Monitoring Single-Stranded DNA Secondary Structure Formation by Determining the Topological State of DNA Catenanes

Xingguo Liang, Heiko Kuhn, and Maxim D. Frank-Kamenetskii
Center for Advanced Biotechnology and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

ABSTRACT Single-stranded DNA (ssDNA) has essential biological functions during DNA replication, recombination, repair, and transcription. The structure of ssDNA must be better understood to elucidate its functions. However, the available data are too limited to give a clear picture of ssDNA due to the extremely capricious structural features of ssDNA. In this study, by forming DNA catenanes and determining their topology (the linking number, Lk) through the electrophoretic analysis, we demonstrate that the studies of catenanes formed from two ssDNA molecules can yield valuable new information about the ssDNA secondary structure. We construct catenanes out of two short (60/70 nt) ssDNA molecules by enzymatic cyclization of linear oligodeoxynucleotides. The secondary structure formed between the two DNA circles determines the topology (the Lk value) of the constructed DNA catenane. Thus, formation of the secondary structure is experimentally monitored by observing the changes of linking number with sequences and conditions. We found that the secondary structure of ssDNA is much easier to form than expected: the two strands in an internal loop in the folded ssDNA structure prefer to braid around each other rather than stay separately forming a loop, and a duplex containing only mismatched basepairs can form under physiological conditions.

INTRODUCTION

Recently, much attention has been paid to single-stranded DNA (ssDNA), which is an essential intermediate in many biological processes that include replication, recombination, repair, transcription, and transposition of DNA (1,2). During these processes, the ssDNA part has to be in a proper conformation allowing or blocking the access of other nucleic acids and proteins that specifically recognize it. The physical characteristics of ssDNA determine its secondary structure and dynamics that are essential for its biological functions. In some cases, the folded ssDNA itself displays enzymatic activities in vitro such as cleavage and ligation of nucleic acids (3–6). Understanding the physical basis underlying the formation of specific ssDNA conformations will therefore significantly contribute to elucidating its biological functions. However, the available data are not sufficient to describe the extremely capricious conformations of ssDNA in solution, which are affected by electrostatic, basepairing, and stacking interactions. Thus, both experimental and theoretical approaches investigating ssDNA conformations are badly needed to be further developed. This is an important issue in an even wider context of single-stranded nucleic acid structures in general, because the secondary structure motifs of ssRNA and ssDNA share many common features. Therefore, the study of ssDNA secondary structures is significant for better understanding ssRNA secondary structures, which play a crucial role in the biological activities of RNA molecules as ribozymes, riboswitches, aptamers, etc. (7–9).

Considering ssDNA as an idealized flexible polymeric molecule without basepairing and stacking, the freely jointed chain model and worm-like chain model have been used to characterize long ssDNA (10,11). By using experimental approaches such as atomic force microscopy (12), transient electric birefringence (13), thermal melting profiles of DNA hairpins (14), and fluorescence resonance energy transfer (15), the persistence length of poly(dT) has been estimated to be in the range of 1.3–4 nm. However, it has been pointed out that ssDNA is usually composed of stacked domains interspaced with random-coil ones under physiological conditions (16). The strength of stacking interactions varies with the identity of bases, although the origin and physical structure of base stacking have not been well understood (17,18). For example, stacking has to be taken into account for ssDNA containing A-tracts for which a significant stacking effect has been observed. For heteropolynucleotides not containing stable secondary structures, stacking effects in single-stranded parts become noticeable at a temperature lower than 10°C (18). For poly(dT), it has been proved that stacking does not occur. In every case, stacking is insensitive to the concentration of salt (18).

Basepairing usually cannot be avoided for ssDNA with a mixed sequence containing all four nucleotides, and the local conformations of ssDNA such as hairpins, loops, and pseudoknots can form (11). To some extent, these secondary structures can be roughly predicted by dynamic programming algorithms, such as the *mfold* program, using database thermodynamic parameters (19). Sometimes, however, predictions are poor because of uncertainties in the thermodynamic parameters. For this reason, *mfold* outputs usually result in multiple suboptimal structures with similar free energy values, and a further analysis is needed to select the

Submitted September 7, 2005, and accepted for publication January 6, 2006.

Address reprint requests to Maxim D. Frank-Kamenetskii, E-mail: mfk@bu.edu.

Xingguo Liang's present address is Dept. of Molecular Design and Engineering, Graduate School of Engineering, Nagoya University, Chikusa, Nagoya 464-8603, Japan.

0006-3495/06/04/2877/13 $2.00

doi: 10.1529/biophysj.105.074104

correct structure. The prediction results can be greatly improved by incorporating into the computation the constraint parameters obtained from experimental data (20,21). However, the relevant thermodynamic data and structural parameters vary widely. New approaches are required and extensive research has to be done for understanding the local conformations of ssDNA.

Another important issue for understanding the ssDNA structure is the formation of a mismatched duplex containing consecutive mismatched basepairs (bps). DNA strands containing $(G_pA_q)_n$ (p, $q = 1-3$) are known to form either a parallel or antiparallel noncanonical duplex by self-association (22–27). These structures are of biological interest because the $(G_pA_q)_n$ sequences are abundant in genomes, including centromeres, where they seem to confer specific conformational properties on DNA (28). Recently, it has been found that tandem, so-called ''sheared'', mismatched bps (G·A, A·A, and A·C) are rather stable and are compatible with flanking Watson-Crick bps to form B-DNA-like duplex structures (29,30). At present, these structures cannot be predicted by the *mfold* program because no systematic thermodynamic data are available.

DNA normally exhibits a highly regular B-form structure when Watson-Crick bps are formed. B-DNA is a regular right-handed double helix (10.5 bp/turn), in which the two strands of DNA braid around each other. One can expect that the linking number between the two DNA strands should be determined solely by the duplex length. Based on this expectation, ssDNA catenanes with a certain linking number have been constructed by cyclizing two linear ssDNAs after hybridizing with each other (31–34). In this work, by forming DNA catenanes and determining their topology through gel electrophoresis, we obtain the structural information on how the DNA strands in the internal loop twist together in the presence of many consecutive mismatched bps. We find that the two strands of the internal loop prefer to braid around each other rather than stay separately forming a loop. Furthermore, the mismatched duplex (consisting of only mismatched bps) proved to be formed through constructing DNA catenanes with specific sequences and determining their linking number.

MATERIALS AND METHODS

Oligodeoxyribonucleotides and enzymes

Oligodeoxyribonucleotides (ODNs) were obtained from Integrated DNA Technologies (Coralville, IA). All ODNs longer than 40 nt were obtained in PAGE-purified form. Sequences of ODNs are given in Table 1. T4 DNA ligase, *Thermus aquaticus* (*Taq*) DNA ligase, restriction enzymes *Hha*I and *Rsa*I, and T7 exonuclease were supplied by New England Biolabs (Ipswich, MA). Restriction enzyme *Nmu*CI was from Fermentas (Hanover, MD), Exonuclease VII from USB (Cleveland, OH), and SYBR Green II from Molecular Probes (Eugene, OR).

