Energy Landscapes of Deoxyxylo- and Xylo-Nucleic Acid Octamers

Daniel J. Sharpe, Konstantin Röder, and David J. Wales*

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

E-mail: dw34@cam.ac.uk

Phone: +44-1223-336354

Abstract

Artificial analogues of the natural nucleic acids have attracted interest as a diverse class of information storage molecules capable of self-replication. In this study, we use the computational potential energy landscape framework to investigate the structural and dynamical properties of xylo- and deoxyxylo-nucleic acids (XyNA and dXyNA), which are derived from their respective RNA and DNA analogues by inversion of a single chiral centre in the sugar moiety of the nucleotides. For an octameric XyNA sequence and the analogue dXyNA, we observe facile conformational transitions between a left-handed helix, which is the free energy global minimum, and a ladder-type structure with approximately zero helicity. The competing ensembles are better separated in the dXyNA, making it a more suitable candidate for a molecular switch, whereas the XyNA exhibits additional flexibility. Both energy landscapes exhibit greater frustration than we observe in RNA or DNA, in agreement with the higher degree of optimisation expected from the principle of minimal frustration in evolved biomolecules.

Introduction

Xeno-nucleic acids (XNAs) are a diverse family of nucleic acid structures derived from DNA or RNA by chemical modification of the sugar moiety of nucleotides.[1] Evolution and heredity have been characterised for these molecules, with associated specific folding and binding properties.[2] XNAs have emerged with important medical applications,[3] for example as aptamers,[4–6] synthetic ribozymes,[7] artificial small interfering RNAs, and antisense oligonucleotides for the targeting of microRNAs.[8,9] The resistance of XNAs to endonucleases, a result of the inability of natural enzymes to recognise the modified nucleic acid structures, is a particularly valuable advantage of XNAs in replacing natural nucleic acids for therapeutic purposes.[10] Recently, it was demonstrated that XNA aptamers are capable of recognising small molecules, further advancing the potential therapeutic use of such artificial nucleic acids.[11]

XNAs are also of current interest in the emergent fields of synthetic biology,[12,13] which demands the development of chemical information storage systems capable of self-replication *in vitro* and *in vivo* for artificial life and biological computation, and in systems chemistry,[14] which requires molecular switches for the control of operations in complex chemical networks.

Other potential applications of XNAs include their use as self-assembling nanomaterials, broadening the possible design scope in DNA nanotechnology, chemical sensors, and catalysts.[15,16] The study of XNAs is also motivated by the fundamental question of the origins of life, where it is important to understand the factors that led to evolution selecting ribofuranosyl nucleic acids as the genetic biopolymer for the basis of life, and where it remains unknown if an alternative nucleic acid was utilised in hypothetical organisms preceding those based on RNA.[7] Despite their importance, there are few reported structures of XNAs, and relatively little is known concerning the structural and dynamical properties of XNAs in atomistic detail.[17] Furthermore, computational studies have been thus far largely limited to molecular dynamics (MD) simulations, without the application of enhanced sampling schemes, which are required to overcome broken ergodicity.[18]

The present study focuses on nucleic acids based on xylose (XyNA) and deoxyxylose (dXyNA), referred to collectively as XyNAs, which represent some of the simplest possible perturbations to the chemical structure of natural nucleic acids. Xylose is derived from ribose by inversion of configuration at the C3′ atom of the sugar moiety (Fig. 1), and likewise deoxyxylose is derived from deoxyribose. Xylose is a thermodynamic product of the formose reaction,[19] the most probable prebiotic route of sugar synthesis,[20] and so XyNAs represent arguably the most credible possibility of a genetic biopolymer adopted by prebiotic organisms that are speculative precursors to RNA-based life forms.

While dXyNA-homoduplexes exhibit similar thermal stability to their DNA equivalents,[21] thermal denaturation studies have demonstrated that for dXyNA:DNA hybrid duplexes the thermodynamic stability is markedly lowered.[22] In this respect, dXyNA exhibits complementary properties to many other XNAs, which can form a stable duplex through hybridisation with DNA or RNA, with strong discrimination against mismatches.[17] While this behaviour precludes the use of XyNAs as an aptamer, and for other applications requiring sequence-specific binding, it is an ideal property if XyNAs are to be utilised alongside and independent of natural nucleic acids as an orthogonal information system.

Circular dichroism (CD) studies have shown that XyNA and dXyNA oligomers may adopt a left-handed helical duplex structure, or a structure with an apparent lack of helicity, and that the observed structure depends on base sequence and sequence length, as well as external factors including temperature and salt concentration.[21,23,24]

MD simulations of XyNA and dXyNA duplexes of length 8, 13 and 29 base pairs (bp) have revealed the existence of an intermediate between right- and left-handed helical conformations.[25,26] The observed timescale for this transition is of the order of tens of nanoseconds. For XyNA duplexes, it was observed that the left-handed helical structure is not stable, but rather that the terminal regions of the duplex undergo oscillatory movements from coiled to uncoiled states that act to repeatedly screw and unscrew the helix, suggesting structural competition between left-handed helical and 'linear' ladder-type structures. This competi-

tion between ensembles makes XyNAs a potential molecular switch, and it is also a property unique to XyNAs among the known XNAs.[17] Other recent MD studies of XyNA duplexes in the presence of a carbon nanotube have demonstrated fast spontaneous unzipping as a consequence of the strained backbone, highlighting the potential of XyNAs in gene delivery for therapeutic purposes.[27,28]

