4242–4249 Nucleic Acids Research, 1994, Vol. 22, No. 20

Tertiary base pair interactions in slipped loop-DNA: an NMR and model building study

Nikolai B.Ulyanov, Karl D.Bishop+, Valery I.Ivanov[1] and Thomas L.James*
Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA and [1]The Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 117984 Moscow B-334, Russia

Received June 8, 1994; Revised and Accepted August 15, 1994

ABSTRACT

Short direct repeat sequences are often found in regulatory regions of various genes; in some cases they display hypersensitivity to S1 nuclease cleavage in supercoiled plasmids. A non-standard DNA structure (Slipped Loop Structure, or SLS) has been proposed for these regions in order to explain the S1 cleavage data; the formation of this structure may be involved in the regulation of transcription. The structure can be generally classified as a particular type of pseudoknot. To date, no detailed stereochemical model has been developed. We have applied one-dimensional ^{1}H NMR spectroscopy to study a synthetic DNA, 55 nucleotides in length, which cannot fold as a standard hairpin but which may favor the SLS formation. AT base pairs were identified, consistent only with the formation of an additional, tertiary miniduplex in the SLS. An all-atom stereochemically sound model has been developed for the SLS with the use of conformational calculations. The model building studies have demonstrated that the tertiary miniduplex can be formed for one of the plausible SLS isomers, but not for the other.

INTRODUCTION

Deoxyribonucleic acids are known to be polymorphic; in addition to 'classical' A and B double helices, DNA molecules with special sequences can assume 'unusual' forms, such as Z-, H-forms, cruciforms, etc. (1,2). As a rule, such forms do not constitute the ground state (lowest energy) conformation, but they can be stabilized either by an appropriate solution environment (high ionic strength or low pH) or superhelical stress in circular covalently closed plasmids. Unusual DNA structures often include nucleotides not involved in base pairs, which can be detected by nucleases or chemical probes specific to single-stranded regions; for a review, see (3).

Short direct repeats are often found in eukaryotic and prokaryotic genomes, including regulatory regions. It has been reported for some of these sequences that a superhelical stress induces their reactivity to single strand-specific S1 nuclease [reviewed by Yagil (4)]. To explain the S1 cleavage data for the 5′-flanking region of the hsp70 gene in *Drosophila* and for the adenovirus late promoter, it has been suggested that two complementary strands of DNA can be shifted relative to each other producing two loops, one in each of the opposing strands (5,6). We will refer to such a structure as a Slipped Loop Structure (SLS), or Slipped Loop-DNA (figure 1). Furthermore, it has been noticed that the loops themselves have a potential to interact with each other due to the complementarity of their sequences (5). Ivanov and co-workers have proposed a three-dimensional scheme for such an interaction, according to which the slipped loops form a secondary miniduplex (7,8). For the experimental study of base pair interactions in the SLS, they designed a sequence (SLS-1, figure 2a) that can potentially form any of the SLS isomers (figures 2c–f) but not the usual hairpin structure. S1 nuclease and chemical modification studies showed that fragments 2, 4, 8, and 10 of the SLS-1 (see notations in figure 2) are protected from chemical modification and S1 cleavage, while fragments 3 and 9 are accessible to chemical reagents and S1 nuclease (7). Those findings were consistent with the formation of a miniduplex between the complementary sequences of the two loops of the SLS (figures 2e,f).

In the present study we have used NMR to demonstrate directly the formation of Watson–Crick pairs between adenine and thymine bases in fragments 2 and 8 of the SLS-DNA. For this purpose, the SLS-1 sequence was modified in such a way that AT pairs can be formed if, and only if, the fragments 2 and 8 form a miniduplex. The resulting sequence, SLS-2 (55 nucleotides) is shown in figure 2b. In addition, we have carried out molecular mechanics calculations of the SLS and showed that formation of the additional miniduplex is stereochemically possible for the SLS isomer I (figure 2e) but not for isomer II (figure 2f).

METHODOLOGY

Sample preparation

The oligodeoxyribonucleotide SLS-2 (figure 2b) was synthesized and purified by the Farminvestprom Company (Moscow, Russia),

*To whom correspondence should be addressed

+Present address: T_m Technologies, Woburn, MA 01801, USA

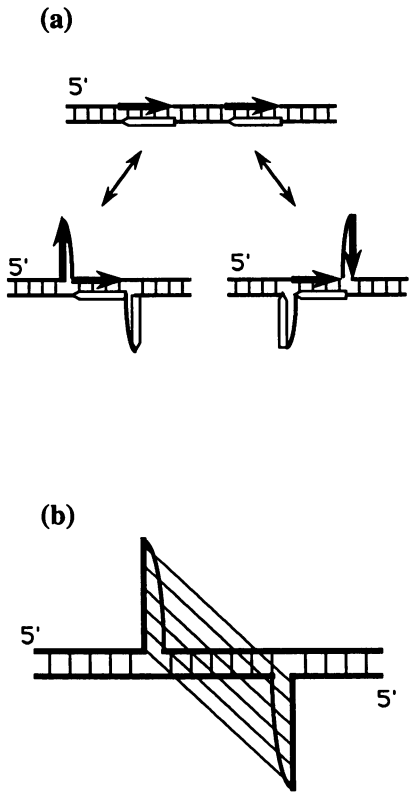

Figure 1. A scheme illustrating SLS formation in a DNA double helix containing direct repeats. **(a)** Two possible results of a 'slippage event'. Direct repeats are shown by filled and open arrows. **(b)** A secondary structure of interloop interaction in SLS. Tertiary base pairs are shown by slanted lines.