Formation of DNA catenanes

Preparation of circular DNA

5′-Phosphorylated ODN (0.5 μM) of 60 or 70 nt was incubated with 1.0 μM splint ODN (60lig30, 70lig30, etc.) for 10 min at 37°C in 1X T4 DNA ligase reaction buffer (50 mM Tris-HCl (pH 7.5 at 25°C), 10 mM $MgCl_2$, 10 mM

TABLE 1 Sequences of ODNs used for constructing catenanes

ODNs	Sequence*
60A	5′-P-GACAATGATTTTTTTTTTTTT*tt*gacaccg**TACTTTAG**TTTTTTTTTTTTTTTACACAGT-3′
60B	5′-P-GACAATGAGATTATGAACTTCCAgacaccg**TACTTTAG**TTTTTTTTTTTTTTTACACAGT-3′
60C	5′-P-GACAATGAGATTATGAACTTCCAgacaccg**TACTTTAG**TTGCATATTATACTTACACAGT-3′
60D	5′-P-GACAATGATTTTTTTTTTTTTTTTTcaccg**TACTT**CATTTTTTTTTTTTTTTTACACAGT-3′
70A	5′-P-cggtgtTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTC**AAAGTA**-3′
70B	5′-P-cggtgtcTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**CTAAAGTA**-3′
70C	5′-P-cggtgtc*aa*CTCTAGATTGTTTTGTATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**CTAAAGTA**-3′
70D	5′-P-cggtgtc*aa*CTCTAGATTGTTTTGTATGTTTCAAGTAAGCGCAGGATTTGTTTGTATATTTT**CTAAAGTA**-3′
70E	5′-P-cggtgtcTGGAAGTTCATAATCTCATTGTCTTTTTTTTTTACTGTGTAAGTATAATATGCAA**CTAAAGTA**-3′
60lig30	5′-GTCTTTCTCATTGTCACTGTGTAAGAAAAA-3′
70Alig30	5′-AAAAAAAAAACACCGTACTTTGAAAAAAAAA-3′
70lig30	5′-TTGAACCAGACACCGTACTTTAGTTCCCAC-3′
60M1	5′-P-TACTTT AG T_{24} GAGGAGGAGGAGGAGGAGGAG GACCACG-3′
70M1	3′-ATGAAATC T_{34} AGGAGGAGGAGGAGGAGGAGG CTGGTGC-P-5′
60M2	5′-P-TACTTTAG T_{25} AGAGAGAGAGAGAGAGAGAG GACCACG-3′
70M2	3′-ATGAAATC T_{35} AGAGAGAGAGAGAGAGAGAG CTGGTGC-P-5′
60M3	5′-P-TACTTTAG T_{25} AGGAGAGGAGAGGAGAGGAG GACCACG-3′
70M3	3′-ATGAAATC T_{35} AGAGGAGAGGAGAGGAGAGG CTGGTGC-P-5′
60Mlig30	5′-TTTTTAACTAAAGTACGTGGTCCTCTTTTT-3′

*In ODNs 70A–E, the two termini are complementary to a central part of ODNs 60A–D. The 5′ termini of 70A–E are complementary to the left part in the middle of 60A–D (shown in lower case), and the 3′ termini of 70A–E are complementary to the right part in the middle of 60A–D (shown in bold face). In 60A and 70A, the complementary parts are underlined. The underlined parts in 60C and 70E are also complementary. The termini of 60M1, 60M2, or 60M3 are complementary to the termini of 70M1, 70M2, or 70M3, respectively. The sequences of $(GGA)_7$ in 60M1 and 70M1, $(GA)_{10}$ in 60M2 and 70M2, and $(GGAGA)_4$ in 60M3 and 70M3 are underlined. ''P'' at the 5′ terminus indicates the phosphate required for ligation. The following ODNs of 30 nt are used to cyclize ODNs of 60 nt or 70 nt: 60lig30 for 60A–D, 70Alig30 for 70A, 70lig30 for 70B–D; 60Mlig30 for 60M1–M3.

dithiothreitol (DTT), 25μg/ml bovine serum albumin, and 1.0 mM ATP). The reaction mixture (100 μL) was then cooled to 25°C, and 1.0 μL T4 DNA ligase (400 NEB units) was added. After incubation overnight at 25°C, the reaction mixture was extracted by phenol/chloroform/isoamylalcohol (25:24:1) and chloroform/isoamylalcohol (24:1), ethanol precipitated, and redissolved in 1X TE buffer (10 mM Tris-HCl (pH 7.4 at 25°C) containing 0.1 mM EDTA); ~80% yield of monocircularized DNA was obtained based on denaturing polyacrylamide gel electrophoretic analysis (Supplemental Fig. 1). The circular DNA was excised from the gel, eluted, extracted by phenol/chloroform/isoamylalcohol (25:24:1) and chloroform/isoamylalcohol (24:1), and ethanol precipitated. The concentration of the PAGE-purified circular DNA was determined with an estimated error of 30% in a denaturing polyacrylamide gel in comparison with known quantities of linear ODN with the same sequence (35).

Formation of DNA catenanes by method A

The mixture of a linear 70-nt-long ODN (0.2 μM) and a circular 60-nt-long ODN (0.2 μM) in 1X T4 DNA ligase reaction buffer (20 μL) was heated to 65°C and incubated for 5 min. The solution was cooled to either 10°C, 25°C, 37°C, or 45°C and kept for 10 min at the corresponding temperature before 0.5 μL T4 DNA ligase (200 units) was added. After incubation for 2 h, the reaction was stopped by inactivating the ligase at 65°C for 10 min. The reaction solution (5 μL) was then analyzed directly in a denaturing polyacrylamide gel. In the case of *Taq* DNA ligase, ligation was carried out at 45°C or higher temperature in 1X *Taq* DNA ligase reaction buffer (20 mM Tris-HCl (pH 7.6 at 25°C), 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1.0 mM nicotinamide adenine dinucleotide (NAD), 0.1% Triton X-100). If needed, linear molecules present after the ligation reaction were digested by treatment with a mixture of T7 exonuclease and Exonuclease VII for 1 h at 25°C and for 1 h at 37°C.

Formation of DNA catenanes by method B

The mixture of a circular 70-nt-long ODN (0.4 μM) and a linear 60-nt-long ODN (0.2 μM) in 1X T4 DNA ligase reaction buffer (20 μL) was heated to 65°C and incubated for 5 min. After cooling to 10°C and keeping for 10 min at 10°C, splint ODN (0.4 μM final concentration) was added and incubated for 3 min at 10°C. Subsequently, 0.5 μL T4 DNA ligase (200 units) were added and the ligation was carried out for 2 h at 10°C. After the ligation reaction, the ligase was inactivated by heating the sample for 10 min at 65°C.

Digestion of DNA catenanes by restriction enzymes

Three restriction enzymes, *NmuC*I (its recognition sequence, GT(G/C)AC, was incorporated once within the 60-nt-long circle), *Hha*I (its recognition sequence, GCGC, was incorporated once within the 70-nt-long circle made from ODN 70D), and *Rsa*I (its recognition sequence, GTAC, was incorporated in both 60- and 70-nt-long circles), were used. The individual species of DNA catenanes were gel purified, and the DNA catenanes (0.2 μM) were digested by incubation with 5 units of the corresponding restriction enzyme for 2 h at 37°C in 20 μL reaction buffer. In the case of *Hha*I and *NmuC*I, 30-nt-long ODNs (0.4 μM) were added to form a duplex (as substrate for digestion) with the complementary sequence in the catenane.

