

Energy Landscape and Pathways for Transitions Between Watson-Crick and Hoogsteen Base Pairing in DNA

Debayan Chakraborty^{*,†,‡} and David J. Wales^{*,†}

*†Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2
1EW, United Kingdom*

*‡Department of Chemistry, The University of Texas at Austin, 24th Street Stop A5300,
Austin, Texas 78712, USA*

E-mail: dc550@cam.ac.uk; dw34@cam.ac.uk

Abstract

The recent discovery that Hoogsteen (HG) base pairs are widespread in DNA across diverse sequences, and positional contexts, could have important implications for understanding DNA replication and DNA-protein recognition. While evidence is emerging that the Hoogsteen conformation could be a thermodynamically accessible conformation of the DNA duplex, and provide a means to expand its functionality, relatively little is known about the molecular mechanism underlying the Watson-Crick to Hoogsteen transition. In this Perspective, we describe pathways and kinetics for this transition at an atomic level of detail, using the energy landscape perspective. We show that competition between the duplex conformations results in a double funnel landscape, which explains some recent experimental observations. The interconversion pathways feature a number of intermediates, with a variable number of Watson-Crick and Hoogsteen base pairs. The relatively slow kinetics, with possible deviations from two-state behavior, suggest that this conformational switch is likely to be a challenging target for both simulation and experiment.

DNA is one of the fundamental building blocks of life. Inside cells, DNA serves as the storehouse of genetic information, and transfers that information from one generation to the next, thereby providing a molecular basis for heredity and inheritance.¹⁻³ DNA is not a static molecule, and its diverse range of functions are often accomplished via highly orchestrated motions of the helix, or large scale deviations from the canonical helical structure.⁴⁻⁸ The deformability of the DNA duplex is a key factor in determining nucleosome positioning,^{9,10} packaging of DNA within the cell nucleus,^{11,12} and sequence-specific recognition by proteins and ligands.¹³⁻¹⁵ There is a growing database highlighting the rich polymorphism exhibited by DNA, and strong dependence of the equilibrium structure of the double helix on sequence and environment.^{3,16} Repetitive sequences, in particular, have a high propensity to exhibit different duplex geometries,^{3,16} and assemble into higher order structures, such as triplexes¹⁷ and quadruplexes.¹⁸ Under specific environmental conditions, poly(G-C) sequences form Z-DNA, a left-handed duplex consisting of canonical Watson-Crick (WC)

base pairs.¹⁹ Sequences rich in A-T repeats are known to exist in different polymorphic forms. For example, many crystal structures of A-T oligomers exhibit a B-type conformation, while poly(A-T) fibers can undergo a transition to C-type or D-type conformations, at low hydration and temperature conditions, respectively.^{20,21}

Many of the alternate helical forms, and higher order structures, are often stabilized by base pairs that exhibit a different mode of hydrogen bonding from the original WC scheme. One such hydrogen bonding pattern, which was first reported by Hoogsteen in the 1960s from X-ray crystallography,²² results from the flipping of the purine base around the glycosidic torsion, changing an *anti* to *syn* orientation. The Hoogsteen (HG) base pairing mode is now known to be quite common for the A-T base pair.^{23,24} In fact, previous computational studies²⁵⁻²⁷ suggest that the WC pairing may not represent the optimal mode of interaction for A-T, and the HG orientation can sometimes be energetically more favorable. HG pairing is less common for the G-C base pair, as it requires protonation of the cytosine.^{23,28} The switch from the canonical WC to the HG orientation substantially modifies the chemical environment around the base pair in question, which can have major implications for DNA-protein recognition,^{29,30} damage repair,^{31,32} and replication.^{33,34} A prominent example includes the replication of DNA by Human DNA polymerase.³⁵⁻³⁷ In two separate studies, Satya Prakash and coworkers^{35,36} provide structural and biochemical evidence suggesting a replication mechanism dominated by HG base pairing. The authors argue that this mechanism helps to minimize interference during replication, circumventing lesions that may appear on the WC edge.

The realization that HG base pairing may provide an energetically viable route for key biological processes has fueled great interest in understanding the relative abundance of HG base pairs in diverse sequence and positional contexts,³⁸ as well as the molecular mechanism underlying the WC to HG switch.³⁹⁻⁴² A systematic survey of the DNA crystal structures deposited in the PDB database reveals that the repertoire of HG bonds is more extensive than previously thought, and the formation of HG base pairs can induce significant DNA

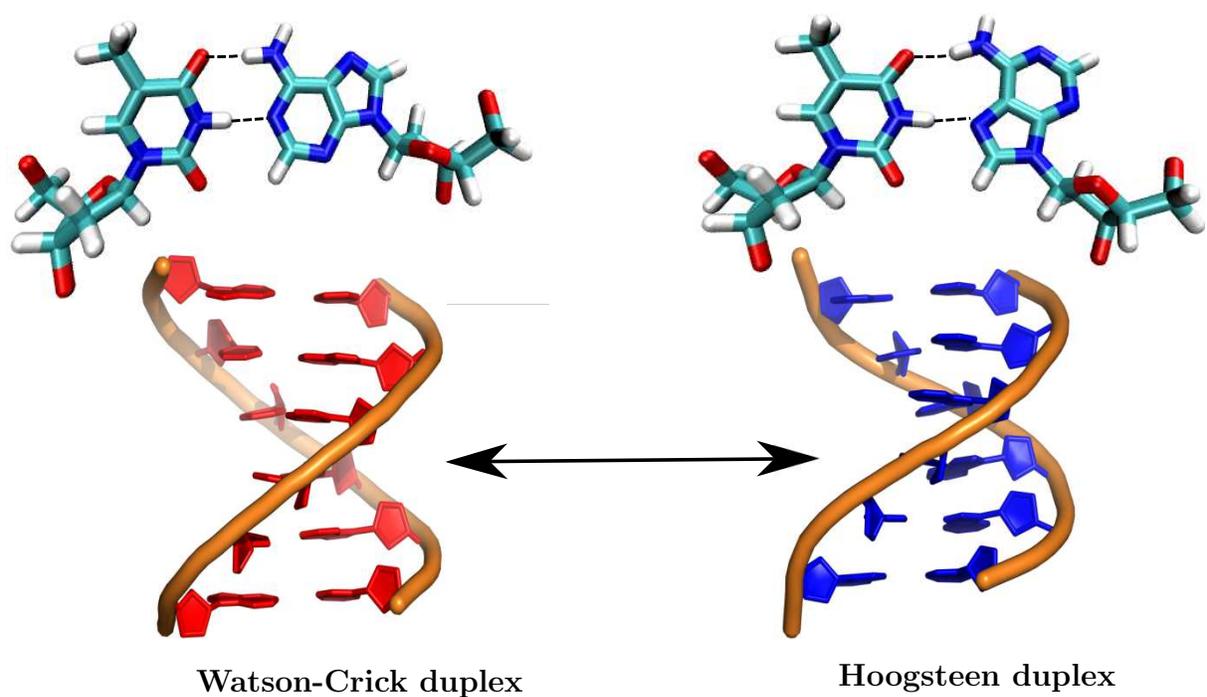


Figure 1. Representative structures for the Watson-Crick (WC) and Hoogsteen (HG) duplexes considered in this study. The WC base pairs are colored red, and the HG base pairs are colored blue. This coloring scheme is used in all the following figures. Also shown are the WC and HG orientations for the adenine-thymine base pair.

bending.²⁴ In another study, Al Hashimi and coworkers employed NMR relaxation dispersion to show that the kinetics and HG base pair formation exhibit a rather complex dependence on sequence and position.³⁸ They further predict that structurally stressed regions within DNA, such as kinks and turns, are likely to be hotspots for HG base pair formation.

Although HG pairing is more widespread within damaged or distorted DNA,^{31,32,43} as well as DNA-protein,^{30,44} and DNA-antibiotic complexes,^{45,46} during the last few years several reports of undistorted DNA and bare DNA duplexes containing HG base pairs have emerged.⁴⁷⁻⁵⁰ Subirana and coworkers were the first to report the crystal structure of a bare poly(A-T) duplex, in which all the base pairs exhibit HG base pairing.⁴⁷ Interestingly, in solution, the poly(A-T) sequence exclusively formed the canonical B-DNA conformation.⁵¹ It was suggested that crystal packing forces may play a role in stabilizing the HG duplex. Although the overall topology of the HG duplex resembles the canonical B-form, it exhibits a narrower minor groove due to the shortening of the C1'-C1' distances across the base pairs.^{22,23} The experimental results were corroborated in subsequent work by Orozco and coworkers.⁵² Using molecular dynamics simulations they suggested that both the HG and B forms are thermodynamically accessible states for the poly(A-T) duplex, and are remarkably similar in terms of structure and dynamics. In a followup study, Orozco and coworkers predicted that chimeras of HG duplexes embedded in B-DNA and HG-WC junctions are also thermodynamically stable under physiological conditions, and retain a strong memory of the shape, and the recognition properties of the original helices.⁵³

The work of Nikolova *et al.* provided a significant advance in understanding of the WC to HG transition.³⁹ Employing NMR relaxation dispersion spectroscopy, they demonstrated that WC base pairs in a DNA duplex could reshuffle due to thermal motion, and transiently adopt the HG conformation. These 'excited states' have an equilibrium population of around 0.1%, and are short-lived, with a lifetime of approximately 5 ms. The authors speculate that the WC to HG equilibrium is intrinsically encoded within the DNA double helix, and could be exploited by cellular machinery to expand the functionality of DNA beyond what is accessible

with WC base pairing. The conjugate peak refinement method (CPR)⁵⁴ was used to provide atomistic insight into the WC to HG transition pathway. Based on a two-dimensional free energy surface, Nikolova *et al.* predict that the *anti* to *syn* transition of the purine base involves base-flipping along the major groove, through small opening angles. In recent work, Pak and coworkers proposed that the WC to HG transition can occur via multiple routes, with contributions from both major and minor groove flipping.⁴² The switch from WC to HG conformations could also be triggered by chemical modifications. For example, Isaksson *et al.*⁵⁵ showed that substituting the O4' oxygen atom of the sugar ring to a CH₂ moiety results in a dynamic equilibrium between the WC duplex conformation, and a partial HG duplex exhibiting embedded Hoogsteen base pairs at the sites of chemical modification.