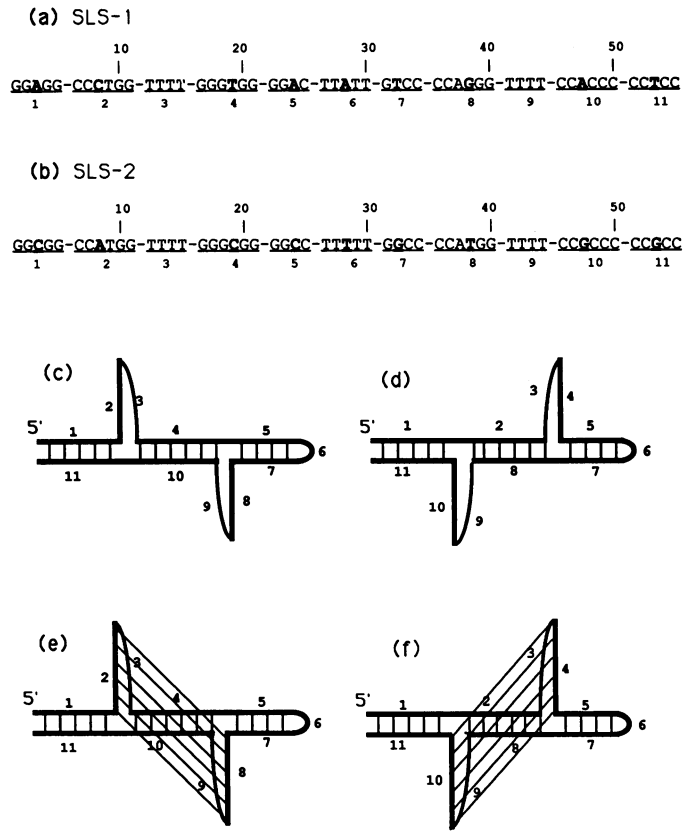

Figure 2. (a) SLS-1, an oligonucleotide sequence studied by Gorgoshidze *et al.* (7). **(b)** SLS- 2, the sequence used in the present study. Differences between the two sequences are shown in bold face. The numbers below the sequence define the fragment numbers which are used in (c–d) and throughout the text. **(c)** SLS isomer I with unpaired loops. **(d)** SLS isomer II with unpaired loops. **(e)** SLS isomer I with loops forming a tertiary mini-helix. **(f)** SLS isomer II with a tertiary mini-helix.

and used without additional purification. Approximately 2 mg of the sodium salt of the oligonucleotide was dissolved in 400 μl of a buffer containing 100 mM NaCl, 10 mM sodium phosphate, 1.0 mM sodium cacodylate, 0.2 mM EDTA and 10% by volume D_2O at pH 7.1.

NMR analysis

The one-dimensional, 500 MHz, proton NMR spectra were acquired on a General Electric GN500 NMR spectrometer using the jump-return method (9) for suppression of the H_2O signal with 2000 acquisitions at a sweep width of 16000 Hz and data size of 512 points. Acquisition time was 32 ms, and the relaxation delay was 0.5 s; the excitation profile had the excitation maximum near 10.7 ppm. A short homospoil pulse of 5 ms, applied just before the first 90° pulse of the pulse sequence, was used to further minimize the residual H_2O signal. Spectra were recorded at 5, 15, 25, 35, 45 and 55°C. These spectra were acquired in the order shown. Another series of spectra was acquired in the reverse order at the same temperatures. A delay of 10 min between the completion of one data set and the start of the next was used to allow for temperature equilibration.

NMR spectra were processed using the NMR1 software package from New Methods Research, Inc. (Syracuse, NY). DC offset correction and exponential line broadening factor of 10 Hz were applied to the FID prior to Fourier transformation without zero filling. The baseline of the imino portion of spectra was flattened with a third order polynomial. The chemical shifts were calibrated using the temperature-dependent H_2O signal (referenced to TSP).

Model building and molecular mechanics calculations

The SLS model building was carried out with the use of a modified version of the DNAminiCarlo program (10), Fitparam program (NBU, A.A.Gorin and V.B.Zhurkin, unpublished) on Sun SPARCstations in our laboratory, and the MIDASPlus program (11,12) on Silicon Graphics IRIS workstations in the Computer Graphics Laboratory at UCSF.

The DNAminiCarlo program performs conformational calculations (energy minimization and Metropolis Monte Carlo simulation) of DNA molecules, using internal coordinates as independent variables. The set of internal coordinates was designed specifically for nucleic acids and includes generalized helical parameters describing the relative position of bases in space (13); the aromatic bases are treated as rigid bodies with idealized geometry. A one-parameter model is used for flexible sugar rings; the backbone conformations are calculated by a chain-closure algorithm (14). The conformational energy of DNA structures is calculated *in vacuo* with a distance-dependent dielectric constant using empirical atom–atom potential functions (15,16). Because of the simple geometric meaning of the

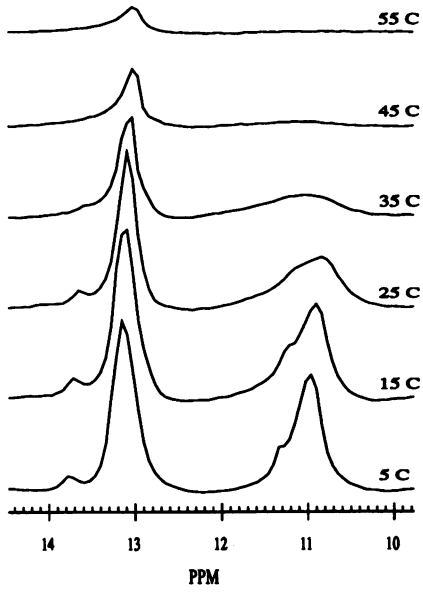

Figure 3. Temperature profile of imino proton NMR spectra of the SLS-2 oligonucleotide.

