JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Visualizing “Alternative Isoinformational Engineered” DNA in A- and B-Forms at High Resolution

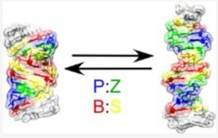
Shuichi Hoshika, Madhura S. Shukla, Steven A. Benner, and Millie M. Georgiadis*

Cite This: *J. Am. Chem. Soc.* 2022, 144, 15603−15611

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A fundamental property of DNA built from four informational nucleotide units (GCAT) is its ability to adopt different helical forms within the context of the Watson−Crick pair. Well-characterized examples include A-, B-, and Z-DNA. For this study, we created an isoinformational biomimetic polymer, built (like standard DNA) from four informational “letters”, but with the building blocks being artificial. This ALternative Isoinformational ENgineered (ALIEN) DNA was hypothesized to support two nucleobase pairs, the P:Z pair matching 2-amino-imidazo-[1,2*a*]-1,3,5-triazin-[8*H*]-4-one with 6-amino-3-5-nitro-1*H*-pyridin-2-one and the B:S pair matching 6-amino-4-hydroxy-5-1*H*-purin-2-one with 3-methyl-6-amino-pyrimidin-2-one. We report two structures of ALIEN DNA duplexes at 1.2 Å resolution and a third at 1.65 Å. All of these are built from a single self-complementary sequence (5′-CTSZZPBSBSZPPBAG) that includes 12 consecutive ALIEN nucleotides. We characterized the helical, nucleobase pair, and dinucleotide step parameters of ALIEN DNA in these structures. In addition to showing that ALIEN pairs retain basic Watson−Crick pairing geometry, two of the ALIEN DNA structures are characterized as A-form DNA and one as B-form DNA. We identified parameters that map differences effecting the transition between the two helical forms; these same parameters distinguish helical forms of isoinformational natural DNA. Collectively, our analyses suggest that ALIEN DNA retains essential structural features of natural DNA, not only its information density and Watson−Crick pairing but also its ability to adopt two canonical forms.



INTRODUCTION

Much effort has been devoted to understanding the molecular properties of DNA essential to store, transfer, and support the evolution of biological information, all thought to be central to the phenomenon that we call “life”. Despite its elegance and familiarity, standard Terran DNA has proven not to be the only molecular system able to do all three of these, not even in the Terran biosphere.[1−4] Accordingly, synthesis of new biopolymers with structures different from natural DNA/RNA has proven critical to developing our understanding of the contributions of the phosphodiester backbone,[5] the deoxyribose/ribose sugar,[6,7] and the nucleobase structures[8] to the ability of biopolymers in general to support Darwinian evolution. Some of these “xeno nucleic acids” (XNAs) have dramatically different molecular recognition architectures.[9,10] At the same time, novel features of some XNAs confer additional functionality and/or altered properties on these molecules, allowing them to support new technologies with applications ranging from diagnostics[11] to therapy.[12]

To date, most of this synthetic biology has been done within the context of standard Watson−Crick pairs, adding only a few unnatural nucleotides to DNA molecules that retain large numbers of standard pairs. For example, unnatural base pairs that rely solely on hydrophobic interactions, from pioneering work of Kool,[13] Romesberg,[14] and Hirao,[15] have been added only at a few sites within DNA molecules that are otherwise entirely standard. Even when hydrogen bonding is retained in non-standard nucleotides within a Watson−Crick context, where independent replication is ensured by alternative arrangements of hydrogen bond donor and acceptor groups, most synthetic biology has just added these to the standard four nucleotides in DNA “alphabet” to create six[16] or eight[17] “letter” DNA alphabets that retain large numbers of natural nucleotides and increase the information density of the polymer. In one case, the pairing retains the same acceptor−donor patterns even though members of the pair (benzo-homologues) are unnatural.[18]

Here, we ask whether a nucleic acid analogue that has the same information density as standard DNA can be supported within a Watson−Crick pairing architecture, but with *entirely* unnatural synthetic components. This is termed an “ALternative Isoinformational ENgineered” (ALIEN) DNA.

We started with two pairs of non-standard nucleotides that should form two non-standard Watson−Crick pairs: the P:Z pair, between 2-amino-8-(1′-*β*-D-2′-deoxyribofuranosyl)-imidazo-[1,2*a*]-1,3,5-triazin-[8*H*]-4-one and 6-amino-3-(1′-*β*-D-2′-deoxy ribofuranosyl)-5-nitro-1*H*-pyridin-2-one, and the B:S pair, between 6-amino-9-(1′-*β*-D-2′-deoxyribofuranosyl)4-hydroxy-5-(hydroxymethyl)oxolan-2-yl]-1*H*-purin-2-one and 3-methyl-6-amino-5-(1′-*β*-D-2′-deoxyribofuranosyl)-pyrimidin-2-one. The hydrogen-bonding patterns joining the P:Z and S:B

Received: May 17, 2022
Published: August 15, 2022

https://doi.org/10.1021/jacs.2c05255

pairs are chemically orthogonal from each other (Figure 1) and from those joining standard nucleobase pairs.

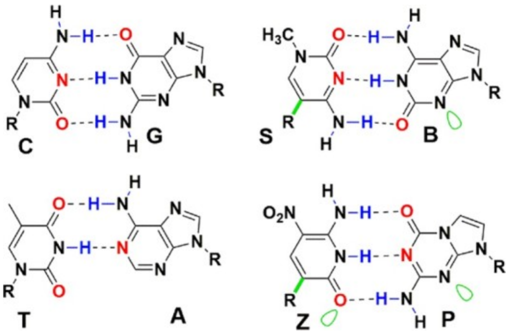


Figure 1. Natural nucleobase pairs T:A and C:G are shown along with unnatural pairs S:B and Z:P that comprise ALIEN DNA. Hydrogen-bonding interactions are shown as dashed lines. Green bonds are *C*-glycosides. Green lobes indicate electron density presented to the minor groove.

To achieve this non-standard hydrogen-bonding pattern, the heterocycle on Z must be attached to its carbohydrate by a carbon−carbon bond (a *C*-glycoside, Figure 1, green bond), rather than the carbon−nitrogen bond (N-glycosidic) that attaches natural nucleotides to their carbohydrate backbones. In this implementation of the S hydrogen-bonding pattern, a similar C-glycosidic linkage is used in the design.