The WC to HG transition is characterized by high potential and free energy barriers, which makes it relatively difficult to obtain microscopic insight using unbiased simulations.^{23,39,42} Although previous computational studies^{39,42} suggest transition pathways, the use of predefined reaction coordinates restricts the configuration space. In previous contributions^{39,42,55} the WC to HG transition has been investigated in the context of a single target base pair, embedded in a B-DNA duplex. To the best of our knowledge, the present contribution may be the first to provide mechanistic and kinetic insight into the complete switch from the WC to the HG helix conformations including all the nucleic acid degrees of freedom.

In this Perspective, we discuss how the computational energy landscape framework,⁵⁶ based on geometry optimization with post-processing using tools of statistical mechanics and unimolecular rate theory,^{57,58} can provide unprecedented mechanistic and kinetic insight into the complete switch in helix morphology from WC to HG. As a prototypical example, we consider a DNA duplex with the sequence (ATTAAT)₂, whose crystal structure was reported in recent work by Campos and coworkers.⁵⁹ This sequence adopts the HG conformation in the crystal lattice, and does not require additional ligands to maintain stability. Interestingly, no trace of the WC duplex was found under the experimental conditions.⁵⁹ Here, we explain

this key result in detail in terms of the underlying energy landscape.

The computational energy landscape framework: In chemical physics, the potential energy landscape perspective serves as a unifying theme in the study of structure, thermodynamics and kinetics.⁵⁶ For biomolecules, the landscape can be quite complex. The presence of competing structures often leads to features in the heat capacity profile, and to characterize interconversion pathways usually requires tools of rare event dynamics.^{60–65} For many practical applications, it is often sufficient to consider a coarse-grained description of the energy landscape, in terms of stationary points (i.e. potential energy minima and transition states). This coarse-graining eliminates the need for explicit dynamics: stationary points on the underlying landscape are located using geometry optimization techniques, and thermodynamic properties and kinetics are predicted using established tools (and approximations) of statistical mechanics and unimolecular rate theory. The resulting simulation framework is largely independent of the magnitude of the corresponding barriers and time scales, and unaffected by the presence of kinetic bottlenecks.^{56,66}

The potential energy landscape framework has been developed as both a conceptual and computational tool. Usually the first step in this analysis is basin-hopping global optimization,^{67–69} which has proved effective in structure prediction for a diverse range of systems, spanning atomic and molecular clusters, glass formers, and biomolecules. For quantitative prediction of thermodynamic properties we have described basin-sampling⁷⁰ and nested-sampling^{71,72} as alternatives to parallel tempering.^{73–77} The discrete path sampling (DPS) technique^{62,78} is used to systematically construct databases of stationary points, corresponding to kinetic transition networks.^{66,79–82} This approach has been applied to problems where the presence of high energy barriers and broken ergodicity makes the exploration of the landscape, and hence the prediction of pathways and kinetics, challenging.^{83–86} Our efforts largely complement the recent advances in the area of rare event sampling made by other groups, which include innovative schemes based on explicit dynamics.^{87–93} Reviewing these methods is beyond the scope of this Perspective, and readers are referred elsewhere for such

details. Here, we focus on how various observable properties of interest can be extracted from stationary point databases constructed from DPS simulations.

The databases of stationary points constitute a kinetic transition network,^{66,79,80,82} which can be analyzed to extract global thermodynamics and kinetics using statistical mechanics and unimolecular rate theory. The superposition approach⁹⁴⁻⁹⁷ enables us to define global equilibrium thermodynamic properties, such as the heat capacity and equilibrium populations, in terms of contributions from local minima. Here we often employ the harmonic approximation, using normal mode analysis to estimate the canonical partition function corresponding to the catchment basin of each minimum.^{96,98} This approach provides us with local free energies, or values for states defined by lumping, and anharmonic as well as quantum corrections can be included.⁹⁹⁻¹⁰⁵ The ensemble of transition paths in the transition network, which are described in terms of interconnected minimum-TS-minimum triples, reflect the connectivity of the landscape and encode the global dynamics. Each minimum-to-minimum rate constant is usually determined employing harmonic transition state theory.^{57,106,107} However, other formulations, such as Kramer’s theory,^{108,109} can be used. The kinetically relevant pathways on the landscape, which make the most significant contributions to the global dynamics, are extracted from the network using Dijkstra’s shortest path algorithm with appropriate edge weights.^{62,110,111} Phenomenological reactant to product rate constants are obtained from the kinetic transition network using the new graph transformation (NGT) method.¹¹² Compared to other approaches NGT scales better with temperature as well as the size of the network, and is robust in terms of numerical precision.¹¹³

In experiments, the relaxation time scales usually correspond to transitions between ensembles, rather than isolated minima. To compare with experimental kinetics, we employ a self-consistent regrouping scheme,^{83,114} which exploits the contrasting time scales between local equilibration within states that can interconvert via low barriers, and the slower conformational transition of interest. In the regrouping, minima separated by free energy barriers below a certain threshold are lumped into a single macrostate. This approach is similar in

spirit to various kinetic clustering schemes that have been recently developed.^{115–118}

Disconnectivity graphs^{119–121} play a key role in conceptualizing energy landscapes, hence providing connections between apparently unrelated systems. The disconnectivity graph construction segregates the landscape into disjoint sets of minima, known as superbasins,^{119,120} at a regular series of threshold energies. The minima within each superbasin are mutually accessible at a given threshold, whereas transitions between superbasins involve higher potential or free energy barriers. In contrast to representations relying on low-dimensional projections of the energy landscape onto structural order parameters, principal components, or reaction coordinates,^{115–117,122–124} disconnectivity graphs capture the full complexity of the landscape, and preserve kinetic information. This construction has proved to be very insightful. For example, efficient self-organization is associated with a ‘palm tree’ structure in the disconnectivity graph,¹²¹ while double funnel landscapes,^{121,125} which support competing morphologies, are associated with broken ergodicity^{70,97,126–131} and rare event dynamics.^{62,78,132}

Conformational ensembles of the WC and HG duplexes: To explore the thermodynamic stability of the WC and HG DNA duplex conformations, molecular dynamics simulations in explicit water were first carried out at 300 K for 200 ns. The duplexes were modeled using a properly symmetrized version¹³³ of the AMBER99bsc0 force field,¹³⁴ employing the χ OL4 corrections for DNA, as suggested by Šponer and coworkers.¹³⁵ Further details regarding the preparation of the initial structures and the simulation protocol are included in the supporting information.

The key results from the simulations are summarized in Figure 2. The evolution of the root-mean-square-deviation (RMSD) with respect to the initial conformation (supporting information, Figure S1) indicate that the WC and HG duplexes are thermodynamically stable, with no major structural changes occurring on the time scale of the simulation. In the WC duplex, the terminal adenines undergo a *anti*→*syn* transition about the glycosidic torsion angle at around 20 ns, after fraying out of the helix. Our observation is consistent

with previous experiments^{136,137} and simulations,^{138–140} which indicate that base pair fraying is particularly common at the helix terminals. After the χ torsional transition, the adenines flip back into the helix, establishing a HG base pair with the corresponding thymines. The HG base pairs at the duplex terminals remain stable throughout the rest of the trajectory, and do not revert back to the canonical WC orientation. It is likely that our specific choice of force field, in particular the χ parametrization, affects the *anti* \leftrightarrow *syn* equilibrium, as observed in previous work.^{138,141} In contrast to the WC duplex, no fraying at the helix terminals was observed for the HG duplex, and all adenines exclusively sampled the *syn* region.

To evaluate force field effects on the conformational dynamics of the DNA duplexes, we carried out additional molecular dynamics simulations with the parmbsc1 force field, a relatively new parametrization introduced by Orozco and coworkers.¹⁴² We find that the terminal adenines in the WC duplex undergo multiple transitions about the glycosidic torsion angle (supporting information, Figure S2). Between 10 and 40 ns the terminal adenines undergo *anti* \rightarrow *syn* transitions, and switch to the Hoogsteen orientation. The bases revert back to the Watson-Crick geometry after around 100 ns. In contrast to the results for the χ OL4 parametrization, the HG duplex undergoes significant distortion along the trajectory (supporting information, Figure S3). Strand slippage occurs at around 40 ns, and the terminal adenines flip to the WC orientation after 160 ns. In their original report describing the parmbsc1 force field,¹⁴² Orozco and coworkers suggest that such geometric distortions arise due to the metastable nature of the HG conformation. They also found that the equilibrium configuration for the HG duplex displayed significant deviation with respect to the experimental structure. As the insight provided in our study critically hinges upon a proper description of the experimental structure,⁵⁹ our focus in this Perspective is on the χ OL4 parametrization.

