

Energy Landscapes of Mini-Dumbbell DNA Octanucleotides

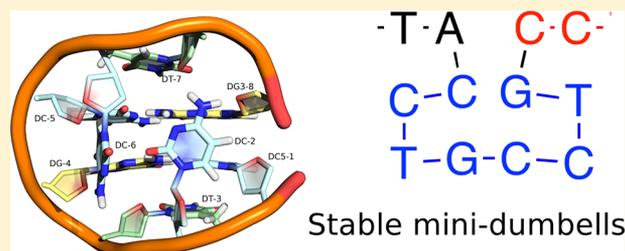
James S. Klimavicz,^{†,‡} Konstantin Röder,[†] and David J. Wales^{*,†}

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

[‡]Department of Entomology, Iowa State University, Ames, Iowa 50011, United States

Supporting Information

ABSTRACT: Single-stranded DNA structures play a significant role in biological systems, in particular during replication, translation, and DNA repair. Tracts of simple repetitive DNA are associated with slipped-strand mispairing, which can lead to genetic diseases. Recent NMR studies of TTTA and CCTG repeats have shown that these sequences form mini-dumbbells (MDBs), leading to frameshift mutations. Here we explore the energy landscapes of $(\text{CCTG})_2$ and $(\text{TTTA})_2$, which are currently the smallest known molecules to form MDBs. While $(\text{CCTG})_2$ MDBs are stable, $(\text{TTTA})_2$ exhibits numerous other structures with lower energies. A key factor identified in the stabilization of MDB structures is the bonding strength between residues 1 and 4, and 5 and 8.



1. INTRODUCTION

In addition to the canonical B-DNA structure, the right-handed double-helical DNA conformation described by Watson and Crick,¹ numerous other stable configurations have been identified, including other helical structures and a plethora of non-helical forms.^{2–5} Although helical DNA mainly exhibits Watson–Crick base pairing, other hydrogen-bonding patterns have been identified,^{6–9} stabilizing nonhelical DNA structures that support the increased flexibility observed for DNA molecules in solution.¹⁰ These alternative base pairings and hydrogen-bondings have significant effects on the double-helical structures,^{11,12} resulting in changes in the size of the major and minor grooves in the double helix, potentially affecting protein binding and DNA translation and replication.^{13,14}

While double-helical B-DNA is by far the most abundant conformation in cells, single-stranded DNA (ssDNA) is also present, particularly during DNA replication, translation, and repair, as the DNA is unwound. Particularly interesting are tracts of simple repetitive DNA (srDNA), which consist of repeats of short base sequences.^{5,15} These srDNAs are associated with slipped-strand mispairings (SSMs), which can promote repeat expansion mutations and are associated with genetic diseases, such as myotonic dystrophy, Huntington's disease, and fragile X syndrome.^{16,17} SSMs typically lead to looped-out bases and hairpins in ssDNA, which upon replication may lead to an expanded DNA sequence (see Figure 1). This process tends to accelerate, with longer srDNA domains being more prone to the formation of SSMs.¹⁵ Alternatively, looping can lead to sequence contractions if bases in the loop are not replicated.¹⁸ Recently, mini-dumbbell (MDB) structures have been proposed for the srDNA sequences $5'-(\text{TTTA})_n-3'$ and $5'-(\text{CCTG})_n-3'$, resulting in a new stable, non-B DNA conformation capable of forming SSMs.^{19,20} As both of these repeat units have four bases, repeat expansions in a coding sequence due to these SSMs will result in

frameshift mutations. Indeed, CCTG repeats in intron 1 of the *ZNF9* gene have been shown to cause myotonic dystrophy type 2 in humans,²¹ while TTTA repeats in the *icaC* gene in *Staphylococcus aureus* result in decreased biofilm production due to the creation of a premature stop codon.²² Other studies have provided evidence that MDB structures are also formed in longer tracts of repeating TTA and CCTG repeat sequences.^{19,23} Clearly, the MDB structures constitute a mechanism leading to disease, and a detailed understanding of the underlying processes is highly desirable.

The use of NMR spectroscopy in combination with computational structure fitting showed that the dimers $(\text{TTTA})_2$ and $(\text{CCTG})_2$ adopt MDB structures and at present are the smallest known molecules to do so.²⁰ These sequences therefore provide interesting test systems to probe the formation of MDBs and gain insight into factors determining their stability.