Gel electrophoresis

All reaction products (DNA circles, catenanes, and digestion fragments) were analyzed on 4%–12% (w/v) denaturing polyacrylamide gels (acrylamide/bis-acrylamide, 29:1, w/w). Unless stated otherwise, the gels were run at ambient temperature in 1X TBE (90 mM Tris (pH 8.0), 90 mM boric acid, 1 mM EDTA) containing 7.5 M urea. In some cases, 30% formamide was also added as a denaturing agent. DNA band patterns were visualized by SYBR Green II staining and detected by the charge-coupled device camera with IS-1000 digital imaging software (Alpha Innotech, San Leandro, CA).

Mfold modeling of secondary ssDNA structure

Mfold (version 3.1, online: http://www.bioinfo.rpi.edu/applications/mfold/old/dna/) was used for the modeling of secondary structure of ssDNA. Salt concentrations were set to 10 mM $MgCl_2$ and 10 mM NaCl. The folding temperature was set to 10°C or 37°C (19).

RESULTS

Formation of ssDNA catenanes and determination of their topology

Fig. 1 *A* shows two approaches for constructing the DNA catenanes we employ in this study. In a 60-nt-long ODN, two termini of which can hybridize with a short splint ODN and be ligated to form a circle, the central part is designed to be partially complementary to a 70-nt-long ODN. The two termini of the 70-nt-long ODN hybridize with the 60-nt-long ODN to form a duplex with a nick (see Table 1 for sequences). Thereafter, this duplex part is designated as a linking duplex (LD), and its length is designated as LD length. After both ODNs are closed and linked to each other, an ssDNA catenane is formed. The reason that two ODNs were designed to have different length (60 and 70 nt) is they can be easily distinguished by gel electrophoresis. In method A, the 60-nt-long ODN is first cyclized, then the linear 70-nt-long ODN hybridizes with it and the nick is sealed by DNA ligase. In method B, a central segment of the linear 60-nt-long ODN hybridizes with the premade circular 70-nt-long ODN before its two ends are ligated together on a short splint ODN (Fig. 1 *A*). Two factors were considered for designing the sequences of ODNs. One is the length of the complementary part between the 60-nt-long and the 70-nt-long ODN, which affects the linking number based on the periodicity of the B-DNA helix (10.5 bp/turn). The other factor is the sequence of the noncomplementary part, which forms various secondary structures depending on sequences and conditions. Regions of single-stranded poly(dT) have been shown to be very flexible and lacking any secondary structure in the absence of poly(dA) tracts (36). In the first series of experiments, we used ODNs such as 60A and 70A (Table 1), which contain complementary regions interspersed between poly(dT) regions.

Fig. 2, *A–C*, shows denaturing polyacrylamide gel electrophoresis patterns of the ligation products. The samples were loaded in the same order on 4%, 6%, and 11% gels for characterizing the synthesized DNA catenanes. Lanes 1 and 2 present the results when ODNs 60A and 70A were used to construct DNA catenane by method B or method A, respectively. The LD length between these two ODNs is 12 bp. In lane 1 of Fig. 2 *A* (6% gel), a new band with the same

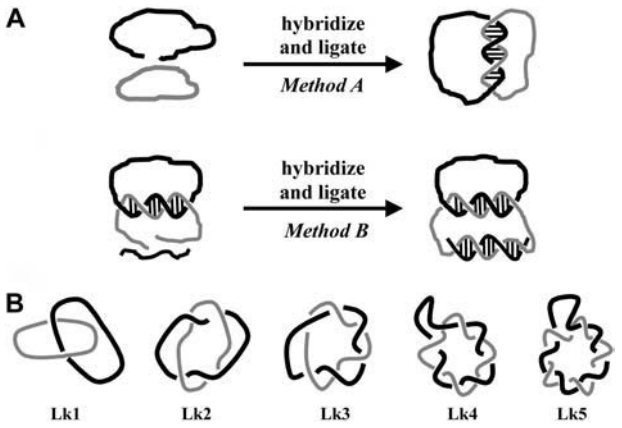

FIGURE 1 Formation and topology of DNA catenanes. (*A*) Two methods were employed in this study to form DNA catenanes. In both methods, a linear ODN is ligated on a circular one. In method A, the two termini of the linear ODN hybridize first to the circular ODN, then the nick is sealed by T4 DNA ligase. In method B, the central part of the linear ODN hybridizes with the circular ODN first, then two ends of the linear ODN are ligated and cyclized after hybridizing with a short splint ODN. The short splint ODN can be removed by exonuclease digestion or denaturing gel purification. (*B*) Topology of DNA catenanes with a different linking number.

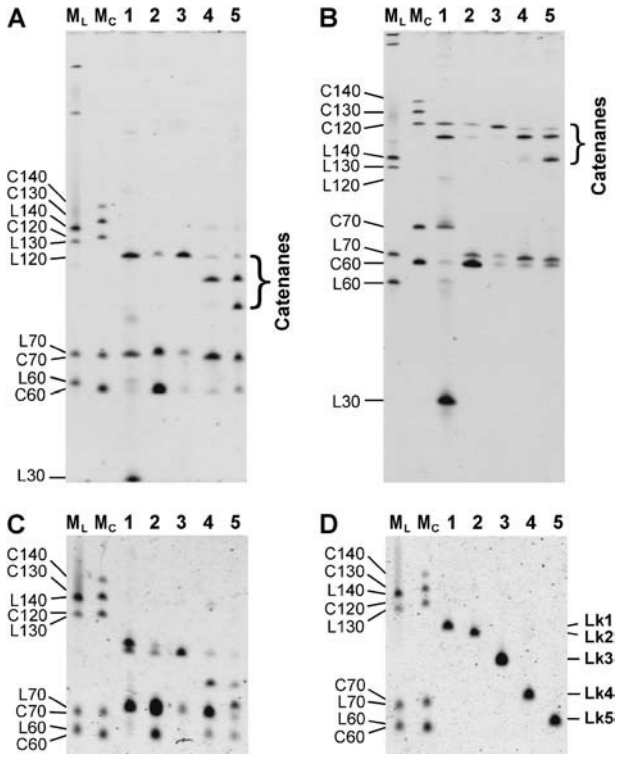

FIGURE 2 Formation of DNA catenanes. Ligation products are resolved, in the same order, by either 6% (*A*), 11% (*B*), or 4% (*C*) denaturing PAGE. Lanes 1 and 2 contain DNA catenanes made from 60A and 70A (12-bp-long LD) using method B (lane 1) or method A (lane 2), respectively; lanes 3–5 contain DNA catenanes made from 60A and 70B, 60A and 70C, or 60A and 70D, respectively (15-bp-long LD, method A). ODNs 70B-D differ from each other in the sequence that is noncomplementary to ODN 60A. In 70B, poly(dT) is used; in 70C, half is poly(dT) and the other half is a random sequence; in 70D, only a random sequence is used. Here and below, lanes M_L denote linear ssDNA markers of 60, 70, 130, and 140 nt, and lanes M_C are circular ssDNA markers of 60, 70, 120, 130, and 140 nt. (*D*) Purified DNA catenanes of Lk1 (60A and 70A), Lk2 (60A and 70B), Lk3 (60A and 70C), Lk4 (60A and 70D), and Lk5 (60C and 70D) are resolved in a 4% gel.

mobility as a 120-nt-long linear ssDNA appeared after ligation. The ligation product was characterized as a DNA catenane using exonuclease digestion methods first applied by Chen et al. (31). The ligated ODN could not be cleaved by Exonuclease VII, which digests ssDNA from both the 5′ and 3′ ends. Similarly, it could not be cleaved by T7 exonuclease, which digests dsDNA in the 5′ to 3′ direction (data not shown). Furthermore, these products could be cleaved by restriction enzyme *Rsa*I, which recognizes the duplex sequence GTAC formed between the 60-nt-long ODN and 70-nt-long ODN (data not shown). After digestion, two bands with similar intensity were observed corresponding to 60- and 70-nt-long linear ODNs. The ligation products could also be digested by restriction enzyme *Nmu*CI (recognition site: GTGAC, only in 60A) after hybridization with the short ODN 60lig30. This digestion resulted in the appearance of a 60-nt-long linear ODN and a 70-nt-long circular ODN. Thus, the ligation product proved to be the DNA catenane consisting of 60A and 70A.