The distributions depicted in Figure 2 indicate that HG pairing requires a constriction of the C1'–C1' distances across the corresponding base pairs. On average, the C1'–C1' distances

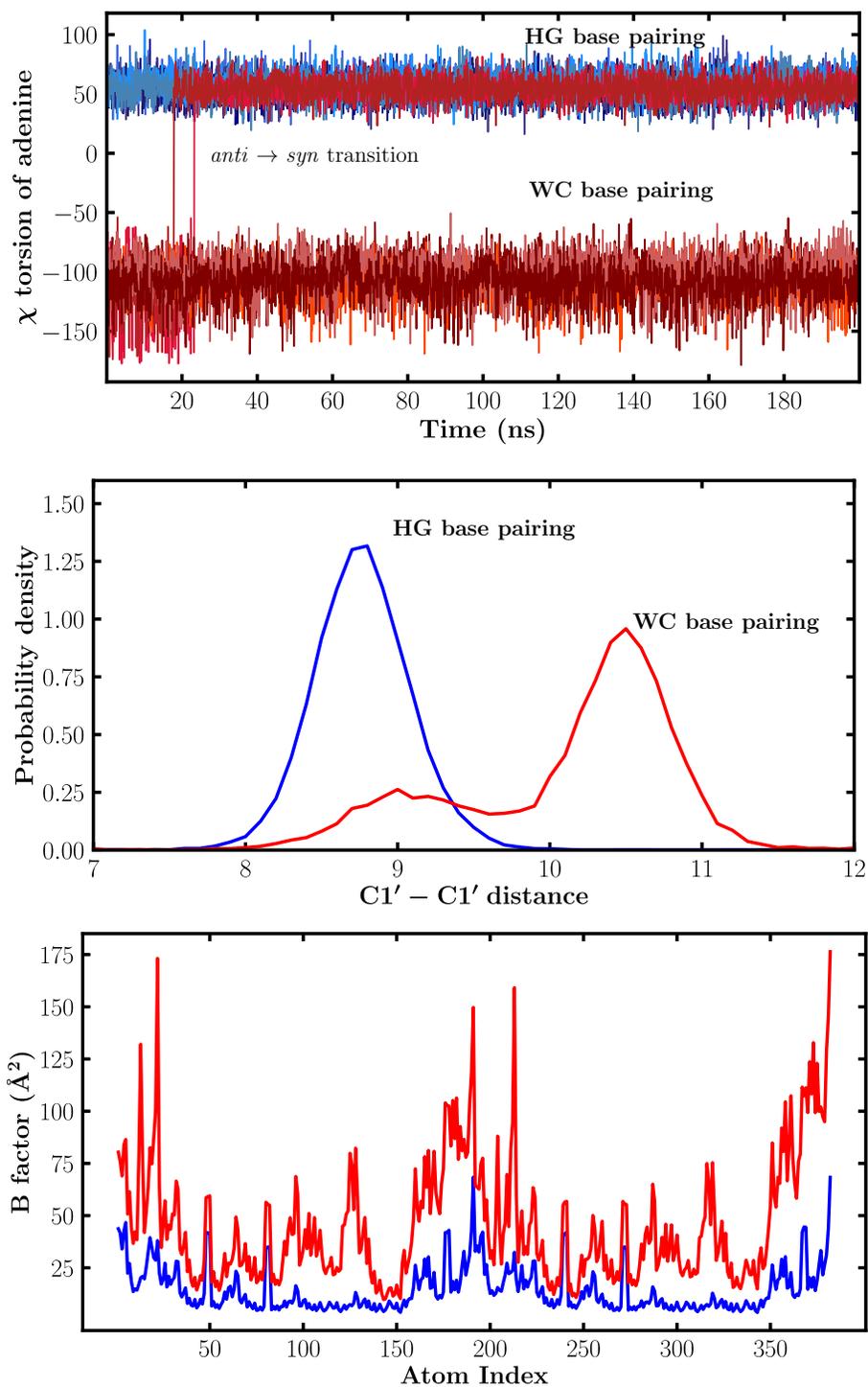


Figure 2. Summary of the key results from the molecular dynamics simulations. Top panel: evolution of the χ torsions for the adenine bases corresponding to the HG duplex (blue lines), and the WC duplex (red lines). Middle panel: distribution of the C1'-C1' distances across the base pairs for the HG duplex (blue), and the WC duplex (red). Bottom panel: the estimated B factors for the HG duplex (blue), and the WC duplex (red).

in the HG duplex are shorter by approximately 2 \AA , compared to the WC duplex. In the distribution corresponding to the WC duplex ensemble, a minor peak appears due to the HG base pairing at the helix terminals.

The relatively high B factors indicate that the WC duplex is intrinsically more flexible than the HG duplex. The entropic stabilization of the WC conformation under physiological conditions, compared to the HG form, has been noted in earlier studies,⁵² and it is thought to play a key role in modulating DNA-protein recognition.

Configurations from the molecular dynamics trajectories were saved every 10 ps, and subsequently locally minimized after removing the solvent molecules and the counterions. This quenching was performed to generate initial samples of minima from which to construct the WC and HG duplex ensembles. The lowest potential energy minimum corresponding to the WC ensemble includes HG base pairs at the helix terminals. Several low-lying minima were also identified in which WC base pairs are found throughout the helix. The lowest energy structure from this ensemble is destabilized by approximately 4 kcal/mol, relative to the putative global minimum for the WC duplex. For the HG ensemble, the lowest energy minimum exhibits a near perfect alignment with respect to the X-ray structure,⁵⁹ as expected from the limited conformational variability of the duplex on the MD time scale.

The local minima identified from MD were systematically used to construct stationary point databases (kinetic transition networks), employing the discrete path sampling (DPS) technique.^{62,78} The network was expanded using standard database refinement schemes, until the phenomenological rate constant for the WC \longleftrightarrow HG transformation converged to within an order of magnitude, with respect to the addition of new stationary points. The technical details of the DPS procedure and the refinement process are included in the supporting information.

In the following sections, we discuss the organization of the free energy landscape and its emergent features, along with the interconversion mechanism and kinetics underlying the WC \longrightarrow HG conformational switch.

Free Energy Landscape and Equilibrium Thermodynamics: The harmonic superposition approach,^{94–97} in conjunction with self-consistent regrouping,^{83,114} was used to construct the free energy landscape at 300 K; the corresponding disconnectivity graph is shown in Figure 3. To obtain this graph, we recursively lumped together minima separated by free energy barriers less than 4 kcal/mol. This regrouping threshold seems appropriate, as free energy graphs exhibiting qualitatively similar topological features were produced for a range of values around this selection, consistent with the ungrouped results (data not shown).

Each branch of the disconnectivity graph is colored according to the number of adenine nucleobases adopting the *syn* orientation in the corresponding free energy minimum. This particular order parameter enables us to separate the two prominent funnels on the landscape precisely, which involve duplex structures rich in HG and WC base pairs, respectively. The inability of geometrically defined order parameters to faithfully represent the complexity of the landscape is well documented, and here this failure is manifested by intermixing of different colored branches in some regions of the disconnectivity graph. In a previous study, Wales and Head-Gordon show that using the radius of gyration R_g as an order parameter for protein L can result in misclassification of folded and unfolded states, leading to artificially high folding rates.¹⁴³ We stress that the use of an order parameter in our analysis is merely to aid visualization, and we calculate transition rates between states defined by exploit kinetic criteria based on recursive lumping.

The free energy global minimum corresponds to an ensemble of duplexes, in which the complementary bases are engaged in HG base pairing. Although this ensemble is largely homogeneous in terms of overall helicity, in some of the constituent structures one of the terminal bases is frayed out of the helix, adopting a *syn* orientation about the glycosidic torsion angle (see supporting information, Figure S4). The terminal adenines have a higher propensity to flip out, and we could identify only one structure within this ensemble in which a terminal thymine was displaced out of the helix.

Due to the coexistence of HG and WC base pairs within the DNA helix, a combinatorial

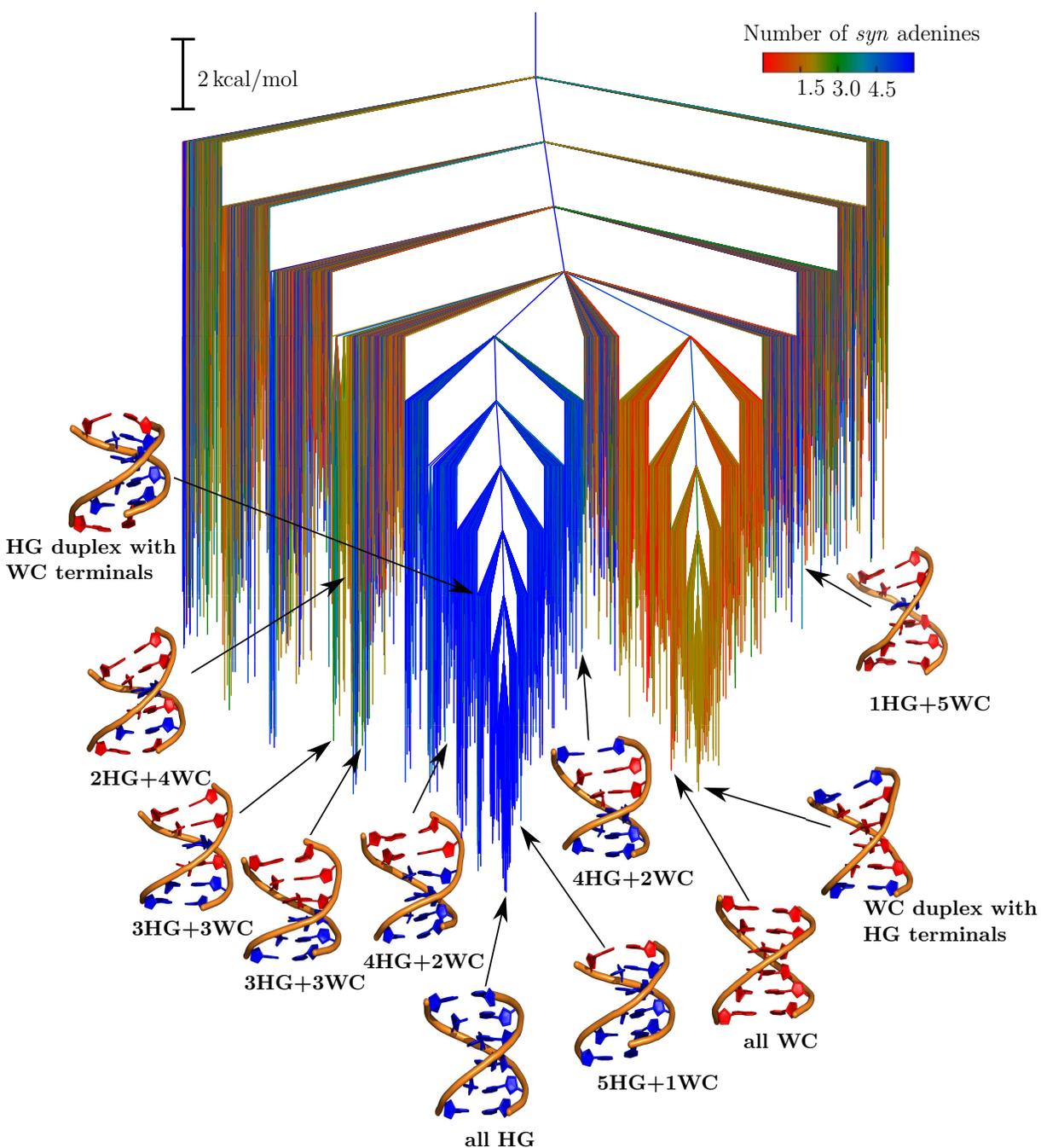


Figure 3. Free energy landscape obtained for a regrouping threshold of 4 kcal/mol at 300 K. The branches are colored according to the number of adenines in *syn* orientation (from red to blue, with red representing structures in which all the adenines are in the *anti* orientation, and blue representing structures in which all the adenines are in the *syn* orientation). Some representative snapshots from the different conformational ensembles are shown. The HG base pairs are colored in blue, and the WC base pairs in red, following the same coloring scheme as in Figure 1.