As the underlying potential energy landscape of a biomolecule encodes all information needed to understand structural, thermodynamic, and kinetic properties,^{24–30} an exploration of the potential energy landscapes for both molecules is desirable, not only to understand the formation of MDBs but also to study the pathways that connect them to other structures. As the experimentally observed multiple time scales for biological processes correspond to high energy barriers between distinct structural ensembles,³¹ enhanced sampling schemes are necessary. Here we employ the potential energy landscape framework,³⁰ which has been successfully applied to a wide range of problems in molecular science, including the transformation between different helical DNA configurations^{32,33} as well as to G-quadruplexes.³⁴

Received: March 13, 2018

Published: May 24, 2018

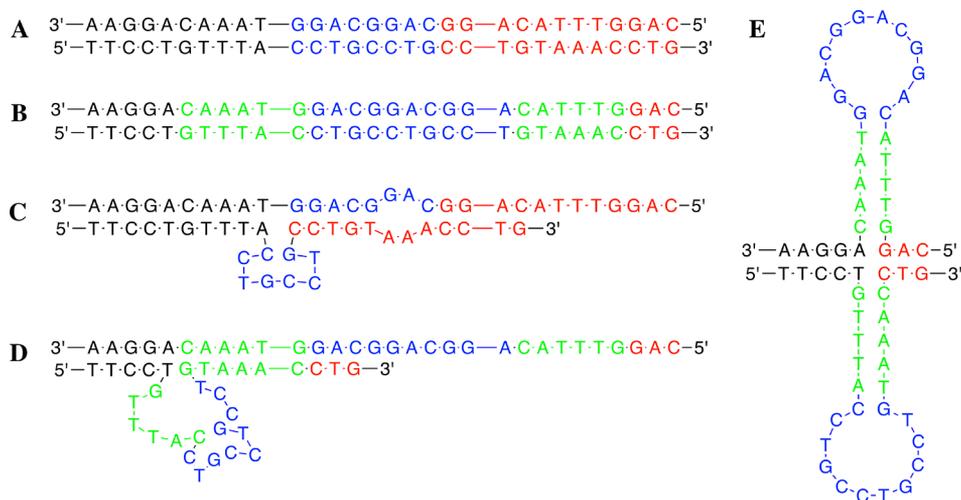


Figure 1. (A) B-DNA sequence with the nucleotides in blue reported to form mini-dumbbells, (B) the same sequence with nucleotides in green forming hairpin loops, (C) SSM with the formation of a mini-dumbbell, (D) larger slip with tertiary structure, and (E) the formation of a DNA dumbbell.

2. METHODOLOGY

2.1. System Description. The initial MDB structures were taken from the Protein Data Bank; the identifiers are SGWQ and SGWL for the TTTA and the CCTG dimers, respectively.²⁰ Single-stranded B-DNA was constructed using the Nucleic Acid Builder in AMBER³⁵ for both sequences. As the experimental work that derived the MDB structures did not use terminal phosphate groups, they are likewise excluded in the present work.

The ff99³⁶ force field with the Barcelona α/γ backbone modification³⁷ and χ modification tuned for DNA³⁸ was employed with a generalized Born solvation model ($igb = 2$)^{39,40} and an effective monovalent salt concentration of 0.1 M. The use of implicit solvent models reduces computational costs, but it introduces additional approximations in the potential model, which can alter the landscapes.⁴¹ Subsequently, the stability of the lowest energy structures identified was probed using molecular dynamics (MD) simulations in explicit solvent. The DNA structures were solvated in TIP3P water within a truncated octahedral bounding box with a solvent hull of at least 10 Å. The system was neutralized with Na⁺ ions, and additional sodium and chloride ions were added to give an effective salt concentration of approximately 0.1 M using appropriate salt parameters.⁴² After initial solvent relaxation, the system was heated to 300 K with harmonic constraints applied to the octanucleotides. The constraints were lowered stepwise and the system equilibrated at 300 K in a thermostated NVT ensemble,⁴³ followed by a pressure equilibration in NVP. The production runs were carried out with a time step of 2 fs using Langevin dynamics.⁴³ The MD simulations were run for 1 μ s, using CUDA-parallelized AMBER on GPUs,^{44,45} providing some measure of stability on short time scales.

2.2. Structure Searches. Basin-hopping global optimization^{46–48} was employed to locate low-energy configurations for both octanucleotides. Group rotation moves^{49,50} were used, including rotations of the nucleobases, as well as rotations around bonds in the phosphate–desoxyribose backbone, as shown in Figure 2.

