1906–1912 *Nucleic Acids Research, 2000, Vol. 28, No. 9*

Structural basis for uracil DNA glycosylase interaction with uracil: NMR study

Mahua Ghosh, N. Vinay Kumar[1], Umesh Varshney[1] and K. V. R. Chary*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India and [1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India

Received January 11, 2000; Revised and Accepted March 10, 2000

PDB accession no. 1DGO

ABSTRACT

Two dimensional (2D) NMR and molecular dynamics simulations have been used to determine the three dimensional (3D) structure of a hairpin DNA, d-CTA-GAGGATCC-TUTT-GGATCCT (22mer; abbreviated as U2-hairpin), which has uracil at the second position from the 5′ end of the tetraloop. The ^{1}H resonances of this hairpin have been assigned almost completely. NMR restrained molecular dynamics and energy minimization procedures have been used to describe the 3D structure of U2-hairpin. This study establishes that the stem of the hairpin adopts a right-handed B-DNA conformation, while the T_{12} and T_{15} nucleotides stack upon 3′ and 5′ ends of the stem, respectively. Further, T_{14} stacks upon both T_{12} and T_{15}. Though U_{13} partially stacks upon T_{14}, no stacking interaction is observed between U_{13} and T_{12}. All the individual nucleotide bases belonging to the stem and T_{12} and T_{15} of the loop adopt '*anti*' conformation with respect to their sugar moiety, while the U_{13} and T_{14} of the loop are in '*syn*' conformation. The turning phosphate in the loop is located between T_{13} and T_{14}. This study and a concurrent NMR structural study on yet another hairpin DNA d-CTAGAGGAATAA-TTTU-GGATCCT (22mer; abbreviated as U4-hairpin), with uracil at the fourth position from the 5′ end of the tetraloop throw light upon various interactions which have been reported between *Escherichia coli* uracil DNA glycosylase (UDG) and uracil containing DNA. The ε of T_{12} and α, β, γ, ε and ζ of U_{13} and γ of T_{14}, which partially influence the local conformation of U_{13} in U2-hairpin are all locked in '*trans*' conformation. Such stretched out backbone conformation in the vicinity of U_{13} could be the reason as to why the U2-hairpin is found to be the poor substrate for its interaction with UDG compared to the other substrates in which the uracil is at first, third and fourth positions of the tetraloop from its 5′ end, as reported earlier by Vinay and Varshney. This study shows that UDG actively promotes the flipping of uracil from a stacked conformation and rules out the possibility of UDG recognizing the flipped out uracil bases.

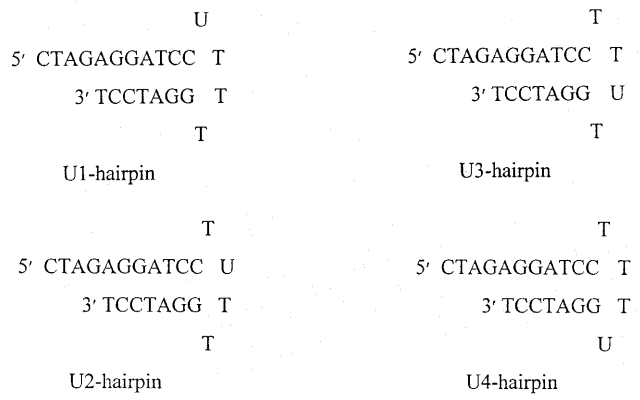
INTRODUCTION

DNA has to face constant challenges to its genomic integrity not only from the internal process of replication but also from the external agents such as chemicals and ionizing radiation (1). This results in the production of a variety of modified bases, such as uracil (U), a constituent of RNA. This U can form as good a Watson–Crick base-pair as T does with A in DNA. Uracil can occur in DNA by either (i) misincorporation of dUMP during replication by DNA polymerases, or (ii) as a product of spontaneous hydrolytic deamination of cytosine residues in DNA. Such incorporation results in G:C→A:T transition mutations in two rounds of replication, unless U is repaired back to C before the first round of replication. Such transition mutations are prevented by uracil DNA glycosylase (UDG), which recognizes U and excises it from DNA so that the correct base is reinserted before the first round of replication. Recently, the structural basis for the recognition of U by various UDGs has been demonstrated (2). The structures show that these enzymes can accommodate non-specific DNA sequences along a channel, which has an active site pocket, tailored to admit only U. Based on substrate specificity studies and X-ray structural analyses of various DNA–UDG co-crystals, the phosphates flanking the U are found to be important in establishing contacts with the enzyme (3). These studies have shown clearly that the UDG scans the DNA backbone from the minor groove side, and the highly conserved Pro–Ser loops of the enzyme establish direct hydrogen bonds thus leading to a compression of the sugar–phosphate backbone by ~4 Å (from 12 to 8 Å). This compression leads to a kink in the backbone, as well as extrahelical protrusion of the sugar and the base into the major groove. This appears to be the first step in the mechanism of UDG action, and is non-specific with regard to the presence or absence of U in the DNA sequence. When the U is encountered during this scanning process, it is taken into the active site pocket ('pull') of the enzyme to form a productive enzyme substrate complex. A Leu residue (272 in human UDG) penetrates this U into the minor groove and occupies the void created in the base stack as a result of extra-helical localization of the U

*To whom correspondence should be addressed. Tel: +91 22 215 2971; Fax: +91 22 215 2110; Email: chary@tifr.res.in

residue, and may even facilitate the step of extra-helical localization of the U ('push'). However, this poses fundamental questions regarding the recognition process (2). One of which is whether UDG locates Us by scanning DNA in a 'one dimensional' process or by a bimolecular collision. The second is whether the enzyme actively promotes the flipping of the U from a stacked conformation, or recognizes the flipped-out bases.

