

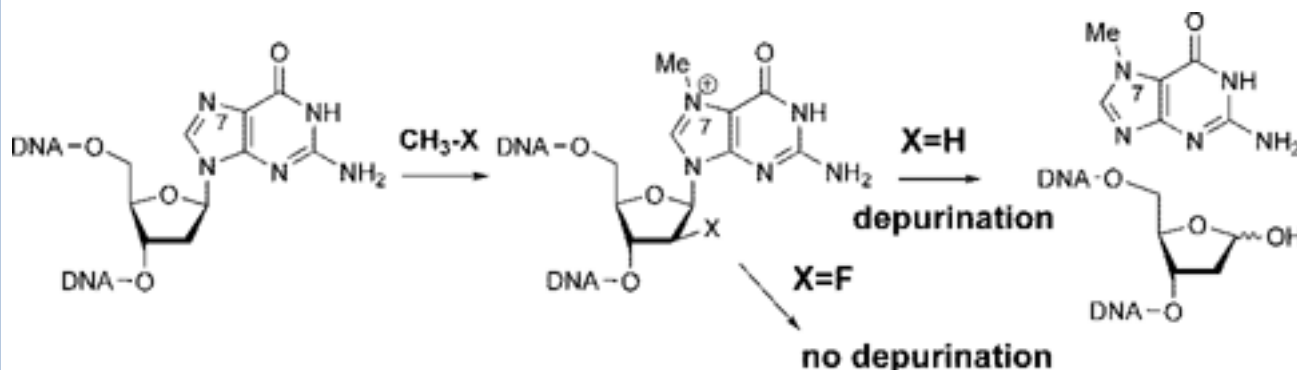
Published in final edited form as:

J Am Chem Soc. 2008 September 3; 130(35): 11570–11571. doi:10.1021/ja8025328.

Synthesis and Structure of Duplex DNA Containing the Genotoxic Nucleobase Lesion N7-Methylguanine

Seongmin Lee, Brian R. Bowman, Yoshihito Ueno, Shuyu Wang, and Gregory L. Verdine*
 Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Spontaneous methylation of DNA by endogenous methylating agents generates a variety of genotoxic adducts, the most prevalent of which is N7-methylguanine (m^7G).¹ Apart from its occurrence as a natural lesion, m^7G generated via deliberate DNA methylation has seen widespread use as a probe of protein–DNA interactions² and as a key component of the Maxam–Gilbert DNA sequencing method, enabled the first experimental determination of gene sequences. Most organisms express DNA glycosylase enzymes that locate m^7G residues and excise them (eq 1); the most well-studied of these are *E. coli* AlkA³ and human Aag,⁴ enzymes that are functionally equivalent but structurally dissimilar.



The molecular basis for recognition of m^7G residues amid a million-fold excess of normal G, and for the initiation of base excision repair by AlkA and Aag, is poorly understood, due in no small part to the lack of an efficient, scalable method by which to incorporate m^7G into DNA; the extreme instability of m^7G residues toward both the basic and acidic conditions employed in solid-phase DNA synthesis has heretofore precluded use of that method. Our laboratory previously developed a chemoenzymatic method for introducing m^7G into DNA⁵, but that method is not applicable to the scale and sequence requirements of crystallographic studies. Here we report site-specific incorporation of m^7G into DNA via solid-phase synthesis, made possible through a combination of 2'-fluorination on its 2'-deoxyribose moiety to provide acid stability, and mild nonaqueous deprotection to provide base stability. Using this synthetic method in combination with a recently developed crystallographic procedure for obtaining

E-mail: verdine@chem.harvard.edu.

Supporting Information Available: Synthetic procedures, ¹H and HRMS data for all new compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

structures of lesions in naked DNA, we report the first structures that reveal the nature of the m⁷G:C base pair in duplex DNA.

