Conformation and Pairing Properties of an O^6-Methyl-2′-deoxyguanosine-Directed Benzimidazole Nucleoside Analog in Duplex DNA

Andrew H. Kellum, Jr.,[⊥] Pradeep S. Pallan,[⊥] Arman Nilforoushan, Shana J. Sturla, Michael P. Stone,* and Martin Egli*

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | 🅢🅘 Supporting Information

ABSTRACT: O^6-Methyl-2′-deoxyguanosine (O^6-MeG) is one of the most common DNA lesions and arises as a consequence of both xenobiotic carcinogens and endogenous methylation by S-adenosylmethionine. O^6-MeG frequently causes G-to-A mutations during DNA replication due to the misincorporation of dTTP and continued DNA synthesis. Efforts to detect DNA adducts such as O^6-MeG, and to understand their impacts on DNA structure and function, have motivated the creation of nucleoside analogs with altered base moieties to afford a more favorable interaction with the adduct as compared to the unmodified nucleotide. Such analogs directed at O^6-MeG include benzimidazolinone and benzimidazole nucleotides, as well as their extended π surface analogs naphthimidazolinone and napthimidazole derivatives. These analogs form a more stable pair with O^6-MeG than with G, most likely due to a combination of H-bonding and stacking. While extending the π surface of the analogs enhances their performance as adduct-directed probes, the precise origins of the increased affinity between the synthetic analogs and O^6-MeG remain unclear. To better understand relevant conformational and pairing properties, we used X-ray crystallography and analyzed the structures of the DNA duplexes with naphthimidazolinone inserted opposite G or O^6-MeG. The structures reveal a complex interaction of the analog found either in an anti orientation and stacked inside the duplex, either above or below G or O^6-MeG, or in a syn orientation and paired opposite G with formation of a single H-bond. The experimental structural data are consistent with the stabilizing effect of the synthetic analog observed in UV melting experiments and calculations and moreover reveal that the origin of these observations appears to be superior stacking between O^6-MeG and the extended π system of the synthetic probe.
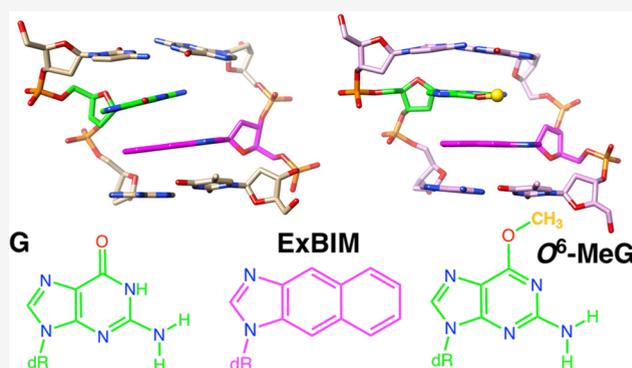
INTRODUCTION

Among the many forms of chemical modification of the canonical bases in DNA, alkylation of purine and pyrimidine bases is ubiquitous, with methylation being the most common.[1−3] Thirteen positions in DNA bases can be spontaneously methylated, and guanine constitutes the major target.[4] Exogenous sources of methylation include anticancer drugs such as Temozolomide, N-methyl-$N′$-nitro-N-nitrosoguanidine, and dacarbazine[3] and compounds present in food, beverages, and tobacco smoke.[5] Methylation of DNA also occurs by the endogenously produced compound S-adenosylmethionine.[6] Despite the fact that methylation at N^7 of dG (N^7-MeG) dominates over methylation at O^6 (O^6-MeG),[7] the latter is the most miscoding and mutagenic methylation adduct, stimulating strong interest in understanding its structure−function relationships and advancing strategies to detect its occurrence.[1]

Although O^6-MeG is especially repaired by O^6-methylguanine-DNA methyltransferase (MGMT[8−10]), other repair pathways include base excision repair, BER, mismatch repair, MMR, and trans-lesion synthesis, TLS,[3,4,11] the replication of DNA templates with residual O^6-MeG leads to insertion of C, the correct base, or T, triggering G:C → A:T transitions.[12] This miscoding is a consequence of the pairing properties of the adduct and the relative stability of its pairs O^6-MeG:C and O^6-MeG:T, as well as the particular active site environment of a DNA polymerase (pol).[13] When G is replaced by O^6-MeG in DNA duplexes, melting temperatures (T_m) decreased by 19−26 °C.[13] Interestingly, the O^6-MeG:T pair is less stable, but generally favored by enzymatic incorporation over C.[14,15] O^6-

MeG forms a wobble-like pair with C, as observed in NMR solution studies[16] (Figure 1A), whereas in the active sites of
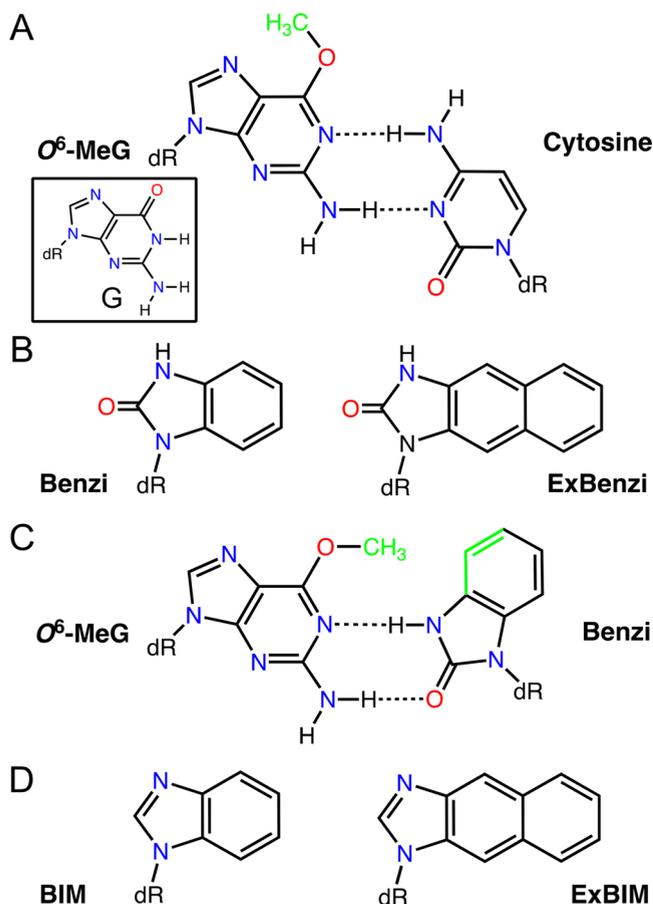
Figure 1. (A) Wobble-like pairing mode between O^6-MeG and cytosine with the O^6-methyl group in a proximal orientation (i.e., turned away from C). The framed inset shows G. (B) Structures of Benzi and ExBenzi. (C) O^6-MeG:Benzi pair observed at the active site of KlenTaq DNA polymerase with the O^6-methyl group in a distal orientation (i.e., turned toward Benzi). (D) Structures of BIM and ExBIM. dR is 2′-deoxyribose; nitrogen and oxygen atoms and the O^6-methyl group are colored in blue, red, and green, respectively; and bonds of the Benzi six-membered ring directed toward the O^6-methyl group are highlighted in green.

polymerases, base-pair geometries can change. For example, for the replicative pol I from *Bacillus stearothermophilus*, T is inserted over C by a factor of 10, and O^6-MeG:C was isosteric with G:C; the authors invoked a rare C tautomer to explain their findings.[17] However, at the active site of the Y-family TLS pol Dpo4 from *Sulfolobus solfataricus*, which preferentially inserts C relative to T by a factor of 4, these same base pairs displayed geometries similar to those found by NMR in solution, that is, wobble-like (O^6-MeG:C) and Watson−Crick-like (O^6-MeG:T), with the O^6-methyl group pointing away from N^4(C) (Figure 1A) and O^4(T), respectively.[18] Human TLS pol ι favors T over C insertion by a factor of 6, and both O^6-MeG:C and O^6-MeG:T adopted a Hoogsteen pairing mode at its active site.[19] Human TLS pol η, which is more proficient at replicating past the adduct than are the other pols, inserts dCTP and dTTP to almost equal degrees.[20] Crystal structures of pol η complexes showed different degrees of shearing between O^6-MeG and incoming dCTP (insertion stage) and

dC (extension stage). The corresponding complex structures with O^6-MeG opposite incoming dTTP (insertion stage) and dT (extension stage) revealed a Watson−Crick-like geometry in both cases. In addition to templating insertion of C or T, in studies aimed to create new biotechnologies to detect and amplify O^6-MeG, KlenTaq DNA polymerase and its M747 K mutant catalyze the specific incorporation of a synthetic adduct-directed benzimidazole-derived nucleoside triphosphate analog (BenziTP; Figure 1B). This is based on a pairing mode with Benzi in the syn conformation, with the formation of two H-bonds and the O^6-methyl group in the proximity of the Benzi phenyl ring (Figure 1C).[21]

