Structural polymorphism exhibited by a quasipalindrome present in the locus control region (LCR) of the human β-globin gene cluster

Mahima Kaushik and Shrikant Kukreti*

Nucleic Acids Research Laboratory, Department of Chemistry, University of Delhi, (North Campus), Delhi 110007, India

ABSTRACT

Structural polymorphism of DNA is a widely accepted property. A simple addition to this perception has been our recent finding, where a single nucleotide polymorphism (SNP) site present in a quasipalindromic sequence of β-globin LCR exhibited a hairpin-duplex equilibrium. Our current studies explore that secondary structures adopted by individual complementary strands compete with formation of a perfect duplex. Using gel-electrophoresis, ultraviolet (UV)-thermal denaturation, circular dichroism (CD) techniques, we have demonstrated the structural transitions within a perfect duplex containing 11 bp quasipalindromic stretch (TGGGG(G/C)CCCCA), to hairpins and bulge duplex forms. The extended version of the 11 bp duplex, flanked by 5 bp on both sides also demonstrated conformational equilibrium between duplex and hairpin species. Gel-electrophoresis confirms that the duplex coexists with hairpin and bulge duplex/cruciform species. Further, in CD spectra of duplexes, presence of two overlapping positive peaks at 265 and 285 nm suggest the features of A- as well as B-type DNA conformation and show oligomer concentration dependence, manifested in A → B transition. This indicates the possibility of an architectural switching at quasipalindromic region between linear duplex to a cruciform structure. Such DNA structural variations are likely to be found in the mechanics of molecular recognition and manipulation by proteins.

INTRODUCTION

It is well established that DNA adopts various conformations and all alternative forms of DNA are restricted to a small subset of nucleotide sequences (1,2). The human genome, like other mammalian genomes has very high proportion of repeated DNA sequences (3). Other than the direct repeats within the non-coding and occasionally within coding regions, there are defined ordered sequences that contain various symmetry elements viz. inverted repeats (palindromes) and mirror repeats. The inverted repeat sequences that are not completely symmetrical or that have a center region which is not an inverted repeat is called imperfect palindrome or quasipalindromes. Because of their nature, inverted repeats in DNA and RNA can engage in intra- and intermolecular base pairing forming a variety of structural forms like hairpins, bulges, internal loops, cruciforms, Holliday junctions etc (4–6).

The biological relevance of palindromic and quasipalindromic sequences is clear from their occurrence at functional recognition sites in DNA (7,8). Eukaryotic DNA, in contrast to that of prokaryotes is characterized by having large palindromic regions. Often inverted repeats occur near putative control regions of genes or at origin of DNA replication (9). In the linear form they play important biological roles as binding sites (operator sequences) for dimeric proteins (repressors and activators) (8). Owing to their distinct structure, hairpins may offer binding sites for proteins (10,11). Role of DNA secondary structure in the initiation of viral DNA replication has been reported (12). Experiments showed that one of the strands in the double-stranded molecule at its palindromic stretches adopts a hairpin with AT rich loop and is specially recognized by the origin binding protein. DNA hairpins and cruciforms are determinants for topoisomerase II recognition and cleavage (13,14).

Imperfect inverted repeats undergo spontaneous mutation to more-perfect inverted repeats and this correction is a general mechanism for mutation in prokaryotes. This was confirmed by an analysis, where relative frequency of quasipalindromes and perfect palindromes in more than 100 sequenced prokaryotic genomes was determined and found that perfect palindromes were relatively more frequent than quasipalindromes (15). Mutational hotspots in natural quasipalindrome in *Escherichia coli* are recently been reported

*To whom correspondence should be addressed. Tel: +91 011 27666726; Fax: +91 011 27666605; Email: skukreti@chemistry.du.ac.in

(16,17). The ability to adopt hairpin and cruciform secondary structures by imperfect inverted repeats is associated with frameshift mutations. Several human genetic diseases illustrate inverted repeat mediated mutagenesis (18). The role of hairpin formation in Fredreich's ataxia triplet repeat expansion has recently been reported (19). A recent elegant survey describes some principles for the formation of unusual DNA duplex and hairpin motifs (20).

Locus control regions (LCRs), first defined in the human β-globin locus are operationally defined by their ability to enhance the expression of linked genes to physiological level in a tissue-specific and copy number-dependent manner at ectopic chromatin sites (21,22). The presence of homeodomain protein binding sites, inverted repeats and nuclear matrix attachment regions along the β-globin gene cluster is well documented (23). Dyad symmetry sequences are present close to the four DNase I hypersensitive site (HS) in the β-globin LCR. More recent studies (24), on the spectrum of β-thalassemia mutation and their association with allelic sequence polymorphism at the human β-globin gene cluster have revealed a single nucleotide polymorphism (SNP) at the quasipalindromic sequence of HS4 of the LCR. The studies carried out on Indian population showed an A → G polymorphism in the sequence d-TGGGG(A/G)CCCCA (24). A significant association was observed between the G allele and the occurrence of β-thalassemia. It was concluded that it was an evolutionarily new mutation and was hypothesized that the quasipalindromic stretch might exist in a hairpin form, where the A/G might form single residue loop.

Our recent study (25) which was an attempt to work out the said hypothesis, reported the structure of the quasipalindromic sequence TGGGG(A/G)CCCCA displaying A → G SNP. Interestingly, the multiple sequence alignment of HS4 region in β-globin gene cluster from different organisms (rabbit, mouse, bovine, galago and goat) revealed that the quasipalindromic stretch studied here is unique to *Homo sapiens*. This imperfect palindrome exhibited a hairpin-duplex equilibrium at near physiological solution conditions. Dependence of CD spectra on oligomer concentration manifested in the shift of CD signals from 265 to 285 nm position was interpreted in terms of interconversion of A → B form of DNA, where hairpin species adopts A-form, while B-form is the preferred conformational state of intermolecular duplex.

Among the polymorphic conformations, the structural elements viz. hairpins, bulges, internal loops may be considered important fundamental building blocks in nucleic acid structure. Structural transitions between various forms of DNA would have consequences *in vivo*, and a thorough understanding of their physical and structural properties is consequential. With these concepts in mind and prompted by our previous study (25) and by others work on similar lines (26–29), we undertook structural study on the perfect duplexes [HP(A+T)11/HP(C+G)11] of 11 nt quasipalindrome 5′-TGGGG(A/T/C/G)CCCCA-3′ namely, (HPA11/HPT11/HPC11/HPG11) and the perfect duplexes of its extended versions [HP(A+T)21/HP(C+G)21], the flanked 21 nt long sequences [5′-GCTCTTGGGG(A/G)CCCCAGT-ACA-3′ namely, (HPA21/HPG21)]/[5′-TGTACTGGGG(T/C)CCCCAAGAGC-3′namely (HPT21/HPC21)] present in the HS4 of major regulatory LCR of β-globin gene. Using gel-electrophoresis, ultraviolet (UV)-thermal denaturation,

circular dichroism (CD) techniques, we demonstrate the structural transitions within a perfect duplex to hairpin and bulge duplex/cruciforms. The characterization of such polymorphic sites at palindromic/quasipalindromic sequences is of crucial importance in understanding the biological functions of DNA. Unusual structural motifs may also represent novel targets for pharmaceutical research.

MATERIALS AND METHODS

The oligonucleotides, synthesized in 1 µM scale by Bio Basic Inc., Canada, were received as PAGE purified in the form of lyophilized powder. They were stored at −20°C and were used without further purification. The concentration of the oligonucleotides was determined spectrophotometrically by using the extinction coefficient (ε) calculated by nearest neighbour method (30) and measuring the absorbance at 260 nm at elevated temperature (95°C), following the method described earlier (31). The ε values used for 11mer oligo-sequences d-TGGGGACCCCA (HPA11), d-TGGGGG-CCCCA (HPG11), d-TGGGGTCCCCA (HPT11) and d-TGG-GGCCCCCA (HPC11) were 104 200, 100 900, 100 700 and 98 000 $M^{-1}cm^{-1}$, respectively and for 21mer oligonucleotides d-GCTCTTGGGG$\underline{A}$CCCCAGTACA (HPA21), d-GCTCTTGGGG$\underline{G}$CCCCAG$\underline{T}$ACA (HPG21), d-TGTACT-GGGG$\underline{T}$CCCCAAG$\underline{A}$GAGC-3′ (HPT21) and d-TGTACTGGGG-GCCCC$\underline{C}$CAAGAGC-3′ (HPC21) were 196 700, 193 400, 198 800 and 196 100 $M^{-1}cm^{-1}$, respectively. d-TAAAAAT (SC, ε = 78 800 $M^{-1}cm^{-1}$), d-CTTGAGCTCAAG (PAL12, ε = 113 700 $M^{-1}cm^{-1}$), d-CGCGCGCGCGCGCGCGCGCG (PAL20, ε = 168 300 $M^{-1}cm^{-1}$), were used as size markers in gel assays.