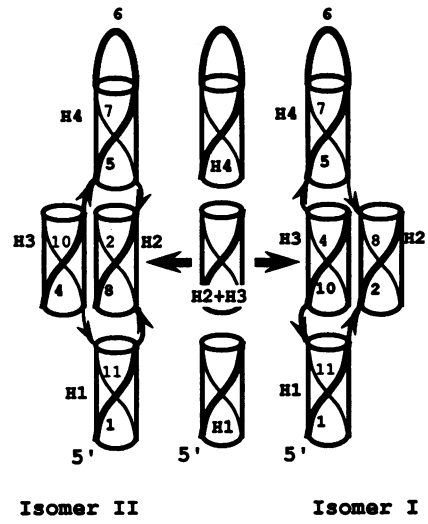

Figure 4. Strategy of the SLS model building (see text). Fragments 1 through 11 are defined in figure 2. Mini-helices H1 through H4 are defined as follows: H1 — 1•11, H2 — 2•8, H3 — 4•10, and H4 — 5•7.

independent variables, i.e., the helical parameters, the DNAminiCarlo program can be used as a modeling tool.

The MIDASPlus program was used for the interactive manipulation with separate domains of the SLS at several stages of model building, as described in detail in the 'Results' section. The Fitparam program was used to extract the internal coordinates of DNA structures produced by MIDASPlus. These internal coordinates unambiguously define the structure, and they were subsequently fed back into the DNAminiCarlo program.

Modeling was carried out for the SLS-2 sequence with deleted fragment 6 (figure 2); this fragment is supposed to have a standard loop conformation, and it does not interact with the essential components of the SLS. All molecular graphics illustrations were prepared with the MIDASPlus program. The atomic coordinates of the final model for the SLS isomer I are available from the authors upon request.

RESULTS

NMR experiments

The temperature profile of the imino proton region of the SLS-2 NMR spectra in H_2O is shown in figure 3. At and below room temperature, the spectra have three distinctive groups of resonances. One group resonates between 10 and 12 ppm, another large group of resonances is centered around 13.1 ppm, and a small peak at 13.7−13.8 ppm constitutes the third.

The tentative assignments of imino protons in these groups were based on their chemical shifts. Resonances between 10 and 12 ppm originate apparently from the imino protons of unpaired thymidines partially protected from exchange with bulk water protons (17). Imino protons in Watson−Crick AT pairs usually resonate down-field from those in GC base pairs; however, this rule is not exact. We have made an extensive search of published assignments of imino protons in DNA duplexes, including those with mismatches and various modifications. Among several hundred base pairs in 30 duplexes checked, the spread of the imino proton chemical shifts in AT pairs was from 13.1 to 14.2 ppm, while that in GC pairs was from 12.2 to 13.3 ppm (our unpublished data). Modifications and various destabilizations of GC base pairs shift the resonances of the guanine imino protons up-field, not down-field. These findings rule out the possibility of assigning the peak at 13.7−13.8 ppm (figure 3) to any of the GC pairs, because the closest chemical shift observed for the GC imino protons is up-field by 0.4 ppm. Therefore, this peak must originate from the AT base pair(s).

The large group of resonances around 13.1 ppm, clearly, must be assigned to the GC pairs of the SLS-2. One cannot exclude that this large peak includes an AT imino resonance as well, however, this is rather unlikely. Indeed, no more than two AT base pairs can be formed in the SLS-2 (see figure 2), and at least one of them was definitely assigned to 13.7−13.8 ppm. Furthermore, the formation of miniduplex 2•8 is a precondition of the existence of any AT base pairs in the SLS-2, and the sequence of the immediate surroundings of the AT pairs in this miniduplex — $(CCATGG)_2$ — possesses dyad symmetry, which suggests that the imino protons in both AT pairs must have nearly identical chemical shifts.

The broadening of peaks due to an increased exchange rate with water starts from 25°C (figure 3). First, unpaired thymidines and then the AT base pairs are affected. Inspection of figure 3 also shows that the melting is not completed yet at 55°C. These results are in accord with the UV melting curves of the SLS-2, which indicate that the melting starts from 20°C and occurs over a broad temperature range (data not shown).

The position of the imino proton resonances in the AT base pairs shifts up-field with the increase in temperature; that peak can be clearly seen up to 35°C. As described in the 'Methodology' section, two spectra were collected at each temperature, i.e., upon heating and upon cooling the sample. No hysteresis was evident in the chemical shifts of individual peaks. Typically, the difference in chemical shifts of the same peak in two experiments at the same temperature was less than 0.05 ppm;