Here, the computational potential energy landscape framework[29,30] is employed to investigate the structural and dynamical properties of XyNA and dXyNA duplexes. Discrete path sampling (DPS)[31,32] uses geometry optimisation to locate transition states and the local minima they connect, enabling exploration of the energy landscape independent of free energy barrier heights. From the resulting databases we can estimate and compare the relative thermodynamic stabilities of the three known major conformations, namely left-handed helical, right-handed helical,, and ladder-type structures. The free energy barriers partitioning these major conformational ensembles determine the applicability of XyNA and dXyNA duplexes as molecular switches, which requires two competing funnels to be separated by a barrier that is surmountable at ambient temperatures. Visualisation of the free energy landscapes will clearly elucidate structural differences between XyNA and dXyNA duplexes, and between them and their naturally evolved counterparts. This computational framework has been successfully applied to a variety of biomolecular systems before,[30,33] including the B- to Z-DNA transition[34] and the formation of DNA mini-dumbbells.[35]

Methods

Force field and initial structures

The xylo- and deoxyxylo-nucleotide monomer units were constructed with the furanose moiety in the C3′-*endo* (*cf.* canonical A-RNA) and C2′-*endo* conformations, respectively, using the LEAP program of AMBER.[37] Force field parameters were taken from the parm99 force field[38] incorporating the bsc0 correction[39] for α and γ backbone torsion angles. This choice

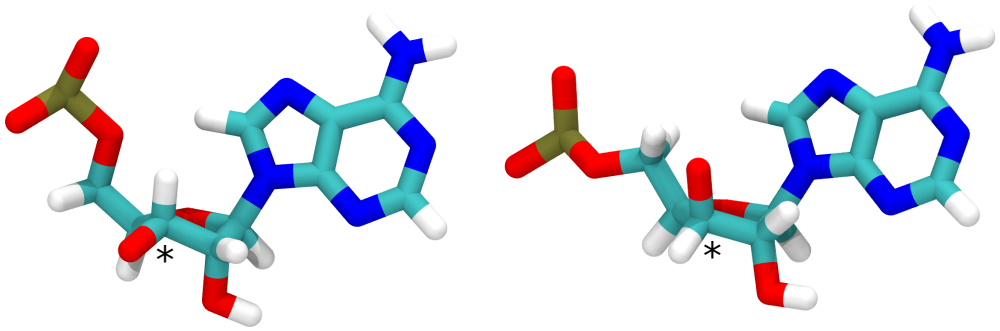


Figure 1: Comparison of a ribonucleotide monomer unit as it appears in canonical A-RNA (left), and a xylonucleotide monomer unit as it appears in the ladder-type conformation of the XyNA1 sequence observed by NMR (right).[36] Xylose is derived from ribose by an inversion of configuration at the C3′ chiral centre, indicated by (∗). Note that the sugar moieties in both units exist in the C3′-*endo* conformation, so that the O3′ atom of XyNA units is axial, and that the glycosidic torsion angle differs between the two units.

of parameters was used in previous MD studies of XyNAs,[25,26] and has been employed in a number of studies of other artificial nucleic acid systems.[40–42] The bsc0 correction is acknowledged to yield a general improvement in the description of the behaviour of both DNA and RNA, including noncanonical structures, and we therefore chose to use it in the present work. As it was previously shown that different parameterisations for the glycosidic torsion angle χ have little effect on the MD trajectories of XyNA sequences,[26] no such reparametrisation was used here. Partial charges were obtained by the two-stage RESP fitting procedure[43] at the HF/6-31G* level of theory using ANTECHAMBER.[44] The potential function was correctly symmetrised.[45,46] To represent an aqueous solution environment, we used a generalised Born implicit solvent model with surface area term[47,48] and an effective monovalent salt concentration of 0.1 M maintained using the Debye-Hückel approximation.[49]

The complete right-handed duplexes of the octameric sequence (5′-3′)[xG-xU-xG-xU-xA-xC-xA-xC-T] (XyNA1) and its deoxyxylose analogue (dXyNA1) were constructed based on a template of canonical B-DNA produced with NAB.[50] Ladder-type structures were obtained from the NMR solution structure of the XyNA oligomer (PDB: 2N4J).[36] From this conformation a left-handed helical structure was obtained by short timescale explicit solvent

MD simulation.

Exploration of the energy landscapes

The three distinct initial structures were used to initiate basin-hopping (BH) global optimisation[51–53] to obtain the lowest-energy structures of the right-handed helical, left-handed helical and ladder-type conformations. We employed group rotation moves[54,55] and simple atomic displacements, and discarded all structures with inverted chirality at any centre.