For simpler understanding, the duplexes formed by mixing equimolar concentration of studied oligonucleotides will be named as HP(A+T)11, HP(C+G)11, HP(A+T)21 and HP(C+G)21.

The stock solutions of the oligomers were prepared by dissolving directly the lyophilized powder in MilliQ water. All other chemicals were of analytical grade. The buffer solution consisted of 20 mM sodium cacodylate (pH 7.4), 0.1 mM EDTA containing 100 mM NaCl.

Non-denaturing gel-electrophoresis

For performing gel assays, oligonucleotide samples were prepared in 20 mM sodium cacodylate buffer (pH 7.4) at desired concentrations. The final volume of the sample in the buffer was 20 µl. Importantly prior to performing gel assays in non-denaturing conditions, the purity of the commercially made oligomers was checked by running them on 20% PAGE using 7 M urea. They migrated as single bands. For non-denaturing gel assays, the samples (20 µl) of total volume, were heat treated at 95°C for 5 min and slowly cooled to room temperature over about 10 h. The oligonucleotides (at 10 µM strand concentrations) samples were incubated at 4°C for 3 h before loading onto 10% polyacrylamide gel pre-equilibrated for 2 h. The gel contained 20 mM sodium cacodylate (pH 7.4) with 100 mM NaCl and 0.1 mM EDTA and the running buffer being 1× TBE with 100 mM NaCl. For simplicity, the salt is designated as Na⁺ cations at appropriate places in the text. Tracking dye consisted of

Orange-G. The gels were run at a constant voltage of 40 V at room temperature (25°C). After electrophoresis the gels were stained with Stains-all (Sigma) solution and finally visualized under white light and photographed by Alphalmager™ 2200 (Alpha Infotech Corp.). We would like to add here that in our previous report (25), gel experiments were performed using radiolabelling method for detecting oligomer structures. Interestingly, we were able to get identical results using staining (stains-all/silver staining) method. However, we found that use of stains did not change the concept of our results.

UV-thermal denaturation

The thermal denaturation experiments were performed on a Varian make CARY-100 Spectrophotometer equipped with a peltier thermo-programmer and interfaced with a Pentium III computer for data collection and analysis. The stoppered quartz cuvettes of 10 and 1 mm optical path length with 1 and 0.35 ml volume, respectively, were used for the measurements. The oligonucleotide samples were prepared by taking their appropriate range of strands concentrations, heating the samples up to 100°C for 5 min followed by slow cooling. The temperature dependence on the absorption value of the DNA was monitored at 260 nm. The temperature of the cell holder was increased from 0 to 100°C at a rate of 0.5°C/min. A teflon-coated temperature probe, immersed directly in a control cuvette, measured the sample temperature. The sample solutions were overlaid with paraffin-oil to prevent evaporation. The melting curves were normalized at lower/higher temperature values. The thermal melting temperature (T_m) was determined from the peak of the computer generated first derivative of the absorbance verses temperature profile. The accuracy of the reported T_m values is ±1°C
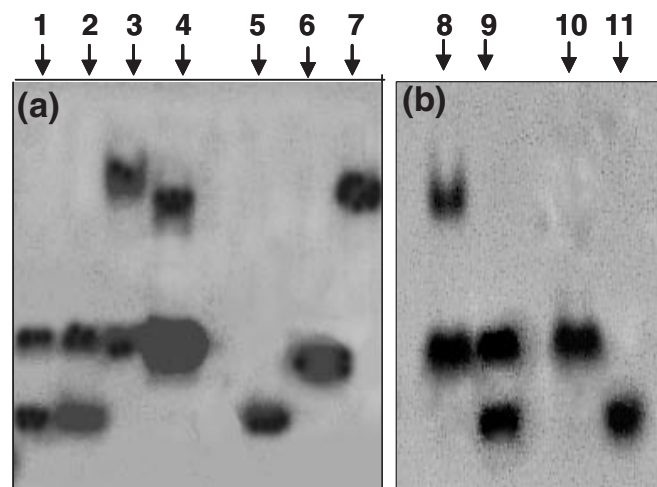
Circular dichroism

For secondary structure determination, CD Spectra were recorded on JASCO-715 Spectropolarimeter interfaced with an IBM PC compatible computer, calibrated with D-Camphor Sulphonic acid. Five scans of the spectrum were collected over the wavelength range of 220–320 nm at a scanning rate of 100 nm/min. The average of multiple scans was used for analysis. The scan of the buffer alone recorded at room temperature was subtracted from the average scans for each DNA duplex. Data were collected in units of millidegrees versus wavelength and normalized to total DNA concentration.

RESULTS AND DISCUSSION

Non-denaturing gel-electrophoresis

Non-denaturing gel studies reported previously (25), revealed that the individual strands HPA11, HPG11 and the extended version of HPG11 (the flanked HPG21), do exist as unimolecular and bimolecular species corresponding to hairpin and bulge duplex conformations, respectively. Further, in this report their duplex forming ability was investigated in presence of their complementary strands (HPT11, HPC11 and flanked HPC21), HPA21 and HPT21. The duplex was prepared in 20 mM sodium cacodylate buffer (pH 7.4)

Figure 1. (**a**: lane 1 → 7) (**b**: lane 8 → 11) 10% Native PAGE mobility pattern of the oligonucleotide sequences. Lane 1: HPC11, d-(TGGGGCCCCCA); lane 2: HPC11 + HPG11, d-(TGGGGCCCCCA) + d-(TGGGGGCCCCA); lane 3: HPC21, d-(TGTACTGGGGCCCCCAAGAGC); lane 4: HPC21 + HPG21, d-(TGTACTGGGGCCCCCAAGAGC) + d-(GCTCTTGGGGGCCCCAGTACA; lanes 5 and 11: SC, d-(TAAAAAT) used as 7mer size marker; lanes 6 and 10: PAL12 d-(CTTGAGCTCAAG) used as 12mer duplex marker; lanes 7: PAL20 d-(CGCGCGCGCGCGCGCGCGCG) used as 20mer duplex marker; lane 8: HPT21, d-(TGTACTGGGGTCCCCAGAAGC); lane 9: HPT11, d-(TGGGGTCCCCA).

containing 100 mM NaCl and 0.1 mM EDTA by mixing equimolar concentrations of 11mer as well as 21mer oligonucleotides with their counterparts at elevated temperature, followed by slow cooling. Such duplex samples were run on a non-denaturing gel at room temperature (∼25°C). The gel was run under constant voltage until the desired separation was achieved. The electrophoretogram of the two duplexes under study is shown in Figure 1a and b.

To our surprise, the gel mobility pattern obtained for various mixtures of oligomers, did not differ much from the ones obtained with individual strands under identical conditions (25). The duplexes HP(C+G)11 and HP(C+G)21 at 10 μM strand concentration (Figure 1, lanes 2 and 4, respectively) and even the individual HPC11, HPC21 (lanes 1 and 3) and HPT21, HPT11 strands (lanes 8 and 9) exhibited two distinct bands. To predict the molecularity of the structural species formed by the oligomers under study, three control strands, a heptamer marker (SC; lanes 5 and 11), a 12 bp (PAL12; lanes 6 and 10) and 20 bp (PAL20; lane 7) palindromic sequences, are used as size markers. In Figure 1, the lower band of HPC11 (lane 1) and HPT11 (lane 9) migrate equivalent to the SC marker (lane 5) indicating migration of HPC11 and HPT11 as folded species (hairpin form), as has been shown for HPG11 and HPA11 (25), while the upper band of HPC11 migrates corresponding to PAL12 (lane 6) and can be assumed a 11 bp bulge duplex. It signifies that individual strands of the short 11mer perfect duplexes are capable of existing in hairpin and bulge duplex forms, independently. Thus it seems conceivable to assume that of the two bands appeared in case of duplex HP(C+G)11 sample (lane 2), the lower band corresponds to the folded (hairpin) form of individual strands, while the bulge duplex and perfect duplex species having identical size, will occupy the slow moving upper

band. The possible structures of hairpin and bulge duplex forms are depicted in Figure 5a.