number of mixed duplex configurations could exist. In the stationary point databases, we identified around 90% of all the possible combinations, corresponding to stable duplex structures exhibiting base pairing between the complementary nucleotides. However, only a small subset of structures are associated with appreciable equilibrium occupation probabilities at 300 K (Figure 3, and supporting information, Figures S5–S9). Snapshots corresponding to some of these minima are shown superimposed on the disconnectivity graph.

Mixed duplexes occupying the low-lying regions of the blue funnel consist of either one or two WC base pairs (denoted 5HG+1WC, and 4HG+2WC, respectively), embedded within a HG base-paired helix. In particular, there is a significant population of HG duplexes exhibiting WC base pairing at the helix terminals. The mixed duplex ensembles within the blue funnel are separated from the HG duplex by low free energy barriers, and are likely to be late-stage intermediates during the WC \rightarrow HG conformational switch. Ensembles of duplexes having an equal number of HG and WC base pairs (denoted as 3HG + 3WC) occupy the low-lying and intermediate regions of the landscape. Interestingly, structures in which the HG and WC base pairs are symmetrically distributed do not exhibit any particular energetic advantage over those in which the different types are more randomly distributed.

Duplexes exhibiting primarily WC base pairing occupy the bottom of the red funnel. The lowest free energy group within this funnel corresponds to an ensemble of duplexes in which the terminal base pairs exhibit a diverse range of conformations (Figure S10). The most prominent members of this group are structures in which either one or both the terminal base pairs adopt a HG orientation. In addition, we also identified structures in which one of the terminal bases (primarily adenines) is frayed out of the helix. Canonical WC duplex structures constitute the second-lowest free energy group within the red funnel. At a regrouping threshold of 6.8 kcal/mol, the two WC duplex ensembles described above merge into a single state, implying that the constituent structures, having different configurations for the terminal bases, would interconvert on a time scale of about 14.4 ns. The predicted time scale is consistent with our initial MD simulations, where the *anti* \rightarrow *syn* torsional transitions

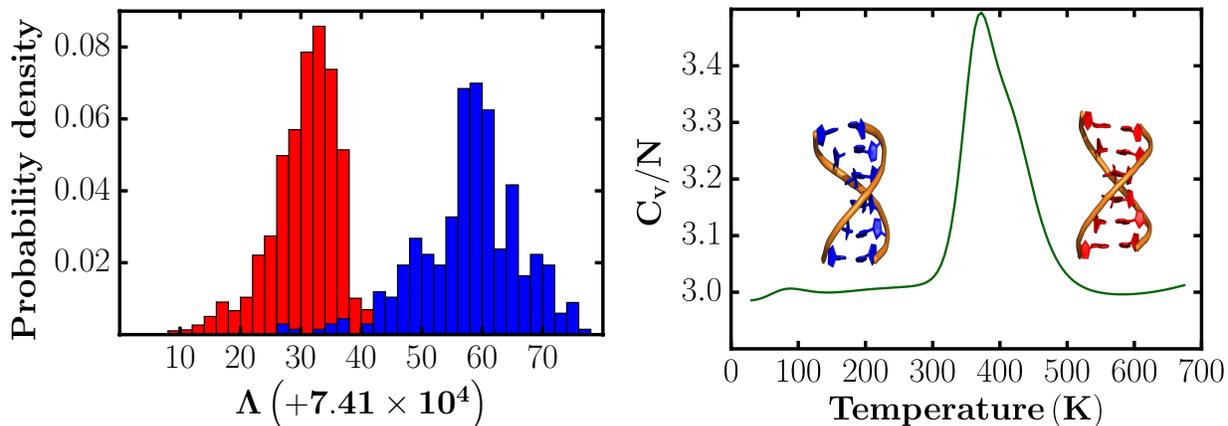


Figure 4. Left: Distribution of Λ values for minima constituting the free energy groups corresponding to the HG duplex (blue) and WC duplex (red) ensembles. Right: The normalized heat capacity profile C_v computed from the database of minima using the harmonic superposition approximation. The normalization involves dividing by the number of atoms (N).

for the terminal adenines were observed at around the 20 ns mark in the trajectory. The presence of one or two embedded HG base pairs (1HG+5WC, and 2HG+4WC, respectively) in the interior of the helix destabilizes the WC duplex structure, and the corresponding minima are located in the intermediate region of the disconnectivity graph.

To compare the contrasting flexibilities of the HG and WC duplexes, we computed the log product of the normal-mode frequencies (Λ) for minima defining the corresponding free energy ensembles. Within the harmonic approximation, vibrational entropy is proportional to $-\Lambda$,^{56,144,145} and therefore flexible structures would be associated with lower (more negative) Λ values. As shown in Figure 4, the Λ values for the WC duplex conformations are lower than most HG duplex structures, implying that the WC duplex is entropically favorable. Although the harmonic approximation does not account for well anharmonicity, which is implicitly included in MD simulations, the predicted trend in flexibilities is in line with the estimated B factors for the WC and HG duplexes.

The WC duplex is destabilized by around 10 kcal/mol compared to the HG duplex in terms of potential energy (supporting information, Figure S11). In contrast, the free energy difference between the two conformations at 300 K is approximately 3 kcal/mol. Hence en-

tropic contributions play a key role in modulating the stability gap (free energy difference) between the HG and WC conformations at physiological temperatures. As discussed above, the presence of soft normal mode vibrations makes the WC conformation favorable in terms of vibrational entropy. Additional stability for the WC form is due to landscape entropy (supporting information, Figure S12), which is determined by the density of high-lying local potential energy minima. The effect of the landscape contribution can be quantified by calculating the heat capacity excluding the vibrational degrees of freedom. The superposition approach includes both the landscape and vibrational contributions, but it is often not possible to characterize sufficient high-lying minima to provide a quantitative description of landscape entropy. In this case an alternative approach, such as the basin-sampling scheme,⁷⁰ is required.

The interplay of enthalpy and entropy results in a heat capacity peak profile (Figure 4), which is typical for solid-solid type phase transitions observed in double funnel systems, such as the LJ₃₈ cluster.^{121,125} At low temperature, the blue funnel is thermodynamically favored, and relaxation down the energy landscape would preferentially lead to the HG duplex. Nonetheless, trapping in the competing red funnel, corresponding to the WC type duplexes, is likely to impede the overall kinetics. At higher temperatures, or under conditions in which entropy dominates, the landscape shifts in favor of the WC duplex. It is likely that the inherent limitations of the harmonic approximation, and incompleteness of the superposition sum, lead to an overestimation of the transition temperature for the HG \longleftrightarrow WC switch.^{56,145} Both effects lead to a systematic underestimation of the entropy, which is greater for the WC ensemble, so a higher temperature is required to shift the equilibrium.

With the parmbsc1 parametrization, some reorganization of minima within the free energy basins is anticipated, but the key topological features should be preserved. To check this intuition we reoptimized some key minima in the red and blue funnels with the parmbsc1 force field. The probability distribution of relative energies (supporting information, Figure S13) reveals that, although the relative ordering of the basins is consistent with the

χ OL4 parametrization, the energy gap between the HG and WC conformations is somewhat reduced. In addition, several distinct potential energy minima located with the χ OL4 parametrization, collapse to a single minimum on the parmbsc1 surface (supporting information, Figure S14). This trend suggests that the parmbsc1 force field results in a smoother potential energy surface, which could have implications in terms of sampling efficiency. A similar trait was reported by Somani *et al.* in their benchmark study on alanine peptides, comparing the AMBER and CHARMM force fields.¹⁴⁶ Furthermore, shifts in the relative potential energies of the minima, with respect to a fixed reference structure (in this case, the free energy global minimum identified with the χ OL4 parametrization), provides clear-cut evidence of reorganization within the funnels. Details are provided in the supporting information. Interestingly, the magnitude of vibrational entropies for the different conformations, as determined by Λ , is in quantitative agreement with the results obtained using the χ OL4 parametrization (supporting information, Figure S15). As a final remark on force field effects, we note that a complete characterization of the key differences between the parmbsc1 and the χ OL4 landscapes, would require systematic benchmarking, not only in terms of structure,¹⁴⁷ but also thermodynamics and kinetics.

The topography of the free energy landscape, and the equilibrium thermodynamics, suggest that at low temperatures, or crystallization conditions that constrain the DNA helix, similar to those employed by Campos and coworkers,⁵⁹ (ATTAAT)₂ preferentially adopts the HG conformation. The high free energy barriers separating the HG and WC duplex configurations will make it challenging to observe transitions between them, and probably preclude the observation of both conformations in a single experiment if the observation time scale is relatively short. The landscape analysis therefore also explains the results of Subirana and coworkers,^{47,51} who isolated different polymorphs for the AT repeat sequence, using X-ray crystallography and NMR.