Hairpin DNA consisting of U in the loop provides a good model system to understand the UDG interaction with DNA. Further, the hairpin loop offers an extra-helical situation, wherein U is sometimes in a 'flipped out' form. Thereby U may be spontaneously recognized by UDG. Besides, hairpin DNAs have a very important role in biology as they often act as regulatory sites in gene transcription and replication.



It has been shown that the excision of U from various hairpin loops by UDG (4,5) is dependent on the U position in the loop. It has been shown that the *Escherichia coli* UDG and U-containing tetraloop hairpin DNA interaction is dependent on the position of U in the loop. The values of K_m and V_{max} for the tetra-T-loops where the dT was systematically substituted with dU (U1-, U2-, U3- and U4-hairpin) were found to be K_m: 39.9, 39.9, 22.7, 2.52 ($\times 10^{-7}$ M) and V_{max}: 132, 15.2, 127.9, 173.5 (10^2 nmol/min/mg protein), respectively. Compared to that of single-stranded DNA (ssDNA) (4), the relative excision efficiencies for double-stranded DNA (dsDNA), U1-, U2-, U3- and U4-hairpins are reported to be 40, 3.60, 0.35, 5.23 and 66.34%, respectively. It means that for a tetralooped hairpin DNA, the excision efficiency is substantially less when U is present at the second position (U2-hairpin) compared to that when present at the fourth position (U4-hairpin) from the 5′ end of the tetraloop. This suggests a structural dependence of the UDG–U interaction.

Recently, we have reported the three dimensional (3D) structure of U4-hairpin DNA (6), which establishes that the stem of the hairpin adopts a right-handed B-DNA conformation, while the T_{12} and U_{15} nucleotides stack upon 3′ and 5′ ends of the stem, respectively. Further, T_{14} stacks upon both T_{12} and U_{15} while T_{13} partially stacks upon T_{14}. Very weak stacking interaction is observed between T_{13} and T_{12}. All the individual nucleotide bases adopt '*anti*' conformation with respect to their sugar moiety. The turning phosphate in the loop is located between T_{13} and T_{14}. The stereochemistry of U_{15} mimics the situation wherein U would stack in a B-DNA conformation. This led to the conclusion that the excision efficiency is not much reduced in the case of U4-hairpin compared to ssDNA.

In order to provide a structural basis for the discrepancy seen in the excision efficiencies of U from U2- and U4-hairpin, we have initiated the elucidation of the 3D structure of the U2-hairpin DNA. This paper reports the intricate details of the 3D structure of U2-hairpin, as derived from two dimensional (2D) NMR data and molecular dynamics simulation, and a comparison of its structure with U4-hairpin. This study provides an insight into the interaction of *E.coli* UDG with U in the loop of hairpin DNA.

MATERIALS AND METHODS

DNA synthesis and purification

Two DNA oligonucleotides (22mers; U2- and U4-hairpins) were designed such that a minimum of 7 bp form the stem of the hairpins with 4 nt in the loops. The 4-nt overhanging at the 5′ end of the hairpins was used to facilitate ^{32}P-labeling by end-filling with Klenow polymerase. The oligonucleotides were custom made by Ransom Hill Bioscience, Inc. (Ramona, CA), and purified from 18% polyacrylamide 8 M urea gels (5), desalted on Sep-Pak (Millipore Corporation, MA) columns and lyophilized. Purified oligomers were examined using gel electrophoresis, which reveals the existence of oligos as monomers. Though the overhang at the 5′ end could trigger the formation of a dumbbell, the single hairpins were favoured by the efficient end filling experiments (4). Cooperative thermal dissociation curves (not shown) are observed for both the hairpins with a melting point (T_m) around 45°C, indicating that the oligos adopt a distinct and ordered conformation below the T_m.

In order to rule out the formation of a dimer containing a double-stranded segment of DNA with an internal bubble, oligomers were chilled rapidly after heating to 80°C, thus favouring only intramolecular structures. Further, in an independent experiment a parent oligo containing an extended stem region was 5′-phosphorylated and subjected to ligation. Predominant population (>90%) stopped at dimer formation (cohesive ends), indicating the formation of a dumbbell (4). If it were to be bubble structures, we would have got trimers and tetramers, etc., as seen in case of a control oligo with intermolecular antiparallel base-pairing. Another reason for us to believe that the DNA hairpins are in stem–loop form comes from the kinetics of U excision from the four looped substrates. If it were the bubble, we would not expect substantial differences in the excision efficiencies.