We originally assumed the linking number (the Lk value) for the obtained catenane to be 1 as the LD length was only 12 bp (hereafter the DNA catenane with the Lk value of *N* is designated as Lk*N*). Surprisingly, when the same samples were analyzed in an 11% gel, two bands corresponding to DNA catenanes appeared (lane 1 in Fig. 2 *B*). The weaker band has the same mobility as a 120-nt-long circular ssDNA, whereas the stronger band has much higher mobility (between L140 and C120). Even more interestingly, two bands corresponding to DNA catenanes also appeared in the 4% gel (lane 1 in Fig. 2 *C*), but the mobility order of the two bands was reversed in comparison with the 11% gel. We conclude that these two bands correspond to Lk1 and Lk2 (we assign Lk1 and Lk2 to particular bands below). When method A was used instead of method B, we also observed one band in the 6% gel and two bands in the 11% and 4% gels, respectively (lanes 2 in Fig. 2, *B* and *C*). The yield of catenanes with method B is ~90%, which is much higher than the 20–30% yield with method A. This difference is caused by the different stability of the LD in these two methods. For method A, there is a nick in the middle of the 12-bp-long LD, whereas there is no nick for method B. Note that the yield is high (>90%) in the case of the 15-bp-long LD when method A is used (lanes 3 in Fig. 2, *A–C*). In the case of the 12-bp-long LD, another difference in the ligation results is that the intensity ratio between the lower mobility band and the higher one is 1:3 for method B and 2:1 for method A.

In method B, a part (~20 nt) of the 60-nt-long ODN strand must hybridize to the splint ODN for ligation (Fig. 1 *A*). If

this part formed a secondary structure with the 70-nt-long ODN in the absence of the splint ODN, both ligation efficiency and topology (Lk) of the obtained DNA catenanes might be affected so that the results would not reflect the secondary structure formation correctly. Therefore, we mainly employed method A for forming DNA catenanes. Lanes 3 of Fig. 2, *A–C*, show the ligation results of 60A and 70B using method A. Here, the LD length is 15 bp (Table 1). In the 6% gel, the DNA catenane has the same mobility as the catenane obtained from 60A and 70A (Fig. 2 *A*). In the 11% gel, only one main band (>95%, lane 3, Fig. 2 *B*) with a mobility similar to the slower bands in lanes 1 and 2 was observed. In the 4% gel, however, the band corresponding to the DNA catenane from 60A and 70B (lane 3, Fig. 2 *C*) has a similar mobility as the faster band for the DNA catenane from 60A and 70A (lanes 1 and 2, Fig. 2 *C*). Since a 15-bp-long LD must obviously yield more products with larger Lk than a 12-bp-long LD, we assign the band in lanes 3 as Lk2. Therefore, the band with higher mobility in lanes 1 and 2 of Fig. 2 *B* (11% gel) must be assigned as Lk1 and the slower one as Lk2. Conversely, the band with higher mobility in lanes 1 and 2 of Fig. 2 *C* (4% gel) is Lk2.

Besides the torus catenanes shown in Fig. 1 *B*, more complex catenanes such as 5_1^2, 6_2^2, and 7_1^2 can potentially form (see Supplemental Fig. 2). To some extent, the formation of nontorus catenanes from a closed DNA circle and an open DNA strand is similar to the formation of nontrivial knots out of a DNA strand. According to Monte Carlo calculations, the knotting probability for a 70-nt-long ssDNA during the random cyclization must be <0.1% (37,38). Consequently, the yield of nontorus catenanes must be extremely low (see the Supplemental Fig. 2 legend).

Let us turn now to the question of how the linking number changes with noncomplementary sequences in both circular ODNs, from which we hope to obtain the information about ssDNA secondary structure. When the poly(dT) regions in the 70-nt-long ODN were partly (70C) or completely (70D) replaced by a ''random'' sequence containing all four natural nucleotides (see Table 1 for the sequences), the ligation products showed higher electrophoretic mobility (Fig. 2, *A–C*, lanes 4 and 5). According to the digestion by restriction enzymes and the analysis in polyacrylamide gels of various concentrations, these products are characterized as catenanes with larger Lk values (data not shown). For the combination of 70C and 60A, Lk3 (>90%) was mostly obtained; for 70D and 60A, both Lk4 (50%) and Lk3 (45%) were main products. Note that the LD region in the above two cases was only 17 bp long, which should mostly give Lk2. The formation of DNA catenanes with larger Lk values indicated that a secondary structure formed between the noncomplementary parts during ligation. Actually, when the combination of 70E and 60A was used, in which a 17-bp-long duplex with a dT_{10} loop (TCA TTG TCT TTT TTT TTT ACT GTG TAA/TTA CAC AGT GAC AAT GA) was present besides the 17-bp-long LD, 75% of the obtained DNA catenanes were Lk5 and the remaining 25% were Lk4 (Supplemental Fig. 3, *A* and *B*, lane 7).

Interestingly, the Lk1 catenane had a similar mobility as the Lk3 catenane in 11% gel, but a large difference in mobility between Lk1 and Lk3 was observed in 6% or 4% gel (Fig. 2, *A* and *C*, lanes 1 and 4). In 15% gel, moreover, the mobility of Lk1 was close to Lk4, and Lk2 had a similar mobility as Lk3 (Supplemental Fig. 3 *C*, lanes 3 and 6). After the DNA catenanes with various linking numbers were gel purified and analyzed in 4% gel, a direct correlation between the linking number (Lk1–Lk5) and the mobility was observed (Fig. 2 *D*). Although our assignment is mainly based on the relative mobility of various species, we do not see a possibility for any alternative interpretation of the data.

The linking number of the obtained catenanes does not change much with the sequence of the 60-nt-long ODN, in contrast to a great change with the sequence of 70-nt-long ODN. When one of the two poly(dT) regions in the 60-nt-long ODN was replaced by a random sequence (60B, Table 1), only Lk2 was obtained for its combination with 70B (Fig. 3 *A*, lane 3). In the case of 70C and 60B, similarly to the combination of 70C and 60A (Fig. 3 *A*, lane 2), more than 90% of the ligation products are Lk3 (Fig. 3 *A*, lane 4). Furthermore, for 60C in which no poly(dT) is present, only Lk2 (>95%) was obtained for its combination with 70B (data not shown).

Subsequently, we used random sequences rather than poly(dT) in both strands of 60C and 70D for the formation of DNA catenanes (Table 1). Unexpectedly, DNA catenanes with even larger linking numbers (Lk4 (~25%) and Lk5 (~70%)) were obtained, indicating that a long range of secondary structures formed (Fig. 3 *B*, lane 1). Note that only a 15-bp-long LD could form between 60C and 70D. These ligation products were also resistant toward exonuclease digestion, proving again that these ligation products were DNA catenanes (Fig. 3 *B*, lane 2).