Various adduct-directed synthetic nucleosides exist and are capable of discriminating between G and O^6-alkylguanine adducts, including O^6-MeG, in polymerase active sites as well as in duplex DNA.[22−24] Four such analogs include the aforementioned Benzi (derived from benzimidazolinone) and BIM (derived from benzimidazole) and their corresponding naphthyl derivatives, ExBenzi and ExBIM, respectively (Figure 1B,D). Interestingly, Dpo4 was able to extend from Benzi opposite O^6-MeG, but not from the analog paired with G.[25,26] The interactions between the partners in the latter pair are distinct from those in the O^6-MeG:Benzi pair as the N^7 hydrogen of Benzi sits opposite the N^1 hydrogen on G. Further extension studies with human pol ζ from base pairs that contained O^6-MeG also supported the importance of H-bonding, for example, a higher catalytic efficiency with base pairs featuring three H-bonds versus only two or two H-bonds versus only one, although there were some exceptions and stacking seemed to also play a role.[27] Via use of the KlenTaq mutant KTqM747 K that was particularly specific, Benzi was incorporated more efficiently opposite O^6-MeG than G, and was also well incorporated opposite two other O^6-alkylG adducts, O^6-benzylguanosine (O^6-BnG) and O^6-carboxyme-thylguanosine (O^6-CMG).[23] While interactions of base pairs in pol active sites are highly influenced by the particular polymerase used, understanding the innate hybridization properties of DNA modifications is important to understand the impacts of chemical modifications on DNA and to devise hybridization-based probe strategies for detecting DNA adducts in particular sequences.

Structural studies of modified duplex oligonucleotides revealed the formation of a wobble-like pair between O^6-MeG and C using NMR in solution[16] (Figure 1A), and both solution NMR and X-ray crystallography provided evidence for the existence of a Watson−Crick-like geometry of the O^6-MeG:T pair in duplex DNA.[16,28] While structural data for O^6-MeG placed opposite synthetic nucleoside probes remain elusive, a structure of O^6-BnG paired with a perimidinone-derived synthetic nucleoside revealed the potential for interstrand stacking interactions as a basis for increasing duplex stability when pairing DNA adducts with synthetic probes.[29] Indeed, for pairs of O^6-MeG with canonical bases, H-bonding interactions dominate, and in the NMR structure of O^6-MeG:T, there is an H-bond between the exocyclic amine of the adduct and the O^2 keto group of T, plus a somewhat long H-bond between N^1(O^6-MeG) and the imino proton of N^3. In related crystal structures, bifurcated H-bonds and shearing were observed between O^6-EtG and C in B-form DNA,[30] and a Watson−Crick-like O^6-MeG:C pair formed in Z-DNA.[31] Oligonucleotides with O^6-MeG placed opposite C or T exhibit dramatically reduced thermal stability as compared to the corresponding oligos with G:C or G:T pairs, respectively.[13−15]

Conversely, placing Benzi and BIM opposite O^6-MeG caused T_m values to increase by 4 °C relative to the corresponding duplexes with the analogs opposite G.[32] The apparent stacking interaction between the aromatic ring of Benzi and the O^6-benzyl group of the opposing guanine motivated the creation of ExBenzi and ExBIM (Figure 1B,D), with the intent to increase the π surfaces and improve their function as probes of DNA adducts.[32,33]

The extension of the π surface area of the analogs indeed enabled the development of new methods for detecting guanine alkylation in DNA. For example, melting temperatures (T_m) of duplexes between the KRAS codon 13 13mer ("Cod13", Figure 2) containing either G or O^6-MeG at the

A   Dickerson Drew Dodecamer: **"DDD"**

     5'-C¹G²C³G⁴A⁵A⁶T⁷T⁸C⁹G¹⁰C¹¹G¹²-3'

B   DDD with ExBIM (X⁹) opposite G⁴: **"GX-DDD"**

     5'-C¹G²C³G⁴A⁵A⁶T⁷T⁸X⁹G¹⁰C¹¹G¹²-3'

C   DDD with X⁹ opposite O^6-MeG⁴ (G*⁴): **"G*X-DDD"**

     5'-C¹G²C³G*⁴A⁵A⁶T⁷T⁸X⁹G¹⁰C¹¹G¹²-3'

D   KRAS Cod13 sequence: **"Cod 13"**

     5'-C¹T²G³G⁴T⁵G⁶G⁷C⁸G⁹T¹⁰A¹¹G¹²G¹³-3'

E   DDD with X⁹ and Cod13 sequence context for G⁴: **"GGX-DDD"**

     5'-C¹G²G³G⁴C⁵A⁶T⁷G⁸X⁹C¹⁰C¹¹G¹²-3'

F   DDD with X⁹ and Cod13 sequence context for G*⁴: **"GG*X-DDD"**

     5'-C¹G²G³G*⁴C⁵A⁶T⁷G⁸X⁹C¹⁰C¹¹G¹²-3'

Figure 2. DNA oligonucleotide sequences (A) based on the Dickerson Drew dodecamer, DDD, and (B, C, E, and F) used for X-ray crystallographic studies. (D) KRAS Cod13 sequence used for solution NMR (13mers modified with EXBIM = X or X and O^6-MeG = G* are not shown). G or G* as well as X and flanking residues are highlighted in red. The abbreviated names for the oligonucleotides used throughout the text and based on the underlined residues (sequences B, C, E, and F) are shown in boldface and quotation marks.

central position and complementary gold nanoparticles decorated with oligonucleotides containing ExBenzi or ExBIM placed opposite native or adducted G in target sequences revealed an advantage of O^6-MeG relative to G, which was coupled with a colorimetric change due to the resulting change in aggregation of the gold nanoprobes.[34] Thus, the T_m of ExBenzi:O^6-MeG was 3 °C higher than that of ExBenzi:G, and the T_m of ExBIM:O^6-MeG was 3.3 °C higher than that of ExBIM:G. Standard T_m measurements using Cod13 with its complementary strand and ExBIM:G or ExBIM:O^6-MeG in the center of the duplex are consistent with the data using nanoprobe aggregates. The T_m of the duplex with ExBenzi:O^6-MeG was 3 °C higher than that of the duplex with ExBenzi:G (55 °C vs 52 °C, respectively). The T_m of the duplex with ExBIM:O^6-MeG was 2.3 °C higher than that of the duplex with ExBim:G (54.0 °C vs 51.7 °C, respectively).[34] For comparison, the T_m of the 13mer duplex with a central C:G pair was 10.8 °C higher than that of the duplex with a C:O^6-MeG pair (66.0 °C vs 55.2 °C, respectively).

To date, no structures of DNA duplexes containing the ExBenzi or ExBIM adducts opposite either G or O^6-MeG have

been determined. Thus, it is not known how they discriminate between G and the methylated adduct. Moreover, the precise origins of the increased stability seen with DNA duplexes carrying these analogs remain unclear. In particular, the relative contributions of H-bonding and stacking are currently not known. Possible pairing modes entail a syn orientation of ExBIM with the formation of a bifurcated H-bond between N^7 of ExBIM and N^1H and N^2H of G and a single H-bond between O^6-MeG and ExBIM (N^2H⋯N^7) as well as a close approach between the O^6-methyl group and the edge of the ExBIM π system, as seen in MD simulations of modified duplexes.[33] To visualize the pairing modes of ExBIM opposite G and O^6-MeG, we used X-ray crystallography to study the 3D structures of various DNA duplexes carrying the ExBIM analog opposite either G or O^6-MeG (Figure 2).

■ EXPERIMENTAL SECTION

Oligonucleotide Synthesis and Purification. The ExBIM nucleoside (1'-β-[1-naphtho[2,3-d]imidazole]-2'-deoxy-D-ribofuranose) and the corresponding nucleoside phosphoramidite 1'-β-[1-naphtho[2,3-d]imidazole]-3'-2-cyanoethyl-tetraisopropylphosphitidyl-5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-D-ribofuranose were prepared as described in ref 29. All oligonucleotides were synthesized by the solid-phase phosphoramidite approach[35] and purified by reversed phase HPLC, followed by characterization by ESI−MS as reported previously.[34]

UV Melting Experiments. T_m measurements for the GX- and G*X-DDDs (Figure 2B,C) were carried out using a Cary 100 Bio UV−vis spectrophotometer (Agilent Technologies Inc., Santa Clara, CA). The conditions were 9 μM DNA and 1 M NaCl at pH 7. Absorbance versus temperature profiles were acquired at 260 nm between 5 and 80 °C with a ramp rate of 1 °C per minute. A_{260} values were measured at 0.2 °C intervals. T_m values were extracted as the maxima of the first derivatives of smoothened melting curves with the Cary WinUV software (Version 3.0).

Expression and Purification of *Bacillus halodurans* RNase H. The C-terminal fragment of *B. halodurans* RNase H (*Bh*RNase H) encompassing amino acids M58 to K196 and featuring a D132N mutation was cloned into the PET15b vector with an N-terminal His₆ tag and a thrombin cleavage site, expressed in *E. coli* BL21 cells, and purified following published procedures.[36−38] The protein solution was concentrated to ca. 20 mg/mL.