NMR

About 8 mg of purified oligomers were dissolved in 0.6 ml of appropriate solvent (~5 mM strand concentration or 60 mM in nucleoside residues) with no buffer. For experiments in ^2H_2O, the DNA was lyophilized three times from ^2H_2O to deprotonate all the exchangeable protons, prior to its dissolution in 0.6 ml of 99.9% ^2H_2O. For experiments in H_2O, a mixture of 90% H_2O and 10% ^2H_2O was used. ^{1}H NMR experiments were carried out on Varian Unity+ 600 and Bruker AMX 500 spectrometers. The spectra in a mixed solvent of 90% H_2O + 10% ^2H_2O include 1D ^{1}H NMR spectra recorded with P1 $\overline{1}$ pulse sequence (7) and 2D nuclear Overhauser enhancement spectroscopy (NOESY) (8) with P1 $\overline{1}$ detection pulse sequence. The 2D experiments in ^2H_2O include exclusive correlation spectroscopy (E-COSY) (9), clean total correlation

spectroscopy (clean TOCSY) (10) and a set of NOESY spectra with different mixing times (50, 80, 100, 150, 200, 250 and 300 ms). A temperature of 32°C was used in most NMR experiments, though 1D ^{1}H NMR experiments were carried out in the range of 15–55°C. In all the experiments, the ^{1}H-carrier frequency was kept at water resonance. In 2D experiments, time domain data points were 512 and 4096 along t_1 and t_2 dimensions, respectively. The data were multiplied with sine bell window functions shifted by $\pi/4$ and $\pi/8$ along t_1 and t_2 axes, respectively, and zero-filled to 1024 data points along t_1 dimension prior to 2D-FT.

Starting structure and structural restraints

The starting structures with the required sequences were generated using the molecular modeling package INSIGHT-II (MSI, San Diego, CA) on Iris (Indigo II) workstation, as discussed earlier (6). The protruding stems at the 5′ end were however ignored for dynamics and energy minimization calculations, primarily because this stretch is ill-defined with few nOe constraints. The interproton distances have been estimated from a set of NOESY spectra recorded with different mixing times, ranging from 50 to 300 ms as discussed earlier (6,11–13). The lower and upper bounds for the dihedral angle restraints were fixed as discussed earlier (6). They include two sugar ring torsion angles namely -C2′-C3′-C4′-O4′- and -C1′-C2′-C3′-C4′-, one of the backbone dihedral angles, -C5′-C4′-C3′-O3′- (δ) and the glycosidic dihedral angles (χ).

Molecular dynamics and energy minimization methods

Molecular dynamics simulations were performed with DISCOVER (MSI). AMBER force field was used to calculate the energy of the system. Electrostatic interactions were calculated using Coulomb's law with point charges (6–31G* standard ESP charges) (14). Distance-dependent dielectric constant of '1*r' was used. Van der Waals contributions were calculated with a 6-12 Lennard-Jones potential. No distance cutoff was used for calculating Van der Waals interactions. A time step of 1 fs was used. Initial random velocities were assigned in accordance with a Maxwell-Boltzmann distribution. To obtain the starting structure, an initial steepest descent minimization of 100 steps was performed on the initial structure followed by conjugate gradient minimization of 1000 steps. The good-fit structure thus obtained was followed by restrained molecular dynamics simulations. Initial random velocities were assigned with a Maxwell-Boltzmann distribution for a temperature of 600 K. Two hundred structures were collected at 1 ps intervals along the restrained molecular dynamics trajectory. These structures were significantly different from each other as evident from their pairwise root-mean-square deviations (RMSDs). Each of these structures was cooled to 300 K in steps of 50 K. After each temperature step, the system was allowed to equilibrate for 10 ps. This was followed by 500 steps of steepest descent minimization and 1000 steps of conjugate gradient minimization for monitoring the convergence and structure analysis. In the event of any constraint violation another round of dynamics was performed by varying initial temperature as well as the weight of the restraint. The molecule was then cooled to 300 K and energy minimized as mentioned before. This procedure was repeated three times, until well converged structures were obtained with zero violations. In these calculations, as discussed earlier (6), the NMR derived

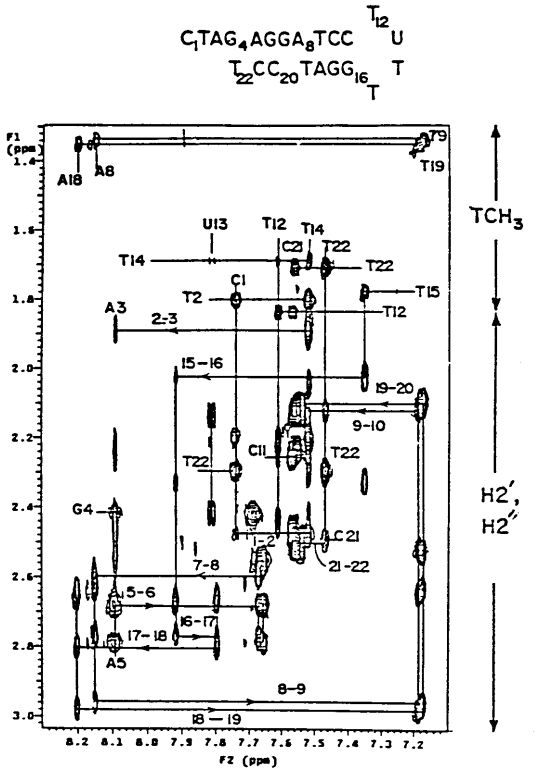


Figure 1. Selected region of NOESY spectrum of U2-hairpin recorded with a mixing time of 250 ms, in 99.9% ^2H_2O at 32°C and pH 7. This shows nOe connectivities from T(CH_3)/H2′/H2″ to H6/H8.

distance restraints were applied all along with the upper and lower bounds ±0.05 nm and with force constants of 25 kcal.mol^{-1}°A^{-2} for all nOes involving non-exchangeable protons, 10 kcal.mol^{-1}°A^{-2} for all nOes involving exchangeable protons and the atoms involved in H-bonds. For the dihedral angle restraints a force constant of 20 kcal mol^{-1}rad^{-2} was used.