The initial structure search for both octanucleotides consisted of two sets of 50 runs with 50 000 basin-hopping steps for each sequence, starting from the MDB structure and B-DNA, respectively. Group rotations were attempted every four steps with a probability of 0.025 for the nucleobase rotations and 0.01 for

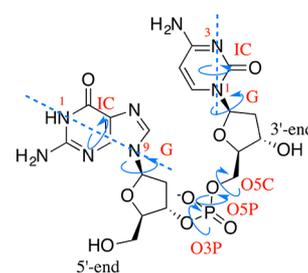


Figure 2. Group rotation moves employed in basin-hopping global optimization. IC and G are the rotations of the bases. O3P, O5P, and OSC are the rotations of the backbone.

the backbone rotations.⁴⁹ The rotations for the bases were unrestricted, while the magnitude of backbone rotations was limited to $\left[-\frac{\pi}{10}, \frac{\pi}{10}\right]$. This restriction was chosen to ensure that the new configurations before quenching were not too distorted by steric clashes, while still allowing for large steps. The minimization used a customized L-BFGS optimizer^{51,52} with a convergence condition for the rms gradient of 10^{-4} kcal mol⁻¹ for the initial search and 10^{-6} kcal mol⁻¹ for the final quenches, which we have found to be sufficient to distinguish structures in the past.⁵³ The resulting structures were tested for chirality inversions and rejected if the chirality changed from the original state.

2.3. Exploration of the Energy Landscapes. The exploration of the energy landscapes was performed by creating a kinetic transition network^{54,55} using discrete path sampling (DPS).^{56,57} Within the DPS scheme, transition state candidates were located using the doubly nudged elastic band (DNEB) algorithm,^{58–60} followed by hybrid eigenvector-following (HEF)⁶¹ to ensure accurate convergence. The corresponding minima for each transition state were characterized by approximate steepest-descent pathways with a convergence criterion of 10^{-6} kcal mol⁻¹ on the rms force. Sampling was initialized from the lowest energy structures found in the global optimizations, and subsequent sampling was carried out to remove artificial kinetic traps⁶² and high energy barriers⁶³ and to improve the connectivity of the landscape. Visualization of the energy landscape uses disconnectivity graphs,^{64,65} and structural representations were created with PyMOL.⁶⁶ The free energy landscapes were obtained from the potential energy landscapes using the harmonic superposition approach.⁶⁷

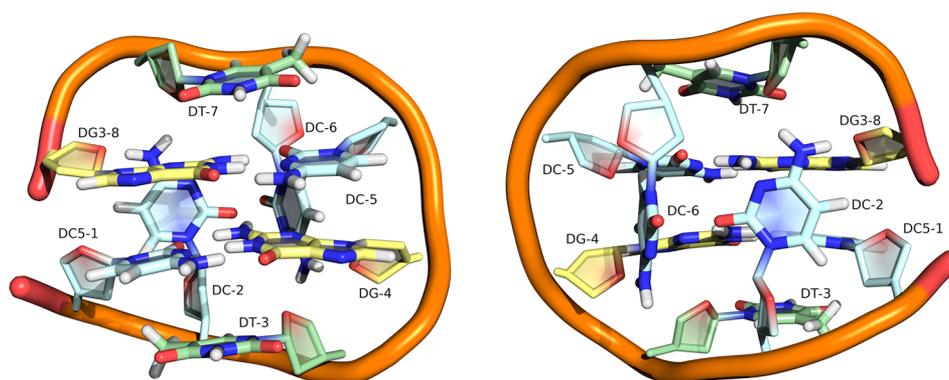


Figure 3. Mini-dumbbell structure for $(\text{CCTG})_2$ viewed from the front (left) and back (right) with labels numbering the nucleotides. In the front view the bonding between 1 and 4 and 5 and 8 is visible; from the back the proximity of nucleotides 2 and 6 is clear.