Similarly, of the two bands exhibited by HPC21 (Figure 1a, lane 3) and HPT21 (Figure 1b, lane 8) where the core HPC11 sequence was flanked by naturally occurring 5 nt non-complementary stretch on both the 5′- and 3′-sides, the lower band moving equivalent to PAL12 (Figure 1; lane 6) corresponds to the folded form of 21mer oligonucleotide, as the hairpin form of HPC21 and HPT21 with unpaired flanking ends. Electrophoretic mobility of single-stranded DNA becomes almost the same as that of half-length double-stranded DNA when it makes a hairpin structure (26) and hairpins migrate faster than their corresponding unstructured single-strands (32,33). The upper slow migrating band of HPC21 with equivalent mobility to the 20 bp palindromic duplex (PAL20, lane 7), is interpreted in terms of a bulge duplex species of HPC21 and HPT21 containing the unpaired flanking sequences (Figure 5b). It seems reasonable to mention here that an identical behavior was displayed by the HPG21 strand, whose highly retarded upper band was regarded due to an extended bulge/cruciform like structure. Based on the observed status of HPG21 studied previously (25) and of HP(C+G)21 in the present study under similar experimental conditions, one can anticipate that the inherent tendency of forming stable secondary structures by each complementary strand may interfere in the formation of the perfect duplexes. Interestingly, such an assumption is reflected in the gel pattern, where a 1:1 mixture HP(C+G)21, of the two complementary strands i.e. HPG21 and HPC21 exhibited two bands (Figure 1; lane 4), the fast moving lower band corresponds to the hairpin form of the HPC21 and HPG21 with unpaired flanking ends and the slow moving upper band possibly might correspond to the perfect duplex or a compact cruciform structure containing two hairpins on each G and C strands (Figure 5b). Interestingly the gel mobility patterns similar to HPC21 discussed above, was also displayed by HPT21 (Figure 1b, lane 8) which further confirms the formation of two structural species by this strand. Thus it can be concluded that the component strands of the short 11mer and long 21mer perfect duplexes, are capable of existing in hairpin and bulge duplex forms, independently (Figure 5).

Worth mentioning is the difference observed in the mobility of upper bands of duplex HP(C+G)21 (Figure 1a, lane 4) and individual strand HPC21 (lane 3). The bulge duplex or cruciform like structure formed by HPC21 oligonucleotide has two non-complementary 5 nt flanking ends, while the possible cruciform structure generated in the duplex [HP(C+G)21] contains two hairpins and paired flanking duplex region, producing overall a compact cruciform structure (Figure 5b). Apparently this compact form of [HP(C+G)21] moves faster than the bulge duplex form of HPC21, giving a different mobility pattern in the gel. Importantly at this stage, the possibility of formation of perfect linear duplex due to Watson–Crick pairing of HPC21 and HPG21 strands cannot be ruled out.

The native PAGE experiments were also run in cold room (8–10°C). Except a longer electrophoresis run time, no difference was found with the results of electrophoretic mobility pattern of the same oligomers, run at room temperature (∼25°C).

More than a decade ago, McMurray's study (27) on human enkephalin gene has demonstrated reversible conformational transition from a 23 bp duplex containing the enhancer, to two individual hairpin structures, each formed from one strand of the duplex. A structural model was suggested describing that DNA secondary structure within the enhancer region plays an active role in cAMP-inducible activation via the formation of cruciform structures. Later study (34) using structural methods concluded that each oligonucleotide strand exists primarily as a hairpin structure over a wide range of oligomer concentration and temperature. Role of cruciform structure in transcription regulation of enkephalin gene was discussed.
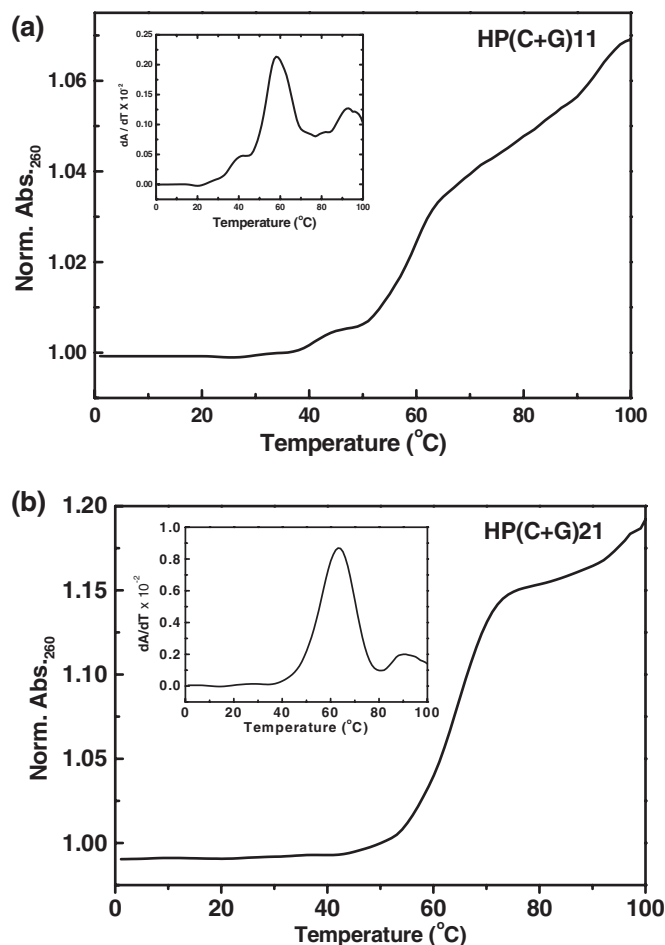
Reports on structural polymorphism including duplex–tetraplex equilibrium (35,36), structural competition between the G-quadruplex, i-motif and Watson–Crick duplex (37) again indicate the importance of elucidation how alternative DNA structures with biological implications form.

In the following section a correlation between the gel studies and UV-thermal denaturation profiles further establishes the presence of various structural possibilities of the studied sequences.

UV-thermal denaturation experiments

The absorbance versus temperature melting profile shown in Figure 2a for the HP(C+G)11 duplex at 1 μM concentration, containing equimolar quantities of HPC11 and its counter part 11mer, HPG11 did not display the characteristic monophasic sigmoidal curve, generally expected for a perfect duplex. Surprisingly, a triphasic melting profile with three transitions was obtained. The upper transition could not attain a proper plateau. The derivative plot of the melting profile shown as inset shows a major peak depicting a T_m of 58°C for the middle transition, while the T_m for upper transition was ∼90°C. A small hump on the derivative plot centered around 43°C cannot be ignored as it seem to be originated from the lower portion of the melting profile. The occurrence of triphasic melting profiles can be interpreted as presence of more than one structural species in the solution. Thus it can be assumed that equimolar mixture of HPC11 and HPG11 also generated structures, other than a perfect duplex species. The interpretation of UV-melting profile present in Figure 2a can be simplified by considering the status of individual quasipalindromic strands. In our previous study (25), HPG11 has been shown to exist in hairpin-duplex equilibrium under identical solution conditions used in present study. Thus the lower, middle and higher temperature transitions correspond to the melting of bulge duplex, perfect duplex and hairpin forms, respectively.
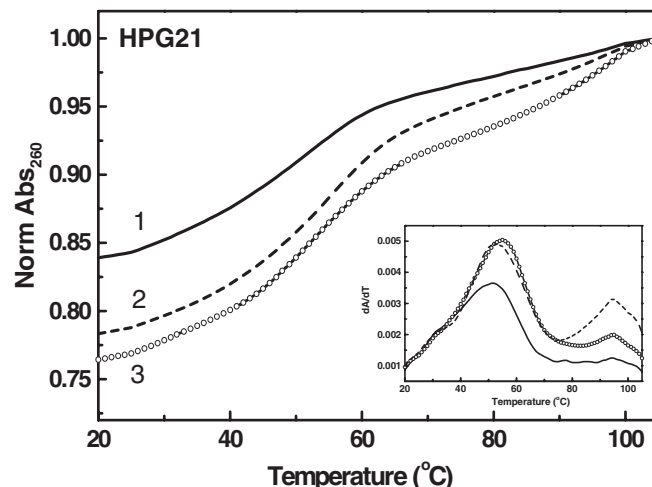
Further, to study the effect of the flanking sequences on the hybridization capabilities of complementary quasipalindromic sequence, thermal denaturation experiments were performed on HP(C+G)21, containing equimolar concentrations of complementary HPC21 and HPG21 strands. HPC21/HPG21 are the 5′- and 3′- 5 nt extensions of HPC11/HPG11. To our surprise, the UV-melting profile of HP(C+G)21 (2 μM) shown in Figure 2b, also displayed a pattern, moderately identical to the melting of HP(C+G)11. The two transitions corresponding to two peaks in the derivative plot (inset) indicated the presence of two structural species

Figure 2. (**a**) Thermal denaturation profile of d-TGGGGCCCCCA (HPC11) + d-TGGGGGCCCCA (HPG11) in 20 mM sodium cacodylate buffer (pH 7.4), 0.1 mM EDTA, containing 100 mM NaCl at 260 nm, [Inset shows the derivative plot of curve] (**b**) Thermal denaturation profiles of d-TGTAC-TGGGGCCCCCAAGAGC (HPC21) + d-GCTCTTGGGGGCCCCAGTACA (HPG21) in 20 mM sodium cacodylate buffer (pH 7.4), 0.1 mM EDTA, containing 100 mM NaCl at 260 nm, [Inset shows the derivative plot of curve].

