Published online 28 December 2012 *Nucleic Acids Research, 2013, Vol. 41, No. 4* ***2689–2697***
doi:10.1093/nar/gks1316

Structure, stability and function of 5-chlorouracil modified A:U and G:U base pairs

Amritraj Patra[1], Joel Harp[1], Pradeep S. Pallan[1], Linlin Zhao[1], Mikhail Abramov[2], Piet Herdewijn[2,3] and Martin Egli[1,*]

[1]Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA, [2]Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium and [3]Institute of Systems and Synthetic Biology, University of Évry-Val-d'Essonne, 5 Rue Henri Desbruère, 91030 Évry Cedex, France

Received October 31, 2012; Revised November 16, 2012; Accepted November 19, 2012

ABSTRACT

The thymine analog 5-chlorouridine, first reported in the 1950s as anti-tumor agent, is known as an effective mutagen, clastogen and toxicant as well as an effective inducer of sister-chromatid exchange. Recently, the first microorganism with a chemically different genome was reported; the selected *Escherichia coli* strain relies on the four building blocks 5-chloro-2′-deoxyuridine (ClU), A, C and G instead of the standard T, A, C, G alphabet [Marlière,P., Patrouix,J., Döring,V., Herdewijn,P., Tricot,S., Cruveiller,S., Bouzon,M. and Mutzel,R. (2011) Chemical evolution of a bacterium's genome. *Angew. Chem. Int. Ed.*, 50, 7109–7114]. The residual fraction of T in the DNA of adapted bacteria was <2% and the switch from T to ClU was accompanied by a massive number of mutations, including >1500 A to G or G to A transitions in a culture. The former is most likely due to wobble base pairing between ClU and G, which may be more common for ClU than T. To identify potential changes in the geometries of base pairs and duplexes as a result of replacement of T by ClU, we determined four crystal structures of a B-form DNA dodecamer duplex containing ClU:A or ClU:G base pairs. The structures reveal nearly identical geometries of these pairs compared with T:A or T:G, respectively, and no consequences for stability and cleavage by an endonuclease (EcoRI). The lack of significant changes in the geometry of ClU:A and ClU:G base pairs relative to the corresponding native pairs is consistent with the sustained unlimited self-reproduction of *E. coli* strains with virtually complete T→ClU genome substitution.

INTRODUCTION

Uracil analogs with halogen substitution at the 5-position represent an important class of compounds with regard to their mutagenic activity (1). Such analogs were first synthesized in the 1950s as potential anti-tumor agents (2,3). The 5-fluorouracil (FU) analog is a well-known anti-cancer drug for treatment of human malignancies (4). 5-Chlorouracil and 5-bromouracil (ClU and BrU, respectively) are associated with inflammation and are considered to be carcinogenic (5). 5-Iodouracil (IU) was shown to have lethal and mutagenic effects on bacteriophage T4 (6).

Halogenated uracil residues are expected to exhibit base pairing properties in double-stranded nucleic acids that are closely related to those of thymine, thus involving complementary pairs with A or wobble pairs with G that are stabilized via Watson–Crick (W–C) hydrogen bonds. However, substitution at the 5-position of uracil can substantially alter the physical (electronic) and chemical properties of the nucleobase, as evidenced by changes in the UV spectra and the values of the pK_a (7). The effects of incorporation of FU into DNA on the structure and dynamics of the latter have been studied quite extensively, arguably as a result of the interesting pharmacological properties of this U analog as an anti-cancer agent (8–10). Interest in other halouracil analogs was primarily focused on their use as radiosensitizing agents in human cancers. BrU and IU were shown to be more effective in killing tumor cells using ionizing radiation (11,12). However, ClU, though an effective mutagen, clastogen and toxicant, as well as an effective inducer of sister-chromatid exchange, is not as sensitive to ionizing radiation as other thymine analogs (3). The relative lack of interest in ClU may explain the fact that a thorough investigation of the impact of the replacement of T with ClU on the base pairing geometry and duplex conformation using structural tools is presently lacking.

*To whom correspondence should be addressed. Tel: +1 615 343 8070; Fax: +1 615 322 7122; Email: martin.egli@vanderbilt.edu

Marlière *et al.* (13) recently evolved genomic DNA composed of the three canonical bases A, C and G and the artificial base ClU in an *Escherichia coli* strain lacking thymidylate synthase and requiring exogenous T. Selection over 25 weeks in a specially developed cultivation device yielded descendants that grew essentially with only ClU instead of T. The DNA of adapted bacteria contained 90% ClU and 10% T and this residual fraction was forced to <2% by disrupting the *trmA* gene for tRNA U54 methyltransferase.

The above pioneering study prompted us to analyze the base pair geometries of ClU with A and G in more detail. To date, no crystallographic study of a DNA duplex comprising ClU:A or ClU:G base pairs has been reported. A nuclear magnetic resonance (NMR) investigation by Theruvathu *et al.* (14) did not bring to light any substantial difference in the geometries of ClU:A and T:A base pairs. Also, the geometries of ClU:G and T:G wobble pairs exhibited similar geometries (15). In another study, the stacking patterns of a halogenated (F, Cl or Br) uridine overhang at the 3′-terminus of an octamer RNA:DNA hybrid duplex were analyzed in presence of rhodium or iridium hexamine salt (16). The crystal structures with ClU and BrU were similar in that the dangling ends were located atop the terminal base pair, whereas FU was ejected from the helical stack.

To compare the geometries of ClU:A and ClU:G base pairs with those of the corresponding T:A and T:G pairs, respectively, and to examine possible effects of these artificial pairs on the conformation of duplex DNA, we determined crystal structures of four Dickerson–Drew Dodecamer (DDD) B-form duplexes containing two ClU:A pairs ($[d(CGCGAA\{ClU\}TCGCG)]_2$, referred to as ClU7 here, and $[d(CGCGAAT\{ClU\}CGCG)]_2$; ClU8), four ClU:A pairs ($[d(CGCGAA\{ClUClU\}CGCG)]_2$; ClU7/8), or two ClU:G base pairs ($[d(CGCGAATT\{ClU\}GCG)]_2$, ClU9) in complex with *Bacillus halodurans* RNase H (*Bh*RNase H) at resolutions between 1.5 and 1.7 Å. Crystals grown for the duplexes alone were not of diffraction-quality and we therefore resorted to using RNase H as a scaffold. DNA duplexes act as inhibitors of RNase H (17) and the enzyme binds double-stranded DNA non-specifically and without perturbing the structure of the central T:A region of the duplex (18), thus not compromising our geometric analysis of ClU:A or ClU:G pairs (the latter being adjacent to the four central T:A pairs).