Fluorine substitution at the 2'-position in 2'-deoxy nucleosides is well-known to slow or even abrogate enzymatic catalysis of N-glycosidic bond cleavage, presumably via destabilization of the obligate oxocarbenium ion intermediate.⁶ Since both enzymatic and nonenzymatic depurination of m⁷G 2'-deoxyribose (m⁷dG) proceed through such an intermediate, we reasoned that 2'-fluorination of m⁷dG would most likely retard both processes, providing a useful means to incorporate the lesion into DNA and study its enzymatic repair. We chose to 2'-fluorinate m⁷dG in the arabino (α) configuration because this configuration was believed not to alter sugar puckering in DNA.

The synthesis of 2'-deoxy-2'-fluoro-N7-methylguanosine (Fm⁷dG) phosphoramidite **7** started with the fluorination of commercially available ribose derivative **1** with (diethylamino) sulfur trifluoride (DAST) followed by bromination to give the known glycosyl bromide **2** (Scheme 1).⁸ N-Glycosidation of the bromide **2** with the sodium salt of 2-amino-6-chloropurine gave mainly the desired β -anomer **3**, which was treated with NaOBn to install the 6-benzyloxy group with concomitant benzoate deprotection. Phe-noxyacetylation of **4** under conditions of transient TMS protection, followed by reductive removal of the benzyl moiety, afforded **5**. Regioselective alkylation of **5** with iodomethane provided 2'-deoxy-2'-fluoro-N7-methyl guanosine **6** in near quantitative yield. Trityl-ation of the 5'-OH followed by phosphitylation of the 3'-OH yielded the desired Fm⁷dG phosphoramidite **7**. Resin-bound Fm⁷dG-containing oligonucleotides **8** were released and deprotected by treatment with K₂CO₃ in MeOH and were purified by urea-PAGE gel electrophoresis. A 12-mer version of **8**, 5'-GACATGA(Fm⁷dG)-TGCC-3', was used for crystallographic studies, and a 25-mer, 5'-CGATAGCATCCT (Fm⁷dG)CCTTCTCTCCAT-3', for biochemical assays. The mass of the 12-mer and the 25-mer were verified by MALDI-TOF analysis (the 12-mer: calculated 3701, observed 3703; the 25-mer: calculated 7560, observed 7565), and nucleoside composition analysis on both oligonucleotides revealed the presence of the Fm⁷dG in the expected ratio relative to the other four bases (see Supporting Information).

We hybridized the Fm⁷dG-containing 25-mer with its complementary strand and investigated the ability of AlkA to catalyze glycosidic bond cleavage of this 2'-fluorinated DNA substrate analogue (see Supporting Information). The glycosylase reaction catalyzed by AlkA generates an abasic site, which is degraded by treatment with aqueous NaOH to effect site-specific DNA strand scission. The cleavage product migrates faster than full-length on a polyacrylamide gel. We tested the Fm⁷dG 25-mer alongside a well-validated AlkA substrate, a 25-mer containing hypoxanthine (HX) in place of Fm⁷dG; both contained cytosine opposite the Fm⁷dG or HX. While the HX duplex DNA was rapidly processed by AlkA, the Fm⁷dG duplex DNA was inert to AlkA, showing no sign of cleavage even after 18 h.

To obtain a structure of Fm⁷dG-containing DNA, we employed a host-guest complex crystallization (HGC) system⁹ in which four molecules of AlkA bind the ends of two duplex DNA molecules.¹⁰ AlkA prefers to crystallize in this end-bound mode even with DNA containing cognate lesions in the center of the duplex, as with the Fm⁷dG-containing 12-mer employed in our studies. AlkA shows no detectable energetic preference to bind DNA containing a cognate lesion versus nonlesion-containing DNA,^{3c} and this therefore provides bias toward crystallization as an end-bound complex, as opposed to a lesion-specific recognition complex. The end-bound crystal packing arrangement leaves the entire middle portion of the duplex nearly devoid of protein contacts, thereby enabling the facile crystallization and structure determination of "naked" lesion-containing DNA molecules, which are ordinarily very difficult to crystallize, especially in the physiologic B-form conformation.