Crystallization Experiments. All crystallizations were performed by the sitting drop vapor diffusion technique. *Bh*RNase H and DNA dodecamer solutions were mixed in a 1:1 ratio in the presence of 5 mM MgCl₂. For GX- and G*X-DDDs, 135 nL of the DNA:RNase H complex was mixed with 135 nL of a crystallization solution containing 0.1 M HEPES sodium pH = 7.5, 10% v/v 2-propanol, and 20% w/v polyethylene glycol 4000. For GGX- and GG*X-DDDs, 135 nL of the DNA:RNase H complex was mixed with 135 nL of a crystallization solution containing 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, and 30% w/v polyethylene glycol 8000. Crystals of the complexes between RNase H and the GX-, G*X-, and GGX-DDDs were mounted in nylon loops, cryoprotected using crystallization solution containing 20% glycerol, and plunged into liquid nitrogen prior to transport to the X-ray synchrotron in a dryshipper.

X-ray Data Collection, Structure Determination, and Refinement. Diffraction data were collected on the 21-ID-G beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL). Data sets were collected for multiple crystals of all three complexes. Crystals were kept at 100 K during data collection, and diffraction data were recorded on a MARCCD 300 detector. Diffraction data were integrated, scaled, and merged with HKL2000.[39] A summary of selected crystal data and data collection parameters is provided in Table 1. The structures were solved by the molecular replacement technique using the program

Table 1. Selected Crystal Data, Data Collection Parameters, and Refinement Statistics[a]

complex	RNase H:GX-DDD	RNase H:G*X-DDD	RNase H:GGX-DDD
Data Collection			
space group	$P1$	$P2_1$	$P2_1$
unit cell lengths: a, b, c [Å]	64.98, 65.83, 95.72	59.51, 66.49, 69.18	36.89, 89.40, 73.89
unit cell angles: α, β, γ [deg]	84.54, 88.20, 62.49	90.00, 107.97, 90.00	90.00, 101.42, 90.00
resolution [Å]	32.73−2.30 (2.38−2.30)	29.74−2.32 (2.36−2.32)	28.59−1.74 (1.81−1.74)
wavelength [Å]	0.97856	0.97856	0.97856
no. of unique reflections	61 160 (5899)	22 274 (1062)	47 425 (4470)
completeness [%]	98.1 (94.4)	99.4 (95.7)	99.3 (93.9)
R-merge	0.133 (1.043)	0.121 (0.622)	0.106 (1.585)
R-pim	0.053 (0.756)	0.059 (0.796)	0.040 (0.716)
$I/\sigma(I)$	3.80 (0.80)	14.63 (1.81)	22.11 (0.63)
redundancy	3.8 (3.1)	4.9 (4.3)	7.8 (5.2)
Refinement			
no. of RNase-H/DNA duplexes per assym unit	8/3	3/1	3/1
resolution	32.73−2.30 (2.33−2.30)	29.74−2.32 (2.42−2.32)	28.59−1.74 (1.77−1.74)
number of reflections	60 645 (3113)	22 161 (1091)	44 462 (2801)
R-work	0.199 (0.295)	0.195 (0.259)	0.202 (0.314)
R-free	0.261 (0.346)	0.265 (0.390)	0.246 (0.364)
no. of protein/nucleotide atoms	8517/1382	3183/498	3212/496
no. of waters/ions/ligands	265/2/20	119/3/9	228/4/8
rms deviations bonds [Å]	0.009	0.008	0.008
rms deviations angles [deg]	1.00	1.04	1.02
avg B-factor, protein-/nucleotide atoms [Å^2]	41.4/49.4	37.1/40.2	32.3/31.2
avg B-factor, H$_2$O/ions/ligands [Å^2]	42.2/53.0/46.7	45.0/61.4/45.3	37.7/31.5/42.0
PDB entry code	8CTY	8CTZ	8CU0

[a]Numbers in parentheses refer to the outermost shell.

MOLREP[40,41] and employing protein alone as the search model (PDB ID 3EY1). In the case of the RNase H:G*X-DDD and RNase H:GGX-DDD complexes (space group $P2_1$), there were three RNase H molecules and one duplex DNA in the crystallographic asymmetric unit. The initial refinement of the protein alone in Refmac5[42] resulted in R_{work} and R_{free} ranges of 0.37−0.39 and 0.39−0.44, respectively, whereby 5% of the reflections were set aside to compute the R_{free}. Inspection of the calculated Fourier $2F_0 − F_c$ sum and the $F_0 − F_c$ difference electron density maps revealed the complete DNA portion of the complex. Map visualization and model rebuilding were performed with the program Coot.[43] After the DNA duplex was placed, additional cycles of refinement were carried out in Refmac5.[40,42] In the next step, modified 2′-deoxynucleotides were incorporated in place of dG (O^6-MeG) or dC (ExBIM) at selected sites. Further isotropic refinement was continued in Refmac5, after the adapted dictionary files were added that were generated by using the program PRODRG.[44] In subsequent refinement cycles, water molecules were added (about 15−25 molecules in each refinement step) on the basis of the Fourier $2F_0 − F_c$ sum and the $F_0 − F_c$ difference electron density maps and were accepted on the basis of standard distance and B-factor criteria. As the refinement progressed, metal ions or acetate and other ligand molecules became clear in the electron density and were added. Final cycles of refinement were carried out using the program Phenix.[45]

In the case of the RNase H:GX-DDD structure (space group $P1$), we followed steps similar to those described above. However, of the three DNA duplexes in the unit cell, only two could be readily traced in the Fourier $2F_0 − F_c$ sum and $F_0 − F_c$ difference electron density maps. The third DNA duplex could only be partially built and refined using the program Refmac5.[40,42] The final refinement was carried out with the program Phenix.[45] Patches of the $2F_0 − F_c$ sum and $F_0 − F_c$ difference densities remain visible in the final structure. Selected refinement parameters and deviations from ideal geometries are summarized in Table 1. Examples of the quality of the final electron density for the three structures are depicted in Figure 3.

Data Deposition. Coordinates and X-ray structure factors for the three complexes were deposited in the RCSB Protein Data Bank with

ID codes 8CTY (GX-DDD:RNaseH complex), 8CTZ (G*X-DDD:RNase H complex), and 8CU0 (GGX-DDD:RNase H complex).

Figure Preparation. Structural figures were prepared with the program UCSF Chimera.[46]

■ RESULTS

Crystallization and Structure Determination of DNAs Containing ExBIM Opposite Either G or O^6-MeG in Complex with RNase H. Our initial efforts to analyze 3D structures of ExBIM-containing oligo-2′-deoxynucleotides with either G or O^6-MeG opposite the naphtho[2,3-d]imidazole moiety were directed at the Cod13 sequence (Figure 2D) and used solution NMR. We observed a temperature-dependent conformational change in which one structure was favored at a low temperature while a different structure was favored at a higher temperature. The temperature dependence was quantified by measuring the difference in chemical shifts for ExBIM protons. We hypothesize that at the lower temperature a looped-out structure dominates in which G or O^6-MeG was pushed out of the duplex, whereas at the higher temperature an arrangement with G or O^6-MeG stacked on ExBIM was preferred. A similar observation was evident for an oligo-2′-deoxynucleotide containing an extra cytidine. An extrahelical conformation was favored at lower temperature, while a stacked conformation was favored at higher temperature.[47] This stacked conformation was favored with ExBIM paired opposite O^6-MeG versus G upon lowering the temperature of the solution. Considering these complications, we did not calculate detailed 3D-structural models on the basis of solution NMR.

Given the heterogeneity of solution conformations, we turned to X-ray crystallography to better understand the
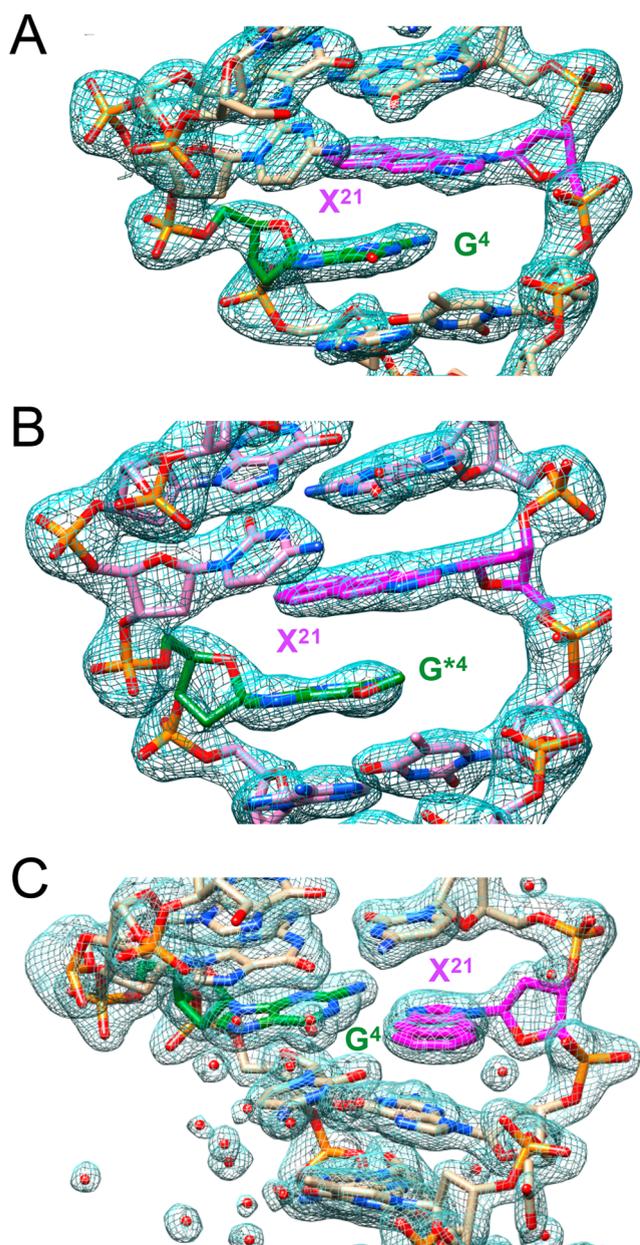
Figure 3. Quality of the final Fourier $2F_o - F_c$ sum electron density drawn at the $\sim 1\sigma$ threshold around a portion of the (A) GX-DDD1, (B) G*X-DDD, and (C) GGX-DDD DNA duplexes. DNA carbon, nitrogen, oxygen, and phosphorus atoms are colored in tan/pink, blue, red, and orange, respectively. Carbon atoms of residues G^4 and O^6-MeG^4 are highlighted in green, and those of ExBIM X^{21} are highlighted in magenta.