The organization of the landscape provides key insight into the chain of events that occur during the WC \rightarrow HG conformational switch. The transition is likely to proceed in a

stepwise fashion, involving mixed duplex intermediates, with a varying number of HG and WC base pairs.

Transition Pathways and Interconversion Kinetics: The diversity of mixed duplex intermediates on the free energy landscape suggests that multiple routes between the HG and WC duplex conformations are feasible. For brevity, we consider only those discrete paths in the network that are associated with the highest statistical weights, which contribute maximally to the flux. The conformational switch takes place via two dominant mechanisms (Figure 5), differing primarily in the nature of the mixed duplex intermediates visited en route to the HG duplex.

Both mechanisms (paths P1 and P2) are essentially characterized by six elementary steps, in which a WC base pair switches to the HG orientation, resulting in a mixed duplex intermediate. Recursive enumeration analysis¹⁴⁸ identifies P1 as the ‘fastest pathway’ in the network, while P2 appears around position one thousand. The pathways ranked in between, in terms of their contribution to the overall interconversion rate, correspond to minor variations of P1, with recrossing of some barriers. Hence we assign them to the same P1 mechanism and pathway ensemble. The highest energy transition states along the pathways correspond to structures in which the adenine that flips breaks away from its partner thymine, and moves out of the helix through small opening angles. Although the order of adenine flips is remarkably different in paths P1 and P2, in both cases base flipping is initiated at the TA step, which is known to stack quite poorly in the WC conformation.^{59,149} The transition from WC to the HG orientation for the central adenines results in significant kinetic bottlenecks, in contrast to the terminal adenines, which are more labile. The characteristic energy barriers for flipping of adenine bases within the center of the helix cover quite a broad range, from 8 to 18 kcal/mol. Our predictions are in excellent agreement with the relaxation dispersion experiments and CPR simulations of Nikolova *et al.*,³⁹ which indicate that the transition state ensemble primarily features flipped-out states in which the purine base unstacks out of the helix via the major groove. A direct comparison with experiment, which provides insight

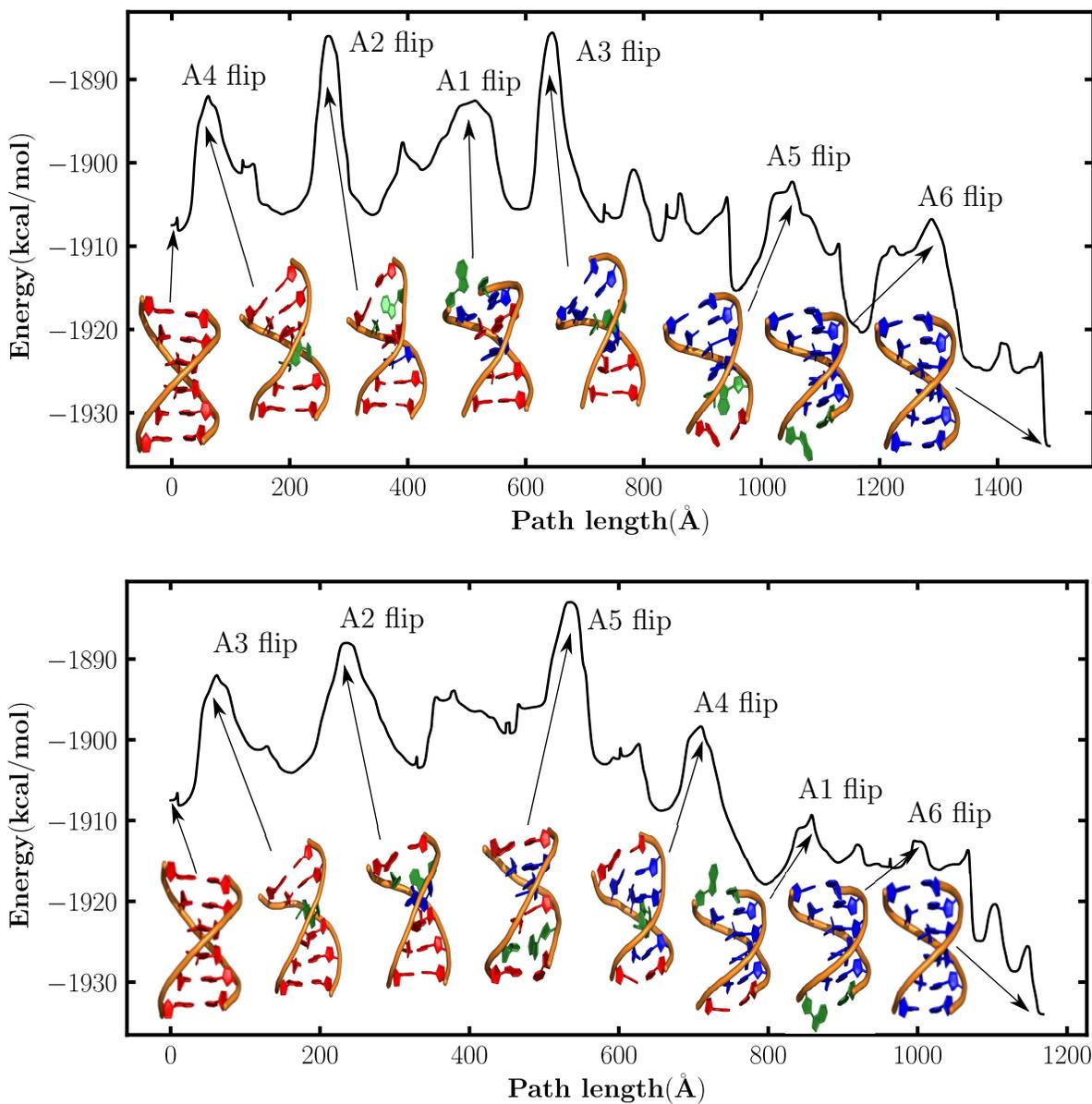


Figure 5. Transition pathways for the WC to HG conformational switch that make the largest contributions to the global dynamics. Some representative structures encountered at different stages of the transformation are also shown. Top panel: Pathway P1 where the order of base flipping is $A4 \rightarrow A2 \rightarrow A1 \rightarrow A3 \rightarrow A5 \rightarrow A6$. Bottom panel: Pathway P2 where the order of base flipping is $A3 \rightarrow A2 \rightarrow A5 \rightarrow A4 \rightarrow A1 \rightarrow A6$. The flipped-out adenine, and its partner thymine are colored green in the transition state structures. The WC base pairs are colored red, and HG base pairs colored blue.

at the single base pair level, is only feasible in this case because both pathways P1 and P2 involve step-wise base flipping, and no signatures of cooperative switching are found. In scenarios where the transformation pathways are dominated by multiple base flips occurring concurrently, a meaningful comparison with the experiment of Nikolova *et al.* would be more difficult.

In the kinetically relevant pathways, we do not find any clear-cut evidence of minor groove flipping, which was suggested by Pak and coworkers as a plausible mechanism of WC to HG transition for a single AT base pair within a DNA helix.⁴² In their analysis the transition pathways were based on a two-dimensional free energy surface. In such representations, projection errors are common, and the true complexity of the landscape can be masked. Energy barriers in orthogonal degrees of freedom may be averaged over, and as a result, the location of dividing surfaces (regions separating two basins) can be problematical. Hence low-dimensional projected surfaces may provide a misleading description of the kinetics; these issues have been discussed in a number of recent contributions.^{115,124} Our results for the free energy landscape, using recursive regrouping based on the calculated barriers, do not employ order parameters or projection, and all degrees of freedom are retained in the thermodynamic and kinetic analysis. Differences of interpretation may therefore originate from the alternative perspectives employed to extract the observables.

Pak and coworkers note that the free energy barrier for minor groove flipping is somewhat higher, compared to the major groove, during the initial stages.⁴² This situation is exacerbated when multiple flipping events are required to completely switch the topology of the duplex from the WC to the HG form. In our kinetic transition network, we do identify transition states in which adenine bases protrude out of the helix along the minor groove. However, the energy barriers connecting these transition states to the adjoining minima are substantial. Hence the corresponding pathways are energetically unfavorable, and are associated with relatively low statistical weights in the transition network. These results suggest that pathways proceeding via minor groove base flipping make little or no contribution to

the global dynamics.

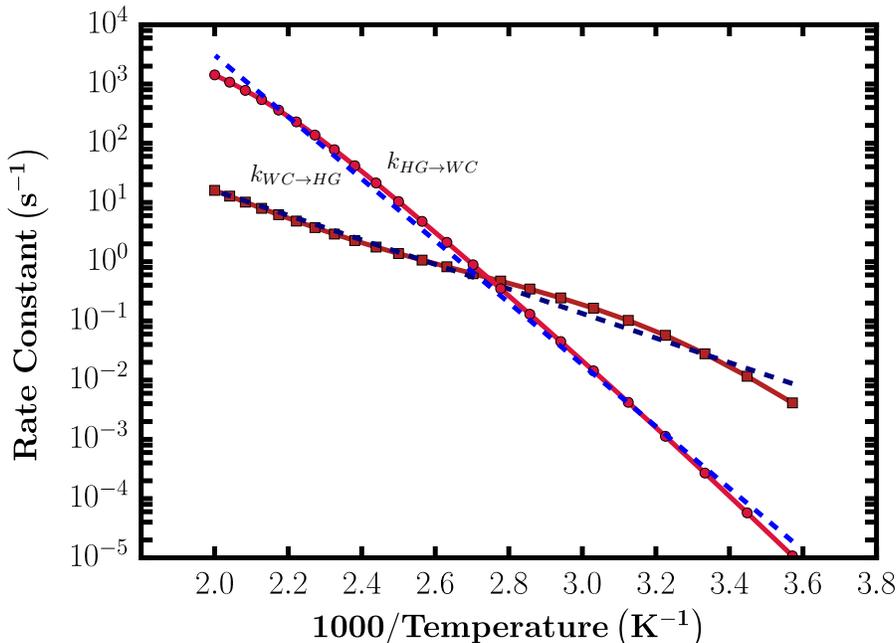


Figure 6. The variation of the rate constants corresponding to the conformational switch from WC to HG duplex ($k_{WC \rightarrow HG}$), and the reverse transformation ($k_{HG \rightarrow WC}$) with reciprocal temperature, are shown as solid lines with symbols. The fits to the Arrhenius equation, $k_{AB}(T) = A \exp\left(\frac{-E_a}{kT}\right)$ are depicted as dashed lines. From the fits, we obtain an activation energy, $E_a = 9.48 \pm 0.2$ kcal/mol, and a prefactor, $A = 2.19 \times 10^5 \text{ s}^{-1}$ for the WC \rightarrow HG transition. The reverse transformation has $E_a = 23.90 \pm 0.3$ kcal/mol, and $A = 8.69 \times 10^{13} \text{ s}^{-1}$.