Our structures reveal nearly identical geometries of the ClU:A and ClU:G base pairs compared with T:A and T:G, respectively. In line with the structural similarities, UV melting experiments of DNA duplexes with ClU pairs uncovered only very minor consequences of the replacement of T by ClU opposite A or G relative to the parent duplexes with T:A or T:G pairs, respectively. These observations regarding structure and stability were mirrored at the level of function. Thus, the restriction endonuclease EcoRI did not display a preference in its ability to recognize and cleave the natural recognition sequence G|AATT C ('|' marks the cleavage site) in the native DDD compared with duplexes featuring the modified recognition sequences G|AA(ClU)TC or G|AAT(ClU)C.

MATERIALS AND METHODS

Protein expression and purification

Bacillus halodurans genomic DNA was purchased from American Type Culture Collection (ATCC, Manassas, VA, USA). The Asp132→Asn mutant of *Bh*RNase H (Met58 to Lys196) was expressed in *E. coli* and purified as described previously (18). The protein solution was concentrated to 25 mg/ml.

Synthesis of 5′-*O*-dimethoxytrityl-5-chloro-2′-deoxyuridine, 3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and incorporation of ClU into oligonucleotides

To a colorless solution of 5-chloro-2′-deoxyuridine (19) (600 mg, 2.28 mmol) in pyridine (20 ml), 4,4′-dimethoxytrityl chloride (930 mg, 2.75 mmol) was added in one portion at room temperature (RT). The reaction mixture was stirred for 12 h and turned to a yellow–orange color. After the starting material had disappeared, the reaction mixture was cooled in an ice bath, methanol (1 ml) was added and the reaction mixture was concentrated and co-evaporated twice with toluene. The residue was dissolved in dichloromethane, washed with H_2O, dried over Na_2SO_4 and purified by column chromatography on silica gel to yield 5′-*O*-dimethoxytrityl-5-chloro-2′-deoxyuridine (1.08 g, 84%). This compound (1.08 g, 1.91 mmol) was dissolved in dichloromethane (10 ml) and cooled in an ice bath. *N*,*N*-Diisopropylethylamine (1.5 ml, 8.76 mmol) and 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite (0.58 ml, 2.6 mmol) were added. The reaction solution was stirred for 30 min at RT. Upon completion, the reaction mixture was concentrated and co-evaporated twice with toluene. The crude material was purified by column chromatography on silica to yield 5′-*O*-dimethoxytrityl-5-chloro-2′-deoxyuridine, 3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (1.0 g, 68%). ^{31}P NMR ($CDCl_3$, 25°C): δ = 149.08, 148.7. High-resolution mass spectrometry (HRMS) calculated for $C_{39}H_{46}ClN_4O_8P$, [MH+] 765.2794, found 765.2820.

All four ClU-modified, high-pressure liquid chromatography-purified DDD oligonucleotides were purchased from Trilink (San Diego, CA, USA). The DNAs were annealed and mixed with the protein at 1.2:1 molar ratio in the presence of 5 mM $MgCl_2$.

Crystallization and structure determination

Crystallization experiments were performed by the sitting drop vapor diffusion technique at 4°C using a sparse matrix screen (Hampton Research, Aliso Viejo, CA, USA) (20). A quantity of 1 µl complex solution was mixed with 1 µl of reservoir solution and equilibrated against 40 µl reservoir wells. Crystals appeared in droplets containing 0.2 M magnesium acetate, 0.1 M sodium cacodylate (pH 6.5) and 20% (w/v) PEG 8000 within 2–3 days. Crystals were mounted in nylon loops, cryo-protected in reservoir solution containing 20% glycerol and frozen in liquid nitrogen. Diffraction data were collected using either a Mar225 or Mar300 CCD detector on the 21-ID F/D beam lines of the Life

Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL, USA). Data were integrated and scaled with the program HKL2000 (21). The structures were determined by the Molecular Replacement technique using the program MOLREP (22,23) and the *Bh*RNase H structure with PDB ID code 3D0P (protein alone) as the search model. Initial refinement was carried out with the program REFMAC (24) and DNA duplexes were then gradually built into the electron density, 2–3 bp at a time, followed by further refinement. Manual rebuilding was performed with the program COOT (25). Water molecules and metal ions were added gradually and isotropic/Translation Libration Screw-motion (TLS) refinement was continued with the program PHENIX (26). A summary of crystallographic parameters is provided in Table 1. Helical parameters were calculated with the program CURVES (27). Illustrations were generated with the program UCSF Chimera (28).

UV thermal melting

Melting of each oligonucleotide (~7.5 μM duplex concentration) was performed in 10 mM sodium phosphate, pH 7.4 and 0.1 mM EDTA in the presence of 150 mM NaCl. Absorbance versus temperature profiles were measured with a Varian Cary 300 spectrophotometer at 260 nm wavelength and 1-cm path length at heating or cooling rates of 0.5°C/min. Melting temperatures were determined with Varian Cary UV T_m Analysis software and are averages of the maxima of the first derivative of the 95-point smoothed curves from heating and cooling experiments. The final T_m values are based on six independent measurements. Hyperchromicities were determined by calculating the difference of absorbance between high- and low-temperature baselines and dividing by the absorbance of the low-temperature baseline.