RESULTS AND DISCUSSION

^{1}H NMR assignments in U2-hairpin

Sequence-specific ^{1}H resonance assignments were achieved through established procedures (15–21). Figure 1 shows an illustrative example of NOESY spectrum of U2-hairpin with H2′/H2″/CH_3-H6/H8 connectivities. Except for the serious overlap seen in the case of H6 resonances belonging to C_{10}, C_{11}, C_{20} and C_{21}, the assignments were straightforward. The degeneracy between these H6 protons could be resolved by the observation of intranucleotide and sequential nOes between their respective CH5 protons and H2′/H2″/CH6. The only missing connectivities were between G_4 and A_5 units. The stereo-specific assignment of individual H2′ and H2″ could be achieved by intensity comparison of the H1′-H2′ and H1′-H2″ cross-peaks in the NOESY spectrum, wherein the latter is found to be stronger than the former. The chemical shifts of all the protons thus obtained have been deposited in the Protein

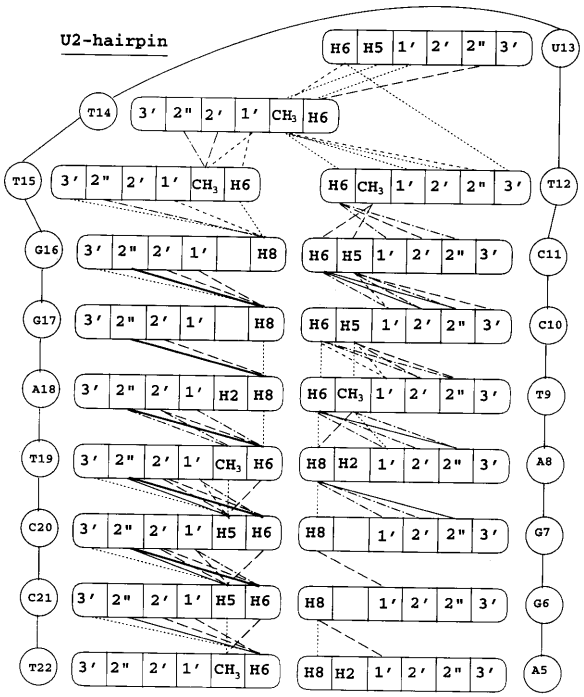


Figure 2. Various internucleotide nOe connectivities seen in 250 ms NOESY spectrum of U2-hairpin. The intensities of the various nOe are depicted as follows: — very strong, —— strong, – – – – – medium, – · – · – medium weak, – · · – · · – weak, · · · · · · · · very weak.

Data Bank (PDB) (PDB accession no. 1DGO; RCSB ID RCSB010071).

Secondary structure of U2-hairpin

Interstrand NOESY cross-peaks G_6(H1)-C_{21}(H41/H42), G_7(H1)-C_{20}(H41/H42), A_8(H2)-T_{19}(H3), T_9(H3)-A_{18}(H2), C_{10}(H41/H42)-G_{17}(1NH) and C_{11}(H41/H42)-G_{16}(H1) establish a hydrogen-bonded base pairing between G_6:C_{21} (G_6 and C_{21}), G_7:C_{20}, A_8:T_{19}, T_9:A_{18}, C_{10}:G_{17} and C_{11}:G_{16} and hence the hairpin conformation. Qualitative analysis of the relative NOESY cross-peak intensities (Fig. 2) establishes that the stem of the hairpin adopts a right-handed B-DNA duplex conformation. The nOe data further confirms the association of A:T and G:C base pairs through Watson–Crick base pairing schemes with almost all the individual bases in the stem of the hairpin adopting *anti* conformation with the glycosidic dihedral angle, χ, ranging from –80 to –120°. This is based on the observation of strong intranucleotide H2′-H6/H8 cross-peaks compared to H2″-H6/H8 cross-peaks, while H1′-H6/H8 cross-peaks are relatively weak or absent. In the case of C_{10}, C_{11}, C_{20} and C_{21} we could not establish the respective χ values because of the severe spectral overlap of H1′/H2′/H2″-H6 cross-peaks. On the other hand, the loop residues, T_{12}, U_{13}, T_{14} and T_{15} show interesting nOe connectivities (Fig. 2). Though most of the expected sequential nOes are seen all along this stretch, the interactions between U_{13}(H6/H5) and T_{12}(H1′/H2′/H2″) are surprisingly absent. The only interaction seen between U_{13} and T_{12} is the T_{12}(H3′)-U_{13}(H6) nOe. Further, T_{14}(CH_3) shows medium intensity nOes to T_{12}(H1′/H2′/H2″/H6/CH_3), indicating a partial stacking interaction between T_{12} and T_{14} bases. These nOe interactions essentially dictate the folding pattern of the loop, which will be discussed later. As for the stem, the T_{12} and T_{15} bases of the loop residues are also found to adopt *anti* conformation. However, the χ values could not be characterized in the case of U_{13} and T_{14}. By the end of the assignment procedure, all the major cross-peaks in the 2D spectra could be assigned uniquely. Although no resonances could be ascribed to another conformer, a lone cross-peak seen between the C_1(H6) and the T_{22}(H2′/H2″) indicated a dumbbell formation. However, no extra imino proton resonance was observed under the conditions of NMR experiments mentioned above to substantiate the formation of dumbbell DNA.