The slow transition between the HG and WC duplex conformations was quantified by estimating the phenomenological rate constants.^{62,112} The WC \rightarrow HG switch is associated with a rate constant ($k_{WC \rightarrow HG}$) of $2.78 \times 10^{-2} \text{ s}^{-1}$ at 300 K. The reverse transformation, HG \rightarrow WC is orders of magnitude slower, with a rate constant ($k_{HG \rightarrow WC}$) of $2.70 \times 10^{-4} \text{ s}^{-1}$. The variation of $k_{WC \rightarrow HG}$, and $k_{HG \rightarrow WC}$ with temperature, depicted in the form of Arrhenius plots (Figure 6), provides further insight into the underlying kinetics. At 300 K, the Arrhenius prefactor associated with the WC \rightarrow HG transition (Figure 6) is several orders of magnitude smaller than the time scale of typical molecular motions. Using the prefactor in standard thermodynamic expressions we obtain a negative value for the entropy of activation ($\sim -0.036 \text{ kcal mol}^{-1} \text{ K}^{-1}$), which strongly suggests that the switch to the HG form is triggered via conformational restriction.

As temperature increases, $k_{HG \rightarrow WC}$ becomes higher than $k_{WC \rightarrow HG}$, indicating that entry into the WC funnel is an entropically driven process. The forward and backward rates are nearly equal at 365 K, which is close to the transition temperature predicted by the heat capacity profile. Interestingly, the temperature dependence of $k_{WC \rightarrow HG}$ becomes non-Arrhenius in the low temperature regime, which is reminiscent of protein and RNA folding.^{150–152} This deviation suggests that the conformational switch cannot be associated with a well defined transition state. At low temperatures, intermediates and kinetic traps that populate the free energy landscape are associated with distinct relaxation time scales. The transition from the WC to the HG conformation is likely to involve off-pathway excursions to trapped states, or periods in on-pathway intermediates, and a multi-state kinetic model seems more appropriate. At higher temperatures, the different basins merge, and an apparent two-state behavior is recovered. In contrast to $k_{WC \rightarrow HG}$, a small curvature is exhibited by $k_{HG \rightarrow WC}$ even at high temperatures, with the estimated rates being slightly lower than those predicted by the Arrhenius relationship. We speculate that this deviation could be due to the entropic stabilization of competing free energy minima close to the WC funnel, and the increased statistical weights of some slower routes in the pathway ensemble.

Conclusions: The potential energy landscape perspective has provided new insight into a wide range of diverse systems, including atomic and molecular clusters, soft and condensed matter, and biomolecules.^{56,66} In this Perspective, we have highlighted how coarse-graining the landscape in terms of stationary points (minima and transition states), provides a powerful alternative to conventional Monte Carlo and molecular dynamics simulations. Equilibrium thermodynamic and kinetic properties can be extracted from the underlying energy landscape using established tools and approximations of statistical mechanics and unimolecular rate theory. The inherently time-independent nature of the geometry optimization procedure makes it possible to probe biomolecular pathways that are not amenable to conventional simulations.¹⁵³ Here, we have discussed one such application in detail, the conformational switch between the Watson-Crick and Hoogsteen forms of a DNA duplex.

The recent discovery that Hoogsteen base pairs are more widespread in DNA than previously thought, expands the repertoire of DNA functionality, with possible implications in DNA-protein recognition.²³ Hence it is important to understand how DNA switches between different base pairing modes. Our results provide new insight at an atomic level of detail into the transition pathways between the WC and HG conformations of DNA, and the associated kinetics. We show that the competition between the two conformations results in a double funnel free energy landscape, which can be tuned as a function of temperature, or other environmental conditions. In particular, we can explain why complementary experiments usually identify only one of the polymorphs.^{47,51} The system exhibits a solid-solid-type heat capacity peak at low temperature, which is a characteristic property of multifunnelled landscapes. In the context of biology, such topographical features could naturally lead to multifunctionality, as we have recently suggested.¹⁵⁴

Our analysis reveals that multiple pathways exist between the HG and WC conformations, via mixed duplex intermediates. The transition state ensemble corresponds to base flipped-out states, in agreement with the experimental work of Nikolova *et al.*³⁹ The interconversion kinetics exhibits non-Arrhenius behavior, suggesting that a two-state description is insufficient to describe the conformational switch. It will be interesting to see if any evidence of intermediate states and multistate kinetics emerge in future experiments. Sequence composition and oligonucleotide length could play a key role in modulating the transition pathways and kinetics. As shown by Orozco and coworkers,⁵² the entropic stabilization of the WC conformation increases with the chain length. On the other hand, energetic stabilization from base-stacking and hydrogen-bonding exhibits a strong linear dependence on length for the HG form. This interplay of competing interactions could provide a means to design a landscape that favors one form or the other, by tuning the environmental conditions. Furthermore, the energetics and kinetics of DNA base-flipping, which is a key step in the transformation between the HG and WC conformations, depends critically on the identity of the proximal bases.^{155,156} Al-Hashimi and coworkers previously suggested the appearance

of a late transition state in the conformational switch from WC to HG, based on the Hammond postulate, and argued that the energetics exhibit a complex dependence on both the sequence and position.³⁸ Similarly, we speculate that oligonucleotide length and sequence effects will play a role in modulating the barrier heights and interconversion rates for the HG \longleftrightarrow WC conformations. These possibilities could be explored further in future studies.

In future work we envisage many new applications of the computational energy landscapes methodology for biomolecules, particularly nucleic acids, whose polyelectrolyte nature, and lower stability gap, makes conventional simulations challenging.^{157,158} Likely targets include biophysical processes of fundamental importance, such as assembly mechanisms for higher order RNAs, and conformational changes in riboswitches.

Supporting Information Available

Details regarding the preparation of initial structures, molecular dynamics protocol, and discrete path sampling. The formalism describing the calculation of free energies, and global dynamics. Figures showing the evolution of the RMSD for the WC and HG duplexes along the MD trajectories, potential energy disconnectivity graph, benchmark calculations with the parmbsc1 force field, plot of landscape entropy, representative structures constituting the HG duplex ensemble, snapshots of mixed duplex intermediates exhibiting appreciable occupation probabilities at 300 K, representative structures constituting the lowest energy WC duplex ensemble. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Acknowledgement

We are grateful to Dr Tristan Cragolini and Dr Rosana Collepardo-Guevara for stimulating discussions. The work was financially supported by the EPSRC. D.C. gratefully acknowledges the Cambridge Commonwealth, European and International Trust for financial

support.

The authors declare no competing financial interest.

References

- (1) Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738.
- (2) Watson, J. D.; Crick, F. H. C. Genetical Implications of the Structure of Deoxyribonucleic Acids. *Nature* **1953**, *171*, 964–967.
- (3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag, New York, 1984.
- (4) Al-Hashimi, H. M. NMR studies of nucleic acid dynamics. *J. Magn. Reson.* **2013**, *237*, 191–204.
- (5) Duzdevich, D.; Redding, S.; Greene, E. C. DNA Dynamics and Single-Molecule Biology. *Chem. Rev.* **2014**, *114*, 3072–3086.
- (6) Pérez, A.; Luque, F. J.; Orozco, M. Frontiers in Molecular Dynamics Simulations of DNA. *Acc. Chem. Res.* **2011**, *45*, 196–205.
- (7) Marko, A.; Denysenkov, V.; Margraf, D.; Cekan, P.; Schiemann, O.; Sigurdsson, S. T.; Prisner, T. F. Conformational Flexibility of DNA. *J. Am. Chem. Soc.* **2011**, *133*, 13375–13379.
- (8) Frank-Kamenetskii, M. D.; Prakash, S. Fluctuations in the DNA double helix: A critical review. *Phys. Life Rev.* **2014**, *11*, 153–170.
- (9) Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thastrom, A.; Field, Y.; Moore, I. K.; Wang, J.-P. Z.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778.
- (10) Xu, F.; Olson, W. K. DNA Architecture, Deformability, and Nucleosome Positioning. *J. Biomol. Struct. Dyn.* **2010**, *27*, 725–739.

- (11) Saiz, L.; Vilar, J. M. G. DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol.* **2006**, *16*, 344–350.
- (12) Felsenfield, G.; Groudine, M. Controlling the double helix. *Nature* **2003**, *421*, 448–453.
- (13) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248–1253.
- (14) Dickerson, R. E.; Chiu, T. K. Helix bending as a factor in protein/DNA recognition. *Biopolymers* **1997**, *44*, 361–403.
- (15) Dervan, P. B.; Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
- (16) Ghosh, A.; Bansal, M. A glossary of DNA structures from A to Z. *Acta Crystallogr. D* **2003**, *59*, 620–626.
- (17) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA Structures. *Annu. Rev. Biochem.* **1995**, *64*, 65–95.
- (18) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- (19) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **1979**, *282*, 680–686.
- (20) Davis, M. A.; Winthrop, S. O.; Stewart, J.; Sunahara, F. A.; Herr, F. New Psychotropic Agents. V. Derivatives of 5-Cyano- and 5-Carboxamidodibenzo[a,d]cycloheptadiene. *J. Med. Chem.* **1963**, *6*, 251–255.
- (21) Brahms, S.; Brahms, J.; Van Holde, K. E. Nature of conformational changes in poly[d(A-T)-d(A-T)] in the premelting region. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 3453–3457.