Figure 3. Thermal denaturation profile of d-GCTCTTGGGGGCCCCCAG-TACA (HPG21) at 10 µM (1), 30 µM (2) and 40 µM (3) strand concentrations in 20 mM sodium cacodylate buffer (pH 7.4), 0.1 mM EDTA, containing 100 mM NaCl at 260 nm, [Inset shows the derivative plot of curve].

and can be interpreted on somewhat similar logics, as was made for HP(C+G)11. On that account, since the sequence HPG21 under identical solution conditions has already been shown to exist in equilibrium with hairpin and bulge duplex forms indicating T_m values of >90 and 40°C for the two forms, respectively (25), the higher temperature transition ($T_m \sim 90$°C) is consistent with the melting of intramolecular hairpin forms of individual HPG21 and HPC21 strands. The lower temperature transition showing higher hyperchromicity than the upper transition, might correspond to the melting of intermolecular bulge duplexes formed by HPC21 and duplexes formed by perfect pairing of HPC21 with HPG21. However, possibilities of other structures cannot be ruled out at this stage.

It is important to mention that the said species could also include the structures generated by association of two hairpin conformers via Watson–Crick base pairing between the flanking nucleotides of complementary strands, finally shaping into cruciforms. Compared to bulge duplexes (with flanking ends) formed by individual HPG21 and HPC21 strands,

cruciform structures [formed by HP(C+G)21], which engage both the complementary strands, are compact structures (such possibilities are displayed in Figure 5b). Though the calculated T_m of 67°C for melting of discrete structures manifested in single broad melting profile, cannot be assigned definitively to one structural form. The differential gel pattern shown by HPC21 and HP(C+G)21 (Figure 1a, lanes 3 and 4) however, reflects in the possibility of compact cruciform structure formed by HPC21 and HPG21 strands. This observation gave us a clue to rationalize that the lower temperature broad transition ($T_m = 67$°C) in the melting curve of HP(C+G)21 most likely corresponds to the disordering of cruciform species. The possibility of involvement of a cruciform intermediate in hairpin-duplex interconversion via a cruciform intermediate has been well suggested (26,38).

Considering the overall interpretation of the thermal melting profiles of HP(C+G)11 and HP(C+G)21 (Figure 2a and b), it can very well be concluded that hybridization of DNA oligonucleotide to their complementary counterparts is complicated by the presence of secondary structures. Presence of palindromic region though intervened by single base in HPG11/HPG21 and HPC11/HPC21 can still facilitate the formation of intra- and intermolecular Watson–Crick base pairing between the complementary ends resulting into hairpin/bulge duplex/bulge duplex with flanking ends or cruciform structures.

For the information on the molecularity of both the structural species detected in native PAGE, a dependence of oligomer concentration of HPG21 sequence on the T_m was carried out at 100 mM NaCl concentration. The melting profiles along with their derivatives (inset) shown in Figure 3, are distinctly biphasic at oligomer concentrations from 10 to 40 µM, again suggesting melting of the two ordered forms. Since both, the lower and higher temperature transitions of the biphasic curves were sufficiently separated, it was possible to extract two T_m's for two ordered forms. The actual T_m's were determined from the first derivative of the observed thermal transition. The biphasic curve (Figure 3) obtained

for HPG21 at 10 µM concentration corresponds to the lower temperature (T_m 50°C) and higher temperature transition (T_m 93°C), whereas at 40 µM concentration, the oligonucleotide exhibited melting at 57 and 93°C for the lower and higher transitions, respectively. As expected for bimolecular (inter-molecularly folded) structures adopted by HPG21, its first transition showed oligomer concentration dependence on T_m, whereas the second thermal transition remains relatively independent of oligomer concentration, revealing the presence of monomolecular (intramolecular duplex) species. It concludes that like HPG11, the oligomer HPG21 also exists in hairpin-duplex equilibria, and it is due to this property that its hybridization with HPC21 is hindered.

More than a decade ago Hirao *et al.* (39,40) in a study on short (heptamers and octamers) DNA fragments reported the formation of extraordinarily stable mini-hairpin structures containing one or three residue loops with the melting temperature as high as 76°C in 100 mM NaCl. The origin of such an unusual stability was revealed by solving the 3D structure of the d-GCGAAGC mini-hairpin by NMR (40). The sequence was found to be folded back on itself between A_4 and A_5 and that all the sugars were in C2′-endo conformation. This compact molecule is stabilized by regular extensive base stacking interactions within each B-form helical strand of $G_1C_2G_3A_4$ and $A_5G_6C_7$ and by two G-C base pairs and one G_3-A_5 base pair (two hydrogen bonds) leaving the single base A_4 in the loop region. These highly stable hairpins also show high resistance to nucleases (40). Nuclease-dependent degradation has suggested that HPG11 is more resistant towards S1 nuclease than HPA11, reflecting the stability of -G- hairpin over -A- hairpin (25).

The generic base sequences of the oligomers used in our study predict that since they are not fully self-complementary there exists a non-self-complementary residue (A or G) in the middle of the sequence intervening the complementary parts of the sequence. Of course the complementary ends of sequence can still interact and may form duplexes either intramolecularly (monomeric hairpin) or intermolecularly (bimolecular bulge duplex). The basis of formation of the bulge duplex structure is the quasipalindromic nature of the oligonucleotide. Since both sides of the central base (A or G) are self-complementary, the resulting intramolecular forms (hairpins) can have a single nucleotide loop and 5 bp stem (one AT and four GC base pairs).

Further, a report from Fermandjian's group (41) on the formation of a DNA hairpin with a single residue loop (closed by a Watson–Crick G-C base pair) supports our prediction of existence of a single residue loop in mini-hairpins. NMR studies on d-AGCTTATC-ATC-GATAAGCT (-ATC-) encompassing the strongest topoisomerase II cleavage site in pBR322 DNA showed that the oligomer exists in hairpin form with a sheared A-C base pair to close the single base T loop. Further structural study using NMR, native PAGE, UV-melting, CD and restrained molecular dynamics on the -GAC- analog of the same oligomer reveal that -GAC- adopts a hairpin structure folded through a single residue loop. The residue A in the loop is closed by Watson–Crick type hydrogen bonds between G and C, however the base pair is not found planar but rather adopts a wedge-shaped geometry with the two bases stacked on top of each other in the minor groove. We predicted a similar situation in our studies
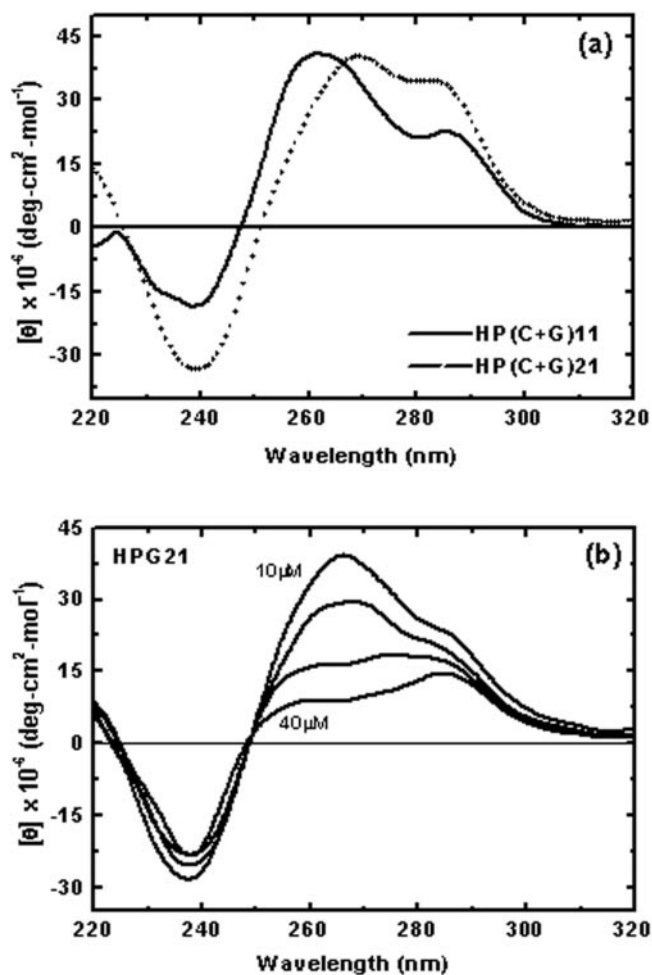
on hairpin formation by HPA11 and HPG11 quasipalindromic oligonucleotides and their extended versions. The oligonucleotide either folds at the 5′- and 3′-sides of the central intervening residue or may be closed by a distorted Watson–Crick G.C base pair, which is further stabilized by extensive base stacking interactions along the 5′- and 3′-sides of the strands. For certain DNA hairpin loops a C.G closing base pair provides enhanced stability (42). Varani (43) has reviewed the structural, functional and thermodynamic aspects of exceptionally stable DNA and RNA hairpins. One of the classes of DNA and RNA hairpins containing teraloops of the GNRA family [sequence G- any nucleotide (N)- purine (R)-A] are highly stable due to the contacts of G and R not to each other but to the phosphates across the loop, and extensive base stacking interactions along the stem. Thus the actual loop is only a single residue (N). The loops are also stabilized by non-Watson–Crick base pairs and base-sugar contacts.