A very important issue for the ssDNA secondary structure is its thermal stability. So far, all ligation products were obtained at 10°C. Our approach allows monitoring any significant changes in secondary structure with temperature. Indeed, for the DNA catenanes forming a secondary structure, the Lk value was found to decrease with increasing the ligation temperature. As described previously, the main product from 60A and 70C was Lk3 when the ligation was carried out at 10°C (Fig. 3 *A*, lane 2). At 37°C, however, Lk2 was the main product, indicating that the secondary structure was not stable at a higher temperature (Fig. 3 *A*, lane 6). The secondary structure formed from 60B and 70C showed a certain stability at a higher temperature. At 37°C, 55% of the obtained catenanes were Lk3 and the remaining 45% were Lk2 in comparison with more than 90% of Lk3 at 10°C (Fig. 3 *A*, lanes 4 and 8). In the cases of 60C and 70D, 20% Lk5, 60% Lk4, 15% Lk3, and 5% Lk2 were obtained at 37°C (70% Lk5, 25% Lk4, and 5% Lk3 at 10°C), showing that some secondary structure domains were fairly stable (Fig. 3 *B*). At

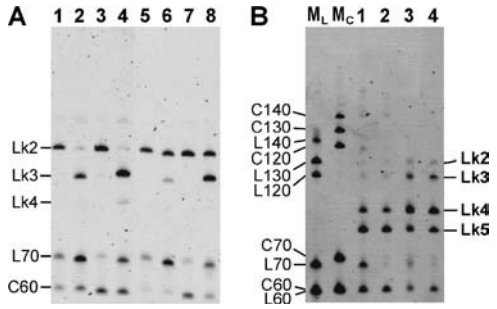

FIGURE 3 Formation of DNA catenanes at 10°C and 37°C. Ligation products are analyzed by 6% denaturing PAGE. (*A*) Lanes 1–4 contain DNA catenanes made from 60A and 70B, 60A and 70C, 60B and 70B, or 60B and 70C, respectively. Ligation was carried out at 10°C; lanes 5–8 are the corresponding ligation results at 37°C. (*B*) Lane 1, catenanes made from 60C and 70D at 10°C; lane 2, digestion of the sample in lane 1 with Exo VII nuclease and T7 exonuclease; lane 3, catenanes made from 60C and 70D at 37°C; and lane 4, digestion of the sample in lane 3 with Exonuclease VII and T7 exonuclease.

45°C, however, almost no catenanes were observed for all the above combinations because even the 15-bp-long LD with a nick was hardly formed during the ligation (data not shown).

Secondary structure analysis by *mfold*

Our next step consisted in using the DNA-*mfold* server to obtain candidate secondary structures of ssDNA catenanes. Since the program deals only with single molecules, it cannot be applied directly to DNA catenanes in which two circular ODNs are linked together. Some operation has to be done before using this program for DNA catenanes. We noted that, if such a catenane were cut at the middle position of the LD, it would look like an intermolecular hybrid formed by two linear ssDNAs consisting of short duplexes at both termini and a large internal loop in the middle (Fig. 4 *A*). If then each of the terminal duplexes were capped with a loop consisting of dT_5, we would obtain a single circular molecule, the secondary structure of which could be predicted by using the *mfold* program. For example, in the case of the DNA catenane made of 60A and 70B, we used as an input the 140-nt-long sequence: 5′-*TTT TT***C GGT GTC TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TCT AAA GTA** *TTT TT*T ACT TTA GTT TTT TTT TTT TTT TAC ACA GTG ACA ATG ATT TTT TTT TTT TTT TGA CAC CG-3′. Herein, the bold sequence is from 70B, the underlined sequence is from 60A, and the two added 5-nt-long caps are shown in italics. The two short hairpin duplexes (GACACCG *TTTTT* **CGGTGTC** and **CTAAAGTA** *TTTTT* TACTTTAG) are constrained to form during *mfold*. The folding temperature is set to 10°C (or 37°C, if specially indicated), and the ionic strength is set to 10 mM Na^+ and 10 mM Mg^{2+}, which are close to ligation conditions. The folded secondary structures are shown in Figs. 4 *B*, 5, and 6. In all these structures, the sequence of the 70-nt-long ODN is positioned from nucleobase 6 to 75, whereas the sequence of the 60-nt-long ODN is from 81 to 140.

For the combination of 60A and 70B, the folding results show that almost no secondary structure can form under the applied conditions (Fig. 4 *B*). Because the TTTT/AATG domain is very unstable, the structure B2 seems hard to form, although the *mfold* modeling shows that structures B1 and B2 have similar stability (Fig. 4 *B*). For the combination of 60A and 70C, also two *mfold* secondary structures with similar free energy values were obtained (Fig. 5 *A*). In the A1 structure, an extra 4-bp-long helix (TTGT/ACAA) is present besides the 17-bp-long LD, whereas the A2 structure has a secondary structure domain consisting of two short helices (ATTGT/ACAAT and TGT/ACA) separated by a 2 × 3 internal loop (TT/GTG). In comparison with the A2 structure shown in Fig. 5 *A*, the secondary structure formed between 60A and 70D has an additional 3-bp-long helix (GTA/TAC, Fig. 5 *B*). Fig. 5 *C* shows the folded secondary structure formed between 60A and 70E, in which another 17-bp-long duplex with a dT_{10} loop is present besides the 17-bp-long LD. Two internal loops without any secondary structure are present. One can see a positive correlation between the linking number of catenanes and the secondary structure predicted by *mfold*.

Some of the *mfold* results for other combinations of various 70-nt-long ODN and 60-nt-long ODN are shown in Fig. 6. In the case of 60B and 70B, a secondary structure may form, however, only in the 60B strand (Fig. 6 *A*). Similarly, a secondary structure could form only in the 60-nt-long ODN in the case of the 60C and 70B combination (Supplemental Fig. 4 *A*). Note that the secondary structure formed in one strand does not contribute to the linking number, and therefore only Lk2 was obtained for 60B and 70B (Fig. 3) as well as for 60C and 70B (data not shown). For 60B and 70C, the *mfold* results show that a secondary structure domain can form consisting of a 3-bp-long helix (CTC/GAG), a 3-nt-long bulge loop (TAG), a 5-bp-long helix (ATTGT/ACAAT), a 2 × 3 internal loop (TT/GTG), and a 3-bp-long helix (TGT/ACA) (Fig. 6 *B*).

For 60C and 70D, the result of *mfold* showed a more complicated secondary structure (at 10°C) consisting of several short duplex parts, four internal loops, and two bulge loops (Fig. 6 *C*). At 37°C, a secondary structure (C2) with much higher free energy value (−2.3 kcal/mol) was obtained, indicating that it is not stable at this temperature according to *mfold*. Several other suboptimal secondary structures with similar free energy values for 60C and 70D were also obtained (Supplemental Fig. 5).