In addition, although Z is size-similar to a pyrimidine, chemical stability with respect to oxidation and epimerization requires the Z heterocycle to have an electron-withdrawing group at its "5-position". In the implementation of the Z hydrogen-bonding pattern used here, a nitro group serves this purpose. However, the nitro group effectively makes this nucleobase larger than standard pyrimidine nucleobases. As another distinction, to achieve its distinctive hydrogen-bonding pattern, the nucleobase on S must present a hydrogen bond donor (rather than acceptor) to the minor groove, in contrast to all standard nucleobases and the other non-standard nucleobases used in this study.

How do the non-standard P:Z and B:S pairs behave relative to standard pairs? In previous studies with 6-letter DNA (A:T, G:C, P:Z) and 8-letter "hachimoji" DNA (A:T, G:C, P:Z, and B:S), rule-based molecular recognition is observed and thermodynamically quantifiable. Furthermore, this information can be transferred by transcription to RNA. ALIEN DNA follows size (shape) complementarity and hydrogen bond complementarity. Therefore, ALIEN DNA can do what natural DNA can, but it can work orthogonally. These properties motivated us to analyze ALIEN DNA's structure.

Crystallographic analysis of such molecules in B-form has been assisted by the ability of a "host" protein, the N-terminal fragment of the reverse transcriptase from Moloney murine leukemia virus, to anchor ends of a 16-mer self-complementary duplex DNA (the "guest").[19] The span connecting these anchors can be observed without any direct contact with the protein. Structures of hachimoji DNA of three different sequences, one with six consecutive P:Z and B:S pairs, were captured in B-form DNA structures using this host−guest system.[17] A DNA-only crystal structure of a sequence including six consecutive P:Z pairs was shown to adopt A-form.[20]

These examples afforded the first information on the properties of consecutive unnatural nucleobases embedded in a natural context. The only other structural characterization of an unnatural base pair in the context of a DNA duplex is an NMR structure reported for dMMO2-d5SICS pair, in which the hydrophobic base pairs violate the hydrogen-bonding and size complementarity rules to give an intercalative stacking mode.[21]

Structural flexibility is exploited in modern Terran biology with standard DNA. It allows standard DNA duplexes to adopt different helical forms that support different biological functions. The ability to adopt different helical forms has been attributed to the properties of the nucleobase pairs and their ability to change slide and roll parameters in order to transition from B to A-form DNA.[22]

Here, we analyze the structures of DNA including 12 consecutive unnatural base pairs, ALIEN DNA. We had no expectation with respect to the helical form(s) that ALIEN DNA might adopt, especially given that natural DNA can

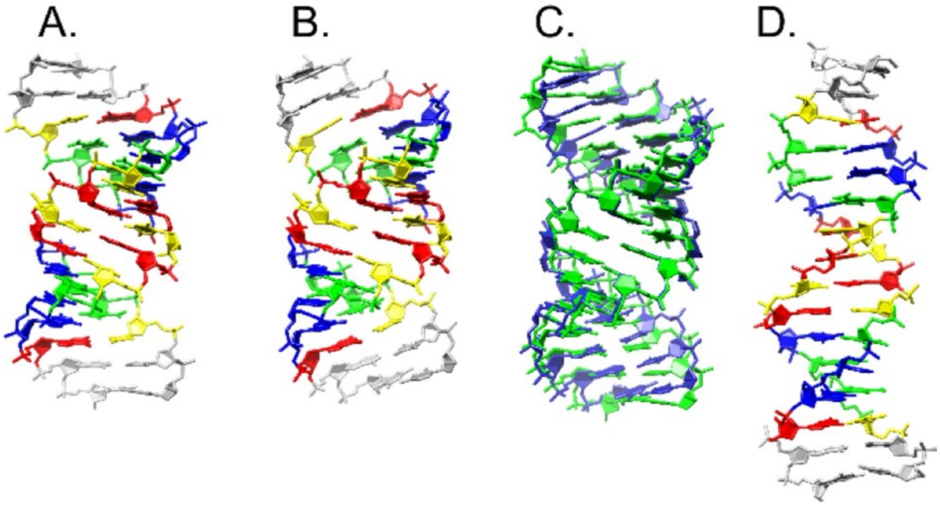


Figure 2. Crystal structures of ALIEN DNA 5′-CTSZZPBSBSZPPBAG in DNA-only and in a host−guest complex. (A) B7 and (B) A3 crystal structures determined at 1.2 Å in DNA-only lattices. Z is shown in green, P in blue, B in red, S in yellow, and A, T, G, and C in light gray. (C) B7 (green) and A3 (blue) superimpose with an rmsd of 1.1 Å for 23 nucleobases. (D) Crystal structure of ALIEN DNA crystallized in our host−guest complex; the full complex is shown in Figure S1.

adopt intermediate forms.[23,24] Nor could we predict any of the inherent parameters associated with non-standard nucleobase pairs in ALIEN DNA. Here, we report three crystal structures of ALIEN DNA, two captured in A- and one in B-form, in which the ALIEN DNA retains many natural features.

RESULTS

Crystal Structures of ALIEN DNA. We crystallized the self-complementary sequence 5′-CTSZZPBSBSZPPBAG in duplex form in two DNA-only lattices (A3 and B7) and as the "guest" in a host−guest complex with the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase serving as the host. The oligonucleotides include two natural base pairs on the ends. These are the first structures to include 12 consecutive unnatural nucleobase pairs, namely P:Z and B:S pairs (Figure 1), the next closest being the host−guest complex of the 1.7 Å hachimoji structure of 5′-CTTATPPSBZZA-TAAG containing 6 consecutive unnatural nucleobase pairs. Both P:Z and B:S pairs are joined by three Watson−Crick hydrogen bonds and are, in this regard, more similar to G:C pairs than they are to A:T pairs.

To phase the 1.2 Å DNA-only structures, we prepared the known phosphoramidite of 7-deaza-7-bromo-isoguanosine (*Q*) to allow for introduction of Br atoms into a crystal of the sequence 5′-CTSZZPBS*Q*SZPP*Q*AG. This supported a Br SAD phasing experiment. All three DNA-only "stand alone" structures crystallized in different lattices (Table S1). We obtained a 1.8 Å experimentally phased map and built a starting model containing natural nucleotides (Figure S1). We used this structure as the search model to determine the structure of B7 by molecular replacement phasing. We modeled the correct sequence and then refined the structure to 1.2 Å (Figure 2A). Both B7 and the brominated duplex crystallized in lattices with half of the duplex in the asymmetric unit. The final DNA-only structure (A3) crystallized in a lattice with the entire 16-mer in the asymmetric unit.