Conformational-dependent characteristic multiplet structures of H2′-H1′ and H2″-H1′ cross-peaks in the E-COSY have been used to estimate values ^{3}J(H1′-H2′) and ^{3}J(H1′-H2″) (6,22–25). Though these Js are error prone, we could conclude accurately as to which one of them is larger for fixing a certain window to the sugar puckers. In the present study, we could estimate the ^{3}J(H1′-H2′) and ^{3}J(H1′-H2″) values for 17 nt units and in all these cases the ^{3}J(H1′-H2′) is clearly found to be larger compared to ^{3}J(H1′-H2″). These J values qualitatively indicate that the corresponding sugar rings adopt conformation in the S domain of the pseudorotational map with P ranging from C1′-exo to C3′-exo (P = 90–198°). The ^{3}J(H2″-H3′) and ^{3}J(H3′-H4′) values which could have helped in further narrowing the domains of sugar puckers could not be estimated from this E-COSY spectrum because of the low intensity of the corresponding peaks.

NMR structure determination of U2-hairpin

Restrained molecular dynamics simulation and energy minimization calculations were performed on the U2-hairpin following the procedure described in Materials and Methods (6). A total of 225 interproton distance constraints (10 involving exchangeable protons and 215 involving non-exchangeable protons) and 64 dihedral angle restraints were used with the force constants described earlier. All these restraints have been deposited in the PDB (PDB accession no. 1DGO; RCSB ID RCSB010071; http://www.pdb.bnl.gov/). Out of the 200 calculated structures, there are eight structures lying within 2.5 kcal/mol above the minimum energy structure. These nine structures are characterized by low all atom pairwise RMSDs, ranging from 0.15 to 0.51 (see Supplementary Material, Table S1). Figure 3 shows the best-fit superimposition of these nine structures. The corresponding PDB files have been deposited in the PDB (PDB accession no. 1DGO; RCSB ID RCSB010071). Average values with standard deviation of all the backbone torsion angles and glycosidic dihedral angles for all the nine structures are listed in Table S2. The stereochemistry of all these nine structures were critically examined for proper hydrogen-bond lengths and angles in the Watson–Crick base-pairs, stereochemical feasibility of the various dihedral angles and any sterically hindered non-bonded inter-atomic distances. All the nine structures satisfied these criteria.

Backbone torsion angles in U2-hairpin

The α (-O3′-P-O5′-C5′-), β (-P-O5′-C5′-C4′-), γ (-O5′-C5′-C4′-C3′-) and ε (-C4′-C3′-O3′-P-) for each nucleotide in the stem of the U2-hairpin DNA of all the nine structures are mostly locked into

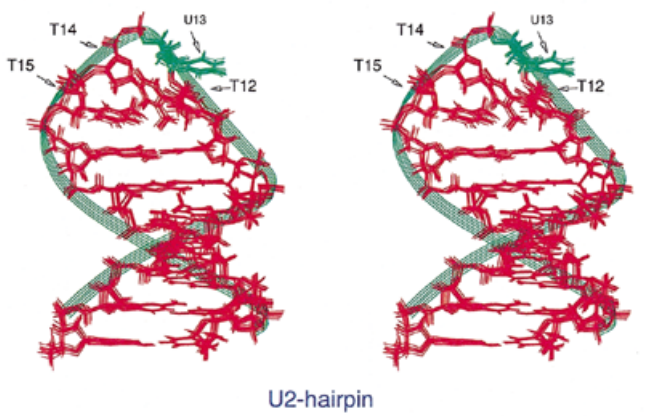


Figure 3. Stereo view showing a best-fit superimposition of the final nine molecular dynamics and energy minimized simulated structures of U2-hairpin.

gauche$^-$ (g$^-$), *trans* (t), *gauche*$^+$ (g$^+$) and *trans* (t) conformations, respectively, similar to those observed in B-DNA. The only exception is the case of G_{16}, which is at 3′ end of the tetraloop. For this, α angle is g$^-$ in five of the nine structures and close to 150° for the rest, while γ is in t conformation in four of the nine structures and g$^+$ for the rest. The ζ (-C3′-O3′-P-O5′-) values adopt –106° on average and range from –97 to 115° for all the residues. The δ (-C5′-C4′-C3′-O3′-) values adopt 136° on average and range from 129 to 145°.