Since, at this stage, we could not look into the structural details of the hairpin or bulge duplex species of the oligonucleotides used in this study, possibility of more than one residue in the hairpin loop cannot be ruled out.

It is crucial here to refer to our earlier observation that none of the oligonucleotides under study showed the formation of G/C-quadruplexes like multistranded structures (25). This could be due to the fact that both the complementary -GGGG- and -CCCC- stretches separated by a base, are positioned on the same strand, leading them a scope of intramolecular or intermolecular base paired structures.

Circular dichroism experiments

Further, CD spectroscopy, known to be extremely sensitive to small changes in mutual orientation of neighboring bases in an ordered or disordered duplex DNA, was used for the secondary structure analysis of the duplexes formed by the quasipalindromic complementary strands and their extended versions. CD spectra of the duplexes HP(C+G)11 and HP(C+G)21, prepared by mixing equimolar concentration (5 µM each) of respective strands in 20 mM sodium cacodylate buffer (pH 7.4) containing 100 mM NaCl and 0.1 mM EDTA are displayed in Figure 4a. Both the spectra are characterized by two prominent overlapping positive bands, which are usually not observed simultaneously for any canonical forms of DNA (44). It is interesting to note that the CD spectrum displayed here by the duplex HP(C+G)11 was found identical to the CD spectrum originated from HPG11 strand reported in our previous study (25). It is characterized by a strong negative band at 240 nm, crossover from negative to positive ellipticity at 247 nm, followed by two strong overlapping positive peaks near 265 and 285 nm. A careful look at the spectra in Figure 4a reveals that the structure formed by HP(C+G)11 show characteristics of both A- and B-type conformations. The spectrum displayed by the duplex HP(C+G)21 containing flanking sequences differs from HP(C+G)11 only at the shifted negative to positive cross over at 251 nm and positive CD peak with a 5 nm shift at 270 nm. Remarkably the second positive band at longer wavelength, belonging to both the duplex sequences occupies the same position at 285 nm. The amplitude of the positive peaks at 265 and 270 nm belonging to

Figure 4. (**a**) CD spectra of d-TGGGGCCCCCA (HPC11) and d-TGGGG-GCCCCA (HPG11) and d-TGTACTGGGGGCCCCCAAGAGC (HPC21) + d-GCTCTTGGGGGCCCCAG TACA (HPG21) in 20 mM sodium cacodylate buffer (pH 7.4) containing 100 mM NaCl and 0.1 mM EDTA at strand concentrations (5 μM each). (**b**) CD spectra of d-GCTCTTGGGGGCCC-CAGTACA (HPG21) in 20 mM sodium cacodylate buffer (pH 7.4) containing 100 mM NaCl and 0.1 mM EDTA at indicated strand concentrations.

HP(C+G)11 and HP(C+G)21 duplex, respectively is almost equal, while it differs substantially at the longer wavelength positive band at 285 nm. The magnitude of the CD change at 285 nm positive band is associated with the sequence context of HP(C+G)11 and HP(C+G)21 duplexes and can be explained on the basis of few recent elegant reports, on the contribution of sequence to the CD profiles of A- and B-form of DNA (45–49).

Coincidently, one of the studied sequences by Lindquist *et al.* (49) is a 10mer self-complementary oligonucleotide (TGGGGCCCCA) differing from our sequence only by 1 nt at the central position [i.e. TGGGG (A/G) CCCCA]. Notably this one base difference is not reflected in the CD spectra of the reported duplex and the duplex under study. Another independent study on similar lines (48) using the self-complementary d-GGGGCCCC sequence has concluded that the unusual spectra with simultaneous presence of A- and B-DNA type positive bands, contains features of A-like stacking of G-bases, and B-like stacking pattern for cytosines.

Since we know that the sequences with guanine tracts forming parallel stranded tetraplexes are characterized by a positive CD signal at ∼260 nm (50), CD signal at 260–265 nm can be attributed to guanine–guanine stacking (46). Duplex sequences [HP(C+G)11, HP(C+G)21] studied by us contain a tract of four guanines, which gave positive CD signal at 265 nm. Following similar explanations given for sequences containing G-tract and C-tracts (45,51), it can be argued that the prefect duplexes formed by the quasipalin-dromic sequences under study have both A-type and B-type base pair segments. The effect of flanking sequences on the CD spectrum of -GGGGCCCC- tract has been reported (49). When the sequence is flanked by 5′-T/A or 5′-A/T, their CD showed a shoulder around 288 nm and a positive peak around 261–262 nm. With 5′-CAT flanking, it shows a weak positive peak at 288 nm, while the intensity of 266–268 nm positive band is somewhat enhanced. Signifi-cantly, the CAT sequence which is framed by 5′-CAT but where central G-tract was replaced by -ATGCAT- showed the spectrum typical of B-form DNA with a negative peak at 250 nm followed by a positive peak at 270 nm.

Our 5′- and 3′-flanks being -GCTCT and -GTACA (in HPA21/HPG21) and -TGTAC and -AGAGC (in HPT21/HPC21), when incorporated in 11mer quasipalindrome core -TGGGG(A/T/G/C)CCCCA- resulting into 21mer sequence (HPA21, HPT21, HPC21 and HPG21) showed a substantial effect on CD spectra of non-flanked sequences (11mers). The 5 nm red shift (265→270 nm) at the positive band of HP(C+G)21 is the consequence of incorporation of flanking sequence but it is still shown to retain A-like CD features (44). Watson–Crick base pairing between the complementary por-tions of HPC21 and HPG21 flanks may generate a small duplex segment. B-DNA stem contributes to the usual maxi-mum at 285 nm (47). This maximum is more pronounced in case of HP(C+G)21 than HP(C+G)11, which concludes that the enhanced magnitude of 285 nm CD band is a result of increase in the 'B' character in HP(C+G)21.

The present study is a follow up of our previous report (25) where the secondary structure of G and A strands of the duplex HP(C+G)11/HP(A+T)11 were investigated and were shown to exist in hairpin-duplex equilibria. The most interest-ing observation was a clear correlation of this hairpin-duplex equilibrium with A → B transition of DNA. The hairpin and duplex conformations exhibited an oligomer concentration dependence generating two different CD profiles of which the hairpin form was interpreted as the A-form, while the bulge duplex form was featured as B-form. Thinking on simi-lar lines, we got interested to see whether the extended versions of duplex strands HPG21 and HPA21 also show oli-gomer concentration dependence on CD spectra as shown by their non-flanked version HPG11 and HPA11. For this pur-pose, CD spectra were recorded for the HPG21 sequence at varied strand concentrations (10–40 μM) (Figure 4b). Such CD profiles depicted a strong oligomer concentration depen-dence, likewise the prominent 265 nm positive peak at 10 μM, disappeared with successive increase of strand con-centrations upto 40 μM. The CD shoulder at 285 nm survived throughout the concentration range. Interestingly the spectra pass through an isoelliptic point at around 258 nm indicating more than one species in equilibrium. It is important to mention here that HPA21 also showed a similar strand

concentration dependence on CD (data not shown). This concludes that like the non-flanked sequences (HPG11/HPA11) their extended versions (HPG21/HPA21) also undergo intramolecular (hairpin) → intermolecular (duplex) structural transition fairly reflected in A → B transition.

The identical behavior shown by both the flanked and non-flanked duplex strands reveal a common structural feature present in HP(C+G)11/HP(C+G)21 sequences. At this stage it is important to mention the extensive work of other groups carried out on similar lines. Since our core sequence [TGGGG (A/T/C/G) CCCCA] of all duplexes share the same 5′- and 3′-terminal pentanucleotide stretches, with the sequence d-TGGGGCCCCA (49) and the -GGGG-/-CCCC-segments with d-GGGGCCCC (48) and d-CCCCGGGG (45,46), conformational similarities are anticipated. In all these reports the sequences have been the perfect palindromes (GGGGCCCC or CCCCGGGG). They were shown to form perfect duplexes with no signatures of hairpin or multi-stranded structures. Moreover the sequence d-TGGGGCC-CCA did not show any oligomer concentration dependence on CD spectra (49). Thus it can be said that seemingly the quasipalindromic nature of our sequences created a space to exist in hairpin-duplex equilibrium. A-form DNA is less flexible than B-DNA and this rigidity is proposed to be of biological significance, since replication can occur at higher fidelity due to the stiffness of A-form structure (52).