ExBIM interaction modes in DNA that afford discrimination by this analog between the O^6-MeG adduct and the native nucleoside. The targets for crystallization experiments were the self-complementary DNA 12mer d(CGCGAATTCGCG) (Dickerson-Drew Dodecamer, DDD; Figure 2A), with $\underline{G}^4$ representing either G or O^6-MeG and ExBIM replacing $\underline{C}^9$ (Figure 2B,C), and the DNA 12mer d(CGGGCATGCCCG) that features G or O^6-MeG at $\underline{G}^4$ flanked by G and C (as in the Cod13 sequence) and ExBIM in place of $\underline{C}^9$ (Figure 2E,F). However, despite extensive trials, we did not obtain viable crystals for X-ray diffraction data collection and structure determination.

We subsequently turned to an approach for solving crystal structures of DDDs with incorporated chemically modified nucleotides that relies on RNase H from *Bacillus halodurans* as a scaffolding protein. The enzyme cleaves the RNA strand in DNA:RNA hybrids, and crystal structures of *B. halodurans* RNase H bound to 12mer hybrids had been reported.[36] We found fortuitously that the enzyme crystallizes with the DDD duplex to form an inhibitor complex[37] and that crystal formation is quite tolerant to chemical modification of the DDD.[38] Both the enzyme alone and complexes with DDDs yield multiple crystal forms with varying unit cell contents. Comparison of the crystal structure of the native DDD with that of the duplex bound to RNase H demonstrated that the protein component in the latter crystal did not alter the DNA conformation to a significant degree.[37] Another advantage of the crystallization scaffold approach is that the crystal structure can be phased with molecular replacement using RNase H as the search model (the conformation of the protein remains basically unchanged in the presence of the DDD). We were able to crystallize three RNase H complexes with the DDD and the "Cod13"-DDD. The first is of the protein with a DDD that contains G opposite ExBIM (X) in both halves of the duplex; we will refer to this duplex as GX-DDD (Figure 2B). The second complex is of the protein with a DDD that contains O^6-MeG (G*) opposite ExBIM (X) in both halves; we will refer to this duplex as G*X-DDD (Figure 2C). The third complex is of the protein with the "Cod13"-DDD that contains G opposite ExBIM (X) in both halves; we will refer to this duplex as GGX-DDD (Figure 2E). We did not obtain diffraction-quality crystals of the corresponding complex with the GG*X-DDD (Figure 2F).

Crystal structures of the three complexes were determined at resolutions between 1.74 and 2.32 Å: 2.3 Å (GX-DDD), 1.74 Å (GGX-DDD), and 2.32 Å (G*X-DDD). Selected crystal data, diffraction data collection statistics, and refinement parameters are listed in Table 1. Examples of the quality of the final electron densities are depicted in Figure 3. The contents of the crystallographic asymmetric units of the three structures are as follows: 8 RNase H molecules and 3 DNA duplexes (GX-DDD), 3 RNase H molecules and 1 duplex (G*X-DDD), and 3 RNase H molecules and 1 duplex (GGX-DDD) (Figure 4). In the first complex, only two duplexes were fully traceable in electron density maps, and we are calling these GX-DDD1 and GX-DDD2. Protein chains and DNA strands are consecutively labeled in the PDB coordinate files, that is, A–H for the 8 protein molecules, I and J for GX-DDD1, and K and L for GX-DDD2 (as well as M and N for the third, partially visible duplex). Similarly, in the G*X- and GGX-DDD complexes, the three protein chains are labeled A–C and the duplex strands are D and E. All three duplexes are self-complementary, but they do not exhibit crystallographic 2-fold symmetry in the structures. Therefore, nucleotides are numbered 1–24, and we refer to modified base pairs in the two duplex halves as G^4:X^{21} and G^{16}:X^9 from here on. In GX-DDD1, G^4 and X^9 are from the I strand and G^{16} and X^{21} are from the J strand. In GX-DDD2, G^4 and X^9 are from the K strand and G^{16} and X^{21} are from the L strand. In G*X-DDD, G^{*4} and X^9 are from the D strand and G^{*16} and X^{21} are from the E strand (Figure 5). Similarly, in GGX-DDD, G^4 and X^9 are from the D strand and G^{16} and X^{21} are from the E strand (Figure 6).

Overall Duplex Conformations and ExBIM Pairing Modes. Side-by-side comparisons of the modified 12mer duplexes reveal two ExBIM intercalation modes in the GX-
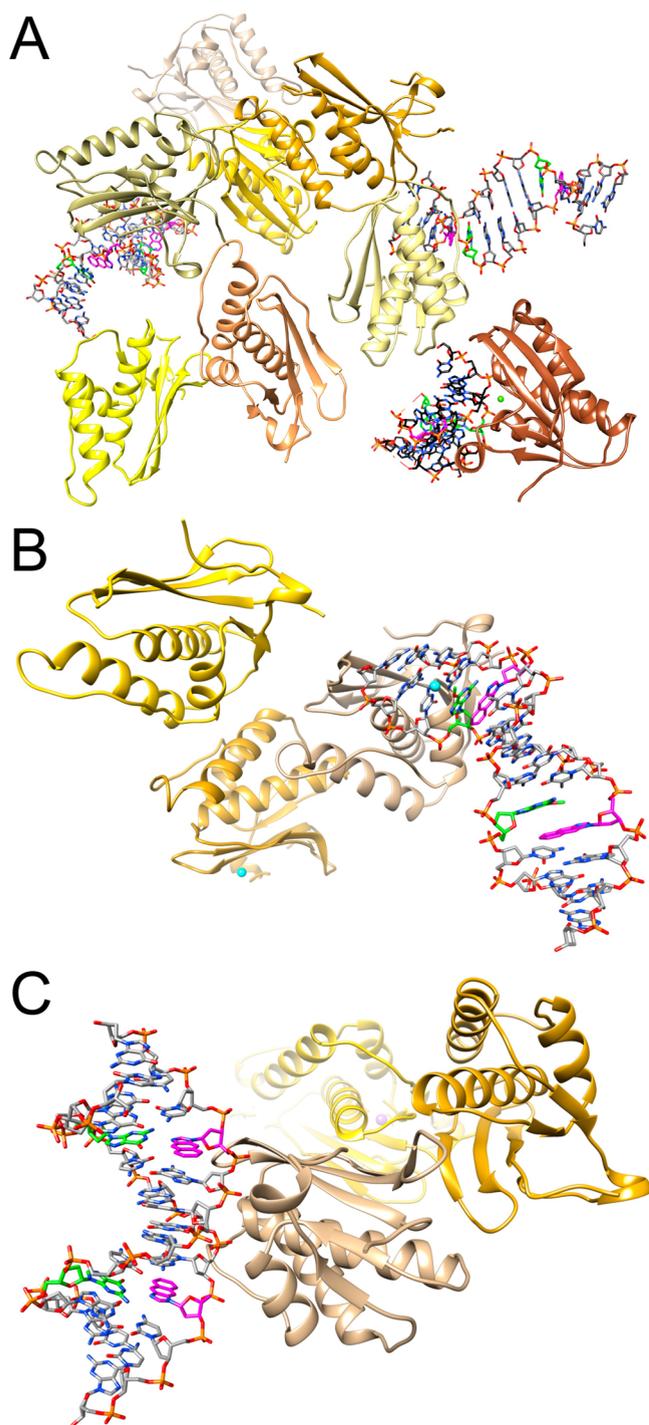
Figure 4. Contents of the crystallographic asymmetric units in the RNase H complexes of the (A) GX-, (B) G*X-, and (C) GGX-DDDs. DNA carbon, nitrogen, oxygen, and phosphorus atoms are colored in gray, blue, red, and orange, respectively, except for residues G^4/G^{16} or O^6-MeG$^4/O^6$-MeG^{16} (carbon atoms highlighted in green) and X^9/X^{21} (carbon atoms highlighted in magenta) in duplex strands.