The potential energy landscapes were explored using discrete path sampling (DPS)[31,32] to create kinetic transition networks (KTNs),[56,57] using the low-energy structures found in BH as starting points. Transition states were located with the doubly-nudged[58,59] elastic band[60,61] algorithm and converged with hybrid eigenvector-following.[62–64] Local minima were then characterised by approximate steepest-descent paths, using a modified version of the L-BFGS algorithm.[65,66] More details of these methods can be found in various reviews.[29,30,33]

After initial pathways were obtained, we continued the sampling to improve the connectivity of the landscape,[67] remove artificial kinetic traps and high energy barriers,[68] and shorten path lengths.[69] Free energy landscapes at 298 K were calculated using the harmonic superposition approximation.[70] A self-consistent recursive regrouping scheme based on a specified free energy barrier threshold was used to lump minima into free energy macrostates.[71]

Analysis of the energy landscapes

We visualise the free energy landscapes using disconnectivity graphs,[72,73] where each leaf corresponds to a free energy group, and is coloured according to the value of an appropriate order parameter for a representative potential energy minimum of that group. The chosen order parameter is the helical handedness H,[74] in the form originally proposed for describing the B to Z transition in DNA duplexes.[75] The magnitude of H is a measure of the number of turns associated with a helix. Values $H > 0$ and $H < 0$ correspond to right- and left-handed double helical turns, respectively, while values $H \approx 0$ indicate approximately zero helicity.

Processing of the data in the KTN was achieved with the CPPTRAJ module[76] of AMBER. The CURVES+ program[77] was used to extract bp-axis, inter-bp and intra-bp geometric parameters.[78] All analyses exclude the terminal base pairs to reduce the influence of terminal effects. Molecular graphics images were produced using VMD[79] and UCSF Chimera.[80]

Results

Free energy landscapes

The free energy disconnectivity graphs at 298 K for XyNA1 and dXyNA1 are shown in Fig. 2 (a) and (b), respectively. For both duplexes, the right-handed helical structures correspond to regions of the landscape with high free energy, so that the occupation probability for right-handed helices is very low. The inversion of chirality in the nucleotide units, with respect to natural nucleic acids, seeds an inversion of preferred helical sense to favour a stable left-handed helix, and XyNAs can be thought of as effective two-state systems, with a dynamic equilibrium between left-handed helical and ladder-type structures.

There are striking differences between the free energy landscapes for XyNA and dXyNA duplexes. In particular, the ideal left-handed helical state of the dXyNA1 duplex is a relatively well-defined global free energy minimum on the landscape, around $5\,\mathrm{kcal\,mol^{-1}}$ more stable than the next lowest minimum, which is the most stable ladder-type structure. The free energy barrier for the conversion of the left-handed helical to the ladder-type structure is around $20\,\mathrm{kcal\,mol^{-1}}$, and for the reverse transition is around $15\,\mathrm{kcal\,mol^{-1}}$. In contrast, the free energy barriers separating the two corresponding states of the XyNA1 duplex are around $10\,\mathrm{kcal\,mol^{-1}}$ in both directions, and the free energy difference is less than $2\,\mathrm{kcal\,mol^{-1}}$.

Another immediately apparent difference between the free energy landscapes of the two systems is in the distribution of values for the helical handedness order parameter. For the XyNA duplex, a much broader range of values for H is observed, from $H \approx 0$ to $H \approx -3.0$ in the low-energy region of the landscape. In contrast, the low free energy region

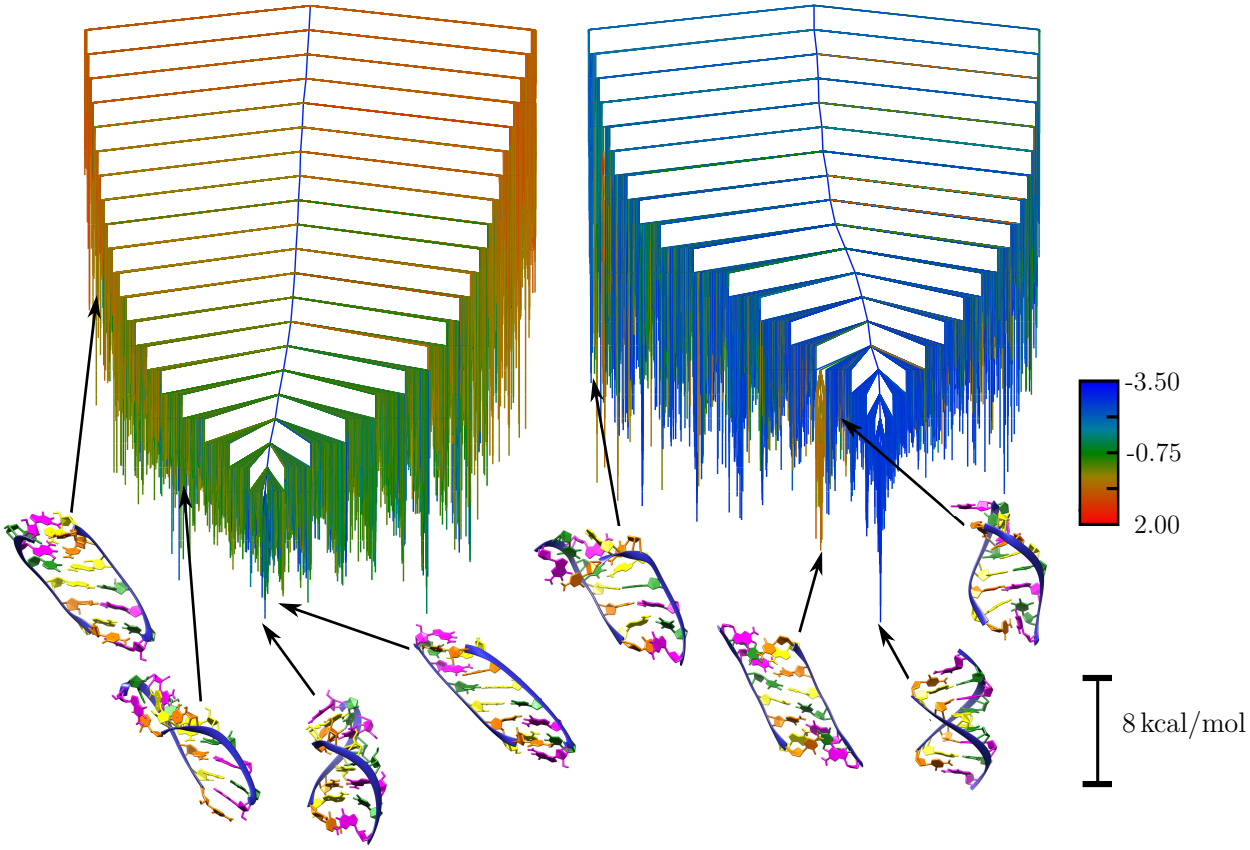
of the landscape for the dXyNA duplex is dominated by left-handed helical structures, with handedness in a narrower range around $H \approx -3.0$. Thus XyNA duplexes are more flexible than their dXyNA analogues.