We refined the structure of the AlkA-Fm⁷dG complex to 2.9 Å resolution and found that it had indeed crystallized as an HGC, with the protein anchoring the DNA ends via the interrogating residue Leu125^{3c} and making limited contacts to the DNA phosphate backbone via its helix–hairpin–helix domain (Figure 1B). The structure reveals, for the first time, the base pairing characteristics of m⁷G:C in DNA (Figure 1C and D).¹¹ The duplex DNA in the complex has a normal B-form conformation throughout, with an average rise of 3.3 Å per residue and no obvious deformations save for a gentle bend of 16°. The base pairing mode of m⁷G:C is similar to that of G:C, with similar degrees of propeller-twisting and base-stacking; the rmsd between the two pairs is 0.30 Å, with the main difference being that the distance between O6 (m⁷dG) and NH₂ (dC) in Fm⁷dG:C is slightly longer than that of G:C (3.4 Å vs ~3.0 Å). The sugar of the Fm⁷dG has an O4'-endo pucker, instead of the more common 2'-endo pucker; we have recently observed this alternate pucker in several other structures of 2'-β-fluorinated nucleosides in DNA,¹⁰ and thus we believe it is the preferred pucker for such analogues.

The absence of any DNA structural distortion induced by m⁷G, together with the observation that the lesion actually has a stabilizing influence on the double helix,⁵ underscores the challenge faced by AlkA and AAG in conducting an efficient search of the genome for this rare lesion. It is possible that these enzymes, like oxoguanine DNA glycosylase,¹² locate intrahelical m⁷G residues only upon performing a muscular interrogation of the DNA helix to expose the lesions hidden therein.

Acknowledgments

This work was supported by the NIH (CA100742). We thank the staff of beamlines 24-ID and 19-ID of the Advanced Photon Source at Argonne National Laboratory for expert assistance with X-ray data collection.

References

1. (a) Lindahl T. *Nature* 1993;362:709. [PubMed: 8469282] (b) Sedgwick B. *Nat Rev Mol Cell Biol* 2004;5:148. [PubMed: 15040447]
2. (a) Maxam AM, Gilbert W. *Methods Enzymol* 1980;65:499. [PubMed: 6246368] (b) Gilbert W, Siebenlist U. *Proc Natl Acad Sci USA* 1980;77:123. (c) Hayashibara KC, Verdine GL. *J Am Chem Soc* 1991;113:5104.
3. (a) Thomas I, Yang CH, Goldthwait DA. *Biochemistry* 1982;21:1162. [PubMed: 7041972] (b) Evensen G, Seeberg E. *Nature* 1982;296:773. [PubMed: 7040984] (c) Labahn J, Schärer OD, Long A, Ezaz-Nikpay K, Verdine GL, Ellenberger TE. *Cell* 1996;86:321. [PubMed: 8706136] (d) Hillis T, Ichikawa Y, Ellenberger T. *EMBOJ* 2000;19:758. (e) O'Brien PJ, Ellenberger TE. *J Biol Chem* 2004;279:26876. [PubMed: 15126496]
4. (a) Lau AY, Scharer OD, Samson L, Verdine GL, Ellenberger T. *Cell* 1998;95:249. [PubMed: 9790531] (b) Lau AY, Wyatt MD, Glassner BJ, Samson LD, Ellenberger T. *Proc Natl Acad Sci USA* 2000;97:13573. [PubMed: 11106395] (c) Saparbaev M, Laval J. *Proc Natl Acad Sci USA* 1994;91:5873. [PubMed: 8016081]
5. Ezaz-Nikpay K, Verdine GL. *J Am Chem Soc* 1992;114:6562.
6. (a) Marquez VE, Tseng CK-H, Mitsuya H, Aoki S, Kelley JA, Ford H Jr, Roth JS, Broder S, Johns DG, Friscoll JS. *J Med Chem* 1990;33:978. [PubMed: 2106581] (b) York JL. *J Org Chem* 1981;46:2171. (c) Schärer OD, Verdine GL. *J Am Chem Soc* 1995;117:10781. (d) Ikeda H, Fernandez R, Eilk A Jr, Marquez VE. *Nucleic Acids Res* 1998;26:2237. [PubMed: 9547286] (e) Sinnott ML. *Chem Rev* 1990;90:1171.
7. Berger I, Tereshko V, Ikeda H, Marquez VE, Egli M. *Nucleic Acids Res* 1998;26:2473. [PubMed: 9580702]
8. Howell HG, Brodfuehrer PR, Brundidge SP, Benigni DA, Sapino C Jr. *J Org Chem* 1998;53:85.
9. Goodwin KD, Lewis MA, Tanious FA, Tidwell RR, Wilson WD, Georgiadis MM, Long EC. *J Am Chem Soc* 2006;128:7846. [PubMed: 16771498]
10. Bowman BR, Lee S, Wang S, Verdine GL. *Structure*. 2008in press