A study from Dickerson group (53) revealed unusual conformation for the dodecamer duplex formed by CATGGGC-CCATG. It lies on a structural continuum along the transition between A- and B-DNA. The structure is an intermediate state of A → B helix transition. All sequences in nucleic acids database containing three or more GpG base steps have been crystallized in A-form (53). The crystal structure of CCCCGGGG possesses similar structural feature to that of GGGGCCCC (54). Thus as expected on the basis of above discussion, our quasipalindromes TGGGG(A/G) CCCCT containing three GpG steps should also adopt geometries closer to A- than B-DNA.

The X-ray study on induction of A → B transition in hexamer duplex d(CCCGGG) by nogalamycin appeared little interesting to us as, the DNA sequence used contains some elements common to our studied sequence (55).

Based on the above mention experimental results, it can be said that since the individual strands of HP(C+G)11 and HP(A+T)11 can adopt intramolecularly folded and intermolecular linear duplex structures, their equimolar mixtures do not only hybridize to produce perfect duplexes but also species of hairpin, bulge duplex conformers.

In case of HP(C+G)21/HP(A+T)21 there rests a fair possibility that both the predicted forms (hairpin and duplex) are present simultaneously, thus formation of an extended bulge duplex or compact cruciform structure cannot be ruled out. Such a possibility is displayed in Figure 5b. It is clear that presence of flanking regions on 21mer sequences after hybridization with their complementary counterparts will generate 5 bp duplex regions on 5′- and 3′-termini of the strands, while the 11 nt core sequences will form the two hairpins (5 bp stem and single base loop) positioned opposite to each other. Such a simultaneous formation of two hairpins and paired flanking duplex regions shape into a cruciform structure. The possibility of such a four junction structure

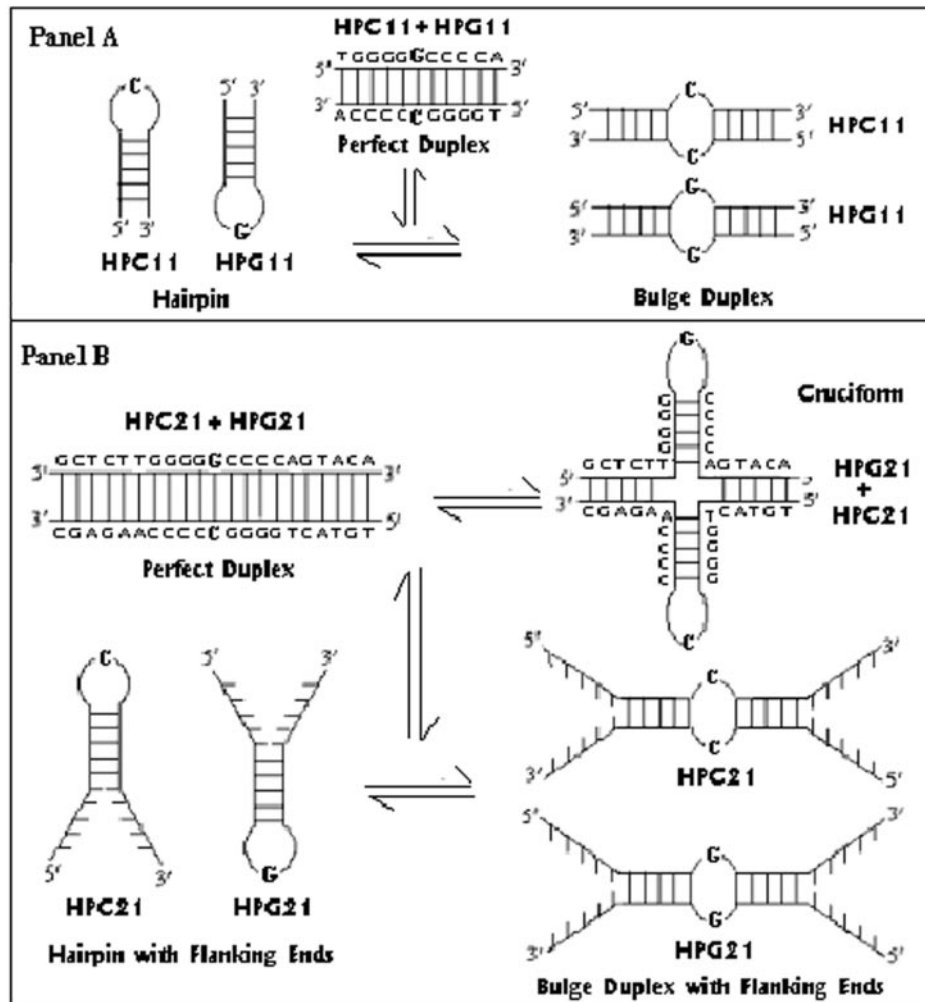in HP(C+G)11/HP(A+T)11 is abridged due to lack of flanking nucleotides.

BIOLOGICAL IMPLICATIONS

DNA sequences are now recognized to be structural determinants modulating the biological activity of genes. Perhaps the most important aspect of DNA structural variations is likely to be found in the mechanics of molecular recognition and manipulation by proteins. Our study presented here, imply that duplex formation at an imperfect inverted repeat sequence of HS4 of the β-globin LCR is restricted by the self-structuration of the relevant complements. As a result, hybridization of 11-base long quasipalindromic sequences generated not only the perfect (hetero-) duplexes but also the imperfect bulge (homo-) duplexes and hairpins via inter- and intramolecular Watson–Crick base pairing, while hybridization products of their extended versions 21mer flanking sequences were found to be in equilibrium with the perfect duplexes/bulge duplexes/cruciforms or hairpin species. The hybridized duplex segments also demonstrated the presence of interconvertible A- and B-DNA structural elements. Correlation of the intramolecular (hairpin) structure with the A-form and intermolecular (duplex) structure with B-form of DNA was concluded by CD signatures. Taken together these observations point out towards an important structural polymorphism, which could be biologically relevant. The conformational switching occurred in an oligomer concentration dependent manner might be considered as a structural motif to act as one of the regulatory elements in LCR. The human β-globin gene LCR, a dominant regulator of globin gene expression, is a contiguous piece of DNA with five tissue-specific DNase I HSs. Since the regions of HSs have a high density of transcription factor binding sites, structural interdependencies between HSs and different promoters may directly or indirectly regulate LCR functions. The DNA sequences of a SNP site studied here may contribute to form stable hairpin or cruciform structures at HS4 region of LCR suggesting that such structural polymorphism may cause variations in the DNase hypersensitivity.

Literature is rich in studies showing that DNA is a dynamic molecule whose structure depends on the underlying nucleotide sequence and is influenced by the environment and the overall DNA topology. Important are the dynamic alterations in the structure of double helix including the generation and removal of non B-structures. Such structural polymorphism may be thought to regulate transcriptional activity in eukaryotic nucleus (56,57).

B-form is the major conformation of physiological DNA, while the A-form is the major conformation of RNA and can exist also for DNA under special conditions. The switch between A- and B-forms with a change in the structural geometries (folded unimolecular → linear biomolecular) is expected to have a spectacular role in the ability of the DNA segment to interact with proteins or other ligands. Implication of such structural transition has recently been identified (58). A minor groove-binding tract (MGBT) structural element of HIV-1 transcriptase is important for both replication frameshift fidelity and processivity. Interestingly the MGBT interactions occur in the DNA minor groove, where

Figure 5. Various structural possibilities adopted by HP sequences with their complements (**a**) Sequences HPC11 + HPG11; (**b**) Sequences HPC21 + HPG21.

the DNA undergoes a structural transition from A-form to B-form DNA. Crystal structure of HIV-1 reverse transcripts complexed with double-stranded DNA also revealed that the template-primer has A-form and B-form regions separated by a significant bend 45° (59). The binding site of the nuclear single-stranded binding factor (NssBF) located in the LTR of the *Drosophila* 1731 retrotransposon mainly adopts two hairpin structures differing loop size, in slow equilibrium at pH 6.0. Two transcription factors bind only to the coding strand within the whole retrotransposon, suggesting that its structural flexibility could be associated with transcription (60). Transition from the B- to A-form of DNA is essential for biological functions as shown by the existence of A-form in many protein–DNA complexes (61).

Our proposed model for the cruciform structure formed by HP(C+G)21 duplex is composed of two hairpins and a four arm junction (Figure 5b). This single structure as discussed above exhibits segments of A- and B-DNA, such a possibility seem to be a mere consequence of nucleotide sequence (48,49,62,63). The formation of cruciforms is mostly favored in DNA sequences with inverted repeat symmetry, producing a discontinuity in regular DNA structure and therefore

increasing the free energy of DNA molecules. However, cruciform extrusion relaxes superhelical strain, lowering the free energy of negative supercoiling (2). The frequency of occurrence of strong cruciform forming sequences has been reported in yeast and humans (64). Most recently Potaman and colleagues have reported (65) specific binding of Poly(ADP-ribose) Polymerase-1 (PARP-1) to cruciform hairpins. Interestingly, this nuclear protein differs from other cruciform-binding proteins by binding to hairpin tips rather than to four way junctions. The same study also indicated that PARP-1 can interact with the gene regulatory sequences by binding to promoter-localized cruciforms.