The ideal ladder-type structure of the XyNA1 duplex compares favourably with the NMR solution structure.[36] As noted by Maiti *et al.*,[24] the ladder-type structures of the XyNA and dXyNA duplexes are stabilised by the interstrand stacking of adjacent bases, which arises due to the strong inclination of bases with respect to the helical axis (Fig. 3). However, the ladder-type structure is predicted to have marginal left-handed helicity, as opposed to the marginal right-handed helicity in the observed structure. The overstabilisation of left-handed helical states of XyNA within the force field parameterisation implemented in this work is also evidenced by the fact that the left-handed helical structure is erroneously predicted to be the global free energy minimum, although it is not far below the idealised ladder-type structure. The force field is apparently more accurate in reproducing the behaviour of dXyNA duplexes, where there appears to be no such bias, and the ladder-type structures are correctly predicted to have marginal right-handed helicity.

The landscapes of both XyNA1 and dXyNA1 duplexes are significantly frustrated,[81,82] with a number of low-energy minima separated by high barriers. This result contrasts with the free energy landscape reported in a study of the analogous B to Z transition in CG-rich DNA sequences, where there is strong funnelling to the native B-DNA state. Frustration in the free energy landscapes of XyNAs is probably attributable to the geometrical frustration of base pairs that prevents formation of a properly extended linear duplex. Instead, necessitates that the ladder-type conformation of XyNA duplexes partly bends inwards on itself to maintain optimal Watson-Crick base-pairing of all base pairs, which drives unwinding and rewinding of the left-handed helix. This geometric frustration may be relieved by the adoption of noncanonical base pairings at one or both of the duplex termini, which then allows the linear structures to form a properly extended conformation.

Evolved biomolecules often appear to follow the principle of minimal frustration.[81–83] The

evolutionary process favours energy landscapes with a distinct bias towards a well-defined native state, and promotes the elimination of kinetic traps on the pathways to this state. This principle, originally formulated for proteins, may also apply to nucleic acids,[84,85] and can be extended to multi-state systems.[86] The fact that the energy landscapes of XyNAs are significantly more frustrated than the landscapes of DNA and RNA represents an evolutionary argument for the adoption of DNA and RNA over XyNAs.



(a) XyNA octamer (b) dXyNA octamer

Figure 2: Free energy landscape for the duplex octamers for (a) XyNA and (b) dXyNA at a temperature of 298 K, calculated using a regrouping threshold of 4 kcal mol^{-1} and an energy increment of 2 kcal mol^{-1}. The branches are coloured according to the helical handedness (H) of a single potential energy minimum representative of each free energy group. Some important representative structures from the different conformational ensembles are shown. The four different nucleotides are coloured as follows: XA and dXA in yellow, XC and dXC in orange, XG and dXG in magenta, and XU and dXT in green. The structural representations were created with UCSF Chimera.[80]

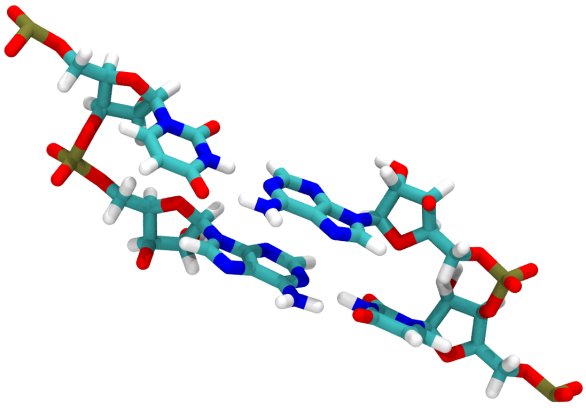


Figure 3: A dinucleotide step in a structure that is part of the lowest free energy group in the conformational ensemble of ladder-type structures for XyNA1. The favourable interstrand stacking of adjacent bases is apparent.