In the case of a tetraloop, it is interesting to note that the α, β, γ and ε of T_{12} and T_{15} nucleotide units get locked into g$^-$, t, g$^+$ and t conformations, respectively, similar to the stem. On the other hand, for T_{12} and T_{15} the ζ is locked into g$^-$ conformation while the δ adopts 124.5 and 132.5° on average, respectively, similar to those observed in B-DNA. As far as T_{14} is concerned, the α, β, γ and ε are locked into g$^+$, g$^-$, t and t conformations, respectively. The most striking observation of the loop conformation concerns the backbone dihedral angles of U_{13} which adopt *trans* conformation except in the case of δ, which takes a value of 91° on average. The dihedral angles that facilitate the loop formation are α, γ and ζ of U_{13} and γ of T_{14}, of which all adopt t conformation and thus stretch the backbone.

Nucleotides have C2′-endo sugar puckers and *anti* conformation

In all the nine structures, the sugar puckers lie in the S domain of the pseudorotational wheel and most of the nucleotides assume a sugar pucker in the range of 124–150°. The exceptions are in the case of sugars of U_{13} and T_{14} which adopt O4′-endo puckers. This is supported by the observation of strong intranucleotide nOes which are expected between the H1′ and H4′ for these nucleotide units (24). A different behavior for these two nucleotides can be expected since these are present in the loop region of the hairpin DNA. As far as the χ is concerned, almost all the nucleotide units are in the *anti* domain, as are evident in the relative intensities of the resolved nOes between the base and the sugar protons. The χ values range from –100 to –127°. The exceptions are again in the case of U_{13} and T_{14} which adopt *syn* conformation, with the χ values as 15.5 and 24.0°, respectively, on average.

Turning phosphate

As mentioned earlier, the ζ and γ for U_{13} and T_{14}, respectively, are characteristically in t conformation. Because of this the backbone takes a sharp swerve near the phosphate linking T_{13} and T_{14}. Similar phosphodiester conformations were found for the turning phosphates in the case of U4-hairpin and -CG TTTT CG- type hairpins (24,26). In the present study, the simulated model reveals that the turning phosphate is indeed between U_{13} and T_{14}. The ^{31}P chemical shifts which could have thrown more light upon this, suffer from extensive overlaps.

Comparison of U2-hairpin structure with U4-hairpin DNA

U4-hairpin has been structurally characterized previously with a combined use of high resolution NMR and molecular dynamics calculations (6). In the light of this, it is interesting to compare the 3D structure of U2-hairpin with that of U4-hairpin. Both the stems of U2- and U4-hairpin are found to contain Watson–Crick base-pairs adopting a right-handed B-DNA conformation. Besides, interesting common features are also noted regarding the conformation of the loop of these hairpins. In both the hairpins, the right-handed backbone continued through the 3′ top of the stem to the 5′ top of the stem, by taking one sharp turn, and the loops are characterized by the stacking of individual bases $(T/U)_d$ (the nucleotide T or U at the position 'd' of the tetraloop forms the 3′ top of the stem), T_c and $(U/T)_b$ over the 5′ end of the stem as seen earlier in the case of -CG TTTT CG- type hairpins (26,27). These are consistent with the observed internucleotide nOes in each case. For example, in the case of U2-hairpin, T_{12} and T_{15} stack upon the top of the stem of the U2-hairpin, T_{12} over C_{11} and T_{15} over G_{16}. As reported earlier, this should have encouraged T_{15} and T_{12} to participate in hydrogen bonding. But NMR data in the present case do not support any formation of T_{12}-T_{15} wobble base-pair. On the other hand, T_{14} shows partial stacking interaction with T_{15} and T_{12} bases. This interaction is responsible for pulling T_{14} base towards the helical stem axis, as is evident from the observation of medium intensity nOes between $T_{14}(CH_3)$ and $T_{12}(H6/CH_3/H1'/H2'/H2'')$. Further U_{13} base is tilted away from T_{12} as evident from the absence of internucleotide base to sugar proton sequential connectivities $[U_{13}(H6)$-$T_{12}(H1'/H2'/H2'')]$. The most striking feature of U2-hairpin loop, however, is the base conformations of U_{13} and T_{14}, which adopt *syn* conformation with respect to the sugar moiety, while T_{12} and T_{15} adopt *anti* conformation. As far the U4-hairpin all the bases in the loop adopt *anti* conformation.

Structural basis for UDG and U containing hairpin DNA interaction

As reported earlier, of all the four hairpin DNAs studied so far, U2-hairpin is found to be the weakest substrate, while U4-hairpin the best (4). This characteristic feature is reflected in the 3D structures of U2- and U4-hairpins.As discussed earlier, all the backbone torsion angles in the stem and the tetraloop of U2- and U4-hairpin are mostly locked in conformations similar to those observed for B-DNA. As reported earlier, the U at the fourth position in the loop of U4-hairpin is fairly stacked with T_{14} and G_{16} on either side (Fig. 4B), with its base in *anti* conformation. Besides, the δ, ε and ζ of T_{14}, and δ, ε and ζ of