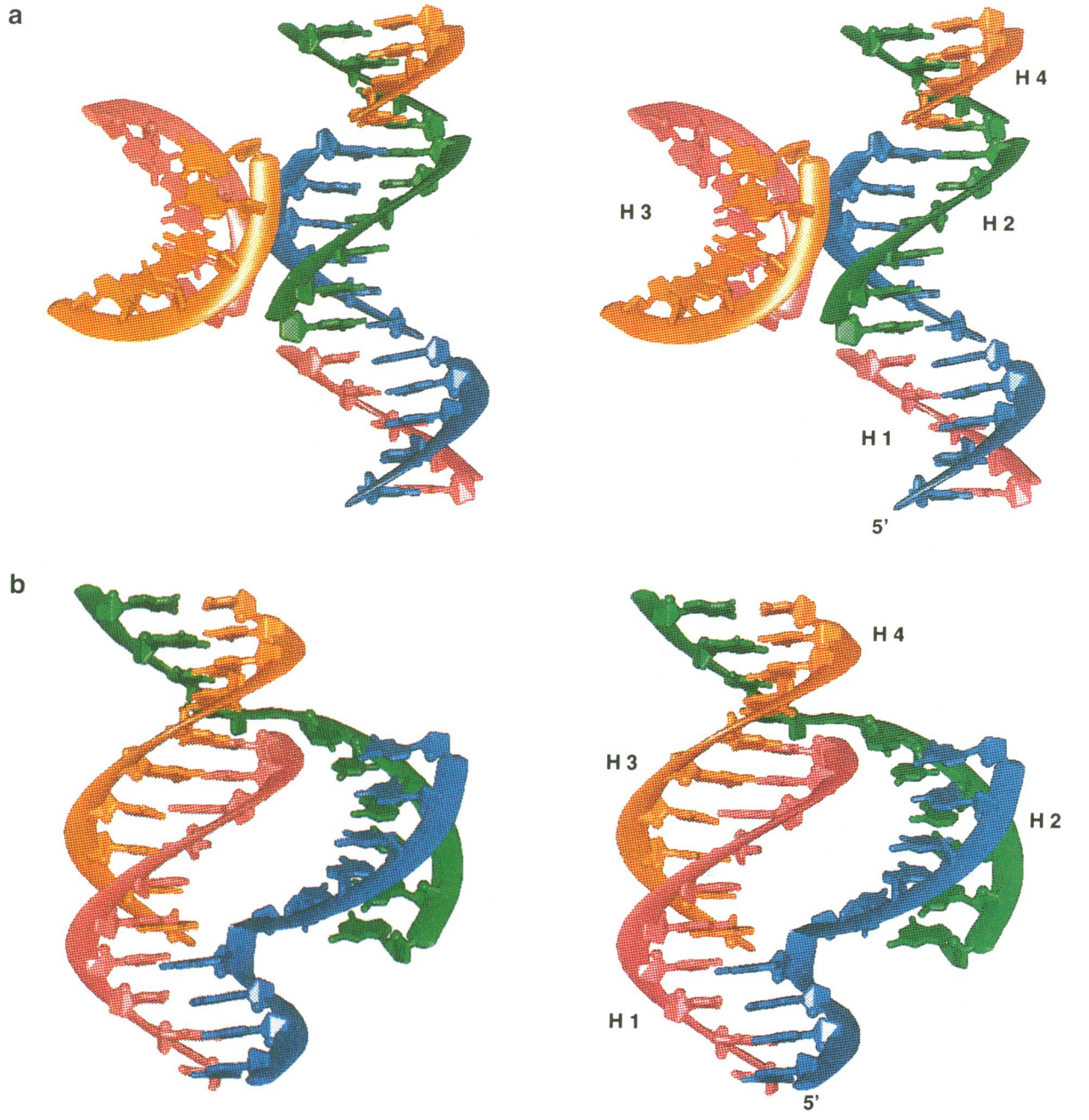

Figure 5. Stereoviews showing the results of intermediate steps in the SLS model building. Fragments 1 and 2 are colored blue, 4 and 5 orange, 7 and 8 green, 10 and 11 magenta. Fragment 6 was omitted from the model building; fragments 3 and 9 were not included in the model at this stage (see text). **(a)** SLS isomer II. A continuation of the model building for this isomer would require a connection of two yellow and two magenta-colored strands which is clearly not possible because the minor grooves of mini-helices H2 and H3 are facing each other. **(b)** SLS isomer I. Note that the major grooves of H2 and H3 are facing each other. Continuation of the model building for this isomer requires placing fragment 3 connecting the blue and orange strands and fragment 9 connecting the green and magenta strands (see figure 6).

the maximum difference, 0.1 ppm, was observed for the imino protons of unpaired thymidines at 25°C: 10.86 ppm in the spectrum collected during the heating of the sample and 10.76 ppm when cooling the sample. However, the line shape of the leftmost shoulder of this group of resonances displayed somewhat higher hysteresis (not shown). The ratio of the number of AT and GC base pairs in the SLS-2 was roughly estimated by integration of the area under the corresponding peaks with the use of a curve-fitting procedure in the data processing software NMR1. The spectra collected at 5, 15 and 25°C during the cooling of the sample were used for the integration, assuming that the DNA structure was better equilibrated after the annealing procedure. The resulting ratios were 1:16, 1:18 and 1:12 at 5, 15 and 25°C, respectively.

Model building

A strategy for SLS model building with the SLS-2 sequence is schematically illustrated in figure 4; it involved manipulations