Pathways for helix transformation

Free energy pathways for helical inversion are shown in Fig. 4 (a) and (b)for the XyNA1 and dXyNA1 sequences, respectively, from a representative right-handed helical structure to a left-handed helix representing the global free energy minimum. For both duplexes, the pathways are overall downhill in energy, and the mechanism features only low barriers, leading to fast kinetics. Furthermore, the interconversion of ladder-type and left-handed helical states occurs by helix winding (unwinding) in left- (right-) handed directions, respectively, propagated inwards from one terminus, rather than from both termini simultaneously.

For the XyNA1 duplex, the helical inversion proceeds via a low-energy ladder-type structure, which subsequently evolves to the left-handed helical state via a transition state ensemble of 'kinked' structures. The corresponding pathway for the dXyNA1 duplex is significantly different. In particular, structures with approximately zero helicity represent a high energy transient state in the early stages of the pathway, and so the mechanism is not mediated by a ladder-type intermediate state, in contrast to the XyNA1 duplex. This result again reflects the greater bias towards left-handed helical over ladder-type structures in dXyNA compared to XyNA duplexes. The transition then continues to progress smoothly with respect to handedness, that is, via a more regular helical structure with partial left-handed

helicity.

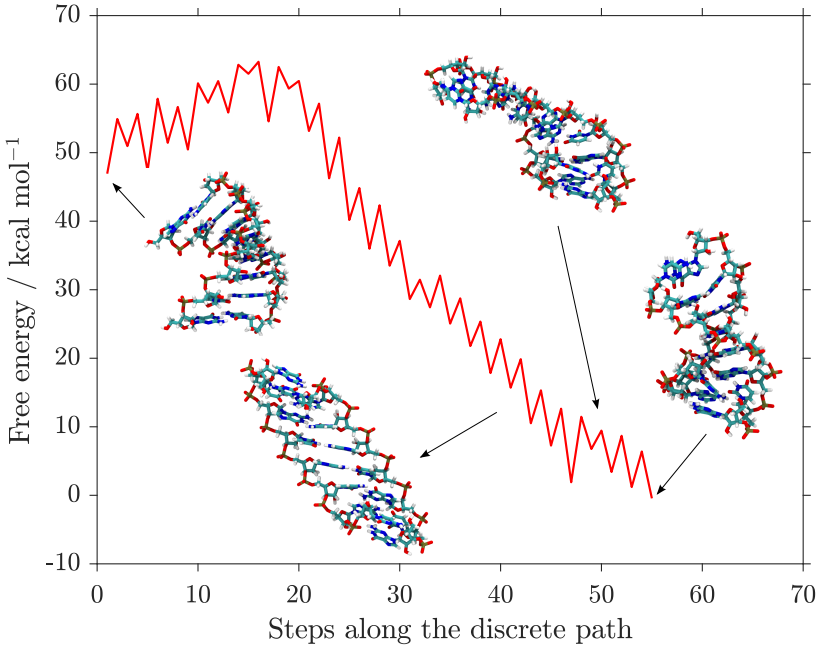
Observed backbone torsional angles

The δ dihedral, with a characteristic value around -40° that effectively defines XyNA and dXyNA duplexes with respect to their natural analogues, remains relatively stable throughout the pathways. The α, β and γ backbone dihedrals for the right-handed helical states of XyNA1 and dXyNA1 adopt values similar to those observed in canonical A-RNA and B-DNA, and undergo sign inversion in the course of the transition to left-handed helical states, seeded by the sign inversion of the δ dihedral with respect to the natural nucleic acid analogues. For the α and β dihedrals, values in the left- and right-handed helical states are of approximately equal magnitude but opposite sign, in both XyNA and dXyNA. The behaviour of the ϵ and ζ dihedrals exhibits less variance along the pathways, although these angles likewise adopt values of opposite sign to the corresponding angles in canonical A-RNA and B-DNA. The glycosidic torsion angle χ takes one of two predominant values in XyNAs, around -160° (*anti*) or -80° (*syn*), and plays an important role in driving the transition.