Observation of cruciform DNA harbouring palindromic sequences and homologous duplex interaction to form four way junctions strengthens further the Gierer's hypothesis proposing that DNA may form branching structures analogous to t-RNA (66). Our comprehension of the role of hairpins, bulge duplexes and cruciform structural features formed at the SNP site in HS4 of the β-globin gene LCR has yet to be revealed. The discoveries described above give rise to a wealth of implications for future investigations relevant to sequence-specific structural heterogeneity within Watson–Crick base paired double helical molecules.

CONCLUSIONS

This paper in general concludes that the genomic quasipalindromic sequences containing G- and C-tracts may produce a discontinuity in regular DNA structure. These regions may fold back on themselves to form hairpins or cruciform structures (comprising of two hairpins). To sum up, the results of gel, UV-melting and CD analysis suggest that due to the propensity of oligonucleotides under study to acquire stable secondary structures other than duplex forms, their hybridization properties are restricted up to some extent. This accounts for the HP(C+G)11 and HP(A+T)11 sequences (expected to be perfect duplexes) to exist in equilibrium with hairpin, bulge duplex and perfect duplex forms. The hairpin forms display A- type CD spectra, while bulge duplexes are found to adopt B-form DNA. The CD spectra displayed (Figure 4a and b) contained both type of CD signatures. Presence of flanking sequences gives a possibility to HP(C+G)21/HP(A+T)21 to exist as a mixture of cruciform / bulge duplex and hairpin forms. Since, hairpins and bulge duplex forms represent A- and B-DNA form, respectively; cruciforms (comprising of two hairpins and duplex regions) render simultaneous presence of A- and B-type structural elements. To the best of our knowledge this is the first report where the possibility of a cruciform structure encompassing the elements of A- as well as B-DNA has been revealed, at a biologically relevant genomic site.

PERSPECTIVES

The usefulness of the work done in the paper, in a broad sense is information based. Structural polymorphism (hairpins, bulge duplexes, cruciforms etc.) and geometrical switching of DNA (A-form $\rightarrow$ B-form) exhibited by quasipalindromic regions within LCR of β-globin gene, highlight a careful understanding of the sequence dependent variations of the DNA structure. The knowledge of sequence-specific structural heterogeneity within Watson–Crick base paired double helical molecules might shed light on the mechanisms involved in transcriptional controlling of gene expression.

Cruciform DNA structures harbouring imperfect G- and C-tract palindromic sequences with simultaneous presence of A-and B-type DNA segments, could be used as a target for structure specific peptides / ligands or may serve as novel targets for pharmaceutical research. A protein could initially recognize a particular sequence from the shape of the DNA it binds to.

It is important to mention here that though both A- and B-forms of DNA have been largely studied by X-ray and NMR methods, the reasons for the preference for one conformation over the other are still unclear. The microscopic mechanism for the A $\rightarrow$ B conversion seem more difficult to understand in a situation like ours, where the A-form is only detected below 50 μM DNA concentration. Structural elucidation for the A-form, detectable only at micro-molar concentrations seems to be a difficult task and a challenging one indeed, for the structural biologists. We believe that such findings emphasize the importance of careful understanding of the molecular switching and a better analysis of the sequence dependent variations of the DNA structure. Our studies should endeavour to uncover structural details of nucleic acid structural transitions occurring at physiological conditions.

REFERENCES

1. Saenger,W. (1984) Principles of nucleic acid structure. Springer-Verlag, NY.
2. Sinden,R.R. (1994) DNA structure and function. Academic press, San Diego.
3. Nowak,R. (1994) Mining treasures from 'Junk DNA'. *Science*, **263**, 608–610.
4. Lilley,D.M. (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc. Natl Acad. Sci. USA*, **77**, 6468–6472.
5. Panayotatos,N. and Wells,R.D. (1981) Cruciform structures in supercoiled DNA. *Nature*, **289**, 466–470.
6. Van Holde,K. and Zlatanova,J. (1994) Unusual DNA structures, chromatin and transcription. *Bioessays*, **16**, 59–68.
7. Engberg,J. and Klenow,H. (1977) Palindromic arrangements of specific genes in lower eukaryotes. *Trends Biochem Sci.*, **2**, 183–185.
8. Weber,I.T., McKay,D.B. and Steitz,T.A. (1982) Two helix DNA binding motif of CAP found in lac repressor and gal repressor. *Nucleic Acids Res.*, **10**, 5085–5102.
9. Boulikas,T. (1996) Common structural features of replication origins in all life forms. *J. Cell Biochem.*, **60**, 297–316.
10. Sierzputowska-Gracz,H., McKenzie,R.A. and Theil,E.C. (1995) The importance of a single G in the hairpin loop of the iron responsive element (IRE) in ferritin mRNA for structure: an NMR spectroscopy study. *Nucleic Acids Res.*, **23**, 146–153.
11. Williams,A.S. and Marzluff,W.F. (1995) The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res.*, **23**, 654–662.
12. Aslani,A., Simonsson,S. and Elias,P. (2000) A novel conformation of the Herpes Simplex virus origin of DNA replication recognized by the origin binding protein. *J. Biol. Chem.*, **275**, 5880–5887.
13. Froelich-Ammon,S.J., Gale,K.C. and Osheroff,N. (1994) Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage. *J. Biol. Chem.*, **269**, 7719–7725.
14. Amir-Aslani,A., Mauffret,O., Bittoun,P., Sourgen,F., Monnot,M., Lescot,E. and Fermandjian,S. (1995) Hairpins in a DNA site for topoisomerase II studied by 1H- and 31P-NMR. *Nucleic Acids Res.*, **23**, 3850–3857.
15. Van Noort,V., Worning,P., Ussery,D.W., Roche,W.A. and Sinden,R.R. (2003) Strand misalignments lead to quasipalindrome correction. *Trends Genet.*, **19**, 365–369.
16. Viswanathan,M., Lacirignola,J.J., Hurley,R.L. and Lovett,S.T. (2000) A novel mutational hotspot in a natural quasipalindrome in *Escherichia coli. J. Mol. Biol.*, **302**, 553–564.
17. Yoshiyama,K. and Maki,H. (2003) Spontaneous hotspot mutations resistant to mismatch correction in *Escherichia coli*: transcription-dependent mutagenesis involving template-switching mechanisms. *J. Mol. Biol.*, **327**, 7–18.
18. Bissler,J.J. (1998) DNA inverted repeats and human disease. *Front. Biosci.*, **3**, d408–d418.