We successfully phased the A3 structure by molecular replacement using the model with the correct sequence obtained for the B7 structure and refined this 1.2 Å structure (Figure 2B). We did not obtain a molecular replacement solution using the model containing natural nucleotides used to solve B7. Although the overall structures of B7 and A3 are similar, they are not identical (Figure 2C). However, the central portions of the structures superimpose well with an rmsd of 1.1 Å for 23 of the 32 nucleobases (rmsd 1.8 Å for all nucleobases). Crystallographic packing interactions likely impact the ends of the structures.

We crystallized the same sequence as a host−guest complex to obtain a different helical form of ALIEN DNA. We solved this structure by molecular replacement using the host protein as the search model and obtained unbiased electron density maps for the DNA. The refined 1.65 Å structure clearly exhibits a helical form different from that observed in the DNA-only lattices (Figure 2D).

ALIEN DNA Adopts A- and B-Helical Forms. To determine the helical form of ALIEN DNA in our crystal structures, we analyzed 31 A-DNA structures reported at resolutions of 2 Å or higher with unmodified nucleobase pairs (Table S2) and 29 B-DNA structures reported at resolutions of 2.35 Å or higher in the same crystal lattice as our new structure (Table S3) using 3DNA. The program 3DNA[25,26] uses Zp and ZpH values defined as the "projection of the phosphorous atom onto the z-axis of the dimer middle frame" and the half distance of this projection, respectively, to distinguish A- and B-forms of DNA. "Dimer" in this context refers to a dinucleotide step; the coordinate frame is explained within the study by Lu et al.[27]

Our plots of Zp versus ZpH, slide versus roll, and x-displacement versus inclination for our reference structures are similar to those previously reported with A- and B-forms separating into distinguishable clusters.[25] Zp values for B-DNA range from −1.39 to 2.11 Å, ZpH values −2.19 to 4.27 Å in our reference set of B-form DNA captured in host−guest structures (Figure 3A). Our new host−guest ALIEN DNA structure has Zp values ranging from −1.13 to 1.18 Å and ZpH values −1.76 to 4.8 Å. The point (1.18, 4.8) is the natural step at the end of the oligonucleotide. For A-DNA structures, Zp values range from 0.49 to 3.01 Å and ZpH values 0.37 to 6.8 Å (Figure 3A). Our two DNA-only structures exhibit Zp values in the range of 2.18 to 2.79 Å with ZpH values of 1.86 to 6.04 Å, which fall well within the reference cluster.

Similar to the analysis for Zp versus ZpH, values for our structures fall within the expected clusters in plots of slide

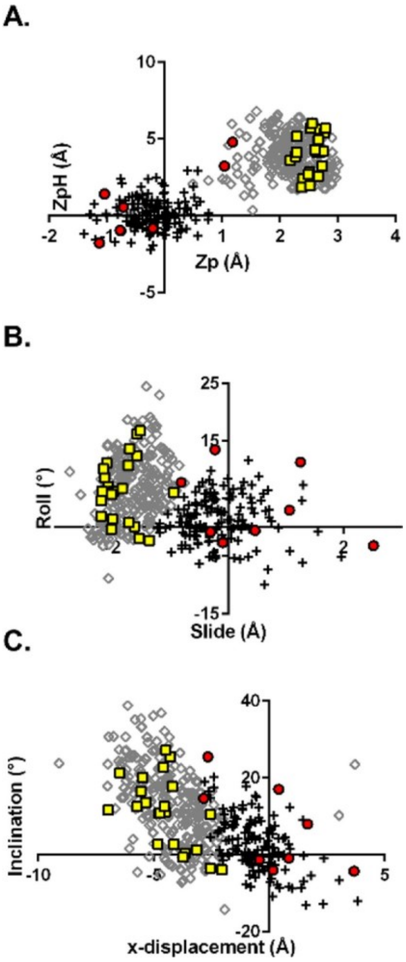


Figure 3. DNA-only ALIEN DNA structures are A-DNA and host−guest DNA complex as B-form in (A) Zp vs ZpH, (B) slide vs roll, and (C) x-displacement vs inclination plots. Black + symbols indicate parameters for B-DNA from our reference data, gray open diamonds indicate the A-DNA reference set, red-filled circles indicate the host−guest ALIEN DNA, and yellow-filled squares indicate DNA-only ALIEN DNA structures.

versus roll (Figure 3B) and x-displacement versus inclination (Figure 3C) for A- versus B-form DNA. Thus, we conclude that the two DNA-only structures are A-form, while the DNA in the host−guest complex is B-form. We refer to these structures as the A- and B-ALIEN DNA structures below.

Nucleobase Pairs in ALIEN DNA and Natural DNA Have Similar Conformational Properties. To compare the properties associated with specific nucleobase pairs and dinucleotide steps in ALIEN versus natural DNA, we analyzed local base pair shear, stagger, buckle, propeller, and opening and local base pair step slide, rise, roll, and twist for all possible nucleobase pairs and dinucleotide A- and B-DNA steps for the selected reference structures. Local base pair parameters impact hydrogen-bonding interactions between paired nucleobases. In the DNA-only structures, all P:Z and B:S nucleobase pairs form three hydrogen bonds with distances ranging from 2.8 to 3.1 Å. For the host−guest complex, the Z:P pair at position 5/12 exhibits significant shearing (−1.88 Å); this positions O6 of P outside of the normal hydrogen-bonding distance to N4 of Z at 3.5 Å (Figure 4A). The negative shear value observed for this Z:P pair is 0.4 Å larger than the most negative value (−1.36 Å) observed in our reference data set for a G-C pair in a natural DNA sequence containing three consecutive Gs (PDB ID: 4XPE). The S:B pair at position 3 forms only two hydrogen bonds as well; N2 of S forms a hydrogen bond with O2 of B and N3 of S with N1 of B (Figure 4B). This nucleobase pair exhibits an opening angle of 9.82° and a propeller angle of −16.9°, which results in positioning of O4 in S and N6 in B 3.7 Å apart. All other unnatural nucleobase pairs form three hydrogen bonds as expected.