DNA catenanes containing mismatched duplexes

As we mentioned above, a continuous stretch of mismatched bps yields an internal loop when *mfold* is applied. At present,

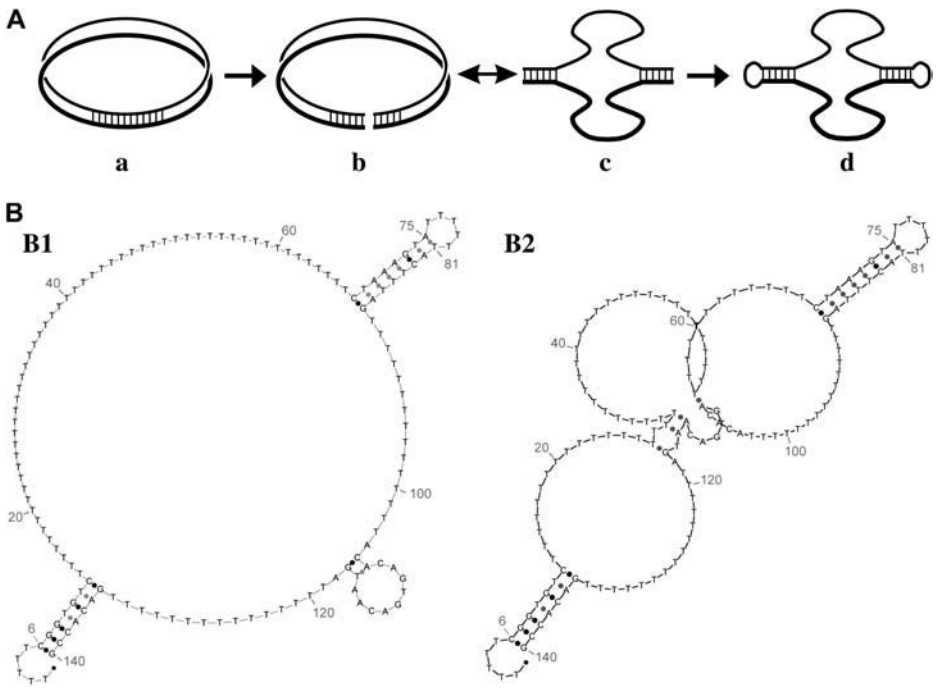

FIGURE 4 (*A*) Transformation of a DNA catenane to a circular ODN which can be used as the input for *mfold* modeling: (*a*) the catenane; (*b*) the catenane is cut at the middle position of the LD; (*c*) after cutting, the catenane is transformed to a duplex with a big internal loop in the middle; and (*d*) after each end is capped with a loop of dT_5, the complex becomes a single circular molecule. (*B*) *Mfold* secondary structures for the combination of 60A and 70B. The sequence of 5′-*TTT TT***C GGT GTC TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TCT AAA GTA** *TTT TT* T ACT TTA GTT TTT TTT TTT TAC ACA GTG ACA ATG ATT TTT TTT TTT TTT TGA CAC CG-3′ was used as the input, and the folding temperature was set to 10°C. Two structures (B1 and B2) with similar free energy values (−5.6 kcal/mol for structure B1, −5.5 kcal/mol for structure B2) were obtained. In these structures, the sequence of the 70-nt-long ODN is located from nucleobase 6 to 75; the sequence of 60-nt-long ODN is from 81 to 140.

the free energy value of this secondary structure can only be estimated according to the length of both strands in the internal loop (19). However, several mismatches such as G·G, G·T, and G·A have a similar stability as an AT bp in some cases. It is difficult to determine the free energy when more than two of these mismatches exist in the internal loop. For example, stable duplexes containing consecutive G·A and/or G·G mismatches have been shown to form either in parallel or in antiparallel manner (22–30). Three consecutive sheared G·A pairs (*trans* Hoogsteen/Sugar edge A·G) have been found to form in an RNA internal loop (UGGA/GAAG) and show high stability (39). These structures are not allowed for in *mfold*. To check whether our approach can detect the formation of these mispaired structures, based on the basepairing patterns of these mispaired duplexes reported by Huertas and Azorin (25) we designed three sets of 60-nt-long and 70-nt-long ODNs containing 20- or 21-nt-long purine-only sequences (Table 1). For example, after hybridization of linear 70M1 with circular 60M1, a 36-bp-long LD might form consisting of a 21-bp-long mispaired helix of $(GGA/GAG)_7$ and a 15-bp-long normal duplex. The nick in the normal duplex part can be sealed by DNA ligase to form a DNA catenane (Fig. 7 *A*).

Indeed, besides Lk2 (15%), catenanes with larger linking numbers including Lk3 (70%) and Lk4 (15%) formed at 10°C for sequence M1 (60M1 and 70M1), indicating that the antiparallel mispaired duplex formed (Fig. 7 *B*, lane 2). Note that only Lk2 forms if the sequence of $(GGA/GAG)_7$ is replaced by poly(dT) in the noncomplementary part (Fig. 2, *A–C*, lane 3). For sequence M3 (60M3 and 70M3), where the mismatched duplex $(GGAGA/GAGGA)_4$ was expected to form, ~85% Lk3 and 15% Lk2 were observed. However, at 37°C, these structures are not stable, and only Lk2 was formed (data not shown). In the case of M2, Lk2 was the main product under any ligation conditions we applied such as lower temperature (0°C) and higher ionic strength (100 mM NaCl), indicating that the mismatched duplex of $(GA/GA)_{10}$ is hard to form.

DISCUSSION

Nontrivial relationship between topology and electrophoretic mobility

It is well known that the electrophoretic mobility of molecules with the same molecular mass and the same overall charge increases with the compactness of the molecules. For knots or catenanes made of relaxed dsDNA (not supercoiled), there is a direct correlation between the electrophoretic mobility (in agarose gels) and the complexity of a knot or the linking number of a catenane (40,41). The sedimentation coefficient of knotted or catenated DNA molecules in solution has been reported to exhibit approximately linear correlation with the average crossing number of ideal configurations of the corresponding knots or catenanes (42). Thus, Lk1 should move slower than Lk2 in the case of very low gel concentrations. As expected, we observed a lower mobility for Lk1 than for Lk2 in 4% denaturing polyacrylamide gels (Fig. 2 *C*, lanes 1–3 and Fig. 2 *D*, lanes 1 and 2). Furthermore, a direct correlation between the linking number (Lk1–Lk5) of short ssDNA catenanes and the gel mobility was observed in the 4% denaturing gel (Fig. 2 *D*). However,