19. Heidenfelder,B.L., Makhov,A.M. and Topa,M.D. (2003) Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J. Biol. Chem.*, **278**, 2425–2431.
20. Chou,S-H., Chin,K.-H. and Wang,A.H.-J. (2003) Unusual DNA duplex and hairpin motifs. *Nucleic Acids Res.*, **31**, 2461–2474.
21. Li,Q., Harju,S. and Peterson,K.R. (1999) Locus control regions: coming of age at a decade plus. *Trends Genet.*, **15**, 403–408.
22. Li,Q., Peterson,K.R., Fang,X. and Stamatoyannopoulos,G. (2002) Locus control regions. *Blood*, **100**, 3077–3086.
23. Boulikas,T. (1993) Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human beta-globin gene complex. *J. Cell. Biochem.*, **52**, 23–36.
24. Kukreti,R., B-Rao,C., Das,S.K., De,M., Talukder,G., Vaz,F., Verma,I.C. and Brahmachari,S.K. (2002) Study of the single nucleotide polymorphism (SNP) at the palindromic sequence of hypersensitive site (HS)$_4$ of the human β–globin locus control region (LCR) in Indian population. *Am. J. Hematol.*, **69**, 77–79.
25. Kaushik,M., Kukreti,R., Grover,D., Brahmachari,S.K. and Kukreti,S. (2003) Hairpin-duplex equilibrium reflected in the A → B transition in an undecamer quasipalindrome present in the locus control region of the human β-globin gene cluster. *Nucleic Acids Res.*, **31**, 6904–6915.
26. Xodo,E., Manzini,G., Quadrifoglio,F., Yathindra,N., van der Marel,G.A. and van Boom,J.H. (1989) A facile duplex-hairpin interconversion through a cruciform intermediate in a linear DNA fragment. *J. Mol. Biol.*, **205**, 777–781.
27. McMurry,C.T., Wilson,W.D. and Douglass,J.O. (1991) Hairpin formation within the enhancer region of the human enkephalin gene. *Proc. Natl Acad. Sci. USA*, **88**, 666–670.
28. Slama-Schwok,A., Brossalina,E., Demchenko,Y., Best-Belpomme,M. and Vlassov,V. (1998) Structural flexibility of a DNA hairpin located in the long terminal repeat of the *Drosophila* 1731 retrotransposon. *Nucleic Acids Res.*, **26**, 5142–5151.
29. Shiflett,P.R., Taylor-McCabe,K.J., Michalczyk,R., Silks,L.A. and Gupta,G. (2003) Structural studies on the hairpins at the 3′ untranslated region of an anthrax toxin gene. *Biochemistry*, **42**, 6078–6089.
30. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions.3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
31. Marky,L.A., Blumenfeld,K.S., Kozlowski,S. and Breslauer,K.J. (1983) Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d (CGCAATTCGCG): evidence for hairpin formation. *Biopolymers*, **22**, 1247–1257.
32. Mitchell,J.E., Newbury,S.F. and McClellan,J.A. (1995) Compact structures of d(CNG)n oligonucleotides in solution and their possible relevance to fragile X and related human genetic diseases. *Nucleic Acids Res*, **23**, 1876–1881.
33. Kejnovska,I., Tumova,M. and Vorlickova,M. (2001) (CGA)(4): parallel, anti-parallel, right-handed and left-handed homoduplexes of a trinucleotide repeat DNA. *Biochim. Biophys. Acta*, **1527**, 73–80.
34. McMurray,C.T., Juranic,N., Chandrasekaran,S., Macura,S., Li,Y., Jones,R.L. and Wilson,W.D. (1994) Hairpin formation within the human enkephalin enhancer region 2. Structural studies. *Biochem.*, **33**, 11960–11970.
35. Miura,T. and Thomas,G.J. Jr (1994) Structural polymorphism of telomere DNA: interquadruplex and duplex-quadruplex conversions probed by Raman spectroscopy. *Biochemistry*, **33**, 7848–7856.
36. Shiber,M.C., Braswell,E.H., Klump,H. and Fresco,J.R. (1996) Duplex-tetraplex equilibrium between a hairpin and two interacting hairpins of d(A-G)10 at neutral pH. *Nucleic Acids Res.*, **24**, 5004–5012.
37. Li,W., Miyoshi,D., Nakano,S. and Sugimoto,N. (2003) Structural competition involving G-quadruplex DNA and its complement. *Biochemistry*, **42**, 11736–11744.
38. Avizonis,D.Z. and Kearns,D.R. (1995) Kinetic and thermodynamic characterization of DNA duplex-hairpin interconversion for two DNA decamers: d(CAACGGGTTG) and d(CAACCCGTTG). *Biopolymers*, **35**, 187–200.
39. Hirao,I., Nishimura,Y., Tagawa,Y., Watanabe,K. and Niura,K. (1992) Extraordinarily stable minihairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res.*, **20**, 3891–3896.
40. Hirao,I., Kawai,G., Yoshizawa,S., Nishimura,Y., Ishido,Y., Watanabe,K. and Miura,K. (1994) Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an

41. El Amri,C., Mauffret,O., Monnot,M., Tevanian,G., Lescot,E., Porumb,H. and Fermandjian,S. (1999) A DNA hairpin with a single residue loop closed by a strongly distorted Watson–Crick G.C base pair. *J. Mol. Biol.*, **294**, 427–442.
42. Nakano,M., Moody,E.M., Liang,J. and Bevilacqua,P.C. (2002) Selection for thermodynamically stable DNA tetraloops using temperature gradient gel electrophoresis reveals four motifs: d(cGNNAg), d(cGNABg), d(cCNNGg), and d(gCNNGc). *Biochemistry*, **41**, 14281–14292.
43. Varani,G. (1995) Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 379–404.
44. Ivanov,V.I., Minchenkova,L.E., Schyolkina,A.K. and Poletayev,A.I. (1973) Different conformations of double stranded nucleic acids in solution as revealed by circular dichroism. *Biopolymers*, **12**, 89–110.
45. Trantirek,L., Stefl,R., Vorlickova,M., Koca,J., Skelenar,V. and Kypr,J. (2000) An A-type double helix of DNA having B-type puckering of the deoxyribose rings. *J. Mol. Biol.*, **297**, 907–922.
46. Kypr,J., Fialova,M., Chladkova,J., Tumova,M. and Vorlickova,M. (2001) Conserved guanine-guanine stacking in tetraplex and duplex DNA. *Eur. Biophys. J*, **30**, 555–558.
47. Kypr,J. and Vorlickova,M. (2002) Circular dichroism spectroscopy reveals invariant conformation of guanine runs in DNA. *Biopolymers*, **67**, 275–277.
48. Stefl,R., Trantirek,L., Vorlickova,M., Koca,J., Skelenar,V. and Kypr,J. (2001) A-like guanine-guanine stacking in the aqueous DNA duplex of d(GGGGCCCC). *J. Mol. Biol.*, **307**, 513–524.
49. Lindquist,M. and Graslund,A. (2001) An FTIR and CD study of the structural effects of G-tract length and sequence context on DNA conformation in solution. *J. Mol. Biol.*, **314**, 423–432.
50. Berova,N., Nakanishi,K. and Woody,R.W. (2000) *Circular Dichroism; Principles and Applications.* Wiley-VCH, NY, Cictiester, UK.
51. Gray,D.M. (1974) A circular dichroism study of poly dG, poly dC, and poly dG:dC. *Biopolymers*, **13**, 2087–2102.
52. Timisit,Y. (1999) DNA structure and polymerase fidelity. *J.Mol.Biol.*, **293**, 835–853.
53. Ng,H.L. and Dickerson,R.E. (2002) Mediation of the A/B-DNA helix transition by G-tracts in the crystal structure of duplex CATGGGCCCATG. *Nucleic Acids Res*, **30**, 4061–4067.
54. Haren,T.E., Shakked,Z., Wang,A.H.-J. and Rich,A. (1987) The crystal structure of d(CCCCGGGG): a new A-form variant with an extended backbone conformation. *J. Biomol. Struct. Dyn.*, **5**, 199–216.
55. Cruse,W.B., Saludjian,P., Leroux,Y., Leger,G., Manouni,D.E. and Prange,T. (1996) A continuous transition from A-DNA to B-DNA in the 1:1 complex between nogalamycin and the hexamer dCCCGGG. *J Biol. Chem.*, **271**, 15558–15567.
56. Wells,R.D. (1988) Unusual DNA structures. *J. Biol. Chem.*, **263**, 1995–1098.
57. Van Holde,K. and Zlatanova,J. (1994) Unusual DNA structures, chromatin and transcription. *Bioessays*, **16**, 59–68.
58. Latham,G.J., Forgacs,E., Beard,W.A., Prasad,R., Bebenek,K., Kunkel,T.A., Wilson,S.H. and Lloyd,R.S. (2000) Vertical-scanning mutagenesis of a critical tryptophan in the 'minor groove binding track' of HIV-1 reverse transcriptase. Major groove DNA adducts identify specific protein interactions in the minor groove. *J. Biol. Chem.*, **275**, 15025–15033.
59. Jacobo-Molina,A., Ding,J., Nanni,R.G., Clark,A.D.,Jr, , Lu,X., Tantillo,C., Williams,R.L., Kamer,G., Ferris,A.L., Clark,P. *et al.* (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0. A resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.
60. Slama-Schowk,A., Brossalina,E., Demchenko,Y., Best-Belpomme,M. and Vlassov,V. (1998) Structural flexibility of a DNA hairpin located in the long terminal repeat of the *Drosophila* 1731 retrotransposon. *Nucleic Acids Res.*, **26**, 5142–5151.
61. Lu,X.J., Shakked,Z. and Olson,W.K. (2000) A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.*, **300**, 819–840.
62. McCall,M., Brown,T. and Kennard,O. (1985) The crystal structure of d(GGGGCCCC). A model for poly(dG).poly(dC). *J. Mol. Biol.*, **183**, 385–396.

extraordinarily stable structure resistant to nucleases and heat. *Nucleic Acids Res.*, **22** (56), 576–582.

63. Gao,Y.G., Robinson,H. and Wang,A.H. (1999) High-resolution A-DNA crystal structures of d(AGGGGCCCT). An A-DNA model of poly(dG).poly(dC). *Eur. J. Biochem.*, **261**, 413–420.

64. Schroth,G.P. and Ho,S.H. (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res.*, **23**, 1977–1983.

65. Potaman,V.N., Shlyakhtenko,L.S., Oussatcheva,E.A., Lyubchenko,Y.L. and Soldatenkov,V.A. (2005) Specific binding of poly(ADP-ribose) polymerase-1 to cruciform hairpins. *J. Mol. Biol.*, **348**, 609–615.

66. Gierer,A. (1966) Model for DNA and proteins interactions and function of the operator. *Nature (London)*, **212**, 1480–1481.