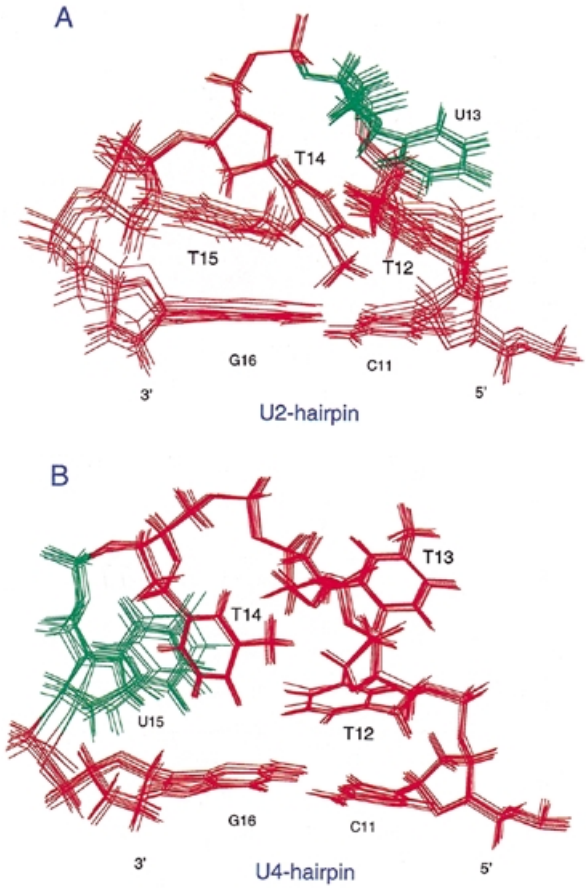


Figure 4. Expanded loop region of the (**A**) U2-hairpin and (**B**) U4-hairpin.

U_{15}, and α, β, γ and δ of G_{16}, which influence the local conformation of U_{15} are all locked in conformations similar to those found for B-DNA. Thus, the stereochemistry of U_{15} is found to mimic the situation in which U would stack in a dsDNA. This led to the conclusion that the excision efficiency is not much reduced in the case of U4-hairpin as compared to ssDNA. On the other hand, solution conformation of U2-hairpin (Fig. 4A) clearly shows that the sugar–phosphate backbone flanking the scissile U glycosidic bond is stretched out into a non B-DNA form. The U at the second position (U_{13}) in the loop of U2-hairpin is found to be in an extra-helical conformation. Though it partially stacks upon T_{14}, no stacking interaction is observed between U_{13} and T_{12}. And the U base adopts a *syn* conformation with respect to the sugar moiety and points into the minor groove side of the helical hairpin stem. Do these structural characteristics reflect on the values of K_m and V_{max} both of which contribute for the poor utilization of U2-hairpin as a substrate by UDG?

While structure of DNA–*E.coli* UDG complex is not available, comparison of the crystal structures of *E.coli* UDG or its complex with Ugi, with that of human UDG or its complex with Ugi, respectively, show enormous similarities, which suggest that the mechanism of U release by the human and the *E.coli* UDGs is mechanistically similar. Thus when we compare the structure of U2-hairpin with that of the DNA bound to UDG, it is clear that this substrate does not present itself for efficient establishment of hydrogen bonds with the Ser–Pro loops of the enzyme because of the stretched out sugar–phosphate conformation of U2-hairpin in the vicinity of U_{13}. As shown in Table S2, the γ of T_{12} and α, β, γ, ε and ζ of U_{13} and γ of T_{14}, which influence the local conformation of U_{13} in U2-hairpin are all locked in t conformation. Such stretched out backbone conformation in the vicinity of U_{13} could be the reason as to why the enzyme is not able to make proper contacts with the backbone. In addition, the extrahelical protrusion of the U towards the minor groove side of the hairpin stem may also lead to steric hindrance in the approach of the UDG to DNA. On the other hand, the U_{15} in the U4-hairpin, which is the best substrate of all the four loop substrates is located in an environment wherein the backbone as well as the base conformation mimic the B-form of DNA. Thus the structural features of U2-hairpin provide a basis for its poor binding with the enzyme. However, this still does not explain why the catalytic rates (V_{max}) for U excision are poor. For productive enzyme substrate complex formation, it is essential that the U which is facing the minor groove side of the stem, and is in *syn* configuration with respect to the sugar, be rotated into the major groove side of the DNA to make appropriate contacts in the active site of the enzyme. Presumably, the time taken for these structural changes to occur before a productive enzyme substrate complex is formed is what results in lower catalytic rates of U release from U2-hairpin. Further, the conformation of dU in U1- and U3-hairpins may be more favourable compared to U2-hairpin for its localization into the active site pocket. This might be the reason as to why the V_{max} is almost 10-fold more in the case of U1- and U3-hairpins compared to U2-hairpin. The 3D strcutures of U1- and U3-hairpins will throw more light on this point. Taken together, these observations support our interpretation that the unfavourable backbone results in poor K_m, whereas the unfavourable nucleotide conformation results in poor V_{max}, and that jointly these parameters make the U2-hairpin the most inefficient substrate for UDG.

Biological implications of poor U release from looped DNA

Presence of hairpin structures in genomes is not unusual. However, based on the structure of a tetraloop hairpin that we present here and as reported by others earlier (28), it is clear that to accommodate a sharp turn in the loop, some of bases are pushed away from the base stack into an exposed configuration. It is known that the deamination rates of cytosine increase when it is not involved in a base pair (29). Clearly, an exposed cytosine will be prone to deamination. However, the enzyme kinetics, and now the solution conformation of the U2-hairpin show that such substrates are an extremely poor base for excision repair. If so, the regions of genome that are prone to extrude as hairpins with short loops are likely to be hotspots for mutation. Have the organisms developed any systems to avoid a high rate of mutations in such regions of DNA? At least, the *in vitro* studies which show that UDGs interact with ssDNA binding proteins, SSBs (30,31) and that inclusion of SSBs lowers the melting temperatures of these hairpin structures and facilitates U release from the looped substrates, suggest that involvement of SSB could be one such system which facilitates the U excision repair pathway and in mutation avoidance in the regions of genome which have a potential to extrude into hairpin structures.