positioned between the C^{15}:G^{10} base pair and G and O^6-MeG. In other words, in the upper half, ExBIM is intercalated on the 3′-side of G and O^6-MeG; in the lower half, it is intercalated on the 5′-side of G and O^6-MeG. In all cases, ExBIM adopts the anti conformation, which leaves no space for the base from the opposite strand to adopt a coplanar orientation with the analog. Accommodating ExBIM in place of C^9 (C^{21} in the second strand) and disrupting base pairing in favor of an intercalation mode results in various changes in the helical parameters, backbone torsion angles, and sugar puckers. Helical parameters as well as backbone torsion and pseudorotation phase cycle P angles were calculated with the 3DNA program[48] and are listed in Tables S1–S12. The GX-DDD and G*X-DDD duplexes can be overlaid with a root-mean-square deviation (rmsd) of 1.8 Å for 566 atom pairs (G^4/G^{*4} and G^{16}/G^{*16} bases were excluded). Interestingly, in the structure of the complex between GX-DDD and RNase H, the GX-DDD2 duplex exhibits a symmetrical intercalation mode; that is, ExBIM is intercalated on the 5′-side of G and O^6-MeG in both halves (Figure 5A; the situation also seen in the lower half of GX-DDD1 and G*X-DDD). The two duplex halves of GX-DDD2 can be overlaid with an rmsd of 0.4 Å for all atom pairs.

In the third complex, G pairs with ExBIM, which adopts the syn conformation in both duplex halves (Figure 6A). These halves can be overlaid with an rmsd of 0.9 Å for all atom pairs. In Figure 6, the GGX-DDD duplex is depicted together with the crystal structure determined for the native DDD in complex with RNase H.[37] The latter duplex exhibits a 2-fold crystallographic symmetry, and the sugar–phosphate backbones of the DDD and GGX-DDD duplexes can be overlaid with an rmsd of 2.7 Å for 242 atom pairs. This relatively large rmsd is not surprising given the different sequences of the two 12mer DNAs (Figure 2). Thus, the GGX-DDD lacks the DDD central A-tract (AATT) that is associated with a narrow minor groove.

Comparison of G-ExBIM and O^6-MeG-ExBIM Interactions in Duplex DNA. The structures of the GX-DDD and G*X-DDD complexes allow a direct comparison of the stacking interactions between ExBIM and G and ExBIM and O^6-MeG, respectively, as well as with the flanking C:G and A:T base pairs. Figure 7 provides an overview of the stacking modes in the upper and lower halves of the two duplexes, viewed either into the major groove or along the base stack. As mentioned above, G/O^6-MeG and ExBIM swap positions inside the $C^3pG^4(G^{*4})pA^5$:$T^{20}pX^{21}pG^{22}$ and $C^{15}pG^{16}(G^{*16})$-pA^{17}:$T^8pX^9pG^{10}$ trimers from the upper and lower halves, respectively, of the GX-DDD1 and G*X-DDD duplexes (Figure 5B,C). Looking into the major grooves of these duplexes, we see clear conformational differences between the G-ExBIM and G*-ExBIM stacks in the respective duplexes (Figure 7A–D). The base planes of G^4 and X^{21} as well as those of G^{16} and X^9 are all inclined relative to the vertical direction of the helix axis in the GX-DDD1 duplex. This results in predominant intrastrand stacking between G^4 and C^3 (Figure 7A,E) and G^{16} and A^{17} (Figure 7C,G), while maintaining a parallel arrangement between the guanine and naphtho[2,3-d]imidazole planes. By comparison, the base planes of G^{*4} and X^{21} as well as those of G^{*16} and X^9 are practically normal to the (vertical) direction of the helix axis in the G*X-DDD duplex. Moreover, the flanking base pairs (C^3:G^{22}/A^5:T^{20} and C^{15}:G^{10}/A^{17}:T^8, respectively) in this duplex exhibit only a minor roll and buckle as compared to the corresponding base

DDD and G*X-DDD duplexes and H-bonded pairing between G and ExBIM in the GGX-DDD duplex (Figures 5 and 6). Comparison of the two duplex halves in the structures of GX-DDD1 and G*X-DDD shows that G/O^6-MeG and ExBIM trade places (Figure 5B,C). In the upper half of these duplexes, G and O^6-MeG are positioned between the C^3:G^{22} base pair and ExBIM. In the lower half of the duplexes, ExBIM is

Figure 5. ExBIM-DNA nearest neighbor interactions in the structures of the GX-DDD:RNase H and G*X-DDD:RNase H complexes: (A) GX-DDD2, (B) GX-DDD1, and (C) G*X-DDD. Carbon atoms of G/O^6-MeG and ExBIM are highlighted in green and magenta, respectively.
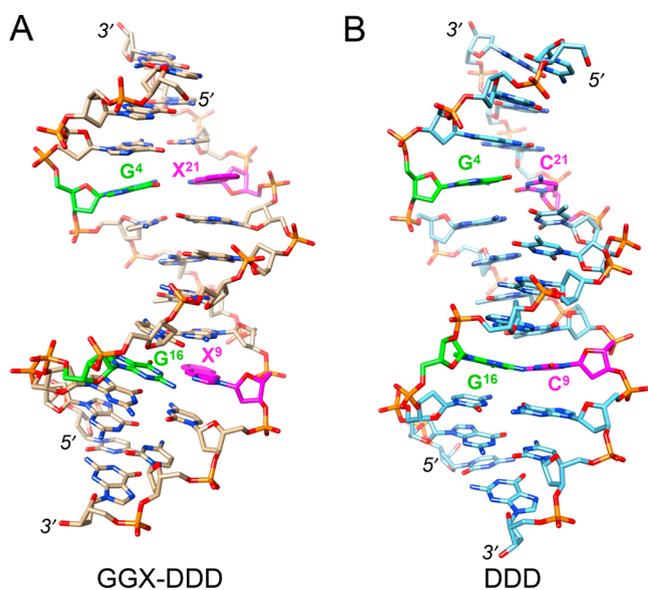
Figure 6. ExBIM-DNA nearest neighbor interactions in the structure of the GGX-DDD:RNase H complex. The conformation of (A) the GGX-DDD duplex from the complex is compared to that of (B) the native DDD bound to RNase H.[31] In the complex with RNase H, the native DDD sits on a crystallographic 2-fold rotation axis, and the two halves therefore adopt identical conformations (PDB ID 3D0P). Carbon atoms of G are highlighted in green, and those of ExBIM and C are highlighted in magenta.

pairs around G and ExBIM in the GX-DDD1 duplex (Figure 7A−D). This leads to enhanced interstrand stacking between both G^{*4} and G^{22} (Figure 7B,F) and G^{*16} and T^8 (Figure 7D,H). This is more easily apparent when looking approximately along the helix axis of the G*X-DDD duplex (Figure 7F,H) and contrasts with the lack of any interstrand stacking between G and neighboring nucleobases in the two halves of the GX-DDD1 duplex (Figure 7E,G). In addition, the adducted G affords better overlap with the ExBIM naphtho-[2,3-*d*]imidazole moiety in the lower half of G*X-DDD (Figure 7H) relative to the situation in GX-DDD1 (Figure 7G). In the upper halves of the GX-DDD1 and G*X-DDD

duplexes, the overlaps between ExBIM and G (Figure 7E) and G* (Figure 7F), respectively, are more similar by comparison. Thus, the O^6-MeG methyl group appears not to be stacked underneath the naphtho[2,3-*d*]imidazole moiety or above the base plane of A^5 there and instead juts into the major groove (Figure 7F).

In the structure of the second RNase H complex with DNA featuring G opposite ExBIM, the GGX-DDD duplex shows ExBIM in syn opposite standard *anti*-G, with the formation of a single H-bond between N^1H of G and N^7 of the analog (3.02 Å distance, Figure 8). The N^2 nitrogen of G is positioned at 3.55 Å from ExBIM N^7. The pairing geometry between G and ExBIM in the other half of the duplex is very similar (Figure 6), with distances of 2.99 and 3.46 Å between N^1 and N^7 and N^2 and N^7, respectively. The guanine and naphtho[2,3-*d*]imidazole planes show minor buckling, with the lower C^5:G^{20} base pair exhibiting moderate propeller twisting to presumably optimize intrastrand stacking with G^4 and X^{21}, respectively (Figure 8A). The view of the trimer along the helix axis demonstrates the predominance of intrastrand stacking between C^5 and G^4 and G^4 and G^3, on one hand, and G^{20} and X^{21} and X^{21} and C^{22}, on the other (Figure 8B). The reduced helical twist at the $G^{20}pX^{21}$ step results in an optimal overlap of guanine and the naphtho[2,3-*d*]imidazole moiety.