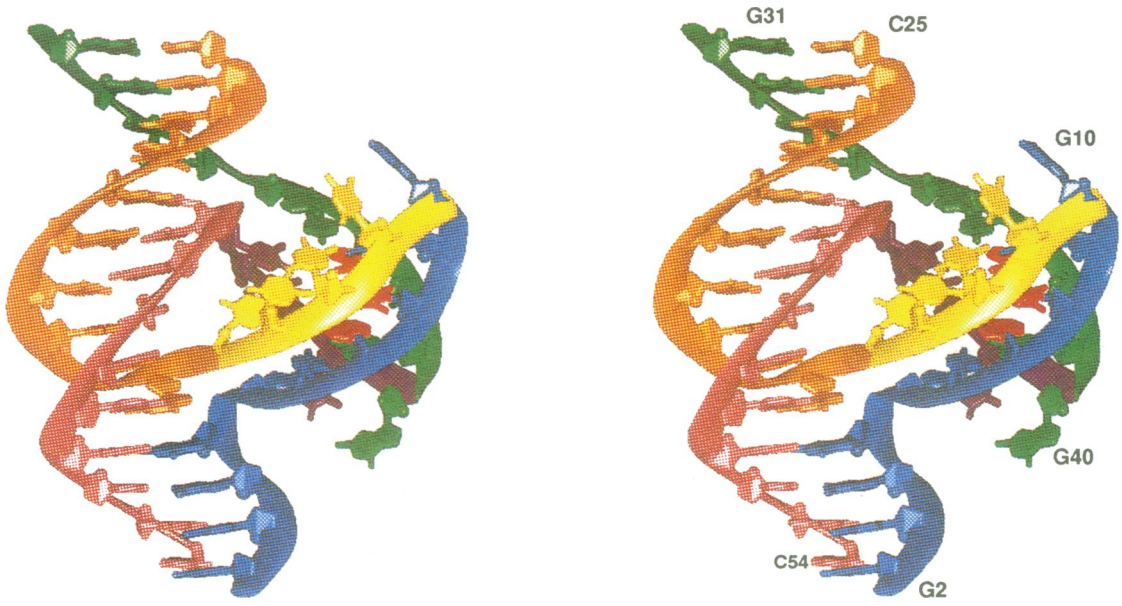

Figure 6. A stereopair of the final model of SLS isomer I with interloop pairing. Fragments 3 and 9 are shown in yellow and purple, respectively; coloring of the rest of the molecule is the same as in figure 5. Fragment 6 (nucleotides T26 through T30) was not included in the model (see text), and the terminal base pair G1•C55 was excluded from this picture. The only AT base pairs in this structure, i.e., A8•T38 and T9•A37, are shown in red. Note that the terminal GC pairs of the H2 mini-helix are disrupted because of the short length of connecting loops 3 and 9.

with several structural elements of the SLS: helices H1 (1•11), H2 (2•8), H3 (4•10), and H4 (5•7), and single-stranded loops 3 and 9; loop 6 was omitted from the model (the SLS-2 fragments 1 through 11 are defined in figure 2b). Each of the helices H1 through H4 is about half-turn of a DNA double helix (four to six base pairs). Our initial plan was to construct the SLS isomers I and II without the T_4 loops (3 and 9), and then add the loops to the model; we assumed that it should not be a major problem to introduce the single-stranded links once the core of the structure is built, provided, of course, that the links are of a sufficient length.

As the first step, we constructed a continuous 15-base pair helix H1+H2+H4, with strand (1+2+5) paired with (7+8+11), and helix H3 (4•10). Then, helix H3 was superimposed with H2, and double-stranded nicks were introduced between the helices H1 and H2 and between H2 and H4 (figure 4, middle). All helices were constructed with the DNAminiCarlo program using helical parameters optimized for the regular B-form of the poly(dG)•poly(dC) polymer with fixed helical twist of 34.5°. The superposition of H3 and H2 was perfect, because both helices had identical regular conformations.

To obtain an approximate geometry for the core of isomer I, all we need is to rotate helix H2 180° about the right edge of an imaginary cylinder representing this helix. As a result of such a rotation, helical elements H1, H3, and H4 will form continuous 1.5 double-helical turns, and H2 will be outside, parallel to H3 (figure 4, right). Likewise, to obtain the approximation for isomer II, we must rotate helix H3 180° about the left edge of its cylinder (figure 4, left). Now, to finish construction of isomers I and II (still without loops 3 and 9), we need to make proper ligations between the nicked strands. Namely, in the case of SLS isomer I, we must connect strands 1 and 2, 4 and 5, 7 and 8, and 10 and 11 (figure 4, right). Of them, the connections between fragments 4 and 5 and between 10 and 11 are trivial, because the elements H1, H3, H4 form a continuous helix, while the other two connections are not trivial. Similarly, in the case of isomer II, we have to ligate the same fragments but in a different structure: the connections 1 with 2 and 7 with 8 are trivial, and the other two are not (figure 4, left).

The rotations of helices H2 and H3 described above were carried out interactively with the MIDASPlus program. The result of the rotation which led to the isomer II configuration is shown in figure 5a. The trivially connected strands are shown in blue (fragments 1 + 2) and in green (7 + 8). To make the non-trivial connections, we have to ligate two magenta-colored strands (fragments 10 and 11) and two orange strands (4 and 5). It is evident in figure 5a that these connections are not possible without major disruptions of the H2 and/or H3 structures. The reason for this is that the minor grooves of the H2 and H3 miniduplexes face each other in the isomer II configuration, and any attempt to decrease the distance between H3 and continuous helix H1 + H2 + H4 would lead to steric clashes. Furthermore, any placement of the single-stranded T4 fragments 3 and 9 would cause additional problems for this isomer. Indeed, loop 3 must connect fragments 2 and 4 (upper blue and lower orange), and loop 9 must span a gap between fragments 8 and 10 (lower green and upper magenta, figure 5a). These loops cannot be accommodated in the space between the small minor grooves of the H2 and H3. Therefore, both loops must extend around the outside of H3. This would require loops 3 and 9 to be significantly larger. We can conclude that if SLS isomer II is to be formed, it will probably correspond to the scheme in figure 2d, and not 2f; in other words, formation of the SLS isomer II with the additional miniduplex H3 is stereochemically unlikely. We have not pursued calculations with this isomer any further.