ACKNOWLEDGEMENTS

The facilities provided by the National Facility for High Field NMR supported by the Department of Science and Technology (DST), Department of Biotechnology (DBT), Council of Scientific and Industrial Research (CSIR) and Tata Institute of Fundamental Research, are gratefully acknowledged. Part of the work was supported by the DBT. N.V. was a CSIR senior research fellow.

SUPPLEMENTARY MATERIAL

See Supplementary Material available at NAR Online.

REFERENCES

1. Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington DC.
2. Krokan,H.E., Standal,R. and Slupphaug,G. (1997) *Biochem. J.*, **325**, 1–16.
3. Parik,S.S., Mol,C.D., Sluppaug,G., Bharati,S., Krokan,H.E. and Tainer,J.A. (1998) *EMBO J.*, **17**, 5214–5226.
4. Kumar,N.V. and Varshney,U. (1997) *Nucleic Acids Res.*, **25**, 2336–2343.
5. Kumar,N.V. (1999) *Ph.D Thesis*. Indian Institute of Science, Bangalore.
6. Ghosh,M., Kumar,N.V., Varshney,U. and Chary,K.V.R. (1999) *Nucleic Acids Res.*, **27**, 3938–3944.
7. Hore,P.J. (1983) *J. Magn. Res.*, **55**, 283–300.
8. Kumar,A., Wagner,G., Ernst,R.R. and Wuthrich,K. (1980) *Biochem. Biophys. Res. Commun.*, **96**, 1156–1163.
9. Griesinger,C., Sorensen,O.W. and Ernst,R.R. (1986) *J. Chem. Phys.*, **85**, 6837.
10. Grieseinger,C., Otting,G., Wuthrich,K. and Ernst,R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.
11. Kumar,A., Wagner,G., Ernst,R.R. and Wuthrich,K. (1981) *J. Am. Chem. Soc.*, **103**, 3654–3658.
12. Wagner,G. and Wuthrich,K. (1979) *J. Magn. Reson.*, **33**, 675–680.
13. Chary,K.V.R., Hosur,R.V., Govil,G., Chen,C. and Miles,H.T. (1988) *Biochemistry*, **27**, 3858–3867.
14. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.
15. Feigon,J., Leupin,W., Denny,W.A. and Kearns,D.R. (1983) *Biochemistry*, **22**, 5943–5951.
16. Scheek,R.M., Boelens,R., Russo,N., van Boom,J.H. and Kaptein,R. (1984) *Biochemistry*, **23**, 1371–1376.
17. Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, NY.
18. Reid,B.R. (1987) *Q. Rev. Biophys.*, **20**, 1–34.
19. van de Ven,F.J.M. and Hilbers,C.W. (1988) *Eur. J. Biochem.*, **178**, 1–38.
20. Chary,K.V.R., Hosur,R.V., Govil,G., Chen,C. and Miles,H.T. (1989) *Biochemistry*, **28**, 5240.
21. Chary,K.V.R. (1991) *Magnetic Resonance-Current Trends*. Narosa Publishing House, New Dehli, pp. 71–104.
22. Chary,K.V.R., Hosur,R.V., Govil,G., Tan,Z. and Miles,H.T. (1987) *Biochemistry*, **26**, 1315–1322.
23. Hosur,R.V., Ravikumar,M., Chary,K.V.R., Seth,A., Govil,G., Tan,Z. and Miles,H.T. (1986) *FEBS Lett.*, **205**, 71–76.
24. Rinkel,L.J. and Altona,C. (1987) *J. Biomol. Struct. Dyn.*, **4**, 621–649.
25. Chary,K.V.R. and Modi,S. (1988) *FEBS Lett.*, **233**, 319–325.
26. Hare,D.R. and Reid,B.R. (1986) *Biochemistry*, **25**, 5341–5350.
27. Baxter,S.M., Greizerstein,M.B., Kushlan,D.M. and Ashley,G.W. (1993) *Biochemistry*, **32**, 8702–8711.
28. Kuklenyik,Z., Yao,S. and Marzilli,L.G. (1996) *Eur. J. Biochem.*, **236**, 960–969.
29. Mosbaug,D.W. and Bennet,S.E. (1994), *Prog. Nucleic Acid Res. Mol. Biol.*, **48**, 315–369.
30. Purapatre,K., Handa,P., Venkatesh,J. and Vashney,U. (1999) *Nucleic Acids Res.*, **27**, 3487–3492.
31. Nagelhus,T.A., Haug,T., Singh,K.K., Keshav,K.F., Skorpen,F., Otterlei,M., Bharati,S., Lindmo,T., Benchou,S., Benarous,R. and Krokan,H.E. (1997) *J. Biol. Chem.*, **272**, 6561–6566.