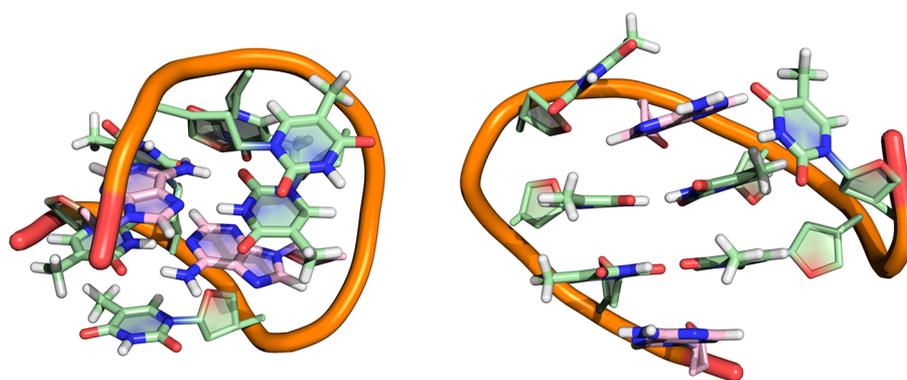


Figure 4. Low-energy structures located for TTTA. Left: MDB-like structure with a Hoogsteen base pair distorting the MDB structure. Right: Typical loop located for TTTA with two thymine–thymine pairs as the key structural element.

2.4. Characterization of MDB Structures. To characterize local minima as MDBs or MDB-like, based on their structure, it is necessary to find suitable diagnostics. We combined several metrics to set up a classification scheme based on machine learning. A structure was considered to be a MDB under the following conditions:

- there is hydrogen-bonding between residues 1 and 4, and 5 and 8;
- residues 3, 1 and 4, 5, and 8, and 7 lie in approximately parallel planes; and
- residues 2 and 6 are close in space, but not hydrogen-bonded.

These criteria are restrictive enough to describe the MDB structures well, while including all previously described consensus structures. An example MDB is shown in Figure 3, including the reference for the nucleotide labeling. For the TTTA dimer, a large number of looplike structures were found; these loops were defined as follows:

- there is hydrogen-bonding between residues 2 and 7 and 3 and 6, with possible hydrogen-bonding between nucleobases 1 and 8;
- the nucleobases pairs 2 and 7, and 3 and 6 lie in parallel planes; and
- residues 4 and 5 are stacked.

These structural requirements can be represented by a set of parameters, namely root-mean-square deviations, Euclidian end-to-end distances, the average diameter of the convex hull of nucleobase atoms, atom deviations from best-fit planes, dihedral angles between nucleobases, and distances between nucleobases.

A training set of 5000 configurations was classified by hand for both of the octamers. A principal component analysis was performed on the parameter space, and components were selected by eigenvalues to account for 95% of the variance observed. Based on these results a bag of decision trees classifier^{68,69} was implemented to assign structures. The initial training set data were verified by the classifier described above.

3. RESULTS

3.1. Low-Energy Structures. The study of low-energy structures located by the basin-hopping runs reveals an important difference between the two sequences. For $(\text{CCTG})_2$ the runs initiated from MDB structures remained in MDB configurations, apart from one run, which produced a slightly distorted MDB with comparable energy. In contrast, the runs for $(\text{TTTA})_2$ did not produce any MDB structures, but instead looped structures were located significantly lower in energy. Examples of both conformational forms are given in Figure 4. Interestingly, some MDB-like structures were found to be similar in energy to the MDBs, with Hoogsteen base pairing between nucleotides 1 and 4, or 5 and 8.

Basin-hopping runs initiated from B-DNA did not produce any minima lower in energy than the MDB structure for $(\text{CCTG})_2$. Furthermore, no low-energy MDB minima were located for $(\text{TTTA})_2$. Overall, these results suggest that for $(\text{CCTG})_2$ MDB structures may be the global minimum conformation whereas for $(\text{TTTA})_2$ this is certainly not the case. This picture is further supported by explicit solvent MD simulations. While the MDB is stable for $(\text{CCTG})_2$ over 1 μs , $(\text{TTTA})_2$ loses the MDB structure at 300 K