11. The DNA duplex containing chains G and H (PDB ID: 3D4V) was chosen for graphical representation and detailed analysis, because it showed more complete electron density and substantially lower B-factors than the other DNA duplex (chains E and F) in the asymmetric unit.
12. Banerjee A, Yang W, Karplus M, Verdine GL. Nature 2005;434:612. [PubMed: 15800616]

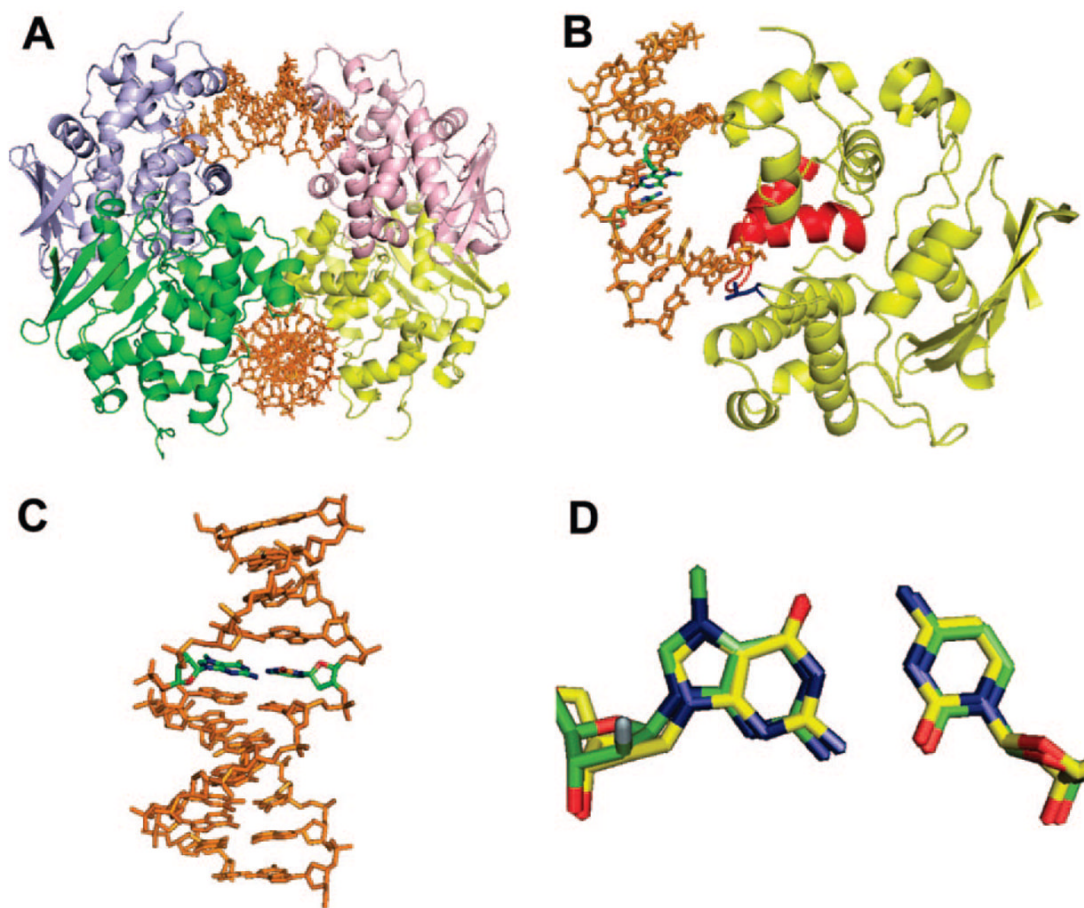
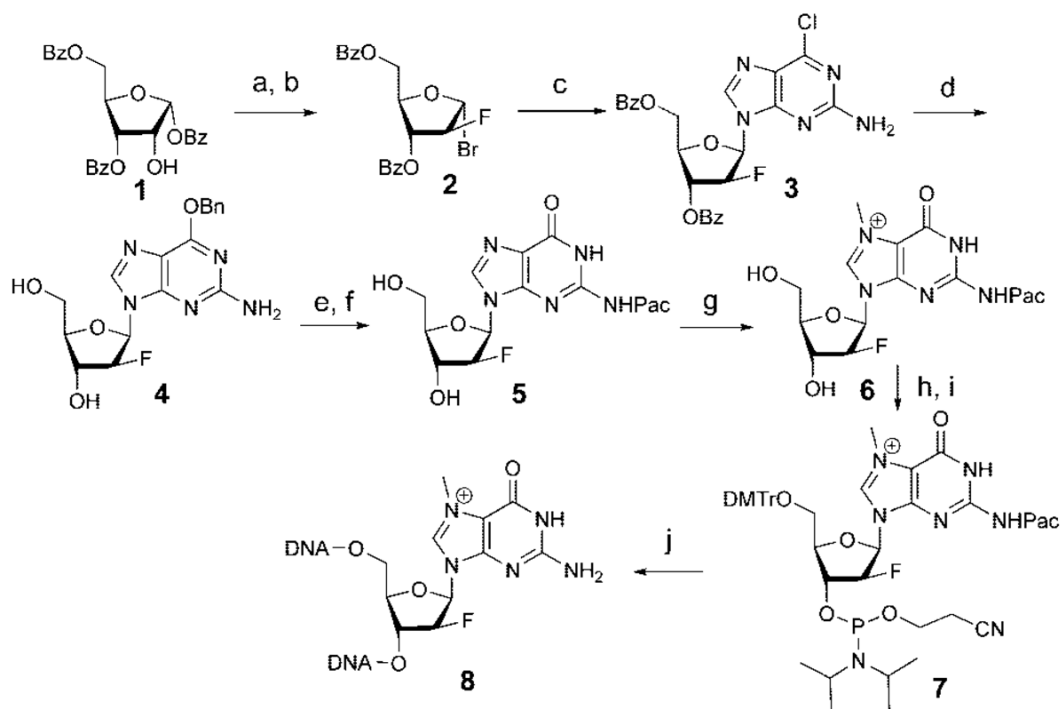