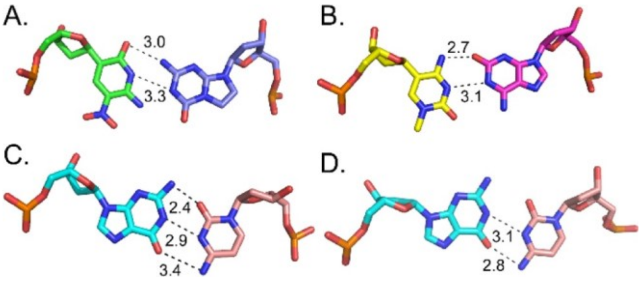


Figure 4. In B-DNA, the Z5:P12 pair (A) shown in green and blue, respectively, and S3:B14 pair (B) yellow and magenta, respectively, form only two hydrogen bonds. Loss of the third hydrogen bond results from shearing of the Z:P pair and from a combination of relatively large opening and propeller angles in the S:B pair. In comparison, the G6:C11 pair (C) exhibits significant shearing but retains all three hydrogen bonds, while G8:C9 (D) has only two hydrogen bonds due to shear and propeller effects.

By way of comparison, we considered whether shearing and loss of full hydrogen-bonding in B-DNA occurred in natural DNA sequences. DNA crystals of sequences including 5′-GGGCCC or 5′-GGGCCG are exclusively A-form DNA, but our host−guest complex including 5′-GGGCCC within the 16 bp ds duplex crystallized in B-form. In this structure, the G:C pair at position 6 exhibits shearing of −1.4 Å but retains all three Watson−Crick hydrogen bonds. The G:C pair at position 8 is less sheared (−1.29 Å) but with a propeller angle of −8.4° compared with that of −5.6° for position 6; this pair has only two hydrogen bonds (Figure 4C,D).

For all parameters analyzed with one exception, nucleobase pairs and dinucleotide step parameters for ALIEN DNA fall within those observed for natural A- or B-DNA structures, and the only parameters distinguishing A- and B-DNA are slide and roll values (Figure 5A−C). Dinucleotide steps in B-DNA exhibit slide values that fall within a range of −1.65 to 1.96 Å for all possible dinucleotide steps, while those in A-DNA range from −2.75 Å to −0.35. The base pair step PZ/PZ exhibits a higher slide value (+2.52 Å) than observed in our reference B-structures (Figure 5B).

One significant chemical difference between natural pyrimidine nucleobases T and C and ALIEN pyrimidine-like nucleobases, Z and S, is the linkage between the nucleobases and the deoxyribose sugar. All natural nucleotides are N-glycosides, while Z and S are C-glycosides (Figure 1), with the C1′ sugar atom covalently bonded to the C1 atom of the nucleobase (the green bonds in Figure 1). We examined C1′−C1′ distances in our ALIEN nucleobase pairs and compared them to those in natural nucleobase pairs (Figure 5D). These distances are tightly clustered in our ALIEN nucleobase pairs and fall within the range observed in natural DNA; however, the average distance is longer in ALIEN DNA. The average C1′−C1′ distance in natural B and A forms is 10.5 and 10.6 Å, respectively; the average in ALIEN B and A is 10.7 and 10.8 Å, respectively.

A second parameter that we examined was the sugar conformation. We found that the distribution of sugar conformations for our ALIEN DNA structures was very similar to those in natural structures with the most common sugar conformation of C2′ endo in B-form and C3′ endo in A-form (Figure 5E). In both ALIEN and natural B-form structures, 56% of sugars are in the C2′ endo conformation. In the very high resolution A-form ALIEN DNA structures, 97% (A3) and 100% (B7) of the sugars are in the C3′ endo conformation, while in the reference A-form 74% are C3′ endo. Thus, by all of these criteria, ALIEN DNA exhibits properties that are similar to natural DNA in the context of both A- and B-helical forms.

Shifts in ALIEN DNA Parameters in B- Versus A-Forms Mirror Those Observed in Natural DNA. In 1984, Calladine and Drew proposed a base-centric model to explain how DNA transitioned between B- and A-forms.[22] In this model, conversion of a B-form helix to that of an A-form helix involves an increase in roll angle of 12° and a negative slide value (−1.5 Å). If we compare the slide values for ALIEN nucleobase pairs in B-form versus A-form (Figure 6A), we observe that all PZBS nucleobase pairs exhibit a significantly more negative slide value in A- than in B-form. With the exception of base step 5 for the PZ/PZ step (+2.52 Å), slide values fall between −0.82 and 1.25 Å in B-form and between −2.2 and −1.58 Å for the same sequence in A-form. With the exception of the third nucleobase pair S:B, roll angles in A-form DNA are significantly larger than in the same nucleobase pairs in A-form (Figure 6B).

The transition between B and A-form DNA was mapped more fully in 13 crystal structures of d(GGCGCC) including brominated and methylated versions that adopt intermediate helical forms. In this analysis, x-displacement, slide, and Zp values vary smoothly (either gradually increasing or decreasing) through intermediate forms distinguishing a transition from B to A-forms.[24] Our study compares the exact same ALIEN sequence in two A-form structures and one B-form structure. Both x-displacement and Zp values distinguish

