**, *140*, 9979–9985.

- (10) Woods, D.; Doty, D.; Myhrvold, C.; Hui, J.; Zhou, F.; Yin, P.; Winfree, E. Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly. *Nature* **2019**, *567*, 366–372.
- (11) Thubagere, A. J.; Li, W.; Johnson, R. F.; Chen, Z.; Doroudi, S.; Lee, Y. L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D. et al. A Cargo-Sorting DNA Robot. *Science* **2017**, *357*, eaan6558.
- (12) SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
- (13) Dirks, R.; Bois, J.; Schaeffer, J.; Winfree, E.; Pierce, N. Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev.* **2007**, *49*, 65–88.
- (14) Wang, C.; Bae, J. H.; Zhang, D. Y. Native Characterization of Nucleic Acid Motif Thermodynamics via Non-covalent Catalysis. *Nat. Commun.* **2016**, *7*, 10319.
- (15) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **1999**, *288*, 911–940.
- (16) Zhang, J. X.; Fang, J. Z.; Duan, W.; Wu, L. R.; Zhang, A. W.; Dalchau, N.; Jordanov, B.; Petersen, R.; Phillips, A.; Zhang, D. Y. Predicting DNA Hybridization Kinetics From Sequence. *Nat. Chem.* **2018**, *10*, 91–98.
- (17) Kafri, Y.; Mukamel, D.; Peliti, L. Why is the DNA Denaturation Transition First Order? *Phys. Rev. Lett.* **2000**, *85*, 4988–4991.
- (18) Garel, T.; Monthus, C.; Orland, H. A Simple Model for DNA Denaturation. *Europhysics Letters (EPL)* **2001**, *55*, 132–138.
- (19) Carlon, E.; Orlandini, E.; Stella, A. L. Roles of Stiffness and Excluded Volume in DNA Denaturation. *Phys. Rev. Lett.* **2002**, *88*, 198101.

- (20) Mergny, J.-L.; Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **2003**, *13*, 515–537.
- (21) SantaLucia, J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics. *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 1460–1465.
- (22) Rauzan, B.; McMichael, E.; Cave, R.; Sevcik, L. R.; Ostrosky, K.; Whitman, E.; Stegemann, R.; Sinclair, A. L.; Serra, M. J.; Deckert, A. A. Kinetics and Thermodynamics of DNA, RNA, and Hybrid Duplex Formation. *Biochemistry* **2013**, *52*, 765–772.
- (23) Ouldrige, T. E.; Šulc, P.; Romano, F.; Doye, J. P. K.; Louis, A. A. DNA Hybridization Kinetics: Zippering, Internal Displacement and Sequence Dependence. *Nucleic Acids Res.* **2013**, *41*, 8886–8895.
- (24) Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J. Coarse-grained Modeling of DNA Oligomer Hybridization: Length, Sequence, and Salt Effects. *J. Chem. Phys.* **2014**, *141*, 035102.
- (25) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. DNA Duplex Formation with a Coarse-grained Model. *J. Chem. Theory Comput.* **2014**, *10*, 5020–5035.
- (26) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. DNA Folding and Melting Observed in Real Time Redefine the Energy Landscape. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 712–716.
- (27) Narayanan, R.; Zhu, L.; Velmurugu, Y.; Roca, J.; Kuznetsov, S. V.; Prehna, G.; Lapidus, L. J.; Ansari, A. Exploring the Energy Landscape of Nucleic Acid Hairpins Using Laser Temperature-Jump and Microfluidic Mixing. *J. Am. Chem. Soc.* **2012**, *134*, 18952–18963.

- (28) Abdollah-Nia, F.; Gelfand, M. P.; Van Orden, A. Three-State DNA Hairpin Conformational Dynamics Revealed by Higher-Order Fluorescence Correlation Spectroscopy. *J. Phys. Chem. B* **2019**, *123*, 1491–1504.
- (29) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **2006**, *314*, 1001–1004.
- (30) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. Uncovering Pathways in DNA Oligonucleotide Hybridization via Transition State Analysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18125–18130.
- (31) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved Through Infrared Spectroscopy. *J. Am. Chem. Soc.* **2016**, *138*, 11792–11801.
- (32) Sanstead, P. J.; Tokmakoff, A. Direct Observation of Activated Kinetics and Downhill Dynamics in DNA Dehybridization. *J. Phys. Chem. B* **2018**, *122*, 3088–3100.
- (33) Wyer, J. A.; Kristensen, M. B.; Jones, N. C.; Hoffmann, S. V.; Nielsen, S. B. Kinetics of DNA Duplex Formation: A-tracts Versus AT-tracts. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18827–18839.
- (34) Araque, J. C.; Panagiotopoulos, A. Z.; Robert, M. A. Lattice Model of Oligonucleotide Hybridization in Solution. I. Model and Thermodynamics. *J. Chem. Phys.* **2011**, *134*, 165103.
- (35) Araque, J. C.; Robert, M. A. Lattice Model of Oligonucleotide Hybridization in Solution. II. Specificity and Cooperativity. *J. Chem. Phys.* **2016**, *144*, 125101.
- (36) Xu, X.; Yu, T.; Chen, S.-J. Understanding the Kinetic Mechanism of RNA Single Base Pair Formation. *Proc. Natl. Acad. Sci. U S A* **2016**, *113*, 116–121.