Cleavage assays

Cleavage assays with EcoRI were conducted following a previous protocol (29) with modifications. Briefly, 5 μM ^{32}P-labeled DNA was incubated at 37°C with EcoRI buffer and 20 U of EcoRI (high fidelity; New England Biolabs, Ipswich, MA, USA) in a volume of 10 μl, with final concentrations of 50 mM potassium acetate, 20 mM Tris–acetate, pH 7.5, 10 mM magnesium acetate and 1 mM dithiothreitol. Reactions were started by adding EcoRI and were stopped by removing an aliquot of 1 μl reaction mixture at various times and adding it to 9 μl of 20 mM EDTA (pH 9.0) in 95% (v/v) formamide, followed by 20% PAGE analysis.

Coordinates

Final coordinates and structure factors for all four dodecamers have been deposited in the Protein Data Bank (http://www.rcsb.org.) The PDB ID codes are 4HUF (ClU7), 4HTU (ClU8), 4HUG (ClU7/8) and 4HUE (ClU9).

RESULTS

Overall structures of the ClU-modified DDD:RNase H complexes

To gain a better understanding of the potential structural change due to the replacement of the thymine 5-methyl group by chlorine, we selected the DDD B-form DNA as a template and synthesized four modified DDDs with ClU either located opposite A (ClU7, ClU8, ClU7/8) or opposite G (ClU9). We used crystals of these modified DDDs in complex with *Bh*RNase H to study the geometries of ClU:A and ClU:G pairs as the crystals of the oligonucleotides alone were not of sufficient quality for a detailed structural analysis. In the crystal of the

Table 1. Crystal data, data collection parameters and structure refinement statistics

Structure/duplex	ClU8	ClU9	ClU7	ClU7/8
Space group		$P2_12_12_1$		
Cell dimensions				
a (Å)	64.08	64.36	64.19	63.83
b (Å)	64.76	64.81	64.75	64.64
c (Å)	116.47	116.62	116.47	116.29
Data collection				
Wave length (Å)		0.97872		
Resolution (Å)	35.88–1.49	33.28–1.56	33.22–1.69	43.23–1.64
Outer shell (Å)	1.52–1.49	1.59–1.56	1.72–1.69	1.67–1.64
Unique reflections (outer shell)	77 514 (3752)	68 234 (3339)	54 372 (2620)	59 446 (2888)
Completeness (outer shell) (%)	96.9 (95.1)	97.3 (96.6)	98.3 (97.3)	99.7 (99.3)
R-merge (outer shell)	0.071 (0.824)	0.082 (0.800)	0.061 (0.675)	0.072 (0.674)
I/σ(I) (outer shell)	27.3 (2.5)	19.8 (2.1)	29.8 (2.9)	26.7 (3.0)
Refinement				
Working set reflections	74 662	68 154	54 301	59 370
Test set reflections	3759	3453	2766	2996
R-work/R-free	0.196/0.223	0.197/0.226	0.202/0.237	0.197/0.220
No. of protein/DNA atoms	2244/896	2203/897	2192/833	2215/833
No. H_2O/Mg^{2+}/ligands	476/2/4	523/2/4	375/2/4	341/2/4
Average B-factors	30.18	33.43	27.80	27.96
RMSD bonds (Å)	0.009	0.009	0.008	0.008
RMSD angles (°)	1.4	1.3	1.3	1.4

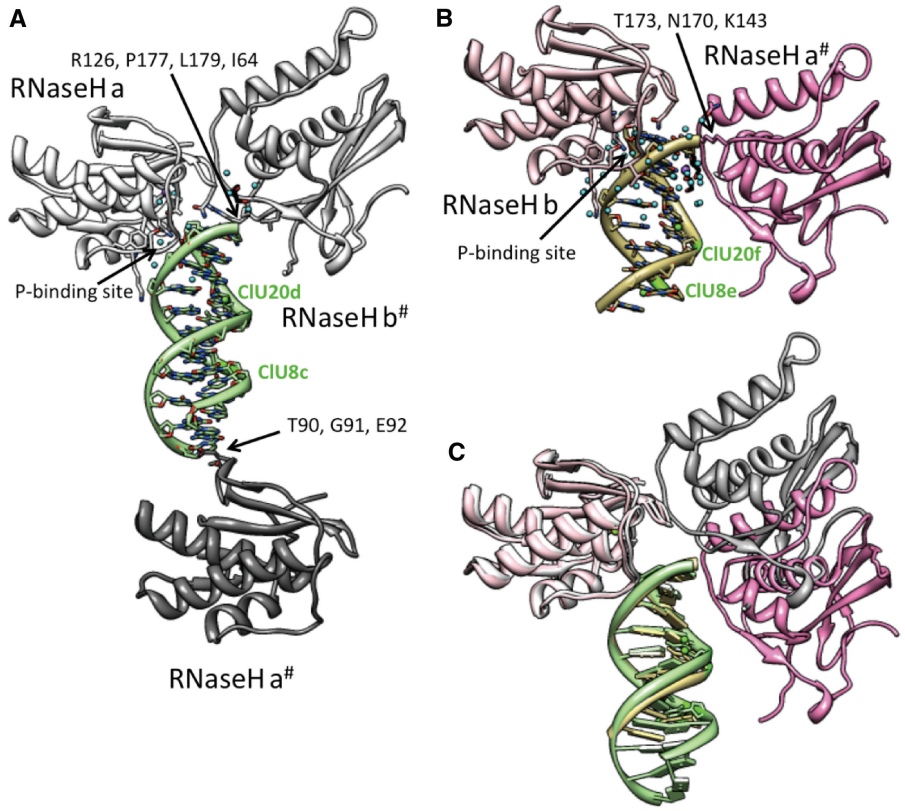

Figure 1. Overall structure of the *Bh*RNase H:ClU8-DDD (ClU8) complex. The asymmetric unit contains two independent complexes, both consisting of a single RNase H molecule bound to a modified DDD duplex. Protein chains in all four structures are labeled a (complex 1) and b (complex 2) and duplex strands are labeled c and d (complex 1) and e and f (complex 2), with nucleotides numbered 1–12 (c/e) and 13–20 (d/f). (**A**) The duplex in the first complex (colored green) is fully resolved in the electron density map and is contacted by three RNase H molecules (colored in white, gray or dark gray, whereby symmetry mates are marked with hash symbol). (**B**) The duplex in the second complex (colored in beige) is contacted by two RNase H molecules (colored in pink or magenta) and four C:G pairs at one end of the duplex are not visible in the electron density map. (**C**) The superimposition of the two independent complexes illustrates that the ClU8-DDDs exhibit similar conformations and virtually identical orientations of the original RNase H molecule relative to the DNA duplex. ClU residues are labeled and chlorine atoms are highlighted as light green spheres, selected water molecules are shown as cyan spheres, RNase H amino acids interacting with the DNA are labeled and the so-called phosphate (P-) binding site is marked by an arrow.