The situation is quite different for the SLS isomer I. After rotation of miniduplex H2 (figure 4, right), the two helices, H2 and H3, have their major grooves facing each other (figure 5b);

this difference between isomers I and II was not immediately clear from the simple scheme shown in figure 4. For isomer I, the strands colored in magenta and orange have trivial connections, while the blue- and green-colored strands have non-trivial connections. Figure 5b shows isomer I (without loops 3 and 9) after the non-trivial ligations have been already made. This structure was calculated in the following way. After the rotation of the H2 helix using the MIDASPlus program, the resulting structure (not shown) was stored, and its internal coordinates (helical parameters, sugar pucker parameters and glycosidic angles) were calculated with the Fitparam program. These internal coordinates determine unambiguously the structure, and they were used as input for the DNAminiCarlo program. The first attempt to close the sugar–phosphate links between the H2 and H1 and H4 helices by energy minimization via DNAminiCarlo was unsuccessful. Consequently, the structure was subjected to the DNAminiCarlo energy minimization with a penalty added for gaps between strands 1 and 2 (blue in figure 5b) and between 7 and 8 (green). The minimization was carried out with respect to helical parameters defining the relative position of helices H1 through H4 and those defining the internal geometry of terminal base pairs involved in the junctions; the rest of the internal coordinates were kept fixed. After the nicks were ligated, the penalty was removed, and the structure was again energy-minimized; this time the internal geometry of the H2 and H3 helices was variable as well. Figure 5b presents the result of this energy minimization.

To finish building the model of SLS isomer I, we must add T_4 stretches between the blue and orange strands (loop 3) and between the green and magenta strands (loop 9). For this purpose, a strategy was chosen similar to that used to build the core of the structure. Different regular single-stranded geometries of the T_4 loops were constructed with DNAminiCarlo which were then interactively docked with the core (figure 5b) with the MIDASPlus. Both the relative position and internal geometry of the loops were variable during this fitting procedure, the internal geometry being changed via interactive varying of torsion angles using MIDASPlus. Several configurations were produced in this manner. For each of them, the internal coordinates were extracted with the Fitparam program. Each configuration's internal coordinates were fed into the DNAminiCarlo program, and a series of energy minimizations were carried out with a penalty added for gaps between the nicked strands. However, not one of these energy minimizations produced a proper chain closure; apparently, the four-nucleotide loops 3 and 9 were too short to span the required distances. Therefore, we decided to disrupt the terminal base pairs in the H2 helix to make the closures possible. Namely, the H-bonds in pairs G11•C35 and G40•C6 were broken. Again, initial adjustments in the structure were made interactively with MIDASPlus, by varying torsion angles in the G10–G11 and G39–G40 dinucleotide steps as well as glycosidic angles in the G11 and G40 residues. The resulting configurations were energy-minimized with the penalty added. This time, the energy minimization led to satisfactory closure of all nicks. As the last step, the SLS isomer I was energy-minimized without penalty; the final structure is shown in figure 6.

DISCUSSION

The very fact that we observed the AT Watson–Crick base pairs in SLS-2 at low and room temperatures implies that the fragments 2 and 8 are paired in the miniduplex H2. Indeed, the SLS-2 sequence contains only two adenines, both in the context of CCATGG (fragments 2 and 8), which can be potentially base paired. In fact, this is the only way how adenines can participate in AT base pairs, because all other thymines are situated inside T_n runs, which cannot pair with CCATGG. Therefore, the unpaired isomer I (figure 2c) can be ruled out. Furthermore, all unpaired isomers are inconsistent with the chemical and enzymatic cleavage experiments performed with the SLS-1 sequence (7); one can expect similar results for SLS-2 as well due to the very close similarity of sequences SLS-1 and SLS-2 (figures 2a,b). However, neither our NMR data, nor protection/modification data can distinguish between the paired SLS isomers I and II (figures 2e,f), because the two structures differ only in the global folding but not in the set of base pairs formed. Fortunately, the paired SLS isomer II (figure 2f) can be confidently excluded by model building studies (*vide supra*). It must be noted that ordinarily model building can be used only to demonstrate the possibility of a particular structure, but it is hard to prove that a structure does not exist by such a method (indeed, if one fails to build a model, it does not necessarily mean that it is impossible to build this model at all). However, in the present case, it is evident that any attempt to form the SLS isomer II would cause severe steric clashes between the H2 and H3 helices (figure 5a). In contrast, minor adjustments of the H2 and H3 miniduplexes were sufficient to construct isomer I (figure 6).

It is interesting that Tran-Dinh *et al.* (18) have carried out an NMR study of a DNA hexamer with a sequence coincident with that of the H2 helix — $d(CCATGG)_2$. The chemical shifts of the imino protons in AT pairs, including their temperature dependence, are exactly the same in this hexamer and in SLS-2 (13.78 ppm in the free hexamer and 13.81 ppm in SLS-2 at 5°C and ca. 13.6 ppm at 25°C; in both cases the resonance is unobservable above 35°C). This remarkable agreement between the two studies confirms our assignment of the AT pairs in the SLS-2; but also it suggests that the structure of the immediate surroundings of the AT pairs in the miniduplex H2 is very similar to that of the free hexamer, i.e., it is not significantly distorted in the SLS.

We have not been able to obtain direct NMR evidence for the formation of the H3 helix at the present stage, partially due to the limited quantity of sample presently in hand. Ideally, the ratio between the number of AT and GC pairs must discriminate between the unpaired isomer II (figure 2d) and any of paired isomers (figures 2e,f). The AT/GC ratio is expected to be 2:13 = 1:6.5 for the unpaired isomer II, and 2:19 = 1:9.5 for both paired isomers (or 2:17 = 1:8.5 if the terminal GC pairs in the H2 helix are broken; see figure 6). The experimental AT/GC ratio — 1:12 to 1:18, based on the ratio of T to G hydrogen bonded imino proton resonance intensities in fully relaxed spectra (*vide supra*) — is less than all the above estimates. One might explain the experimental AT/GC ratio assuming an equilibrium with the unpaired isomer I, which does not have any AT pairs (figure 2c); however, a definite conclusion cannot be drawn at this time. Indeed, the precision of the peak integration is not very high, especially in the case of the small AT peak, because of a rather large line width (figure 3). Besides, such an equilibrium appears to be unlikely in the light of the chemical modification and S1 nuclease-cleavage data for the SLS-1 sequence (7): even a small portion of molecules with unpaired nucleotides would be modified by chemical probes and cut by S1 nuclease. Taking into account all these considerations, the SLS isomer I with the

additional miniduplex H2 appears to be the most probable structure for the SLS-2 oligonucleotide.