■ DISCUSSION

To better understand how physical interactions between the ExBIM nucleobase analog, that features an extended π surface, and G or the adduct O^6-MeG (Figure 1) govern their relative stabilities in duplex DNA, we determined the crystal structures of DNA dodecamer duplexes with ExBIM incorporated opposite either G or O^6-MeG. Crystallization experiments with the four duplexes alone (Figure 2B,C,E,F) did not result in viable crystals. As a consequence, we resorted to an approach that we had used successfully in the past, initially with the native DDD[37] and then with several chemically modified DDDs (ref 38 and cited references), co-crystallization with *B. halodurans* RNase H. This enzyme cleaves the RNA strand of an RNA:DNA hybrid[36] and also produces several crystal forms alone and multiple others in complex with DNA dodecamers. Indeed, crystallization trials with complexes
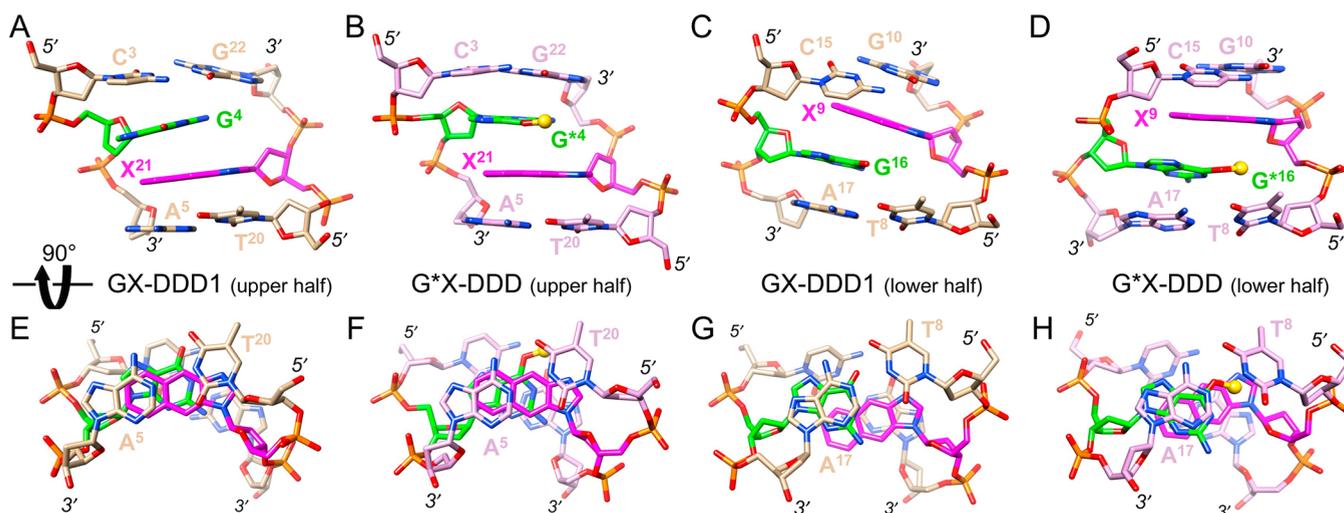
Figure 7. Stacking interactions involving G/O^6-MeG and ExBIM in the GX-DDD1 and G*X-DDD duplexes. Top panels depict views into the major groove of DNA trimers from the upper duplex halves of (A) GX-DDD1 and (B) G*X-DDD, as well as from the lower duplex halves of (C) GX-DDD1 and (D) G*X-DDD. Bottom panels depict views of the above trimers rotated around the horizontal axis by 90°: (E) GX-DDD1 and (F) G*X-DDD (from the upper duplex halves), and (G) GX-DDD1 and (H) G*X-DDD (from the lower duplex halves). The color code matches that in Figure 5, O^6-MeG methyl carbon atoms are highlighted as gold spheres, residues are labeled, and strand polarities are indicated.
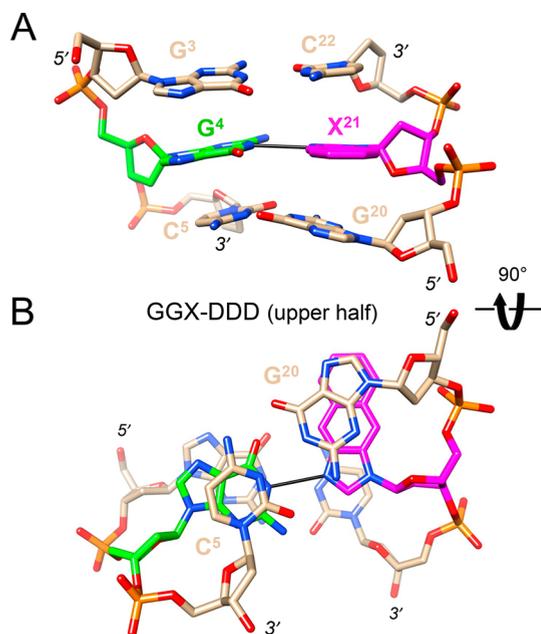
Figure 8. Pairing between G and ExBIM in the GGX-DDD duplex. (A) G^4:X^{21} and the flanking G^3:C^{22} and C^5:G^{20} base pairs (A) viewed into the major groove and (B) rotated by 90° around the horizontal axis and viewed approximately along the normal to the bottom base pair. The color code matches that in Figure 6, residues are labeled, and the H-bond (3 Å) between N^{1}H of G and N^7 of ExBIM is indicated with a thin solid line.

between RNase H and the GX-DDD, G*X-DDD, and GGX-DDD DNA duplexes were successful and produced crystals that diffracted to between 1.7 and 2.3 Å. Crystals were not obtained for the complex between RNase H and the GG*X-DDD duplex. An advantage of the scaffolding approach, that is, using RNase H:DDD complexes to grow crystals instead of targeting modified DNA dodecamers alone, is that the protein component can be used as the search model for phasing the structures by molecular replacement. In crystals of RNase H

alone or containing complexes with substrate (RNA:DNA) or inhibitor (e.g., DDD) duplexes, the enzyme adopts virtually identical conformations.[36−38] However, this is often not the case for the DNA portion, for example, DDDs with chemical modifications and/or altered sequences.

After the crystal structures of the three RNase H complexes were solved, it became evident why crystallization efforts directed at the GX-, G*X-, and GGX-DDDs alone might have been unsuccessful. The GX-DDD and G*X-DDD duplexes exhibit stacked arrangements of G and X and G* and X, respectively, in both halves, thereby converting them to resemble 14-bp duplexes and altering local helical rise and twist values. These changes obviously preclude crystal packing arrangements of the GX- and G*X-DDDs that are identical or closely similar to those seen in the lattice of the native DDD, and evidently others that result in diffracting species (at least within the confines of the crystallization conditions we had screened). The sequence of the GG-DDD differs significantly from that of the native DDD. The most important consequence is the loss of the central "A-tract" (AATT) in the latter that results in the characteristic, narrow minor groove of the DDD and the formation of the iconic water spine.[49−51] These differences are important with regard to duplex packing and crystal lattice formation, even in the absence of the additional introduction of ExBIM (GGX-DDD) or both ExBIM and O^6-MeG (GG*X-DDD). Even by screening a large number of conditions, we did not succeed in obtaining crystals of these duplexes, and, in the case of the GG*X-DDD duplex, the co-crystallization approach with RNase H was also unsuccessful.

The three crystal structures of RNase H:DNA duplex complexes were determined to have different space groups and/or unit cell dimensions, and also packing interactions (Table 1 and Figure 4). Despite these differences, the GX-DDD and G*X-DDD duplexes exhibit similar stacking arrangements between ExBIM and G and O^6-MeG, respectively, as well as with 5′- and 3′-adjacent bases (Figure 5), thereby allowing us to correlate the conformation and pairing stability independent of crystal packing. Moreover, ExBIM-G

and ExBIM-O^6-MeG stacks and their flanking base pairs are not contacted by any protein side chains in the structures of the RNase H:GX-DDD and RNase H:G*X-DDD complexes, respectively. The same applies to the crystal structure of the RNase H:GGX-DDD complex in which G and ExBIM adopt an H-bonded, coplanar arrangement (Figure 6). These observations lead us to the conclusion that the distinct interaction modes between ExBIM and G (and O^6-MeG) seen in the modified DDD duplexes (stacking) and in the modified GG-DDD duplex (H-bonding) are the result of different sequence contexts. As pointed out above, in the GX-DDD and G*X-DDD duplexes, G and ExBIM are located between C:G and A:T pairs. By contrast, in the GGX-DDD duplex, G and ExBIM are located between G:C and C:G pairs. Irrespective of the sequence contexts and observed interaction modes of G or the adduct O^6-MeG opposite ExBIM, the structural data allow a better understanding of the previously established stability differences (e.g., T_m data for modified duplexes) between the G:ExBIM and O^6-MeG:ExBIM pairs. We also believe, based on the structural data, that the usefulness of ExBIM is not limited to the sequences studied here, but that the analog can be applied to target adducted bases in diverse sequence contexts.