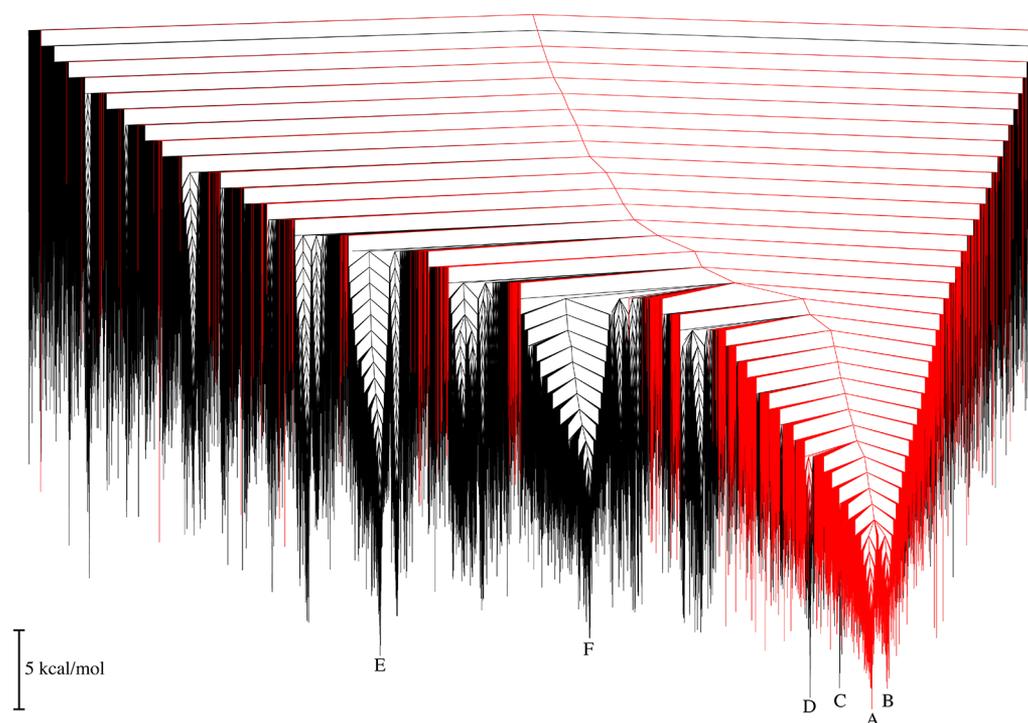


Figure 5. Disconnectivity graph for $(\text{CCTG})_2$, with the MDB structures highlighted in red, showing that the lowest energy funnel contains mainly MDB structures. There are other structures with low energies in distinct funnels, giving the landscape a relatively frustrated appearance.^{26,70} The labels correspond to the structures in Figure 7.

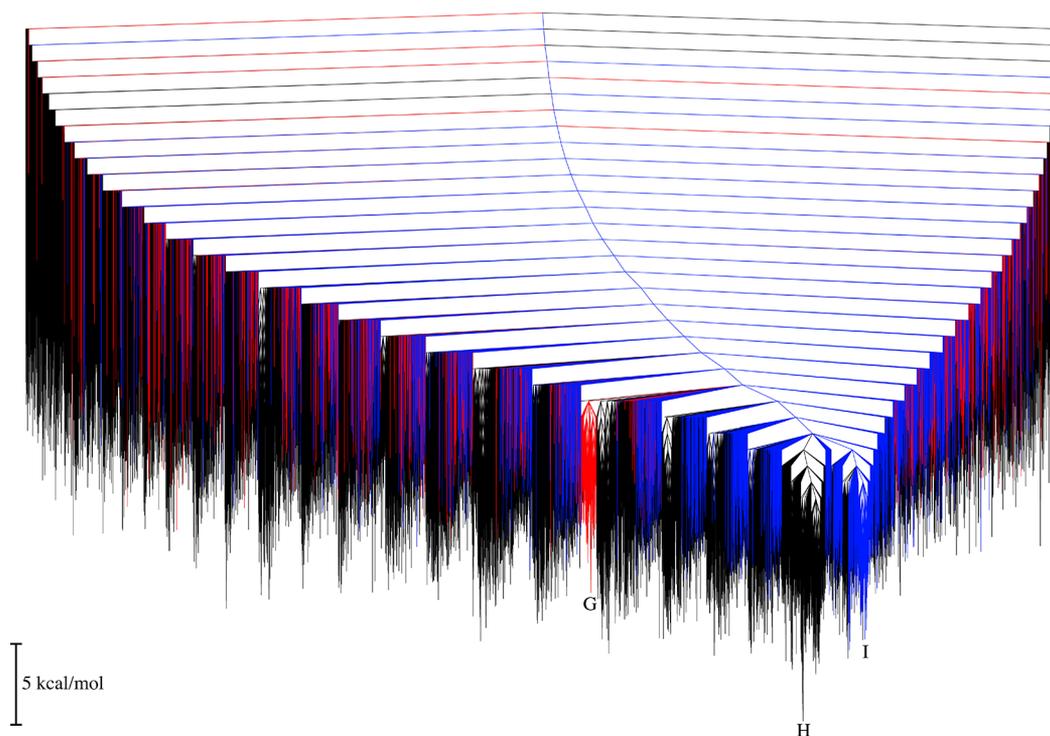


Figure 6. Disconnectivity graph for $(\text{TTTA})_2$, again with MDB structures colored in red, showing that the global minimum structure is not the MDB, consistent with the global optimization results. Minima corresponding to loops are colored in blue. Overall the landscape is less multifunneled than for $(\text{CCTG})_2$, although there is clearly some interesting substructure. The labels correspond to the structures given in Figure 7.