According to our model of SLS (figure 6), the miniduplex H2 can be melted independently of the rest of the molecule. This is consistent with the imino proton melting profile (figure 3), which shows that AT pairs (located in the middle of the H2) disappear at about 35°C [consistent with melting of the free hexamer (18)], while some GC pairs still exist at 55°C.

The SLS-2 sequence designed for this study cannot form a fully paired duplex, so the SLS becomes a feasible conformation for this oligonucleotide. However, in the case of natural DNA with short direct repeats, the SLS is not the lowest-energy conformation, because its formation requires the disruption of about five base pairs (figure 1). Nevertheless, the SLS may be stabilized by supercoiling in circular covalently closed plasmids, because the formation of this structure reduces the effective twisting of DNA (figure 6). We note, however, that the superhelical density must be sufficiently high to make the SLS favorable. The degree of double helix unwinding due to SLS formation depends critically on the length of direct repeat and inter-repeat spacer. For the repeat of 6 nucleotides and inter-repeat spacer of 4 nucleotides, which are modeled by the SLS-2 sequence, the SLS extrusion is topologically equivalent to the melting of one turn of DNA. Evidently, longer spacers (modeled by the thymine loops, 3 and 9) can be easily accommodated in our SLS model (figure 6); however, additional studies are necessary to ascertain the allowed range of the length of direct repeats.

The 'slippage mechanism' for DNA was first proposed by Hentschel (19) in order to explain the S1 cleavage data for the histone gene repeat in sea urchin; the cutting sites were located inside the alternating $(GA)_n$ sequence. Ironically, these data are better explained by formation of another unusual DNA — the H-form of DNA, which must be stable under the acidic conditions required for the S1 cleavage experiment (20). Since that time, the formation of slipped loops in DNA with direct repeats has been repeatedly discussed as a possible mechanism of supercoiling-dependent regulation of transcription (5,6,21,22). The SLS might be implicated in this process because its formation effectively unwinds DNA, similar to other unusual DNA conformations; the unwinding changes the level of supercoiling, which, in its turn, affects the functioning of different promoters [reviewed by Cantor and co-workers (23)].

A characteristic feature of the SLS formed in DNA is the S1 nuclease cleavage or chemical modification of both strands of a spacer between short direct repeats, while the repeats themselves must be protected from cleavage and modification if the slipped loops formed a tertiary miniduplex (figure 1). To the best of our knowledge, the S1 cleavage data consistent with the interloop miniduplex in SLS include 5′-flanking sequences of *Drosophila* hsp70 gene (5), repeated Sp1 sites in SV40 (24), and adenovirus 2 major late promoter (6,25). At the same time, there are a number of examples of direct repeats where the S1 cleavage data are not consistent with SLS formation (4). There may be different explanations for this. First, we do not yet know the exact limitations for the length of direct repeats and a spacer between them. Our current model accounts for repeats spanning about a half-turn of DNA and a spacer not less than 5 nucleotides; other unusual structure(s), yet unknown, may be formed for longer direct repeats (26). Then, as already mentioned, the level of supercoiling must be sufficiently high for SLS extrusion; consequently the plasmid studied must be devoid of sequences which can potentially adopt another competing unusual conformation.

Initially, this study was inspired by the S1 cleavage data for DNA. However, the SLS may be formed for RNA as well. Furthermore, for RNA with special sequences, this structure may be the lowest-energy conformation. A number of natural RNA sequences have been found, which may allow SLS formation (7). In this respect, it is interesting that model building showed that SLS isomer I is stereochemically feasible for the A-DNA as well (V.A.Farutin and V.I.I., unpublished). Moreover, in this case the tetranucleotide loops 3 and 9 each appear to be of sufficient length to make all the proper connections without disruption of the terminal GC pairs in the H2 helix.

In the proposed SLS folding, nucleotides in the extruded loop pair with the complementary sequence which is not in this loop (figure 2). This structure can therefore be classified as a special case of pseudoknot structures (27). However, in contrast to a 'classical' pseudoknot, our pseudoknot is embedded inside a continuous double helix, hence the difference in the tertiary structure of the two types of folding: while two 'pseudoknotted' mini-helices are coaxial in the classical case (28), they have a parallel side-by-side orientation in our SLS model (figure 6).

Knowledge about the principles of SLS formation is necessary to examine the plausible involvement of this structure in the biological phenomena; our present work is the first step in characterizing such principles. Even though the present NMR data alone lack the necessary resolution to provide a high resolution structure for the SLS isomer I with the tertiary miniduplex (figure 2e) for the SLS-2 sequence, these data together with the model building study and published data on structural probing of a similar sequence give very strong evidence in favor of the SLS formation; also, they provide a working model for this non-standard conformation. Clearly, high resolution 2D NMR studies will help to clarify structural details of the SLS. The model oligonucleotide sequence for such a study will be designed with the present results in mind, in particular, the length of the unpaired loops will be increased.

ACKNOWLEDGEMENTS

We are grateful to Drs R.H.Sarma for the help with design of the SLS-2 sequence, E.E.Minyat for verifying the sequence and carrying out UV melting of the oligonucleotide, and V.B.Zhurkin for many useful discussions. We thank Dr C.C.Huang for very useful modifications in the 'ribbon-junior' option of the MIDASPlus program. This work was partially supported by NIH grant GM39247 (to T.L.J.) and by the Russian State Program 'Trends in Genetics' (grant to V.I.I.). We gratefully acknowledge use of the UCSF Computer Graphics Laboratory (supported by NIH Grant RR01081).