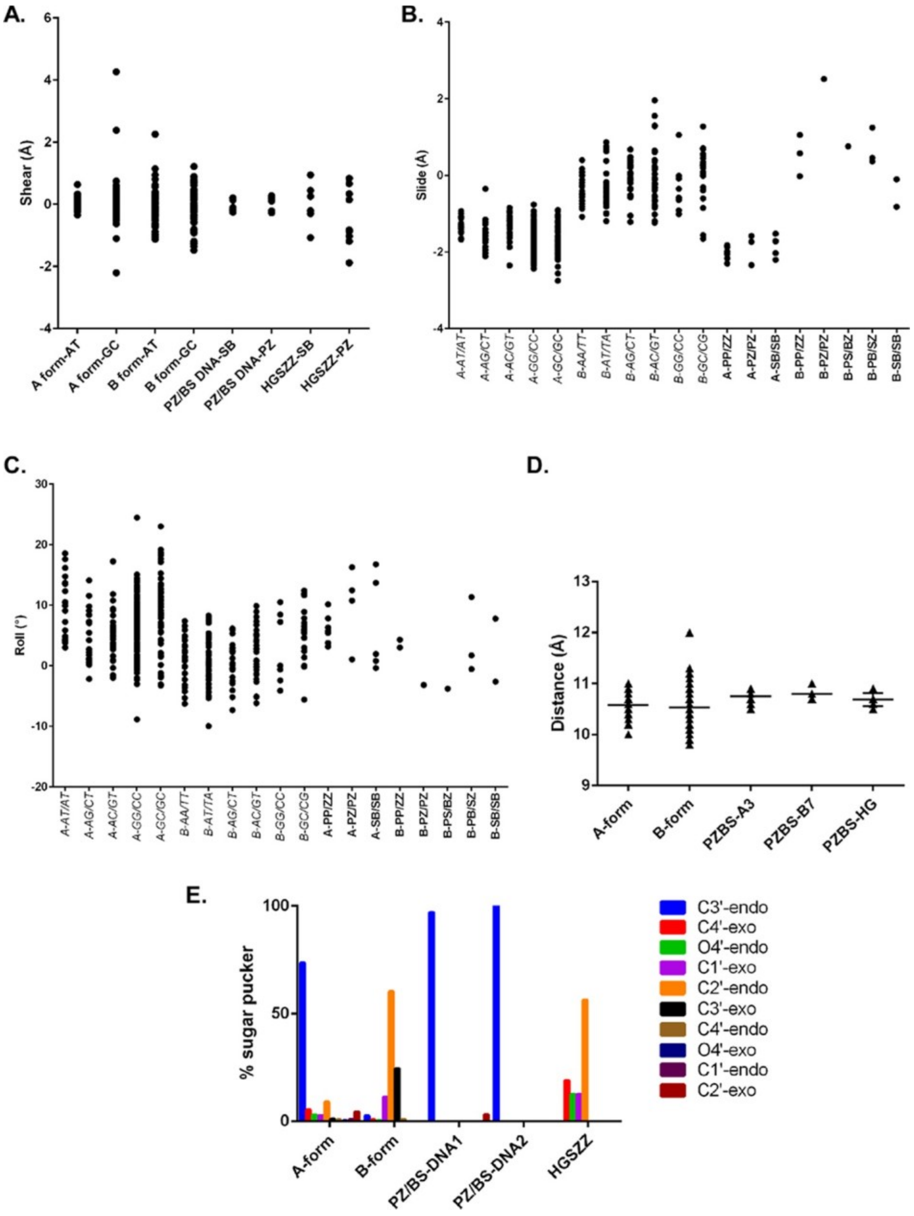


Figure 5. Local base pair and base pair step parameter plots for natural and ALIEN DNA base pairs are shown for (A) shear, (B) slide, (C) roll, (D) C1′−C1′ distances, and (E) sugar puckers arranged by pseudorotation angles. PZBS-A3 and B7 are the DNA-only structures, PZBS-HG is the host guest ALIEN DNA. The most common sugar puckers observed are C3′ endo for A-form and C2′ endo in B-form.

ALIEN DNA in B-form versus A-form (Figure 6C,D) for every ALIEN nucleotide base pair step in our structures. *x*-displacement values are consistently more positive in B- than A-form ALIEN DNA (Figure 6C), while Zp values are consistently more negative (Figure 6D). Thus, ALIEN DNA appears to transition between B- and A-forms through a base-centered mechanism similar to that observed in natural DNA. The only example of a natural sequence, 5′-CCGCCGGCGG, that crystallized in A- and B-form,[28−30] exhibits patterns for differences in slide, roll, *x*-displacement, and Zp values similar to those observed for our ALIEN DNA (Figure S2).

DISCUSSION

Studies of unnatural molecular systems that reproduce the behaviors of natural genetic biopolymers yield information having both scientific and technological values. First, alternative informational biopolymers, such as ALIEN DNA, define the scope of chemistry able to store information. This "synthetic biology" lets us learn whether the nucleic acids delivered by prebiotic chemistry and natural evolution are unique solutions to the problems posed by genetics, or whether other solutions are possible.[5−10]

Technology with alternative genetic systems generally recruits natural enzymes, proteins, and even cells. This has been the case where components of ALIEN DNA have been used in environmental surveillance products (P and Z),[11] in human diagnostics products (B and S),[11] and in therapeutic strategies (P and Z).[12] Such recruitment requires some level of structural compatibility between ALIEN DNA and those recruited systems, which (in turn) interact with natural DNA and RNA as dynamic systems. Here, alternative conformations are exploited, often adaptively, in natural systems. This makes it important to understand the conformational versatility of ALIEN DNA.