native DDD bound to RNaseH studied earlier (18), protein–DNA contacts in the complex were limited to the backbone of the C:G tracts (underlined) at both ends: CGCGAATTCGCG. Thus, we expected base pairing in the central GAATTC region where we replaced T opposite either A or G by ClU not to be affected by binding of RNase H (Figure 1).

Crystals diffracted to resolutions between 1.5 and 1.7 Å and belong to space group $P2_12_12_1$ (Table 1). The structures were solved by the molecular replacement technique using the protein portion of the complex between RNase H and the native DDD (18) as the search model. All four crystals of the complexes with ClU-modified DDDs are isomorphous (Table 1). However, unlike in the structure of the complex between RNase H and the native DDD where the duplex sits on a dyad and the asymmetric unit contained two protein molecules and two DNA single strands, crystals of the complexes with ClU-modified DDDs feature two independent complexes consisting of RNase H bound to a duplex (Figure 1A and B). The first of these displays clear electron density around all nucleotides, whereas in the second, the density around the DNA is only partially resolved and either three or four base pairs at one end of the duplex are missing in the individual complex crystals (Figure 1B and C). Similarly, two (ClU7 crystal structure) and three N-terminal amino acids (ClU7/8 crystal structure) in the RNase H molecules from the second complex could not be resolved in the electron density maps. Examples of the quality of the final electron density are shown in Figure 2 and Supplementary Figures S1–S3.

Inspection of the lattice interactions in the structures of the complexes reveals the origin of the diminished order of four base pairs in the second duplex. Three RNase H molecules interact with the duplex from the first complex, whereby two proteins cradle one end and a loop entailing residues T90, G91 and E92 of a third is stacked against the terminal base pair at the other end (Figure 1A). In contrast, the second duplex only exhibits interactions with RNase molecules at one end, with the other not being stabilized by protein contacts and jutting out into the solvent (Figure 1B). The comparison between the orientations of RNase H molecules around the two independent duplexes reveals similar positions of the

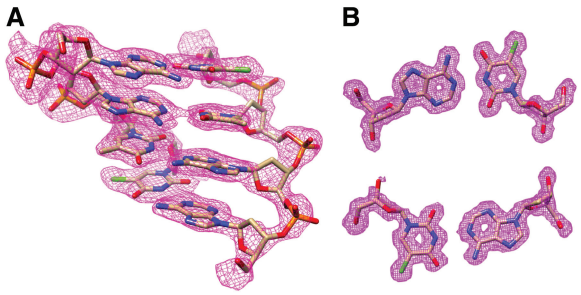

Figure 2. Quality of the final electron density. Fourier ($2F_o$–F_c) sum electron density drawn at the ~1.0 σ threshold (**A**) around the central ApApTpClU tetramer in the ClU8 duplex, and (**B**) around base pair A5:ClU20 (top) and base pair ClU8:A17 (bottom). Atoms are colored beige, red, blue, orange and light green for carbon, oxygen, nitrogen, phosphorus and chlorine, respectively.

protein that binds a phosphate group at its P-binding site but a considerable shift between the second protein molecules contacting the same end of the two duplexes (Figure 1C).

Protein–DNA interactions in the ClU-DDD:RNase H complex structures

The two complexes per asymmetric unit in the four crystal structures exhibit similar interactions between RNase H and the ClU-modified DDD (Figure 1A, white RNase H a and green duplex; Figure 1B, pink RNase H b and beige duplex). In both cases, the phosphate group of C3 is lodged at the phosphate-binding pocket (Figure 3) that harbors a phosphate of a DNA nucleotide separated by 2 bp from the scissile RNA phosphate in the structure of an RNA:DNA hybrid bound to *Bh*RNase H (30). In the crystal structure of the complex between RNase H and the native DDD, the phosphate of the terminal G12 sits at the active site and thus mimics a phosphate from the RNA strand (18). The phosphate of residue G4 from the paired strand is anchored at the phosphate-binding pocket. Compared with these interaction modes between RNA:DNA hybrid and native DDD and RNase H, ClU-modified duplexes thus exhibit a less intimate interaction with the protein as only phosphates from one strand are contacted by amino acids (P3 by T104, S147 and T148 and P4 by W139; Figure 3). As a consequence, only one Mg^{2+} ion (Mg_B) is observed at the protein active site, whereby the ion coordination sphere comprises D71, E109 and four water molecules (Figure 3). In addition to the protein–DNA contacts involving phosphate groups, four residues from a symmetry-related RNase H molecule interact with the terminal base pair of the first duplex. R126 and P177 stack onto the nucleobases of G24 and C1, respectively, and I64 and L179 form hydrophobic interactions with the base and sugar moieties, respectively, of nucleotide C1 (Figures 1A and 3). In the case of the second complex where the symmetry-related RNase H molecule adopts a somewhat different orientation (Figure 1A, gray RNase H b# versus Figure 1B, magenta RNase H a#), amino acids establishing contacts with the terminal base pair include K143, N170 and T173 (Figure 1B). However, neither duplex 1 nor duplex 2 in the four crystal structures exhibits interactions between RNase H and A or ClU/T nucleotides from the central tetramer.