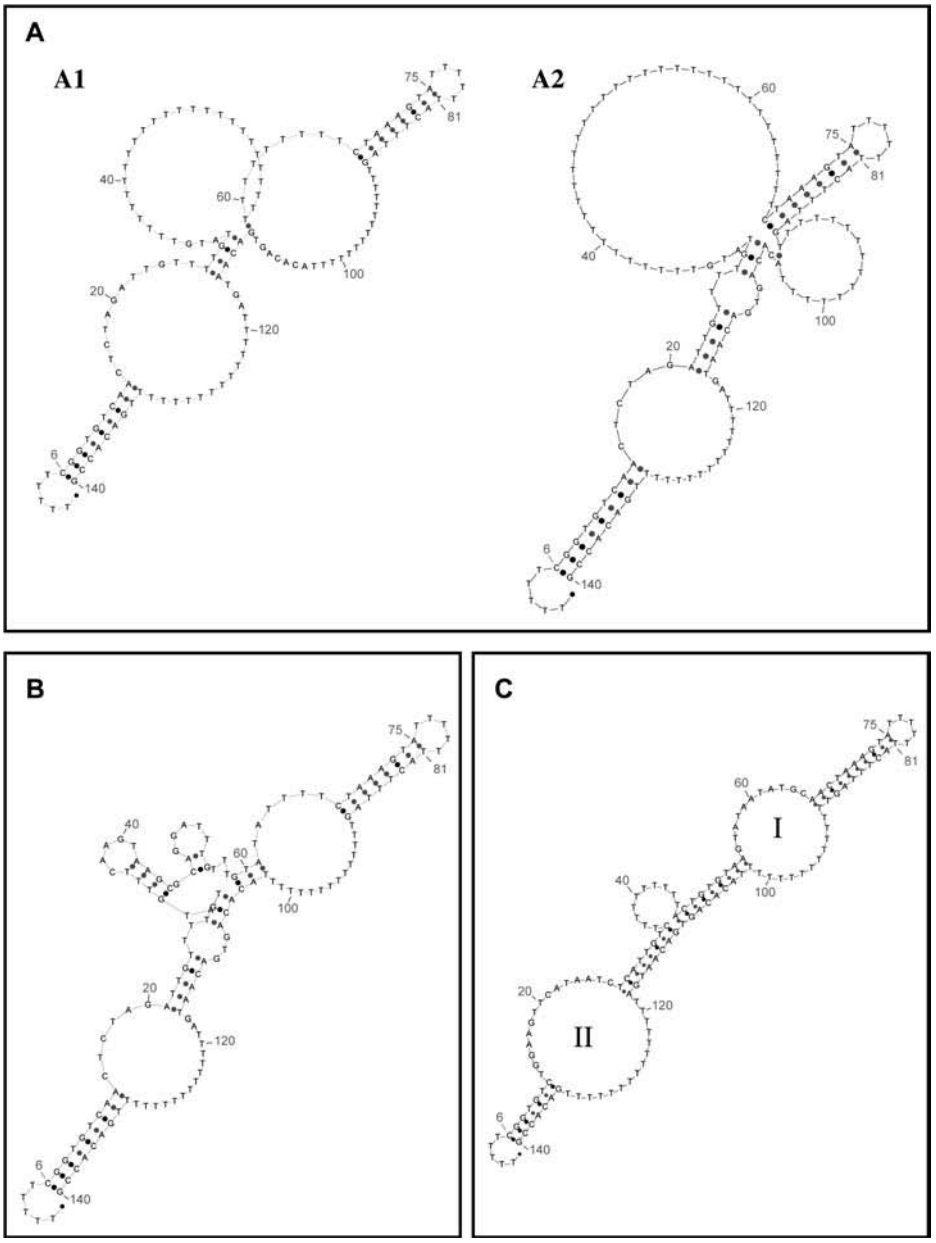

FIGURE 5 Secondary structures obtained by *mfold* for (*A*) 60A and 70C, (*B*) 60A and 70D ($\Delta G = -22.2$ kcal/mol), and (*C*) 60A and 70E ($\Delta G = -37.6$ kcal/mol). For 60A and 70C, two structures (A1 and A2) with similar free energy values (−13.2 kcal/mol for A1, −12.9 kcal/mol for A2) are shown. The folding temperature was set to 10°C.

our data in Fig. 2 *B* clearly show that this correlation between Lk value and the mobility fails at high polyacrylamide gel density.

Bucka and Stasiak also reported that in 12% denaturing polyacrylamide gels (8 M urea) two bands corresponding to Lk1 and Lk2 were observed for the catenanes with an LD length of 12 bp between two 60-nt-long circular ODNs (34). Although the authors did not clarify which of the two bands corresponded to Lk1, they pointed out that Lk3 migrated between Lk1 and Lk2. Similarly, when the DNA catenanes we obtained were analyzed in a 15% denaturing gel, Lk1 moved much faster than both Lk2 and Lk3, whereas Lk2 and Lk3 displayed a similar mobility (Supplemental Fig. 3 *C*). By contrast, Fu et al. reported that the order of the mobility of DNA catenanes in denaturing polyacrylamide gels containing both formamide and urea was Lk3 > Lk2 > Lk1 (33). This order did not change with the gel concentration ranging from 4% to 10%. One may argue that an incomplete denaturation caused the inverse mobility pattern we observed. However, we found that Lk1 also moved faster than Lk2 in an 11% gel and Lk1 moved slower than Lk2 in a 4% gel when both formamide and urea were used as denaturing agents, indicating that this reverse order did not originate from incomplete denaturation (Supplemental Fig. 3, *A* and *B*). Note that the 12-bp-long LD (with 42% GC content) in the DNA catenane we obtained should easily denature under normal denaturing conditions. The different order in gel mobility obtained by Fu et al. may be caused by different

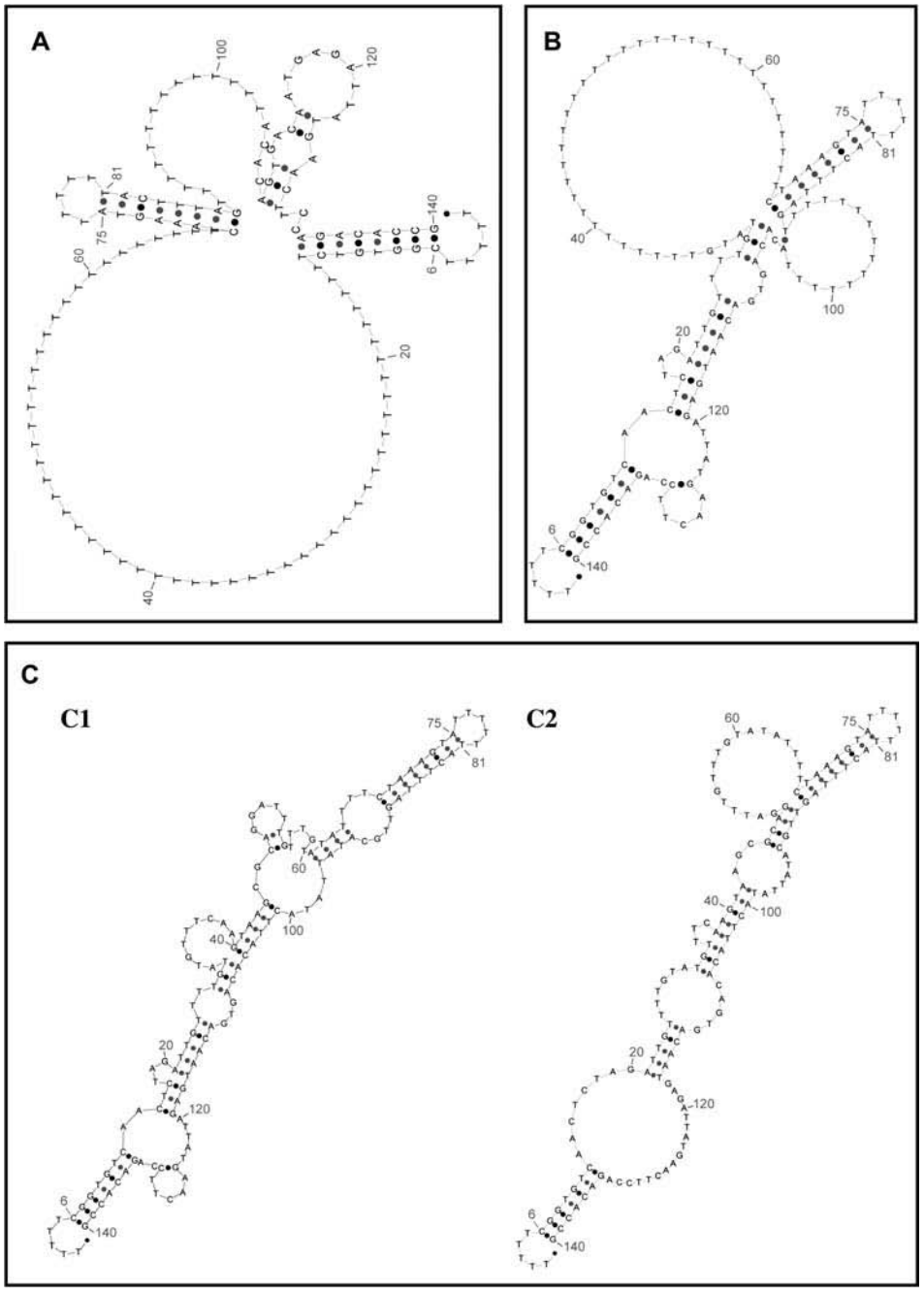

FIGURE 6 Secondary structures obtained by *mfold* for (*A*) 60B and 70B ($\Delta G = -8.0$ kcal/mol), (*B*) 60B and 70C ($\Delta G = -13.2$ kcal/mol), and (*C*) 60C and 70D, at 10°C (C1, $\Delta G = -27.0$ kcal/mol) and 37°C (C2, $\Delta G = -2.3$ kcal/mol).

sequences and different lengths of DNA circles (80 nt) from ours (60 and 70 nt).