REFERENCES

1. Wells,R.D. and Harvey,S.C. (1988) Unusual DNA Structures. Springer-Verlag, New York.
2. Lilley,D.M.J. and Dahlberg,J.E. (1992) DNA Structures. Part A. Synthesis and Physical Analysis of DNA. *Meth. Enzymol.*, **211**. Academic Press, San Diego.
3. Lilley,D.M.J. and Dahlberg,J.E. (1992) DNA Structures. Part B. Chemical and Electrophoretic Analysis of DNA. *Meth. Enzymol.*, **212**. Academic Press, San Diego.
4. Yagil,G. (1991) *Crit. Rev. Biochem. Mol. Biol.*, **26**, 475–559.

5. Mace,H.A.F., Pelham,H.R.B. and Travers,A.A. (1983) *Nature*, **304**, 555–557.
6. Yu,Y.-T. and Manley,J.L. (1986) *Cell*, **45**, 743–751.
7. Gorgoshidze,M.Z., Minyat,E.E., Gorin,A.A., Demchuk,E.Ya., Farutin,V.A. and Ivanov,V.I. (1992) *Molekulyarnaya Biologiya (English translation)*, **26**, 832–838.
8. Ivanov,V., Gorgoshidze,M., Minyat,E. and Voloshin,O. (1991) J. *Biomol. Struct. Dynam.*, **8**, a088.
9. Gueron,M., Plateau,P. and Decorps,M. (1991) *Progress in Nuclear Magnetic Resonance Spectroscopy*, **23 (part 2)**, 135–209.
10. Ulyanov,N.B., Gorin,A.A. and Zhurkin,V.B. (1989) In Kartashev,L.P. and Kartashev,S.I. (eds) Proceedings Supercomputing'89: Supercomputer Applications. International Supercomputing Institute, Inc., St Petersburg, Florida, pp. 368–370.
11. Ferrin,T.E., Huang,C.C., Jarvis, L.E. and Langridge,R. (1988) *J. Mol. Graphics*, **6**, 13–27.
12. Huang,C.C., Pettersen,E.F., Klein,T.E., Ferrin,T.E. and Langridge,R. (1991) *J. Mol. Graphics*, **9**, 230–236.
13. Dickerson,R.E., Bansal,M., Calladine,C.R., Diekmann,S., Hunter,W.N., Kennard,O., Lavery,R., Nelson,H.J.C., Saenger,W., Shakked,Z., Sklenar,H., Soumpasis,D.M., Von Kitzing,E., Wang,A.H.-J. and Zhurkin,V.B. (1989) *EMBO J.*, **8**, 1–4.
14. Zhurkin,V.B., Lysov,Yu.P. and Ivanov,V.I. (1978) *Biopolymers*, **17**, 377–412.
15. Zhurkin,V.B., Poltev,V.I. and Florentiev,V.L. (1981) *Molekulyarnaya Biologiya (English translation)*, **14**, 882–895.
16. Poltev,V.I. and Shulyupina,N.V. (1986) *J. Biomol. Struct. Dynam.*, **3**, 739–765.
17. Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.), NMR of Macromolecules. A Practical Approach. Oxford University Press, New York, pp. 217–288.
18. Tran-Dinh,S., Neumann,J.-M., Huynh-Dinh,T., Allard,P., Lallemand,J.Y. and Igolen,J. (1982) *Nucleic Acids Res.*, **10**, 5319–5332.
19. Hentschel,C.C. (1982) *Nature*, **295**, 714–716.
20. Mirkin,S.M., Lyamichev,V.I., Drushlyak,K.N., Dobrynin,V.N., Filippov,S.A. and Frank-Kamenetskii,M.D. (1987) *Nature*, **330**, 495–497.
21. Wells,R.D., Amirhaeri,S., Blaho,J.A., Collier,D.A., Hanvey,J.C., Hsieh,W.-T., Jaworski,A., Klysik,J., Larson,J.E., McLean,M.J., Wohlrab,F. and Zacharias,W. (1988) In Wells,R.D. and Harvey,S.C. (eds), Unusual DNA Structures. Springer-Verlag, New York, pp. 1–21
22. Ozoline,O.N., Uteshev,T.A., Masulis,I.S. and Kamzolova,S.G. (1993) *Biochim. Biophys. Acta*, **1172**, 251–261.
23. Cantor,C.R., Bondopadhyay,S., Bramachari,S.K., Hui,C.-F., McClelland,M., Morse,R. and Smith,C.L. (1988) In Wells,R.D. and Harvey,S.C. (eds), Unusual DNA Structures. Springer-Verlag, New York, pp. 73–89.
24. Evans,T., Schon,E., Gora-Maslak,G., Patterson,J. and Efstratiadis,A. (1984) *Nucleic Acids Res.*, **12**, 8043–8058.
25. Kilpatrick,M.W., Torri,A., Kang,D.S., Engler,J.A. and Wells,R.D. (1986) *J. Biol. Chem.*, **261**, 11350–11354.
26. Wang,Z.-Y., Qiu,Q.-Q. and Deuel,T.F. (1994) *Biochem. Biophys. Res. Commun.*, **198**, 103–110.
27. ten Dam,E., Pleij,K. and Draper,D. (1992) *Biochemistry*, **31**, 11666–11676.
28. Puglisi,J.D., Wyatt,J.R. and Tinoco,I., Jr. (1990) *J. Mol. Biol.*, **214**, 437–453.