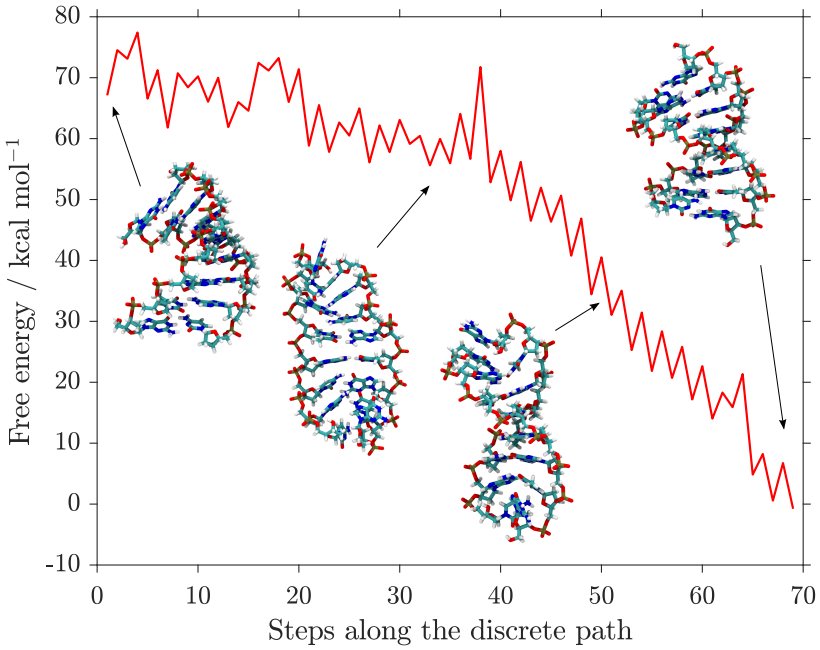
The evolution of inter-bp, intra-bp and bp-axis geometrical parameters[78] along the fastest potential energy pathway for the helical transition of XyNA1 exhibits large scale changes in the bp-axis inclination angle, and in the communicative parameters of inter-bp twist angle, roll angle and slide distance. Values of the helical rise and helical twist are also diagnostic of each of the three major conformations. Plots for the described properties are provided in the supplementary material (Figures S1 and S2).

Conclusions

For XyNA and dXyNA duplexes, we observe an equilibrium between left-handed helical and ladder-type structures. For dXyNA compared to XyNA the global free energy minimum, a left-handed helical structure, is stabilised with respect to the ladder-type structures. One



(a) Fastest pathway energy profile for the helical inversion transition of the XyNA1 duplex. The steps correspond to stationary points, i.e. minimum-transition state-minimum-transition state-$\cdots$-minimum.



(b) Fastest pathway energy profile for the helical inversion transition of the dXyNA1 duplex.

Figure 4: Free energy pathways for the right- to left-handed helical transitions in XyNA and dXyNA duplexes that make the single largest contribution to the steady-state rate constants. Some representative structures that are key intermediates or transition states are included.

possible explanation for this effect is the increased solvation of the C2′ hydroxyl group, present only in XyNAs, in the ladder-type compared to the helical state. The separation between the two ensembles is better defined for dXyNA than for XyNA, both with respect to the magnitude of the free energy barrier separating the basins and with respect to the helical handedness order parameter. Hence the XyNA duplexes are more flexible than dXyNA duplexes, and the latter appear more promising candidates for use as a molecular switch, or as chemical information storage molecules capable of self-replication, where facile unwinding of a helical structure is undesirable.

The free energy landscapes of both XyNA and dXyNA duplexes are significantly frustrated, highlighting an important factor that may have led evolution to select ribofuranosyl nucleic acids, and not the xylose-based analogues, as the genetic basis for life. The origin of the structural competition evident in these systems is the geometrical frustration that prevents Watson-Crick base pairing without inducing strain in the nucleic acid backbone that must be relieved by bending, or else by the adoption of non-canonical base pairing modes at one or both of the duplex termini.

Free energy pathways from a disfavoured right-handed helical state to a left-handed helical state, which is the global free energy minimum, also differ significantly between XyNA and dXyNA duplexes. For dXyNA, extended linear structures represent an early and high free energy transition state along the pathway, which then proceeds via helical structures with a smooth change in the handedness order parameter. For XyNA, the transition is mediated by low-energy ladder-type structures, which transform to left-handed helical structures via a transition state ensemble of 'kinked' structures. Left-handed helix winding and unwinding transitions of XyNA and dXyNA duplexes are driven by the highly flexible terminal base pairs. The inversion of the δ dihedral angle in XyNAs with respect to their natural nucleic acid analogues seeds a direct inversion, not only in the overall helical sense, but also in the backbone dihedral angles. The glycosidic torsion angle also undergoes large scale changes in the course of the transition, as do certain key geometric parameters, most notably the

bp-axis inclination angle.

Future work could address the design and application of XyNA and dXyNA duplexes for molecular devices, for example by investigating the sequence and length-dependence of the propensity for helicity, and how the equilibrium responds to environmental conditions.

Acknowledgement

DJS acknowledges the Cambridge Commonwealth, European and International Trust for a PhD scholarship. KR and DJW acknowledge funding from the Engineering and Physical Sciences Research Council (EP/N035003/1).

Supporting Information Available

Evolution of the bp-axis, inter-bp and intra-bp geometric parameters and backbone torsional angles along the fastest helical inversion pathways for XyNA1 and dXyNA1 duplexes based on potential energy barriers.

The databases, library files and representative structures are available online.[87]

References

(1) Herdewijn, P.; Marlière, P. Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem. Biodivers.* **2009**, *6*, 791–808.

(2) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H. et al. Synthetic genetic polymers capable of heredity and evolution. *Science* **2012**, *336*, 341–344.

(3) Wang, Q.; Chen, L.; Long, Y.; Tian, H.; Wu, J. Molecular beacons of xeno-nucleic acid for detecting nucleic acids. *Theranostics* **2013**, *3*, 395–408.