after about 300 ns. Furthermore, while $(\text{CCTG})_2$ retains MDB characteristics for approximately 500 ns at 350 K, all MDB characteristics are lost within 50 ns for $(\text{TTTA})_2$ at this temperature.

3.2. The Potential Energy Landscapes. The disconnectivity graphs for the potential energy landscapes of $(\text{CCTG})_2$ and

$(\text{TTTA})_2$ are shown in Figure 5 and Figure 6, respectively, with key structures illustrated in Figure 7.

The potential energy landscapes reveal significant differences between the two systems, in particular with respect to the stability of MDB structures and the overall organization.

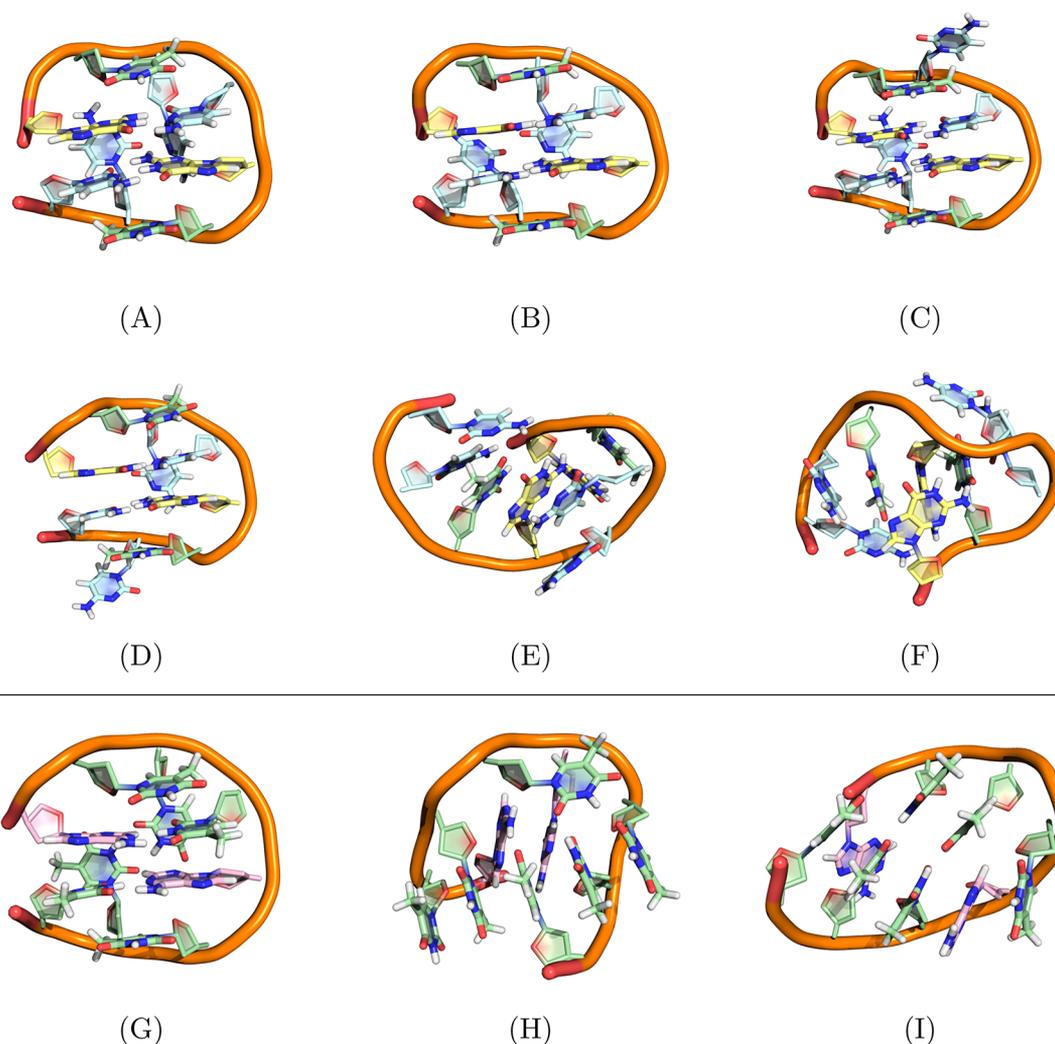


Figure 7. Key structures illustrated in the disconnectivity graphs (Figures 5 and 6) for $(\text{CCTG})_2$ (top and middle) and $(\text{TTTA})_2$ (bottom). A, B, and G are MDB structures, and C and D are MDB-like, with rotations of residues 6 and 2, respectively. E and F are additional low-energy structures located for $(\text{CCTG})_2$. H is the lowest energy minimum located for $(\text{TTTA})_2$. Low-energy structures, such as these three, generally include five π -stacked nucleotides, with hydrogen-bonds to the other three bases. I is one of the loop structures found for $(\text{TTTA})_2$.

The landscape for $(\text{CCTG})_2$ has multifunnel character, with different structures of comparable energies. The MDB and MDB-like conformations (structures A–D) are all in the lowest energy funnel, which contains the global minimum, a MDB conformation. While some flexibility of the nucleotides 2 and 6 is observed, the MDB otherwise exhibits a consistent pattern of interactions based around the two CG base pairs (nucleotides 1 and 4, and 5 and 8). The landscape is multifunneled, with other low-energy structures exhibiting π -stacking of nucleotides (structures E and F). The MDB appears to be the most stable structure and should be kinetically accessible and observable in experiment.

The landscape for $(\text{TTTA})_2$ exhibits less multifunnel character and frustration. While there are a number of small funnels, they contain fewer minima and are not as deep as those found for $(\text{CCTG})_2$. The MDB configurations are not the lowest energy structures, and there are fewer of them on the landscape. Although various loop structures are low in energy, the global minimum (H) is π -stacked. These structures are observed for both sequences, but the stabilization of the MDB structures by the CG base pair in $(\text{CCTG})_2$ appears more favorable than for the TA base pairs in $(\text{TTTA})_2$. This result is probably

a consequence of the greater number of hydrogen-bonds for CG.

4. CONCLUSIONS