Geometry of ClU:A and ClU:G base pairs

A central question that we wanted to address with our structural studies is whether there are any changes in the pairing modes and/or geometries of ClU:A and ClU:G base pairs relative to the native T:A and T:G pairs, respectively. Analysis of the duplexes containing either two or four ClU:A pairs reveals that they are of the standard W–C type with formation of two hydrogen bonds. Comparison between the geometries of ClU:A and T:A pairs by superimposing the ClU7 and ClU8 duplexes (Figure 4A) or the ClU7/8 and native DDD (31) duplexes (Figure 4C) demonstrates that replacing the methyl group of T by chlorine is of little consequence. Except for the methyl group carbon and chlorine, whose positions deviate somewhat as a result of the longer C-Cl bond relative to C-CH_3, the ClU:A and T:A pairs neatly overlap. Similarly, like T:G the ClU:G pair adopts the familiar wobble geometry, with G and ClU being shifted into the minor and major grooves, respectively, under formation of two hydrogen bonds (Figure 4B). As with the ClU:A and T:A pairs, the superimposition of the ClU9 duplex and a DDD featuring T:G pairs (32) illustrates the nearly identical geometries of the two pairs. Moreover, calculated geometric parameters for these duplexes (27), including rise, twist, x- and y-displacement, etc. (see the Supplementary Material) confirm that chlorine attached at the 5-position of thymine in place of a methyl group does not trigger a substantial difference in either pairing behavior or base pair geometry.

The similarities observed at the level of individual base pairs extend to the overall conformations of the dodecamer duplexes. As mentioned above, RNase H molecules in the crystal structures of the complexes with ClU-modified DDDs contact the duplexes exclusively in the CG portions, either at both ends (duplex 1, Figure 1A) or at only one end (duplex 2, Figure 1B). As a result, a hallmark of the DDD, the narrow A-tract minor groove remains largely unaffected by the interactions with the protein (Figure 5). Introduction of T:G mismatches at the border of the A-tract results in a wider minor groove in the ClU9 and G:T DDDs at the location of the mismatch pairs. However, replacement of T by ClU has no obvious effect on the change in groove width. The depth of the minor groove along the entire dodecamer exhibits much less variation compared with the width and neither G:T mismatch pairs nor ClU:A/G pairs cause any significant changes relative to the native DDD (Figure 5).

Thermodynamic stability of ClU-modified duplexes

To examine whether replacement of T by ClU affects duplex stability, we carried out UV melting studies with the native DDD, a DDD with T:G mismatches and all four ClU-modified DDDs (Table 2). The melting temperatures for duplexes with two (ClU7, ClU8) or four ClU:A pairs (ClU7/8) vary only slightly and are similar to the T_m of the native DDD. Insertion of T:G mismatches is

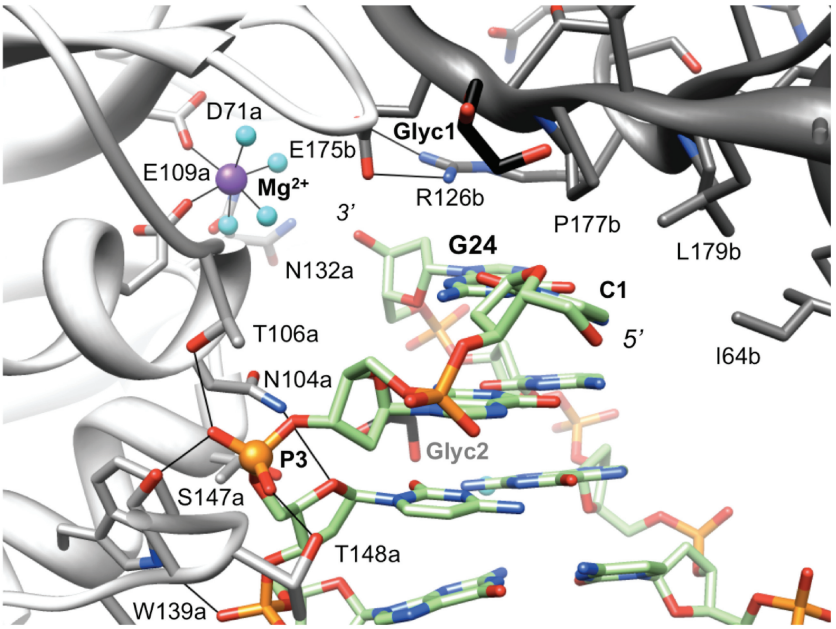

Figure 3. Interactions between RNase H and ClU modified DDDs. The original RNase H molecule (white ribbon) binds phosphates from one DNA strand; the phosphate of C3 (P3) is lodged at the phosphate-binding pocket and W139 forms a hydrogen bond to the phosphate of G4. The active site that normally harbors the scissile phosphate from the RNA strand (18) remains unoccupied and only one of the two Mg^{2+} ions (purple sphere) is present, with some water molecules (cyan spheres) taking the place of phosphate oxygens. Amino acid side chains from a symmetry-related RNase H molecule (I64, L179, P177 and R126) interact with the terminal C1:G24 base pair. Hydrogen bonds and Mg^{2+} coordination sphere are indicated with thin solid lines and two glycerol molecules trapped at the protein–DNA interface are highlighted with carbon atoms colored in black.