- (22) Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1959**, *12*, 822–823.
- (23) Nikolova, E. N.; Zhou, H.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M. A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* **2013**, *99*, 955–968.
- (24) Zhou, H.; Hintze, B. J.; Kimsey, I. J.; Sathyamoorthy, B.; Yang, S.; Richardson, J. S.; Al-Hashimi, H. M. New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res.* **2015**, *43*, 3420–3433.
- (25) Kratochvíl, M.; Šponer, J.; Hobza, P. Global minimum of the adenine-thymine base pair corresponds neither to Watson-Crick nor to Hoogsteen structures. Molecular dynamic/quenching/AMBER and ab initio beyond Hartree-Fock studies. *J. Am. Chem. Soc.* **2000**, *122*, 3495–3499.
- (26) Kabeláč, M.; Hobza, P. Potential Energy and Free Energy Surfaces of All Ten Canonical and Methylated Nucleic Acid Base Pairs: Molecular Dynamics and Quantum Chemical ab Initio Studies. *J. Phys. Chem. B* **2001**, *24*, 5804–5817.
- (27) Gould, I. R.; Kollman, P. A. Theoretical investigation of the hydrogen bond strengths in guanine-cytosine and adenine-thymine base pairs. *J. Am. Chem. Soc.* **1994**, *116*, 2493–2499.
- (28) Portugal, J. Do Hoogsteen base pairs occur in DNA? *Trends Biochem. Sci.* **1989**, *14*, 127–130.
- (29) Kitayner, M.; Rozenberg, H.; Rohs, R.; Suad, O.; Rabinovich, D.; Honig, B.; Shakked, Z. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **2010**, *17*, 423–429.

- (30) Rice, P. A.; Yang, S.; Mizuuchi, K.; Nash, H. A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **1996**, *87*, 1295–1306.
- (31) Patel, D. J.; Shapiro, L.; Kozlowski, S. A.; Gaffney, B. L.; Jones, R. A. Covalent carcinogenic O6-methylguanosine lesions in DNA. Structural studies of the O6 meG X A and O6meG X G interactions in dodecanucleotide duplexes. *J. Mol. Biol.* **1986**, *188*, 677–692.
- (32) Yang, H.; Zhan, Y.; Fenn, D.; Chi, L. M.; Lam, S. L. Effect of 1-methyladenine on double-helical DNA structures. *FEBS Lett.* **2008**, *582*, 1629–1633.
- (33) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Human DNA polymerase ι incorporates dCTP opposite template G via a G.C + Hoogsteen base pair. *Structure* **2005**, *13*, 1569–1577.
- (34) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. An Incoming Nucleotide Imposes an anti to syn Conformational Change on the Templating Purine in the Human DNA Polymerase- ι Active Site. *Structure* **2006**, *14*, 749–755.
- (35) Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K. Replication by human DNA polymerase- ι occurs by Hoogsteen base-pairing. *Nature* **2004**, *430*, 377–80.
- (36) Johnson, R. E.; Prakash, L.; Prakash, S. Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase ι . *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10466–71.
- (37) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase ι . *Nat. Struct. Mol. Biol.* **2006**, *13*, 619–625.

- (38) Alvey, H. S.; Gottardo, F. L.; Nikolova, E. N.; Al-Hashimi, H. M. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat. Commun.* **2014**, *5*, 4786.
- (39) 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.
- (40) Nikolova, E. N.; Gottardo, F. L.; Al-Hashimi, H. M. Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. *J. Am. Chem. Soc.* **2012**, *134*, 3667–3670.
- (41) Nikolova, E. N.; Goh, G. B.; Brooks, C. L.; Al-hashimi, H. M. Hoogsteen Base Pairs in Duplex DNA. *J. Am. Chem. Soc.* **2013**, *135*, 67666769.
- (42) Yang, C.; Kim, E.; Pak, Y. Free energy landscape and transition pathways from Watson-Crick to Hoogsteen base pairing in free duplex DNA. *Nucleic Acids Res.* **2015**, *43*, 7769–7778.
- (43) Singh, U. S.; Moe, J. G.; Reddy, G. R.; Weisenseel, J. P.; Marnett, L. J.; Stone, M. P. Proton NMR of an oligodeoxynucleotide containing a propanodeoxyguanosine adduct positioned in a (CG)₃ frameshift hotspot of *Salmonella typhimurium* hisD3052: Hoogsteen base-pairing at pH 5.8. *Chem. Res. Toxicol.* **1993**, *6*, 825–836.
- (44) Patikoglou, G. A.; Kim, J. L.; Sun, L.; Yang, S.-H.; Kodadek, T.; Burley, S. K. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes. Dev.* **1999**, *13*, 3217–3230.
- (45) Ughetto, G.; Wang, A. H. J.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res.* **1985**, *13*, 2305–2323.

- (46) Mendel, D.; Dervan, P. B. Hoogsteen base pairs proximal and distal to echinomycin binding sites on DNA. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 910–914.
- (47) Abrescia, N. G. A.; Thompson, A.; Huynh-Dinh, T.; Subirana, J. A. Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2806–2811.
- (48) Pous, J.; Urpí, L.; Subirana, J. A.; Gouyette, C.; Navaza, J.; Campos, J. L. Stabilization by extra-helical thymines of a DNA duplex with Hoogsteen base pairs. *J. Am. Chem. Soc.* **2008**, *130*, 6755–6760.
- (49) Raghunathan, G.; Miles, H. T.; Sasisekharan, V. Parallel nucleic acid helices with hoogsteen base pairing: Symmetry and structure. *Biopolymers* **1994**, *34*, 1573–1581.
- (50) Luchi, D. D.; Tereshko, V.; Gouyette, C.; Subirana, J. A. Structure of the DNA Coiled Coil Formed by d(CGATATATATAT). *ChemBioChem* **2006**, *7*, 585–587.
- (51) Abrescia, N. G. A.; González, C.; Gouyette, C.; Subirana, J. A. X-ray and NMR Studies of the DNA Oligomer d(ATATAT): Hoogsteen Base Pairing in Duplex DNA. *Biochemistry* **2004**, *43*, 4092–4100.
- (52) 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.
- (53) Cubero, E.; Luque, F. J.; Orozco, M. Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA. *Biophys. J.* **2006**, *90*, 1000–1008.
- (54) Fischer, S.; Karplus, M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.* **1992**, *194*, 252–261.

- (55) The First Example of a Hoogsteen Basepaired DNA Duplex in Dynamic Equilibrium with a Watson-Crick Basepaired Duplex A Structural (NMR), Kinetic and Thermodynamic Study. *J. Biomol. Struct. Dyn.* **2001**, *18*, 783–806.
- (56) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.
- (57) Forst, W. *Theory of Unimolecular Reactions*; Academic Press: New York, 1973.
- (58) Laidler, K. J. *Chemical Kinetics*; Harper & Row: New York, 1987.
- (59) Acosta-Reyes, F. J.; Alechaga, E.; Subirana, J. a.; Campos, J. L. Structure of the DNA Duplex d(ATTAAT)₂ with Hoogsteen Hydrogen Bonds. *PLoS One* **2015**, *10*, e0120241.
- (60) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236–9245.
- (61) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A Novel Path Sampling Method for the Calculation of Rate Constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774.
- (62) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.
- (63) Passerone, D.; Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Phys. Rev. Lett.* **2001**, *87*, 108302.
- (64) Weinan E.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.
- (65) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestone. *J. Chem. Phys.* **2004**, *120*, 10880–10889.
- (66) Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.

- (67) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611–6615.
- (68) Wales, D. J.; Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (69) Wales, D. J.; Scheraga, H. A. Global optimization of clusters, crystals and biomolecules. *Science* **1999**, *285*, 1368–1372.
- (70) Wales, D. J. Surveying a Complex Potential Energy Landscape: Overcoming Broken Ergodicity Using Basin-Sampling. *Chem. Phys. Lett.* **2013**, *584*, 1 – 9.
- (71) Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Analysis* **2006**, *1*, 833–859.
- (72) Martiniani, S.; Stevenson, J. D.; Wales, D. J.; Frenkel, D. Superposition Enhanced Nested Sampling. *Phys. Rev. X* **2014**, *4*, 031034.
- (73) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607.
- (74) Geyer, G. Markov chain Monte Carlo maximum likelihood. Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. 1991; p 156.
- (75) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Japan* **1996**, *65*, 1604–1608.
- (76) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. Monte carlo study of the interacting self-avoiding walk model in three dimensions. *J. Stat. Phys.* **1996**, *82*, 155–181.
- (77) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

- (78) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.
- (79) Rao, F.; Caffisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (80) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.
- (81) 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.
- (82) Prada-Gracia, D.; Gómez-Gardenes, J.; Echenique, P.; Fernando, F. Exploring the Free Energy Landscape: From Dynamics to Networks and Back. *PLoS Comput. Biol.* **2009**, *5*, e1000415.
- (83) Carr, J. M.; Wales, D. J. Folding Pathways and Rates for the Three-Stranded beta-Sheet Peptide Beta3s using Discrete Path Sampling. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.
- (84) 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.
- (85) Chakraborty, D.; Wales, D. J. Probing helical transitions in a DNA duplex. *Phys. Chem. Chem. Phys.* **2017**, *19*, 878–892.
- (86) Oakley, M. T.; Johnston, R. L. Exploring the Energy Landscapes of Cyclic Tetrapeptides with Discrete Path Sampling. *J. Chem. Theory Comput.* **2013**, *9*, 650–657.
- (87) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

- (88) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for beta Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
- (89) Singhal, N.; Snow, C. D.; Pande, V. S. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **2004**, *121*, 415–425.
- (90) Allen, R. J.; Frenkel, D.; Wolde ten, P. R. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.* **2006**, *124*, 024102.
- (91) Chodera, J. D.; Dill, K. A.; Singhal, N.; Pande, V. S.; Swope, W. C.; Pitera, J. W. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
- (92) Krivov, S. V.; Muff, S.; Caffisch, A.; Karplus, M. One-dimensional barrier-preserving free-energy projections of a beta-sheet Miniprotein: new insights into the folding process. *J. Phys. Chem. B* **2008**, *112*, 8701–8714.
- (93) Kuczera, K.; Jas, G. S.; Elber, R. Kinetics of Helix Unfolding: Molecular Dynamics Simulations with Milestoning. *J. Phys. Chem. A* **2009**, *113*, 7461–7473.
- (94) Stillinger, F. H.; Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **1984**, *225*, 983–989.
- (95) Stillinger, F. H. A Topographic View of Supercooled Liquids and Glass Formation. *Science* **1995**, *267*, 1935–1939.
- (96) 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.