Published melting temperature data for gold nanoprobe aggregates formed by an ExBIM nanoprobe targeting the Cod13 sequence with G or G* at the central position (Figure 2) show that O^6-MeG opposite ExBIM increases the T_m as compared to the pairing with G.[34] ExBIM dangling ends in a DNA hexamer duplex 5′-XCGCGCG-3′:3′-GCGCGCX-5′ resulted in a higher stabilization (ΔG = −11.8 kcal/mol) than Benzi (−10.9), ExBenzi (−11.3) (see Figure 1 for structures), benzene (−9.4), naphthalene (−10.9), or pyrene (−11.3) dangling ends, demonstrating the superior stacking ability of the ExBIM analog.[33] The ExBIM elongated nucleoside also specifically stabilized O^6-MeG relative to G when placed (X) opposite either of them in the center of the duplex 5′-CCTACGG(G*)CACCAG-3′:3′-GGATGCXGTGGTC-5′ (Cod13 and complement). Standard UV melting assays showed a difference in T_m of ca. 10 °C in favor of the duplex with O^6-MeG opposite ExBIM.[33] We also assayed the stability of the GX-DDD and G*X-DDD duplexes using UV melting at 1 M NaCl concentration (pH 7) and established T_m values of 27 and 40 °C, respectively. These data paint a clear picture by showing the higher stability of ExBIM:O^6-MeG relative to ExBIM:G pairing and stacking as the underlying source rather than H-bonding.

The structures of the GX-DDD1 and G*X-DDD duplexes that reveal two shared stacking modes between G and ExBIM and O^6-MeG and ExBIM, respectively, allow a qualitative rationalization of the higher stability afforded by the latter pair. Both ExBIM positioned on the 5′-side of O^6-MeG in one-half of the duplex and on the 3′-side of O^6-MeG in the other exhibit improved overlaps between the π systems and the O^6-methyl group and/or flanking bases relative to the corresponding situations involving ExBIM and G. The G*X-DDD duplex also reveals more cross-strand stacking at the site of ExBIM incorporation as compared to the GX-DDD1 duplex. The conclusion that ExBIM:G and ExBIM:O^6-MeG pairings are associated with different contributions of intra- and interstrand stacking to their overall stabilities is evident from a comparison of trimer duplexes depicted in Figure 7. Thus, the O^6-MeG-ExBIM stacks in the G*X-DDD duplex are oriented more or less normal to the vertical helix axis, thereby resulting in less distorted geometries of base pairs above and below as

compared to the GX-DDD1 duplex. In the latter, G-ExBIM stacks are slanted relative to the helix axis, thereby producing good overlap between guanine and the elongated base analog from opposite strands, but also disrupting optimal stacking across the trimer by introducing deviations from base pair planarity.

Stacking between G and ExBIM and O^6-MeG and ExBIM as observed in the respective DDD duplexes very likely results in higher stability than the H-bonded pairing mode seen in the GGX-DDD duplex. The G:ExBIM pair features only a single H-bond as compared to three in a standard C:G pair. Moreover, the syn oriented ExBIM base does not produce the same degree of stacking in the trimer context (Figure 8) as the sum of the interactions seen in the GX-DDD1 and G*X-DDD duplexes (Figure 7). This is consistent with calculated relative contributions resulting from H-bonding and stacking in the corresponding G:ExBIM and O^6-MeG:ExBIM pairs.[33] Unlike the predicted G:ExBIM pair with H-bonds between both N^1 and N^2 of G and N^7 of ExBIM,[33] the observed pairs in the crystal structure of the GGX-DDD duplex feature only a single H-bond between N^1 of G and N^7 of ExBIM (avg distance 3.0 Å). By comparison, the average distance of 3.5 Å between the N^2 of G and the N^7 of ExBIM is too long for the formation of an effective H-bond (Figure 9A). Unfortunately, we do not
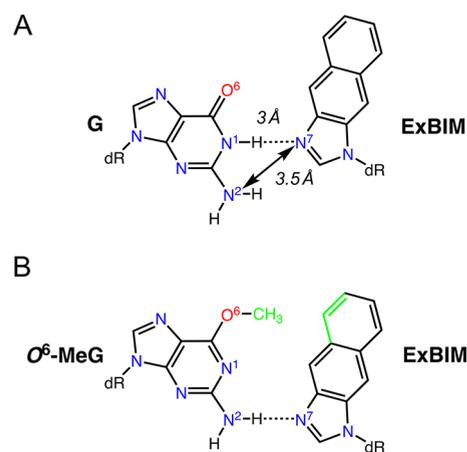
Figure 9. (A) Pairing between G and ExBIM observed in the crystal structure of the GGX-DDD duplex. Average distances between N^1H (G) and N^7 (ExBIM) and N^2H (G) and N^7 (ExBIM) seen in the structure are indicated. (B) Proposed pairing between O^6-MeG and ExBIM with a putative interaction between the distal O^6-methyl group and the edge of the outer ring of the base analog (bonds and atoms are highlighted in green).

have a structure of the GG*X-DDD duplex that would inform us about the geometry of an H-bonded O^6-MeG:ExBIM pair. However, we can speculate that the adducted G could shift slightly toward the major groove, thus allowing the formation of an H-bond between the exocyclic amino group (N^2) and the N^7 of ExBIM (Figure 9B). This would result in an edge-on interaction between the O^6-methyl group and ExBIM, not unlike that seen in the O^6-MeG:Benzi pair (Figure 1C) at the active site of a mutated KlenTaq DNA polymerase.[21] This is based on the assumption that O^6-MeG features the methyl group in the distal orientation (i.e., turned toward the Watson−Crick edge) as seen in the stacked scenarios in the G*X-DDD duplex (Figures 5C and 7B,D). As one would expect on the basis of steric reasons, and in contrast to the

distal orientation of the O^6-methyl group in the proposed pairing mode between O^6-MeG and ExBIM at a polymerase active site (Figure 9B), crystal structures of complexes between replicative[17] and lesion bypass polymerases[18,20] and DNA template:primer duplexes with O^6-MeG opposite incoming dCTP (Figure 1A) or dTTP show the O^6-methyl group in a proximal orientation of (i.e., turned away from the Watson–Crick edge).

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.2c00165.

> Selected helical parameters, backbone and glycosidic torsion angles, and sugar pucker values for the GX-DDD1, GX-DDD2, G*X-DDD, and GGX-DDD DNA duplexes (Tables S1–S12) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Martin Egli − *Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, United States;* ⓞ orcid.org/0000-0003-4145-356X; Email: martin.egli@vanderbilt.edu

Michael P. Stone − *Department of Chemistry, Vanderbilt University, College of Arts and Science, Nashville, Tennessee 37235, United States; Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, United States;* ⓞ orcid.org/0000-0002-0922-0216; Email: michael.p.stone@vanderbilt.edu

Authors

Andrew H. Kellum, Jr. − *Department of Chemistry, Vanderbilt University, College of Arts and Science, Nashville, Tennessee 37235, United States;* Present Address: Department of Chemistry, University of California, 900 University Avenue, Riverside, California 92521, United States

Pradeep S. Pallan − *Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, United States*

Arman Nilforoushan − *Department of Health Sciences and Technology, ETH Zürich, Zurich 8092, Switzerland*

Shana J. Sturla − *Department of Health Sciences and Technology, ETH Zürich, Zurich 8092, Switzerland;* ⓞ orcid.org/0000-0001-6808-5950

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.2c00165

Author Contributions

[⊥]A.H.K., Jr. and P.S.P. contributed equally to this work. CRediT: **Andrew H. Kellum** conceptualization, data curation, formal analysis, methodology, validation, writing-original draft; **Pradeep S. Pallan** conceptualization, data curation, formal analysis, methodology, validation, visualization, writing-original draft; **Arman Nilforoushan** resources; **Shana J. Sturla** conceptualization, resources, writing-original draft, writing-review & editing; **Michael P. Stone** conceptualization, funding acquisition, investigation, methodology, supervision, writing-original draft; **Martin Egli** conceptualization, data curation, formal analysis, funding acquisition, investigation, method-ology, supervision, validation, visualization, writing-original draft, writing-review & editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Loveless, A. Possible relevance of O-6 alkylation of deoxy-guanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. *Nature* **1969**, *223*, 206−207.

(2) Lawley, P. D. Carcinogenesis by alkylating agents. In *Chemical Carcinogens*, 2nd ed.; Searle, C. E., Ed.; American Chemical Society: Washington, DC, 1984; pp 325−484.

(3) Kondo, N.; Takahashi, A.; Ono, K.; Ohnishi, T. DNA damage induced by alkylating agents and repair pathways. *J. Nucleic Acids* **2010**, 543531.

(4) Christmann, M.; Roos, W. P.; Kaina, B. DNA methylation damage: formation, repair and biological consequences. In *Chromosomal Alterations: Methods, Results and Importance in Human Health*; Obe, G., Vijayalaxmi, Eds.; Springer-Verlag: Berlin, Heidelberg, 2007; pp 99−121.

(5) Drabløs, F.; Feyzi, E.; Aas, P. A.; Vaagbø, C. B.; Kavli, B.; Bratlie, M. S.; Peña-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. Alkylation damage in DNA and RNA -repair mechanisms and medical significance. *DNA Repair* **2004**, *3*, 1389−1407.