Figure 1.

(A) Ribbon representation of the AlkA-(Fm⁷dG:C DNA) complex. (B) View of an AlkA monomer interacting with the DNA. The side chain of Leu125 is in blue, the helix-hairpin-helix motif, in red, and the DNA, in gold except the Fm⁷dG:C base pair, which is in green/natural atom. (C) Side view of the DNA duplex, with the Fm⁷dG:C base pair highlighted as in B. (D) Superposition of Fm⁷dG:C with the G:C base pair. The Fm⁷dG:C is in green/natural atom with fluorine in sky blue; the G:C is in yellow/natural atom.

**Scheme 1.**Preparation of Fm⁷dG-Containing DNA^a

^a Reagents and conditions: (a) DAST, CH₂Cl₂, reflux, 15 h, 91%; (b) HBr, AcOH, 25 °C 24 h, 98%; (c) 2-amino-6-chloropurine, NaH, CH₃CN, 50 °C 6 h, 63%; (d) BnOH, NaH, DMF, 25 °C, 78%; (e) (i) TMSCl, pyridine; (ii) (Pac)₂O; (iii) TBAF 92%; (f) Pd/C, H₂, EtOH/THF, 25 °C, 98%; (g) MeI, DMF, 25 °C, 95%; (h) DMTrCl, pyridine, 68%; (i) 4,5-dicyanoimidazole, 2-cyanoethyltetraisopropyl phosphordiamidite, CH₂Cl₂, 25 °C, 1 h, 70%; (j) solid phase DNA synthesis.