(4) Appella, D. H. Non-natural nucleic acids for synthetic biology. *Curr. Opin. Chem. Biol.* **2009**, *13*, 687–696.

(5) Taylor, A.; Arangundy-Franklin, S.; Holliger, P. Towards applications of synthetic genetic polymers in diagnosis and therapy. *Curr. Opin. Chem. Biol.* **2014**, *22*, 79–84.

(6) Röthlisberger, P.; Hollenstein, M. Aptamer chemistry. *Adv. Drug Deliver. Rev.* **2018**, *134*, 3–21.

(7) Taylor, A.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.-Y.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P. Catalysts from synthetic genetic polymers. *Nature* **2015**, *518*, 427–430.

(8) Pinheiro, V. B.; Holliger, P. The XNA world: Progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245–252.

(9) Morihiro, K.; Kasahara, Y.; Obika, S. Biological applications of xeno nucleic acids. *Mol. Biosyst.* **2017**, *13*, 235–245.

(10) Chaput, J. C.; Yu, H.; Zhang, S. The emerging world of synthetic genetics. *Chem. Biol.* **2012**, *19*, 1360–1371.

(11) Rangel, A. E.; Chen, Z.; Ayele, T. M.; Heemstra, J. M. In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Res.* **2018**, *46*, 8057–8068.

(12) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **2016**, *14*, 135–149.

(13) Taylor, A. I.; Houlihan, G.; Holliger, P. Beyond DNA and RNA: The expanding toolbox of synthetic genetics. *CSHr Perspect. Biol.* **2019**, *11*, a032490.

(14) Ashkenasy, G.; Hermans, T. M.; Otto, S.; Taylor, A. F. Systems chemistry. *Chem. Soc. Rev.* **2017**, *46*, 2543–2554.

(15) Pinheiro, V. B.; Holliger, P. Towards XNA nanotechnology: New materials from synthetic genetic polymers. *Trends Biotechnol.* **2014**, *32*, 321–328.

(16) Taylor, A.; Beuron, F.; Peak-Chew, S.-Y.; Morris, E. P.; Herdewijn, P.; Holliger, P. Nanostructures from synthetic genetic polymers. *Chembiochem* **2016**, *17*, 1107–1110.

(17) Anosova, I.; Kowal, E. A.; Dunn, M. R.; Chaput, J. C.; Van Horn, W. D.; Egli, M. The structural diversity of artificial genetic polymers. *Nucleic Acids Res.* **2016**, *44*, 1007–1021.

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

(19) Müller, D.; Pitsch, S.; Kittaka, A.; Wagner, E.; Wintner, C. E.; Eschenmoser, A.; Ohlofjgewidmet, G. N. Chemistry of α-aminonitriles – aldomerisation of glycolaldehyde phosphate to rac-hexose 2,4,6-triphosphates and (in presence of formaldehyde) rac-pentose 2,4-diphosphates - rac-allose 2,4,6-triphosphate and rac-ribose 2,4-diphosphate are the main reaction products. *Helv. Chim. Acta.* **1990**, *73*, 1410–1468.

(20) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 39–99.

(21) Seela, F.; Heckel, M.; Rosemeyer, H. Xylose-DNA containing the four natural bases. *Helv. Chim. Acta* **1996**, *79*, 1451–1461.

(22) Babu, B. R.; Raunak,; Poopeiko, N. E.; Juhl, M.; Bond, A. D.; Parmar, V. S.; Wengel, J. XNA (xylo nucleic acid): A summary and new derivatives. *Eur. J. Org. Chem.* **2005**, *2005*, 2297–2321.

(23) Schöppe, A.; Hinz, H. J.; Rosemeyer, H.; Seela, F. Xylose-DNA: Comparison of the thermodynamic stability of oligo(2'-deoxyxylonucleotide) and oligo(2'-deoxyribonucleotide) duplexes. *Eur. J. Biochem.* **1996**, *239*, 33–41.

(24) Maiti, M.; Siegmund, V.; Abramov, M.; Lescrinier, E.; Rosemeyer, H.; Froeyen, M.; Ramaswamy, A.; Cuelemans, A.; Marx, A.; Herdewijn, P. Solution structure and conformational dynamics of deoxyxylonucleicacids (dXNA): An orthogonal nucleic acid-candidate. *Chem. Eur. J.* **2012**, *18*, 869–879.

(25) Ramaswamy, A.; Froeyen, M.; Herdewijn, P.; Ceulemans, A. Helical structure of xylose-DNA. *J. Am. Chem. Soc.* **2010**, *132*, 587–595.

(26) Ramaswamy, A.; Smyrnova, D.; Froeyen, M.; Maiti, M.; Herdewijn, P.; Ceulemans, A. Molecular dynamics of double stranded xylo-nucleic acid. *J. Chem. Theory Comput.* **2017**, *13*, 5028–5038.

(27) Ghosh, S.; Chakrabarti, R. Unzipping of double-stranded ribonucleic acids by graphene and single-walled carbon nanotube: Helix geometry versus surface curvature. *J. Phys. Chem. C* **2016**, *120*, 22681–22693.