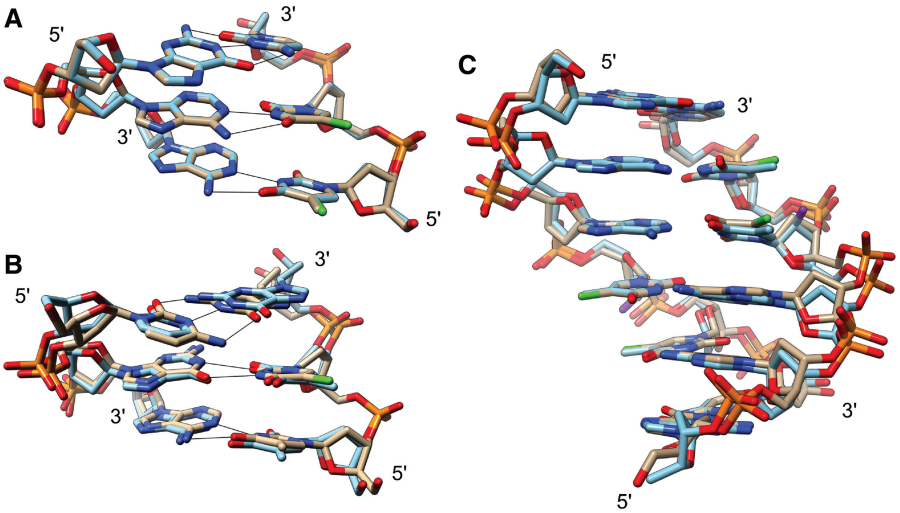

Figure 4. Comparison between the geometries of ClU:A and ClU:G and those of T:A and T:G base pairs, respectively. (**A**) Superimposition of the d(ClU7pT8pC9):d(G16pA17pA18) (light blue carbon atoms) and d(T7pClU8pC9):d(G16pA17pA18) (gray carbon atoms) trimer portions from the ClU7 and ClU8 structures, respectively. (**B**) Superimposition of the d(T8pClU9pG10):d(C15pG16pA17) (gray carbon atoms) and d(T7pT8pC9):d(G16pA17pA18) (light blue carbon atoms) trimer portions from ClU9 and a G:T mismatch-containing DDD structure [PDB ID 113D (32)], respectively. (**C**) Superimposition of the central hexamers from the ClU7/8 structure (gray carbon atoms) and the native DDD [PDB ID 436D (31)] (light blue carbon atoms). Chlorine atoms are highlighted in green, hydrogen bonds are indicated with thin solid lines (omitted in panel C for clarity), and fluorine atoms of two 2′-deoxy-2′-fluoroarabino-Ts in the reference structure are highlighted in purple (panel C).

accompanied by a steep drop in the stability that is similar for the duplex featuring two ClU:G pairs (ClU9). Overall, these data argue against an either stabilizing or destabilizing effect of the ClU nucleotide analog in B-form DNA.

EcoRI cleavage assays with ClU-modified DDDs

The EcoRI endonuclease recognizes the self-complementary hexamer 5′-G|AATTC-3′:3′-CTTAA|G-5′ and cleaves between G and A (|) under formation of

sticky ends with 5′-overhangs. We carried out cleavage assays with the restriction enzyme and ClU-modified DDDs, as well as the native DDD and a DDD with T:G mismatch pairs as reference duplexes (Figure 6). As expected, T:G and ClU:G mismatches abrogated cleavage. Conversely, the native DDD and all three modified duplexes with ClU:A pairs were cleaved by EcoRI without obvious effects of the ClU modification in terms of the time course of cleavage.

DISCUSSION

Our investigation of the structure, stability and function of DNA duplexes with ClU in place of T was motivated by the recently demonstrated genome-wide transliteration of T with ClU in *E. coli* by combining tight metabolic selection and long-term automated cultivation of bacterial populations (13). Among nucleobases, only thymine is unique to DNA and the fact that its metabolism is separated from RNA biosynthesis provides an opportunity to replace it *in vivo* by starvation and exogenous introduction of unnatural alternatives (33). Incorporation of 5-halogenopyrimidines into DNA was established decades ago (34). Among the analogs with fluoro-, bromo-, chloro- or iodo-substituents, ClU exhibits the closest likeness to T (14), is readily converted to the nucleoside triphosphate in the cell (35) and lacks the photolabile behavior of BrU and IU (36). Starting from an *E. coli* strain lacking thymidylate synthase and furnishing exogenous ClU instead of T, 25 weeks of selection in a cultivation device yielded a strain with A, G, C and 90% ClU and 10% T in its genome (13). Through additional disruption of the tRNA U54 methyltransferase gene, the ClU content was further reduced to <2%. This T→ClU transliteration was accompanied by more than 1500 A to G and G to A transitions in a particular culture, whereby the former was about twice as common. The frequency of these transitions suggests that ClU is prone to mispairing with G, although it is not clear whether ClU:G resembles the T:G pair in the wobble configuration with two hydrogen bonds or whether the mismatch pair exhibits a different hydrogen bonding pattern as a consequence of subtle changes in pK_a, dipole moment and/or hydration of the chlorouracil base compared with thymine (15).

We selected the DDD B-form DNA as a template to analyze the base pairing behavior of ClU opposite either A or G by X-ray crystallography. Because crystals of the DNAs alone did not diffract to high resolution, we decided to determine the structures of their complexes with *Bh*RNase H. We previously found that crystals of complexes between the endonuclease and the native DDD or chemically modified DDDs diffract X-rays to resolutions of around 1.5 Å (18,29,37). Indeed, all crystals of complexes with ClU-modified DDDs diffracted to better than 1.7 Å resolution (Table 1). Two of the complexes feature duplexes with two T:A pairs replaced by ClU:A (ClU7 and ClU8) and a third contains a duplex with all four T:A pairs in the central A-tract of the DDD replaced by ClU:A (ClU7/8). A fourth complex is between *Bh*RNase H and a DDD in which two ClU:G pairs

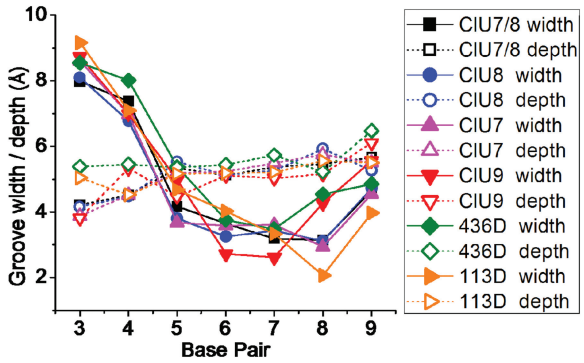

Figure 5. Minor groove widths (solid lines) and depths (dashed lines) in ClU-modified DDD duplexes in complex with RNase H compared with native DDD d(CGCGAATTCGCG) [PDB ID 436D; (31)] and a DDD d(CGCGAATTTGCG) with T:G mismatch pairs [underlined, PDB ID 113D; (32)]. All parameters were calculated with the program Curves (27).