The present results indicate that the MDB conformation is not the most stable for the $(\text{TTTA})_2$ octanucleotide. Many lower energy structures were found using basin-hopping, MD simulations, and discrete path sampling. Additionally, the MDB structures that do exist reside in a small basin, decreasing the overall probability of observing this morphology. In contrast, the $(\text{CCTG})_2$ MDB structures are stable, forming a large funnel on the energy landscape, which includes the global minimum.

The key structural feature of the MDB structures, namely the base pairing of residues 1 and 4, and 5 and 8, seems to be the key factor that determines this difference in stability, suggesting possible constraints for MDB-forming sequences. The stability of MDBs formed by one of the two sequences clearly shows that these structures represent another nonhelical motif, which may be relevant for biological processes. In the future, larger sequences should be considered, along with the formation of MDB structures from B-DNA, and their relative stability compared to other nonhelical motifs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00262.

Further information on the classification of structures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dw34@cam.ac.uk.

ORCID

Konstantin Röder: 0000-0003-2021-9504

David J. Wales: 0000-0002-3555-6645

Notes

The authors declare no competing financial interest.

The research data is available at <http://doi.org/10.5281/zenodo.1197178>.

■ ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Engineering and Physical Sciences Research Council.

■ 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) Wells, R. D. Unusual DNA structures. *J. Biol. Chem.* **1988**, *263*, 1095–1098.
- (3) Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 364–366.
- (4) Mirkin, S. M. Discovery of alternative DNA structures: a heroic decade (1979–1989). *Front. Biosci., Landmark Ed.* **2008**, *13*, 1064–1071.
- (5) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **2010**, *67*, 43–62.
- (6) Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1963**, *16*, 907–916.
- (7) Kang, C. H.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. Crystal structure of four-stranded *Oxytricha telomeric* DNA. *Nature* **1992**, *356*, 126–131.
- (8) Leontis, N. B.; Stombaugh, J.; Westhof, E. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.
- (9) 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.
- (10) Rich, A.; Nordheim, A.; Wang, A. H. J. The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **1984**, *53*, 791–846.
- (11) 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.
- (12) Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893–5909.
- (13) Pabo, C. O.; Sauer, R. T. Protein–DNA recognition. *Annu. Rev. Biochem.* **1984**, *53*, 293–321.
- (14) Zimmer, C.; Wähnert, U. Nonintercalating DNA-binding ligands: Specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* **1986**, *47*, 31–112.
- (15) Levinson, G.; Gutman, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **1987**, *4*, 203–221.