(28) Ghosh, S.; Chakrabarti, R. Spontaneous unzipping of xylonucleic acid assisted by a single-walled carbon nanotube: A computational study. *J. Phys. Chem. B* **2016**, *120*, 3642–3652.

(29) Wales, D. J. Exploring energy landscapes. *Annu. Rev. Phys. Chem.* **2018**, *69*, 401–425.

(30) Röder, K.; Joseph, J. A.; Husic, B. E.; Wales, D. J. Energy landscapes for proteins: From single funnels to multifunctional systems. *Adv. Theory Simul.* **2019**, *2*, 1800175.

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

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

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

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

(35) Klimavicz, J. S.; Röder, K.; Wales, D. J. Energy landscapes of mini-dumbbell DNA octanucleotides. *J. Chem. Theory Comput.* **2018**, *14*, 3870–3876.

(36) Maiti, M.; Maiti, M.; Knies, C.; Dumbre, S.; Lescrinier, E.; Rosemeyer, H.; Ceulemans, A.; Herdewijn, P. Xylonucleic acid: Synthesis, structure, and orthogonal pairing properties. *Nucleic Acids Res.* **2015**, *43*, 7189–7200.

(37) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; K.M. Merz, K. M. et al. `AMBER12`. University of California, San Francisco, 2012.

(38) Cheatham III, T. E.; Cieplak, P.; Kollman, P. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dynam.* **1999**, *16*, 845–862.

(39) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, 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.

(40) Ivanova, A.; Rösch, N. The structure of LNA : DNA hybrids from molecular dynamics simulations: The effect of locked nucleotides. *J. Phys. Chem. A* **2007**, *111*, 9307–9319.

(41) Froeyen, M.; Abu el Asrar, R.; Abramov, M.; Herdewijn, P. Molecular simulation of cyclohexanyl nucleic acid (CNA) duplexes with CNA, DNA and RNA and CNA triloop and tetraloop hairpin structures. *Bioorg. Med. Chem.* **2016**, *24*, 1778–1785.

(42) Verona, M. D.; Verdolino, V.; Palazzesi, F.; Corradini, R. Focus on PNA flexibility and RNA binding using molecular dynamics and metadynamics. *Sci. Rep.* **2017**, *7*, 42799.

(43) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(44) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(45) Małolepsza, 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.

(46) Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Erratum to: Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2012**, *33*, 2209.

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

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

(49) 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.

(50) Macke, T. J.; Case, D. A. In *Molecular modeling of nucleic acids*; Leontes, N., SantaLucia Jr., J., Eds.; American Chemical Society: Washington D.C., 1998; pp 379–393.

(51) 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.

(52) Li, Z.; Scheraga, H. A. Structure and free-energy of complex thermodynamic systems. *J. Mol. Struct.* **1988**, *48*, 333–352.

(53) 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. Chem. Phys. A* **1997**, *101*, 5111–5116.

(54) 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.

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

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

(57) Wales, D. J. Energy landscapes: Some new horizons. *Curr. Opin. Struc. Biol.* **2010**, *20*, 3–10.

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

(59) Sheppard, D.; Terrell, R.; Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **2008**, *128*, 134106.

(60) 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.

(61) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

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

(63) 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.

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

(65) Nocedal, J. Updating quasi-newton matrices with limited storage. *Math. Comput.* **1980**, *35*, 773–782.

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

(67) Röder, K.; Wales, D. J. Energy landscapes for the aggregation of Aβ_{17-42}. *J. Am. Chem. Soc.* **2018**, *140*, 4018–4027.

(68) 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.

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

(70) 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.

(71) Carr, J. M.; Wales, D. J. Folding pathways and rates for the three-stranded β-sheet peptide Beta3s using discrete path sampling. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.

(72) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1998**, *106*, 1495–1517.

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

(74) Moradi, M.; Babin, V.; Roland, C.; Darden, D. A.; Sagui, C. Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20746–20751.

(75) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Res.* **2013**, *41*, 33–43.

(76) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(77) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. CURVES plus web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.* **2011**, *39*, W68–W73.

(78) Lu, X. J.; Olson, W. K. Resolving the discrepancies among nucleic acid conformational analyses. *J. Mol. Biol.* **1999**, *285*, 1563–1575.

(79) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Molec. Graphics* **1996**, *14*, 33–38.

(80) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

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

(82) Wolynes, P. G.; Eaton, W. A.; Fersht, A. R. Chemical physics of protein folding. *Proc. Natl Acad. Sci. USA* **2012**, *109*, 17770–17771.

(83) Tzul, F. O.; Vasilchuk, D.; Makhatadze, G. I. Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1627–E1632.

(84) Thirumalai, D. Native secondary structure formation in RNA may be a slave to tertiary folding. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11506–11508.

(85) Chen, S. J.; Dill, K. A. RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 646–651.

(86) Röder, K.; Wales, D. J. Evolved minimal frustration in multifunctional biomolecules. *J. Phys. Chem. B* **2018**, *14*, 10989–10995.

(87) Sharpe, D. J.; Röder, K.; Wales, D. J. Dataset for Energy landscapes of deoxyxylo- and xylo-nucleic acid octamers. DOI:10.5281/zenodo.3754742, 2020.

Graphical TOC Entry