https://doi.org/10.1021/jacs.2c05255
J. Am. Chem. Soc. 2022, 144, 15603−15611

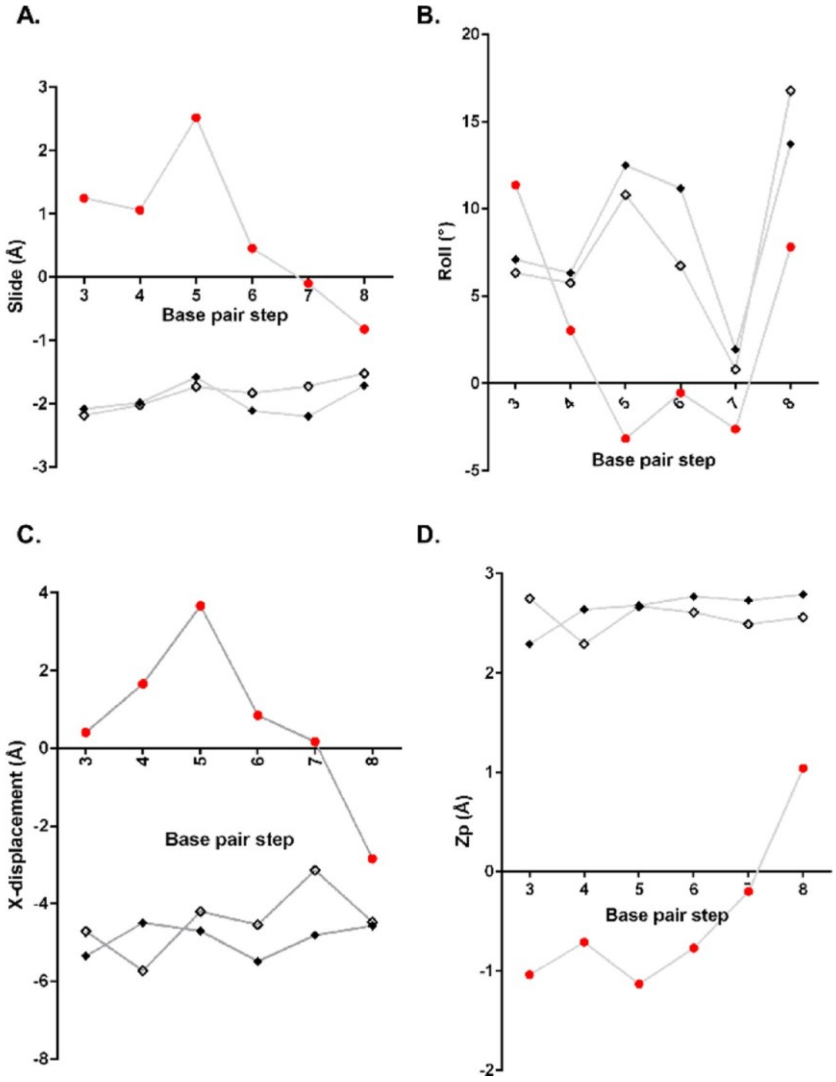


Figure 6. Shifts in base pair step parameters observed in A- vs B-forms of ALIEN DNA. Base pair steps indicated as numbers 3−8 are SZ/PB, ZZ/PP, ZP/ZP, PB/SZ, BS/BS, and SB/SB, respectively, with parameters from DNA-only structures indicated in open diamonds for B7, black diamonds for A3, and red circles for the ALIEN DNA host−guest complex shown for (A) slide, (B) roll, (C) *x*-displacement, and (D) Zp values.

Our host−guest system selects for 16-mer duplex DNA sequences that can adopt B-form.[31,32] The ALIEN 5′-CTSZZPBSBSZPPBAG is the latest of many sequences to be crystallized in B-form using this system. And yet, this same sequence readily crystallized as A-form DNA in DNA-only crystals in precipitant solutions that would preclude formation of a host−guest complex, namely 20% methylpentandiol and 2.0 M ammonium sulfate. To date, this is the only sequence that we have crystallized in both A- and B-form. Crystals of 5′-CTTATPPPZZZATAAG (4XNO) in a DNA-only lattice were also obtained in a condition including 1.7−2.0 M ammonium sulfate,[20] similar to one of the conditions noted above, but did not crystallize in the host−guest system. In a separate study to be published elsewhere, we crystallized three oligonucleotides including 10 consecutive PZBS pairs in host−guest complexes as B-DNA, but none of these sequences crystallized in A-form. Based on these studies to date, we suggest that ALIEN DNA exhibits sequence preferences for A- and/or B-form DNA.

Collectively, our analyses of high-resolution crystal structures of ALIEN DNA suggest that it shares fundamental structural features of natural DNA including the ability to adopt both A and B helical forms. One could argue that these ALIEN nucleobases retain features that are similar to natural nucleobases, and thus, our findings are consistent with the notion that the nucleobases dictate many of the properties associated with DNA. Through characterization of ALIEN DNA, we have the opportunity to evaluate the generality of mechanisms like the Calladine−Drew[22] nucleobase centric model, which might have been regarded as being peculiar to standard DNA. Thus, we might ponder the uniformity of A:T and G:C nucleobase pairs as the genetic material in Terran organisms when other nucleobase pairs would meet many of the criteria deemed critical for Darwinian evolution at a structural level, including the ALIEN DNA described here.

EXPERIMENTAL PROCEDURES

Synthesis and Purification of Oligonucleotides Containing Four Natural Nucleotides (dG, dA, dC, and T) and AEGIS Nucleotides (dZ, dP, dB, dS, and dQ). Standard phosphoramidites (Bz-dA, Ac-dC, dmf-dG, dT and dB: dmf-isodG-CE Phosphoramidite) and CPG having standard residues were purchased from Glen Research (Sterling, VA) and AEGIS phosphoramidites (dZ, dP, and dS) were purchased from Firebird Biomolecular Sciences LLC (Alachua, FL). 7-Bromo-7-deaza-isoG (d*Q*) phosphoramidite was prepared by following the study of Seela et al.[33] All oligonucleotides containing dZ, dP, dB, dS, and 7-bromo-7-deaza-dB were synthesized on an ABI 394 DNA Synthesizer following standard phosphoramidite chemistry and as previously reported.[34] The CPGs having oligonucleotides were treated with 2.0 mL of 1 M DBU in anhydrous acetonitrile at room temperature for 24 h to deprotect the NPE group on the dZ nucleobase. Then, the CPGs were filtered, dried, and treated with concentrated ammonium hydroxide at 55 °C for 16 h. After the removal of ammonium hydroxide, the oligonucleotides containing dZ, dP, dB, and dS were purified on ion-exchange HPLC and then desalted using Sep-Pac Plus C18 cartridges (Waters).

Crystallization of ALIEN DNA. The oligonucleotide 5′-CTSZZPBSBSZPPBAG was dissolved in 10 mM HEPES pH 7.0, 10 mM $MgCl_2$, annealed at 70 °C for 20 min, and then slowly cooled to room temperature at a concentration of 2.5 mM duplex. We screened Natrix 1 crystallization (Hampton Research, Inc) conditions for the DNA sample alone using our Orxy4 crystallization robot in vapor diffusion sitting drop plates with drops including 0.3 μL of DNA and 0.3 μL of the precipitant solution. Crystals of DNA grown in two different conditions from the Natrix screening solutions, A3 and B7, were used for data collection following optimization. A3 crystals were obtained from 0.1 M magnesium acetate, 0.05 M ADA pH 6.5, and 20% methylpentanediol. These crystals were cryocooled directly from the sitting drops. Crystals of the B7 condition were obtained from 0.1 M magnesium sulfate, 0.05 M ADA pH 6.5, and 2.0 M ammonium sulfate. B7 crystals were cryocooled in the precipitant with 20% glycerol. To phase the DNA-only structures, we crystallized 5′-CTSZZPBS*Q*SZPP*Q*AG (Br2), *Q* being 7-Br-7-deazaisoG, in 0.01 M magnesium sulfate, 0.05 M HEPES pH 7.0, and 2.2 M ammonium sulfate. Host−guest crystals of the oligonucleotide were obtained from vapor diffusion hanging drops with 1 μL of a complex containing 0.43 mM N-terminal fragment of Moloney murine leukemia virus reverse transcriptase and 0.86 mM duplex DNA and 1 μL of precipitant solution containing 5 mM magnesium acetate, 0.05 M ADA pH 6.5, and 9% PEG 4000. The drops contained microseeds generated from host−guest crystals of a natural 16-mer duplex sequence diluted in the precipitant solution.