- (97) Sharapov, V. A.; Meluzzi, D.; Mandelshtam, V. A. Low-temperature structural transitions: Circumventing the broken-ergodicity problem. *Phys. Rev. Lett.* **2007**, *98*, 105701.
- (98) Hoare, M. R.; McInnes, J. J. Statistical mechanics and morphology of very small atomic clusters. *Faraday Discuss. Chem. Soc.* **1976**, *61*, 12–24.
- (99) Wales, D. J. Coexistence in Small Inert-Gas Clusters. *Mol. Phys.* **1993**, *78*, 151–171.
- (100) Chekmarev, S. F.; Umirzakov, I. H. An analytic model for atomic clusters. *Z. Phys. D* **1993**, *26*, 373–376.
- (101) Doye, J. P. K.; Wales, D. J. Magic numbers and growth sequences of small face-centred-cubic and decahedral clusters. *Chem. Phys. Lett.* **1995**, *247*, 339–347.
- (102) Doye, J. P. K.; Wales, D. J. Calculation of thermodynamic properties of small Lennard-Jones clusters incorporating anharmonicity. *J. Chem. Phys.* **1995**, *102*, 9659–9672.
- (103) Calvo, F.; Doye, J. P. K.; Wales, D. J. Characterization of Anharmonicities on Complex Potential Energy Surfaces: Perturbation Theory and Simulation. *J. Chem. Phys.* **2001**, *115*, 9627–9636.
- (104) Calvo, F.; Doye, J. P. K.; Wales, D. J. Quantum partition functions from classical distributions: Application to rare-gas clusters. *J. Chem. Phys.* **2001**, *114*, 7312–7329.
- (105) Georgescu, I.; Mandelshtam, V. A. Self-consistent phonons revisited. I. The role of thermal versus quantum fluctuations on structural transitions in large Lennard-Jones clusters. *J. Chem. Phys.* **2012**, *137*, 144106.
- (106) Eyring, H. J. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.
- (107) Miller, W. H. Unified statistical model for complex and direct reaction mechanisms. *J. Chem. Phys.* **1976**, *65*, 2216–2223.

- (108) Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284–304.
- (109) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
- (110) Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271.
- (111) Evans, D. A.; Wales, D. J. Folding of the GB1 hairpin peptide from discrete path sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090.
- (112) Wales, D. J. Calculating rate constants and committor probabilities for transition networks by graph transformation. *J. Chem. Phys.* **2009**, *130*, 204111.
- (113) Stevenson, J. D.; Wales, D. J. Communication: Analysing kinetic transition networks for rare events. *J. Chem. Phys.* **2014**, *141*, 041104.
- (114) Carr, J. M.; Wales, D. J. Refined kinetic transition networks for the GB1 hairpin peptide. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3341–3354.
- (115) Krivov, S. V.; Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14766–14770.
- (116) Krivov, S. V.; Karplus, M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B* **2006**, *110*, 12689–12698.
- (117) Krivov, S. V.; Karplus, M. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13841–13846.
- (118) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002–5011.

- (119) 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.
- (120) Krivov, S. V.; Karplus, M. Free energy disconnectivity graphs: Application to peptide models. *J. Chem. Phys.* **2002**, *117*, 10894–10903.
- (121) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.
- (122) Muff, S.; Caffisch, A. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β -sheet miniprotein. *Proteins: Struct., Func. and Bioinf.* **2008**, *70*, 1185–1195.
- (123) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.
- (124) Wales, D. J. Perspective: Insight into Reaction Coordinates and Dynamics from the Potential Energy Landscape. *J. Chem. Phys.* **2015**, *142*, 130901.
- (125) Doye, J. P. K.; Miller, M. A.; Wales, D. J. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* **1999**, *110*, 6896–6906.
- (126) Neirrotti, J. P.; Calvo, F.; Freeman, D. L.; Doll, J. D. Phase changes in 38 atom Lennard-Jones clusters. I: A parallel tempering study in the canonical ensemble. *J. Chem. Phys.* **2000**, *112*, 10340–10349.
- (127) Calvo, F.; Neirrotti, J. P.; Freeman, D. L.; Doll, J. D. Phase changes in 38 atom Lennard-Jones clusters. II: A parallel tempering study of equilibrium and dynamic properties in the molecular dynamics and microcanonical ensembles. *J. Chem. Phys.* **2000**, *112*, 10350–10357.

- (128) Mandelshtam, V. A.; Frantsuzov, P. A.; Calvo, F. Structural Transitions and Melting in LJ 74 - 78 Lennard-Jones Clusters from Adaptive Exchange Monte Carlo Simulations. *J. Phys. Chem. A* **2006**, *110*, 5326–5332.
- (129) Sharapov, V. A.; Mandelshtam, V. A. Solid-Solid Structural Transformations in Lennard-Jones Clusters: Accurate Simulations versus the Harmonic Superposition Approximation. *J. Phys. Chem. A* **2007**, *111*, 10284–10291.
- (130) Calvo, F. Free-energy landscapes from adaptively biased methods: Application to quantum systems. *Phys. Rev. E* **2010**, *82*, 046703.
- (131) Sehgal, R. M.; Maroudas, D.; Ford, D. M. Phase behavior of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* **2014**, *140*, 104312.
- (132) Picciani, M.; Athenes, M.; Kurchan, J.; Tailleur, J. Simulating structural transitions by direct transition current sampling: The example of LJ 38. *J. Chem. Phys.* **2011**, *135*, 034108.
- (133) Malolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2010**, *31*, 1402–1409.
- (134) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Loughton, 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.
- (135) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Šponer, J. Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.

- (136) Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* **1988**, *200*, 223–238.
- (137) Andreatta, D.; Sen, S.; Lustres, J. L. P.; Kovalenko, S. A.; Ernsting, N. P.; Murphy, C. J.; Coleman, R. S.; Berg, M. A. Ultrafast Dynamics in DNA: Fraying at the End of the Helix. *J. Am. Chem. Soc.* **2006**, *128*, 6885–6892.
- (138) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankaš, F.; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177–3189.
- (139) Dršata, T.; Pérez, A.; Orozco, M.; Morozov, A. V.; Šponer, J.; Lankaš, F. Structure, Stiffness and Substates of the Dickerson-Drew Dodecamer. *J. Chem. Theory Comput.* **2013**, *9*, 707–721.
- (140) Dans, P. D.; Danilne, L.; Ivani, I.; Dršata, T.; Lankaš, F.; Hospital, A.; Walther, J.; Pujagut, R. I.; Battistini, F.; Gelpí, J. L.; Lavery, R.; Orozco, M. Long-timescale dynamics of the DrewDickerson dodecamer. *Nucleic Acids Res.* **2016**, *44*, 4052–4066.
- (141) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; Gonzalez, C.; Orozco, M. How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **2017**, *45*, 4217–4230.
- (142) Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: a refined force field for DNA simulations. *Nat. Meth.* **2016**, *13*, 55–58.
- (143) Wales, D. J.; Head-Gordon, T. Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins. *J. Phys. Chem. B* **2012**, *116*, 8394–8411.

- (144) Stillinger, F. H.; Weber, T. A. packing structures and transitions in liquids and solids. *Science* **1984**, *225*, 983.
- (145) Wales, D. J. Coexistence in small inert gas clusters. *Mol. Phys.* **1993**, *78*, 151.
- (146) Somani, S.; Wales, D. J. Energy landscapes and global thermodynamics for alanine peptides. *J. Chem. Phys.* **2013**, *139*, 121909.
- (147) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.
- (148) Jimènez, J.; Marzal, A. *Algorithm Engineering: 3rd International Workshop, WAE'99, London, UK, July 1999*; Springer, Berlin, 1999; pp 15–29.
- (149) Brown, R. F.; Andrews, C. T.; Elcock, A. H. Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 2315–2328.
- (150) Chan, H. S.; Dill, K. A. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* **1998**, *30*, 2–33.
- (151) Thirumalai, D.; Woodson, S. A. Kinetics of Folding of Proteins and RNA. *Acc. Chem. Res.* **1996**, *29*, 433–439.
- (152) Zhang, W.; Chen, S.-J. RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci. USA* **2001**, *99*, 1931–1936.
- (153) Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring biomolecular energy landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.
- (154) Chebaro, Y.; Ballard, A. J.; Chakraborty, D.; Wales, D. J. Intrinsically Disordered Energy Landscapes. *Sci. Rep.* **2015**, *5*, 10386.

- (155) Banavali, N. K.; MacKerell, A. D. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **2002**, *319*, 141–160.
- (156) Lindahl, V.; Villa, A.; Hess, B. Sequence dependency of canonical base pair opening in the DNA double helix. *PLoS ONE* **2017**, *13*, e1005463.
- (157) Thirumalai, D.; Hyeon, C. In *Non-Protein Coding RNAs*; Walter, N., Woodson, S., Batey, R., Eds.; Springer Series in Biophysics; Springer Berlin Heidelberg, 2009; Vol. 13; pp 27–47.
- (158) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* **1997**, *96*, 14–22.