(6) Barrows, L. R.; Shank, R. C. Aberrant methylation of liver DNA in rats during hepatotoxicity. *Toxicol. Appl. Pharmacol.* **1981**, *60*, 334−345.

(7) Brookes, P.; Lawley, P. D. The alkylation of guanosine and guanylic acid. *J. Chem. Soc.* **1961**, 3923−3928.

(8) Kaina, B.; Christmann, M.; Naumann, S.; Roos, W. P. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst.)* **2007**, *6*, 1079−1099.

(9) Christmann, M.; Kaina, B. O^6-methylguanine-DNA methyl-transferase (MGMT): impact on cancer risk in response to tobacco smoke. *Mutat. Res.* **2012**, *736*, 64−74.

(10) Fahrer, J.; Kaina, B. O^6-methylguanine-DNA methyltransferase in the defense against N-nitroso compounds and colorectal cancer. *Carcinogenesis* **2013**, *34*, 2435−2442.

(11) Fu, D.; Calvo, J. A.; Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **2012**, *12*, 104−120.

(12) Loechler, E. L.; Green, C. L.; Essigmann, J. M. In vivo mutagenesis by O^6-methylguanine built into a unique site in a viral genome. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6271−6275.

(13) Swann, P. F. Why do O^6-alkylguanine and O^4-alkylthymine miscode? The relationship between the structure of DNA containing O^6-alkylguanine and O^4-alkylthymine and the mutagenic properties of these bases. *Mutat. Res.* **1990**, *233*, 81−94.

(14) Gaffney, B. L.; Marky, L. A.; Jones, R. A. Synthesis and characterization of a set of four dodecadeoxyribonucleoside undecaphosphates containing O^6-methylguanine opposite adenine, cytosine, guanine, and thymine. *Biochemistry* **1984**, *23*, 5686−5691.

(15) Gaffney, B. L.; Jones, R. A. Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O^6-methylguanine with reference both to Watson-Crick pairs and to mismatch pairs. *Biochemistry* **1989**, *28*, 5881−5889.

(16) Patel, D. J.; Shapiro, L.; Kozlowski, S. A.; Gaffney, B. L.; Jones, R. A. Structural studies of the O^6-MeG.T interaction in the d(C-G-T-G-A-A-T-T-C-O^6MeG-C-G) duplex. *Biochemistry* **1986**, *25*, 1036−1042.

(17) Warren, J. J.; Forsberg, L. J.; Beese, L. S. The structural basis for the mutagenicity of O^6-methyl-guanine lesions. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19701−19706.

(18) Eoff, R. L.; Irimia, A.; Egli, M.; Guengerich, F. P. *Sulfolobus solfataricus* DNA polymerase Dpo4 is partially inhibited by "wobble" pairing between O^6-methylguanine and cytosine, but accurate bypass is preferred. *J. Biol. Chem.* **2007**, *282*, 1456−1467.

(19) Pence, M. G.; Choi, J.-Y.; Egli, M.; Guengerich, F. P. Structural basis for proficient incorporation of dTTP opposite O^6-methylguanine by human DNA polymerase ι. *J. Biol. Chem.* **2010**, *285*, 40666−40672.

(20) Patra, A.; Zhang, Q.; Guengerich, F. P.; Egli, M. Mechanisms of insertion of dCTP and dTTP opposite the DNA lesion O^6-methyl-2'-deoxyguanosine by human DNA polymerase η. *J. Biol. Chem.* **2016**, *291*, 24304−24313.

(21) Betz, K.; Nilforoushan, A.; Wyss, L. A.; Diederichs, K.; Sturla, S. J.; Marx, A. Structural basis for the selective incorporation of an artificial nucleotide opposite a DNA adduct by a DNA polymerase. *Chem. Commun.* **2017**, *53*, 12704−12707.

(22) Gong, J.; Sturla, S. J. A Synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA. *J. Am. Chem. Soc.* **2007**, *129*, 4882−4883.

(23) Wyss, L. A.; Nilforoushan, A.; Eichenseher, F.; Suter, U.; Blatter, N.; Marx, A.; Sturla, S. J. Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *J. Am. Chem. Soc.* **2015**, *137*, 30−33.

(24) Räz, M. H.; Aloisi, C. M. N.; Gahlon, H. L.; Sturla, S. J. DNA adduct-directed synthetic nucleosides. *Acc. Chem. Res.* **2019**, *52*, 1391−1399.

(25) Gahlon, H. L.; Schweizer, W. B.; Sturla, S. J. Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc.* **2013**, *135*, 6384−6387.

(26) Gahlon, H. L.; Boby, M. L.; Sturla, S. J. O^6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem. Biol.* **2014**, *9*, 2807−2814.

(27) Räz, M. H.; Sturla, S. J.; Gahlon, H. L. Hydrogen bonding interactions at the DNA terminus promote extension from methylguanine lesions by human extender DNA polymerase ζ. *Biochemistry* **2018**, *57*, 5978−5988.

(28) Leonard, G. A.; Thomson, J.; Watson, W. P.; Brown, T. High-resolution structure of a mutagenic lesion in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9573−9576.

(29) Kowal, E. A.; Lad, R.; Pallan, P. S.; Muffly, E.; Wawrzak, Z.; Egli, M.; Sturla, S. J.; Stone, M. P. Recognition of O^6-benzyl-2'-deoxyguanosine by a perimidinone-derived synthetic nucleoside: an unique DNA interstrand stacking interaction. *Nucleic Acids Res.* **2013**, *41*, 7566−7576.

(30) Sriram, M.; van der Marel, G. A.; Roelen, H. L.; van Boom, J. H.; Wang, A. H. Conformation of B-DNA containing O^6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e^6G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342. *EMBO J.* **1992**, *11*, 225−232.

(31) Ginell, S. L.; Kuzmich, S.; Jones, R. A.; Berman, H. M. Crystal and molecular structure of a DNA duplex containing the carcinogenic lesion O^6-methylguanine. *Biochemistry* **1990**, *29*, 10461−10465.

(32) Gahlon, H. L.; Sturla, S. J. Hydrogen bonding or stacking interactions in differentiating duplex stability in oligonucleotides containing synthetic nucleoside probes for alkylated DNA. *Chem.—Eur. J.* **2013**, *19*, 11062−11067.

(33) Dahlmann, H. A.; Berger, F. D.; Kung, R. W.; Wyss, L. A.; Gubler, I.; McKeague, M.; Wetmore, S. D.; Sturla, S. J. Fluorescent nucleobase analogues with textended π surfaces stabilize DNA duplexes containing O^6-alkylguanine adducts. *Helv. Chim. Acta* **2018**, *101*, No. e1800066.

(34) Trantakis, I. A.; Nilforoushan, A.; Dahlmann, H. A.; Stäuble, C. K.; Sturla, S. J. In-gene quantification of O^6-methylguanine with elongated nucleoside analogues on gold nanoprobes. *J. Am. Chem. Soc.* **2016**, *138*, 8497−8504.

(35) Watts, J. K.; Gait, M. J. Synthesis of nucleic acids. In *Nucleic Acids in Chemistry and Biology*, 4th ed.; Blackburn, G. M., Egli, M., Gait, M. J., Watts, J. K., Eds.; Royal Society of Chemistry: Cambridge, UK, 2022; pp 279−323.

(36) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005**, *121*, 1005−1016.

(37) Pallan, P. S.; Egli, M. Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a nonspecific enzyme-dsDNA complex. *Cell Cycle* **2008**, *7*, 2562−2569.

(38) Egli, M.; Pallan, P. S. Generating crystallographic models of DNA dodecamers from structures of RNase H:DNA complexes. In *Nucleic Acid Crystallography: Methods and Protocols, Methods in Molecular Biology*; Ennifar, E., Ed.; Humana Press, Springer Science and Business Media: New York, 2015; Vol. *1320*, Chapter 8, pp 111−126.

(39) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **1997**, *276*, 307−326.

(40) Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760−763.

(41) Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, *30*, 1022−1025.

(42) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355−367.

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

(44) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1355−1363.

(45) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 352−367.

(46) 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**, *13*, 1605−1612.

(47) Kalnik, M. W.; Norman, D. G.; Zagorski, M. G.; Swann, P. F.; Patel, D. J. Conformational transitions in cytidine bulge-containing deoxytridecanucleotide duplexes: extra cytidine equilibrates between looped out (low temperature) and stacked (elevated temperature) conformations in solution. *Biochemistry* **1989**, *28*, 294−303.

(48) Zheng, G.; Lu, X.-J.; Olson, W. K. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W240−W246.

(49) Tereshko, V.; Minasov, G.; Egli, M. The Dickerson-Drew B-DNA dodecamer revisited - at atomic resolution. *J. Am. Chem. Soc.* **1999**, *121*, 470−471.

(50) Tereshko, V.; Minasov, G.; Egli, M. A "hydrat-ion" spine in a B-DNA minor groove. *J. Am. Chem. Soc.* **1999**, *121*, 3590−3595.

(51) Minasov, G.; Tereshko, V.; Egli, M. Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *J. Mol. Biol.* **1999**, *291*, 83−99.