(16) Bacolla, A.; Wells, R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **2004**, *279*, 47411–47414.

(17) Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **2007**, *447*, 932–940.

(18) Strand, M.; Prolla, T. A.; Liskay, R. M.; Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **1993**, *365*, 274–276.

(19) Guo, P.; Lam, S. L. The competing mini-dumbbell mechanism: new insights into CCTG repeat expansion. *Signal Transduct. Target. Ther.* **2016**, *1*, 16028.

(20) Guo, P.; Lam, S. L. Minidumbbell: A new form of native DNA structure. *J. Am. Chem. Soc.* **2016**, *138*, 12534–12540.

(21) Liquori, C. L.; Ricker, K.; Moseley, M. L.; Jacobsen, J. F.; Kress, W.; Naylor, S. L.; Day, J. W.; Ranum, L. P. W. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **2001**, *293*, 864–867.

(22) Guo, P.; Lam, S. L. Unusual structures of TTTA repeats in *icaC* gene of *Staphylococcus aureus*. *FEBS Lett.* **2015**, *589*, 1296–1300.

(23) Blommers, M. J. J.; Walters, J. A. L. I.; Haasnoot, C. A. G.; Aelen, J. M. A.; Van der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. Effects of base sequence on the loop folding in DNA hairpins. *Biochemistry* **1989**, *28*, 7491–7498.

(24) Dobson, C. M.; Karplus, M. The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **1999**, *9*, 92–101.

(25) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.

(26) Wolynes, P.; Onuchic, J.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619–1620.

(27) Wolynes, P. G. Energy landscapes and solved protein-folding problems. *Philos. Trans. R. Soc., A* **2005**, *363*, 453–467.

(28) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.

(29) Wolynes, P. G. Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* **2015**, *119*, 218–230.

(30) Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring biomolecular energy landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.

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

(32) Chakraborty, D.; Wales, D. J. Probing helical transitions in a DNA duplex. *Phys. Chem. Chem. Phys.* **2017**, *19*, 878–892.

(33) 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.

(34) Cragolini, T.; Chakraborty, D.; Šponer, J.; Derreumaux, P.; Pasquali, S.; Wales, D. J. Multifunctional energy landscape for a DNA G-quadruplex: An evolved molecular switch. *J. Chem. Phys.* **2017**, *147*, 152715.

(35) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. *AMBER12*; 2012.

(36) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(37) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, 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.

(38) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, 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.

(39) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(40) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 383–394.

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

(42) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(43) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylananyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.

(44) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(45) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(46) 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.

(47) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 6611–6615.

(48) Li, Z.; Scheraga, H. A. Structure and free energy of complex thermodynamic systems. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 333.

(49) Mochizuki, K.; Whittleston, C. S.; Somani, S.; Kusumaatmaja, H.; Wales, D. J. A conformational factorisation approach for estimating the binding free energies of macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2842–2853.

(50) Oakley, M. T.; Johnston, R. L. Energy landscapes and global optimization of self-assembling cyclic peptides. *J. Chem. Theory Comput.* **2014**, *10*, 1810–1816.

(51) Nocedal, J. Updating quasi-newton matrices with limited storage. *Math. Comput.* **1980**, *35*, 773.

(52) Liu, D.; Nocedal, J. On the limited memory bfgs method for large scale optimization. *Math. Prog.* **1989**, *45*, 503.

(53) Somani, S.; Wales, D. J. Energy landscapes and global thermodynamics for alanine peptides. *J. Chem. Phys.* **2013**, *139*, 121909.

(54) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.

(55) Wales, D. J. Energy landscapes: some new horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.

(56) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.

(57) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(58) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(59) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(60) Trygubenko, S. A.; Wales, D. J. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

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

(62) Strodel, B.; Whittleston, C. S.; Wales, D. J. Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(63) Carr, J. M.; Wales, D. J. Global optimization and folding pathways of selected alpha-helical proteins. *J. Chem. Phys.* **2005**, *123*, 234901.

(64) 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.

(65) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.

(66) Schrödinger, L. L. C. *PyMOL Molecular Graphics System*, version 1.6.x; 2013.

(67) 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.

(68) Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.

(69) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

(70) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524–7528.