Data Collection and Structure Determinations. Data were collected at the Advanced Photon Source, LRL-CAT, beamline 31-ID for A3, B7, and HG at 0.97933 Å and for Br2 at 0.91987 Å (Table S1). We used SHELX (CDE)[35] to identify two Br sites within an 8-mer duplex in the asymmetric unit of this crystal and then calculate phases for the Br2 structure at 1.8 Å. The experimental electron density map was readily interpretable (Figure S1). A polyA model was autobuilt into the electron density map using NAUTILUS.[36] This polyA sequence was mutated in COOT to include two natural pairs on the end, T for S and Z and G for B and P for initial refinement in REFMAC5.[37] The refined model was then used as the search model in molecular replacement calculations performed in PHASER[38] for the B7 structure, which also includes an 8-mer duplex in the asymmetric unit, albeit in a different lattice. The unnatural base pairs were modeled in COOT,[39] and the structure was refined using PHENIX.[40] The refined model for B7 was then used as the search model for molecular replacement in PHASER[38] for the A3 structure. For this structure, the natural model derived from Br2 did not produce a molecular replacement solution. The A3 crystal includes a 16-mer duplex in the asymmetric unit and provides a snapshot of the entire duplex. Adjustments to the model were done in COOT[39] followed by refinement in PHENIX.[40]

The host−guest complex of 5′-CTSZZPBSBSZPPBAG was determined by molecular replacement in PHASER[38] using the protein model alone as the search model. A model for the DNA was built into the unbiased electron density map starting with the first two base pairs and then adding base pairs and iteratively refining the model in PHENIX.[40] The model was built initially as a natural model with T for S or Z and G for B or P. Once the 8 bp duplex had been modeled and refined, the natural base pairs for positions 3−8 were replaced with the correct unnatural base pairs. In the Fo−Fc electron density maps, positive peaks for the oxygen atoms of the nitro groups were apparent for Z nucleobases prior to modeling of the unnatural nucleobase pairs. A parameter file for each ALIEN nucleotide was generated and used for refinement. Following refinement of the 8 bp duplex, a single strand 16 base oligonucleotide was created and the phosphate backbone joining nucleotides at positions 8 and 9 was adjusted. This final model was then refined in PHENIX.[40] Coordinate files have been deposited with RCSB identification codes of 8CRY, 8CRZ, and 8CS0 for A3, B7, and HGSZZ, respectively.

Database for Comparative Analysis of DNA Parameters. We performed an advanced search for A-form DNA within the RCSB database and selected structures deposited between the years 1991 and 2021 having a resolution of 2.0 Å or less (Table S2). We compiled the DNA structures in B-form from host−guest complex structures all crystallized in the same lattice (Table S3). These structures were all determined at resolutions ranging from 1.5 to 2.35 Å. We analyzed the structures using web.x3DNA.org database[26] and compiled all of the structural parameters. From each structure, we omitted the modified natural base pairs such as pairs with methylated or uracil bases, CTT/GAA base pairs interacting with protein, and unnatural base pairing such as purine−purine or pyrimidine−pyrimidine pairs. This yielded a list of 31 structures of A- and 29 of B-form. We compiled a list of structural parameters for all natural base pairs and base pair steps, which includes C1′−C1′ distance, phosphorus positions (Zp and ZpH), local base pair step parameters (slide and roll), local helical parameters (x-displacement and inclination), and sugar puckering.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c05255.

Electron density maps, host−guest complex, DNA parameters for 5′-CCGCCGGCGG in A- and B-form, crystallographic data, database A-form DNA structures, and database B-form DNA structures (PDF)

AUTHOR INFORMATION

Corresponding Author

Millie M. Georgiadis − *Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States;* orcid.org/0000-0003-2976-4576; Email: mgeorgia@iu.edu

Authors

Shuichi Hoshika − *Foundation for Molecular Evolution, Alachua, Florida 32615, United States*
Madhura S. Shukla − *Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States*
Steven A. Benner − *Foundation for Molecular Evolution, Alachua, Florida 32615, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c05255

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by an NSF grant (1939086) to S.A.B. and M.M.G. This research used resources of the Advanced Photon Source; a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly and Company, which operates the facility. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation or the Advanced Photon Source. The Oryx4 crystallization robot used in these studies is maintained by the Chemical Genomics Core Facility at Indiana University School of Medicine.

REFERENCES

(1) Czernecki, D.; Bonhomme, F.; Kaminski, P. A.; Delarue, M. Characterization of a triad of genes in cyanophage S-2L sufficient to replace adenine by 2-aminoadenine in bacterial DNA. *Nat. Commun.* 2021, *12*, 4710.

(2) Pezo, V.; Jaziri, F.; Bourguignon, P. Y.; Louis, D.; Jacobs-Sera, D.; Rozenski, J.; Pochet, S.; Herdewijn, P.; Hatfull, G. F.; Kaminski, P. A.; Marliere, P. Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science* 2021, *372*, 520−524.

(3) Sleiman, D.; Garcia, P. S.; Lagune, M.; Loc'h, J.; Haouz, A.; Taib, N.; Röthlisberger, P.; Gribaldo, S.; Marliere, P.; Kaminski, P. A. A third purine biosynthetic pathway encoded by aminoadenine-based viral DNA genomes. *Science* 2021, *372*, 516−520.

(4) Zhou, Y.; Xu, X.; Wei, Y.; Cheng, Y.; Guo, Y.; Khudyakov, I.; Liu, F.; He, P.; Song, Z.; Li, Z.; Gao, Y.; Ang, E. L.; Zhao, H.; Zhang, Y.; Zhao, S. A widespread pathway for substitution of adenine by diaminopurine in phage genomes. *Science* 2021, *372*, 512−516.

(5) Benner, S. A.; Hutter, D. Phosphates, DNA, and the Search for Nonterrean Life: A Second Generation Model for Genetic Molecules. *Bioorg. Chem.* 2002, *30*, 62−80.

(6) Zhang, S.; Switzer, C.; Chaput, J. C. The resurgence of acyclic nucleic acids. *Chem. Biodiversity* 2010, *7*, 245−258.