- (37) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. Atomistic Understanding of Kinetic Pathways for Single Base-Pair Binding and Unbinding in DNA. *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 13922–13927.
- (38) Markegard, C. B.; Gallivan, C. P.; Cheng, D. D.; Nguyen, H. D. Effects of Concentration and Temperature on DNA Hybridization by Two Closely Related Sequences via Large-Scale Coarse-Grained Simulations. *J. Phys. Chem B.* **2016**, *120*, 7795–7806.
- (39) 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.
- (40) Joseph, J. A.; Roder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring Biomolecular Energy Landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.
- (41) 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.
- (42) Wales, D. J. Exploring Energy Landscapes. *Ann. Rev. Phys. Chem.* **2018**, *69*, 401–425.
- (43) Wales, D. J. Decoding the Energy Landscape: Extracting Structure, Dynamics and Thermodynamics. *Trans. Phil. Roy. Soc. A* **2012**, *370*, 2877–2899.
- (44) Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.
- (45) 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.

- (46) 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. *Biophys. J.* **2007**, *92*, 3817–3829.
- (47) 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.
- (48) 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.
- (49) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (50) Zhang, D. Y.; Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **2009**, *131*, 17303–17314.
- (51) 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.
- (52) 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.
- (53) 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.
- (54) Carr, J. M.; Wales, D. J. Refined Kinetic Transition Networks for the GB1 Hairpin Peptide. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3341–3354.

- (55) Ng, H.-L.; Dickerson, R. E. Mediation of the A/B-DNA Helix Transition by G-tracts in the Crystal Structure of Duplex CATGGGCCCATG. *Nucleic Acids Res.* **2002**, *30*, 4061–4067.
- (56) 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.
- (57) Stillinger, F. H.; Weber, T. A. Packing Structures and Transitions in Liquids and Solids. *Science* **1984**, *225*, 983–989.
- (58) 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.
- (59) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. Theoretical Study of a New DNA structure: The Antiparallel Hoogsteen Duplex. *J. Am. Chem. Soc.* **2003**, *125*, 14603–14612.
- (60) Cheng, Y. K.; Pettitt, B. M. Hoogsteen Versus Reversed-Hoogsteen Base Pairing: DNA Triple Helices. *J. Am. Chem. Soc.* **1992**, *114*, 4465–4474.
- (61) Goldsmith, G.; Rathinavelan, T.; Yathindra, N. Selective Preference of Parallel DNA Triplexes Is Due to the Disruption of Hoogsteen Hydrogen Bonds Caused by the Severe Nonisostericity Between the G*GC and T*AT Triplets. *PLOS ONE* **2016**, *11*, 1–29.
- (62) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269–271.
- (63) Evans, D. A.; Wales, D. J. Folding of the GB1 Hairpin Peptide From Discrete Path Sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090.
- (64) Sharpe, D. J.; Wales, D. J. Identifying Mechanistically Distinct Pathways in Kinetic Transition Networks. *J. Chem. Phys.* **2019**, *151*, 124101.

- (65) Wales, D. J. Some Further Applications of Discrete Path Sampling to Cluster Isomerization. *Mol. Phys.* **2004**, *102*, 891–908.
- (66) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.
- (67) Pörschke, D. Model Calculations on the Kinetics of Oligonucleotide Double Helix Coil Transitions. Evidence for a Fast Chain Sliding Reaction. *Biophys. Chem.* **1974**, *2*, 83–96.
- (68) Jin, R.; Maibaum, L. Mechanisms of DNA Hybridization: Transition Path Analysis of a Simulation-informed Markov Model. *J. Chem. Phys.* **2019**, *150*, 105103.
- (69) Frigioni, D.; Marchetti-Spaccamela, A.; Nanni, U. Fully Dynamic Algorithms for Maintaining Shortest Paths Trees. *J. Algorithms* **2000**, *34*, 251–281.
- (70) 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.
- (71) Wales, D. J. Calculating Rate Constants and Commitor Probabilities for Transition Networks by Graph Transformation. *J. Chem. Phys.* **2009**, *130*, 204111.
- (72) Zhang, J. X.; Fang, J. Z.; Duan, W.; Wu, L. R.; Zhang, A. W.; Dalchau, N.; Jordanov, B.; Petersen, R.; Phillips, A.; Zhang, D. Y. Predicting DNA Hybridization Kinetics From Sequence. *Nat. Chem.* **2018**, *10*, 91–98.
- (73) Chen, C.; Wang, W.; Wang, Z.; Wei, F.; Zhao, X. S. Influence of Secondary Structure on Kinetics and Reaction Mechanism of DNA Hybridization. *Nucleic Acids Res.* **2007**, *35*, 2875–2884.
- (74) Pörschke, D.; Eigen, M. Co-operative Non-enzymatic Base Recognition III. Kinetics of the Helix-coil Transition of the Oligoribouridylic · Oligoriboadenylic Acid System and of Oligoriboadenylic Acid Alone at Acidic PH. *J. Mol. Biol.* **1971**, *62*, 361 – 381.

- (75) Pörschke, D.; Uhlenbeck, O. C.; Martin, F. H. Thermodynamics and Kinetics of the Helix-coil Transition of Oligomers Containing GC Base Pairs. *Biopolymers* **1973**, *12*, 1313–1335.