Table 2. UV-melting temperatures of the native DDD, ClU-modified DDDs and a DDD with T:G mismatches

Duplex	DDD	ClU7	ClU8	ClU7/8	ClU9	T:G DDD
T_m (°C)	62.6 ± 0.9	61.8 ± 1.2	64.0 ± 0.5	63.8 ± 0.6	35.4 ± 0.7	36.5 ± 0.5
Hyperchromicity (%)	15.5	14.3	15.7	13.5	15.6	16.3

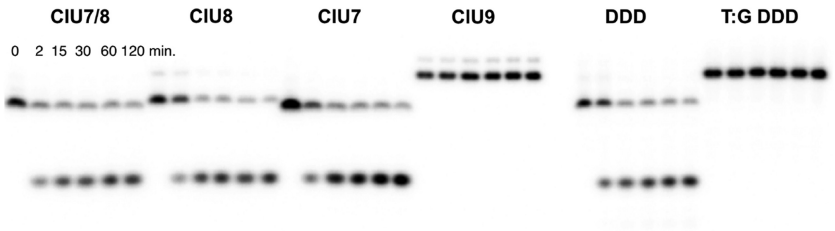

Figure 6. PAGE assay of EcoRI cleavage experiments with ClU-modified DDDs, the native DDD d(CGCGAATTCGCG) and the DDD d(CGCGAATTTGCG) with T:G mismatch pairs (underlined).

bracket the central T:A tract. The first three structures illustrate that ClU:A and T:A pairs adopt nearly identical configurations with two hydrogen bonds (Figures 2, 4A and 4C), thus confirming the earlier observations based on the structure of a B-form DNA duplex with ClU:A pairs analyzed by solution NMR (14). Similarly, the geometry of ClU:G mismatch pairs in the structure of the fourth complex closely resembles that of the T:G pair in the structure of the B-form duplex of the same sequence (32) (Figure 4B). Crystals of the RNase H complexes were grown at near neutral pH (6.5) and the wobble configuration with two hydrogen bonds confirms previous reports of the similarity of the ClU:G and T:G pairs as analyzed by NMR in solution (15).

The similarities at the level of conformation indicate that chlorine appears to closely mimic the methyl substituent of thymine. Indeed, when both bond length and van der Waals radius are taken into account, the two substituents exhibit very similar sterics. The length of the C(5)-Cl bond is clearly longer than that of the C(5)-CH_3 bond (1.73 Å versus 1.5 Å, respectively), but the van der Waals radius of a methyl group (2 Å) exceeds that of chlorine (1.75 Å) by roughly the same amount. Thus, the two substituents will be nearly indistinguishable for a protein probing the major groove of a B-form duplex. Indeed, we demonstrate here that EcoRI cuts the ClU7-, ClU8- and ClU7/8 DDDs with two or more Ts in the recognition site replaced by ClU (Figure 6). Conversely, replacing the C:G pairs in the target sequence with either T:G or ClU:G mismatches abolishes cleavage by EcoRI. Chlorine not only mimics the native methyl substituent of thymine, thus allowing ClU:A pairs to evade recognition by enzymes, but replacement of T opposite A by ClU does not lead to any obvious changes in the thermodynamic stability of B-form DNA (Table 2). Also, T:G and ClU:G mismatch pairs result in very similar losses of stability. The stability data support the notion that chlorine and methyl at the 5-position of uracil are basically interchangeable. Although chlorine can, in principle, participate in halogen bonds (38), the observed distances between chlorine and phosphate oxygens in the ClU-modified DDD structures are all clearly above the sum of the van der Waals radii for Cl and O (3.27 Å): minimum 5.09 Å, maximum 6.23 Å and average 5.49 Å. Chlorine atoms are also poorly hydrated and the shortest contact between a chlorine and water in our structures is 3.71 Å and thus not indicative of a hydrogen bond. Similar melting temperatures established for the native DDD and the ClU-modified DDDs are also inconsistent with differential polarizations of the thymine and chlorouracil nucleobases as this should affect stacking and thus stability.

The similarity between ClU and T in terms of pairing properties, conformation and duplex stability helps rationalize the successful outcome of the *in vivo* evolution of *E. coli* strains that rely on a ClU, A, C, G alphabet instead of T, A, C, G (13). It is unlikely that a similar success could be achieved with the FU or BrU analogs that exhibit more significant steric and stereoelectronic deviations from T compared with ClU. Although we found in one instance (EcoRI) that an enzyme appeared to ignore the switch from a methyl group to chlorine in the major groove, T → ClU base substitution changes groove geometry and electrostatics in a subtle fashion and may affect nucleic acid recognition by proteins involved in replication and transcription. Indeed, *in vitro* kinetic assays of DNA polymerization using human polymerase β and *E. coli* DNA polymerase I exo$^-$ Klenow fragment with templates containing either ClU or T showed facilitated incorporation of dGTP opposite ClU compared with T, particularly at increased values of pH (i.e. 9.0) (39). A high mutation rate is important for the evolutionary process and the massive number of A → G mutations seen in the adaptation of *E. coli* to ClU in place of T is consistent with facile formation of the ClU:G wobble pair.

ACCESSION NUMBERS

4HUF (ClU7), 4HTU (ClU8), 4HUG (ClU7/8), and 4HUE (ClU9).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–6 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

P.H. is indebted to KU Leuven, FWO and ERC for financial support. We are grateful to Dr. Z. Wawrzak, Northwestern University, for assistance with X-ray diffraction data collection.

FUNDING

Funding for open access charge: US National Institutes of Health [R01 GM55237 to M.E.].

Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
2. Heidelberger,C., Chaudhuri,N.K., Danneberg,P., Mooren,D. and Griesback,L. (1957) Fluorinated pyrimidines, a new class of tumor-inhibitory compounds. *Nature*, **179**, 663–666.
3. Morris,S.M. (1993) The genetic toxicology of 5-fluoropyrimidines and 5-chlorouracil. *Mutat. Res.*, **297**, 39–51.
4. Longley,D.B., Harkin,D.P. and Johnston,P.G. (2003) 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nat. Rev. Cancer*, **3**, 330–338.
5. Henderson,J.P., Byun,J., Takeshita,J. and Heinecke,J.W. (2003) Phagocytes produce 5-chlorouracil and 5-bromouracil, two mutagenic products of myeloperoxidase, in human inflammatory tissue. *J. Biol. Chem.*, **278**, 23522–23528.
6. Byrd,D.M. and Prusoff,W.H. (1975) Multiplicity reactivation of 5 iodouracil substituted, nonviable bacteriophage T4td8. *Antimicrob. Agents Chemother.*, **8**, 558–563.
7. Jang,Y.H., Sowers,L.C., Cagin,T. and Goddard,W.A. III (2001) First principles calculation of p*K*a values for 5-substituted uracils. *J. Phys. Chem. A*, **105**, 274–280.

8. Coll,M., Saal,D., Frederick,C.A., Aymami,J., Rich,A. and Wang,A.H.-J. (1989) Effects of 5-fluorouracil/guanine wobble base pairs in Z-DNA: molecular and crystal structure of d(CGCGFG). *Nucleic Acids Res.*, **17**, 911–923.
9. Lukin,M. and de los Santos,C. (2006) NMR structures of damaged DNA. *Chem. Rev.*, **106**, 607–686.
10. Parker,J.B. and Stivers,J.T. (2011) Dynamics of uracil and 5-fluorouracil in DNA. *Biochemistry*, **50**, 612–617.
11. Kinsella,T.J. (1996) An approach to the radiosensitization of human tumors. *Cancer J. Sci. Am.*, **2**, 184–193.
12. Saif,M.F., Cheng,T.-C., Berk,G. and Kinsella,T.J. (2007) IPdR: A novel oral#radiosensitizer. *Expert. Opin. Investig. Drugs*, **16**, 1415–1424.
13. Marlière,P., Patrouix,J., Döring,V., Herdewijn,P., Tricot,S., Cruveiller,S., Bouzon,M. and Mutzel,R. (2011) Chemical evolution of a bacterium's genome. *Angew. Chem. Int. Ed.*, **50**, 7109–7114.
14. Theruvathu,J.A., Kim,C.H., Rogstad,D.K., Neidigh,J.W. and Sowers,L.C. (2009) Base-pairing configuration and stability of an oligonucleotide duplex containing a 5-chlorouracil-adenine base pair. *Biochemistry*, **48**, 7539–7546.
15. Theruvathu,J.A., Kim,C.H., Darwanto,A., Neidigh,J.W. and Sowers,L.C. (2009) pH-dependent configurations of a 5-chlorouracil-guanine base pair. *Biochemistry*, **48**, 11312–11318.
16. Cruse,W., Saludjian,P., Neuman,A. and Prangé,T. (2001) Destabilizing effect of a fluorouracil extra base in a hybrid RNA duplex compared with bromo and chloro analogues. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 1609–1613.
17. Lima,W.F. and Crooke,S.T. (1997) Binding affinity and specificity of *Escherichia coli* RNase H1: impact on the kinetics of catalysis of antisense oligonucleotide-RNA hybrids. *Biochemistry*, **36**, 390–398.
18. Pallan,P.S. and Egli,M. (2008) Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. *Cell Cycle*, **7**, 2562–2569.
19. Rui,E.K. and Kim,J.N. (1989) The oxidative chlorination of pyrimidine and purine bases and nucleosides using acyl chloride-dimethyl-formamide-*m*-chloroperbenzoic acid system. *Nucleosides, Nucleotides*, **8**, 43–48.
20. Jancarik,J. and Kim,S.H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins *J. Appl. Cryst.*, **24**, 409–411.
21. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.
22. Vagin,A. and Teplyakov,A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**, 1022–1025.
23. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
24. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
25. Emsley,P. and Cowtan,K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
26. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
27. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
28. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera, a visualization system for exploratory research and analysis. *J. Comp. Chem.*, **25**, 1605–1612.
29. Pallan,P.S., Prakash,T.P., Li,F., Eoff,R.L., Manoharan,M. and Egli,M. (2009) A conformational transition in the structure of a 2′-thiomethyl-modified DNA visualized at high resolution. *Chem. Comm.*, 2017–2019.
30. Nowotny,M., Gaidamakov,S.A., Crouch,R.J. and Yang,W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, **121**, 1005–1016.
31. Tereshko,V., Minasov,G. and Egli,M. (1999) The Dickerson-Drew B-DNA dodecamer revisited - at atomic resolution. *J. Am. Chem. Soc.*, **121**, 470–471.
32. Hunter,W.N., Brown,T., Kneale,G., Anand,N.N., Rabinovich,D. and Kennard,O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.*, **262**, 9962–9970.
33. Kornberg,A. and Baker,T. (1992) *DNA Replication*, 3rd edn. W. H. Freeman, New York.
34. Dunn,D.B. and Smith,J.D. (1957) Effects of 5-halogenated uracils on the growth of *Escherichia coli* and their incorporation into deoxyribonucleic acids. *Biochem. J.*, **67**, 494–506.
35. Munch-Petersen,A. (ed.), (1983) *Metabolism of Nucleotides, Nucleosides and Nucleobases in Microorganisms*. Academic Press, New York.
36. Willis,M.C., Hicke,B.J., Uhlenbeck,O.C., Cech,T.R. and Koch,T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, **262**, 1255–1257.
37. Pallan,P.S. and Egli,M. (2009) The pairing geometry of the hydrophobic thymine analog 2,4-difluorotoluene in duplex DNA as analyzed by X-ray crystallography. *J. Am. Chem. Soc.*, **131**, 12548–12549.
38. Auffinger,P., Hays,F.A., Westhof,E. and Ho,P.S. (2004) Halogen bonds in biological molecules. *Proc. Natl Acad. Sci. USA.*, **101**, 16789–16794.
39. Kim,C.H., Darwanto,A., Theruvathu,J.A., Herring,J.L. and Sowers,L.C. (2010) Polymerase incorporation and miscoding properties of 5-chlorouracil. *Chem. Res. Toxicol.*, **23**, 740–748.