(7) Zhang, S.; Yu, H.; Chaput, J. C. Synthesis of threose nucleic acid (TNA) triphosphates and oligonucleotides by polymerase-mediated primer extension. *Curr. Protoc. Nucleic Acid Chem.* 2013, *4*, 4.54.

(8) Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Fajardo, D.; Moussatche, P. Alternative Watson-Crick Synthetic Genetic Systems. *Cold Spring Harbor Perspect. Biol.* 2016, *8*, a023770.

(9) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. New base pairing motifs. The synthesis and thermal stability of oligodeoxynucleotides containing imidazopyridopyrimidine nucleosides with the ability to form four hydrogen bonds. *J. Am. Chem. Soc.* 2003, *125*, 9970−9982.

(10) Hoshika, S.; Singh, I.; Switzer, C.; Molt, R. W., Jr.; Leal, N. A.; Kim, M. J.; Kim, M. S.; Kim, H. J.; Georgiadis, M. M.; Benner, S. A. "Skinny" and "Fat" DNA: Two New Double Helices. *J. Am. Chem. Soc.* 2018, *140*, 11655−11660.

(11) Glushakova, L. G.; Bradley, A.; Bradley, K. M.; Alto, B. W.; Hoshika, S.; Hutter, D.; Sharma, N.; Yang, Z.; Kim, M. J.; Benner, S. A. High-throughput multiplexed xMAP Luminex array panel for detection of twenty two medically important mosquito-borne arboviruses based on innovations in synthetic biology. *J. Virol. Methods* 2015, *214*, 60−74.

(12) Zhang, L.; Wang, S.; Yang, Z.; Hoshika, S.; Xie, S.; Li, J.; Chen, X.; Wan, S.; Li, L.; Benner, S. A.; Tan, W. An Aptamer-Nanotrain Assembled from Six-Letter DNA Delivers Doxorubicin Selectively to Liver Cancer Cells. *Angew. Chem., Int. Ed.* 2020, *59*, 663−668.

(13) Kool, E. T.; Morales, J. C.; Guckian, K. M. Mimicking the Structure and Function of DNA: Insights into DNA Stability and Replication. *Angew. Chem., Int. Ed.* 2000, *39*, 990−1009.

(14) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R., Jr.; Romesberg, F. E. A semi-synthetic organism with an expanded genetic alphabet. *Nature* 2014, *509*, 385−388.

(15) Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* 2013, *31*, 453−457.

(16) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A. Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem., Int. Ed.* 2010, *49*, 177−180.

(17) Hoshika, S.; Leal, N. A.; Kim, M. J.; Kim, M. S.; Karalkar, N. B.; Kim, H. J.; Bates, A. M.; Watkins, N. E., Jr.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J., Jr.; Georgiadis, M. M.; Benner, S. A. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 2019, *363*, 884−887.

(18) Gao, J.; Liu, H.; Kool, E. T. Assembly of the complete eight-base artificial genetic helix, xDNA, and its interaction with the natural genetic system. *Angew. Chem., Int. Ed.* 2005, *44*, 3118−3122.

(19) Cote, M. L.; Yohannan, S. J.; Georgiadis, M. M. Use of an N-terminal fragment from moloney murine leukemia virus reverse transcriptase to facilitate crystallization and analysis of a pseudo-16-mer DNA molecule containing G-A mispairs. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2000, *56*, 1120−1131.

(20) Georgiadis, M. M.; Singh, I.; Kellett, W. F.; Hoshika, S.; Benner, S. A.; Richards, N. G. Structural basis for a six nucleotide genetic alphabet. *J. Am. Chem. Soc.* 2015, *137*, 6947−6955.

(21) Malyshev, D. A.; Pfaff, D. A.; Ippoliti, S. I.; Hwang, G. T.; Dwyer, T. J.; Romesberg, F. E. Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chemistry* 2010, *16*, 12650−12659.

(22) Calladine, C. R.; Drew, H. R. A base-centred explanation of the B-to-A transition in DNA. *J. Mol. Biol.* 1984, *178*, 773−782.

(23) Ng, H. L.; Kopka, M. L.; Dickerson, R. E. The structure of a stable intermediate in the A ↔ B DNA helix transition. *Proc. Natl. Acad. Sci. U.S.A.* 2000, *97*, 2035−2039.

(24) Vargason, J. M.; Henderson, K.; Ho, P. S. A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A.* 2001, *98*, 7265−7270.

(25) Lu, X. J.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003, *31*, 5108−5121.

(26) Li, S.; Olson, W. K.; Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* 2019, *47*, W26−W34.

(27) Lu, X. J.; Shakked, Z.; Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* 2000, *300*, 819−840.

(28) Timsit, Y.; Moras, D. DNA self-fitting: the double helix directs the geometry of its supramolecular assembly. *EMBO J.* 1994, *13*, 2737−2746.

(29) Mayer-Jung, C.; Moras, D.; Timsit, Y. Hydration and recognition of methylated CpG steps in DNA. *EMBO J.* 1998, *17*, 2709−2718.

(30) Hays, F. A.; Teegarden, A.; Jones, Z. J.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102*, 7157−7162.

(31) Richards, N. G. J.; Georgiadis, M. M. Toward an Expanded Genome: Structural and Computational Characterization of an Artificially Expanded Genetic Information System. *Acc. Chem. Res.* 2017, *50*, 1375−1382.

(32) Goodwin, K. D.; Long, E. C.; Georgiadis, M. M. A host-guest approach for determining drug-DNA interactions: an example using netropsin. *Nucleic Acids Res.* 2005, *33*, 4106−4116.

(33) Seela, F.; Peng, X.; Li, H. Base-pairing, tautomerism, and mismatch discrimination of 7-halogenated 7-deaza-2'-deoxyisoguanosine: oligonucleotide duplexes with parallel and antiparallel chain orientation. *J. Am. Chem. Soc.* 2005, *127*, 7739−7751.

(34) Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A. Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res.* 2006, *34*, 6095−6101.

(35) Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, *66*, 479−485.

(36) Cowtan, K. Automated nucleic acid chain tracing in real time. *IUCrJ* 2014, *1*, 387−392.

(37) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 1997, *53*, 240−255.

(38) Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2001, *57*, 1373−1382.

(39) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004, *60*, 2126−2132.

(40) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr., Sect. D: Struct. Biol.* 2019, *75*, 861−877.

https://doi.org/10.1021/jacs.2c05255
J. Am. Chem. Soc. 2022, 144, 15603−15611