Why does Lk1 have a higher mobility than Lk2 in polyacrylamide gels of higher concentration? When the pore size is small enough, we believe that not only the compactness but also the topological structure itself can cause variation of the mobility. The pore size of 8% polyacrylamide gels has been determined to be ~13 nm (43). This value is close to the results obtained by electron microscopy and theoretical approaches (44,45). In our case, considering ssDNA as an ideal polymer chain with a persistence length of 2 nm (15, 46), the mean size of 60–70-nt-long circular DNA is estimated to be ~9 nm (see Supplemental Material 7). The size of corresponding DNA catenanes should be in the range of 9–18 nm. The concentration dependence of the pore size generally lies in the range $c^{-0.5}$–$c^{-0.8}$ (c is the concentration of polyacrylamide) (43,45). Thus, the pore size of 4% gels is in the range of 18–22 nm. In gels of higher concentration (>11%), the rigidity and shape of a DNA catenane may also affect its mobility as its size is within a similar range as the pore size (<10 nm). Lk1 moves faster than Lk2 because it has a higher flexibility to pass through small pores. Note that

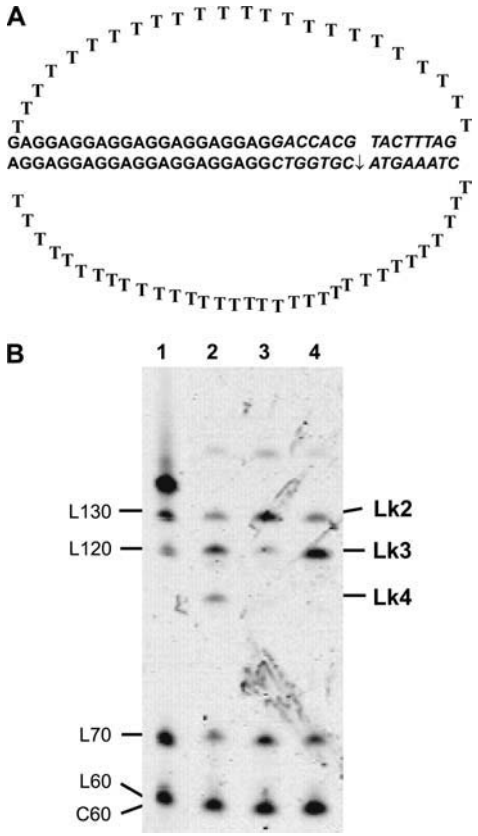

FIGURE 7 Formation of DNA catenanes containing mismatched duplex. (*A*) Illustration of the DNA catenane containing the mismatched duplex $(GGA/GAG)_7$. Method A is used for the catenane formation. The position of the nick after hybridization of linear 70M1 to circular 60M1 is indicated with an arrow. (*B*) Ligation results analyzed by 6% denaturing PAGE. Lane 1, linear ssDNA markers of 60, 70, 120, 130, and 140 nt; lanes 2–4, DNA catenanes formed from 60M1 and 70M1, 60M2, and 70M2, or 60M3 and 70M3, respectively. Ligations were carried out as follows: the 70-nt-long linear ODN (0.2 μM) was hybridized with the 60-nt-long circular ODN (0.3 μM) for 16 h at 10°C in a volume of 20 μL containing 1× T4 DNA ligase reaction buffer; then 0.5 μL of T4 DNA ligase (200 units) was added, and the samples were incubated for 2 h at 10°C.

the two circles in Lk1 can move more freely with respect to each other than in Lk2 (Fig. 1 *B*).

Relationship between the LD length and the linking number

To obtain secondary structure information from linking number values for DNA catenanes, the relationship between DNA structure and the Lk values has to be clarified. Specifically, we need to understand why the 12-bp-long LD yields a significant amount of Lk2, especially when method A is used (Fig. 2 *B*, lane 2). One of the reasons may be the rigidity of ssDNA, which has a persistence length >1.5 nm (2 nt) under our conditions. It has been reported that the persistence length of oligo-dT is 3 nm at 25 mM and 1.5 nm at 2 M NaCl (15). As a result, the single strands may prefer to keep their directions as they protrude from the duplex ends. Interestingly, even in the case of a 10-bp-long LD and using method B for catenane formation, Lk2 (>10%) was observed (Supplemental Fig. 3 *B*, lane 8). In this case, method A cannot be used because only a nicked duplex longer than 11 bp can serve as a substrate for ligation (47). In the case of the 15-bp-long LD, only Lk2 was obtained (Fig. 2, *A–C*, lanes 3). Accordingly, an 11–14-bp-long LD should give a mixture of Lk1 and Lk2, whereas virtually no Lk1 should be observed for a 15–20-bp-long LD (34).

For a short LD (<15 bp), the linking number of the DNA catenane also depends on the formation method. More Lk2 was obtained with method A than with method B in the case of the 12-bp-long LD (Fig. 2, *B* and *C*). In method A, the DNA ligase has to bind the LD and seal the nick to form the catenane; in method B, however, the 60-nt-long ODN is cyclized on a splint ODN (Fig. 1 *A*). When method A is used, the enzyme binding may help the single strands to keep their directions after they protrude from the duplex ends, thus leading to formation of Lk2. For a longer LD (>15 bp), the linking number does not change much with the formation methods. In the case of 60A and 70B (15-bp-long LD), for example, Lk2 is the main product for both method A and method B.

Improving ssDNA secondary structure prediction using information obtained from DNA catenane formation

By combining our results of determining Lk values for DNA catenanes with the *mfold* modeling results (see Table 2), the secondary structure of ssDNA can be more precisely predicted. Most of DNA catenanes made of 70B and 60A are Lk2 (mixed with only ~5% of Lk1, see Fig. 2) because almost no secondary structure can form besides the 15-bp-long LD (Fig. 4 *B*). The 55-nt-long single-stranded part in 70B and the 45-nt-long single-stranded part in 60A are hard to twist around each other, probably due to the unfavorable entropy and the electrostatic repulsion between them. The fact that no Lk3 was observed also implies that the B2 structure did not form due to its low stability, although the *mfold* results showed that structures B1 and B2 have similar free energies (Fig. 4 *B*).

For the DNA catenanes formed from 60A and 70C, more than 90% were Lk3 (Fig. 2, *A–C*, lanes 4), indicating that noncomplementary parts braid around each other besides the LD. It is hard to imagine that the A1 structure (Fig. 5 *A*) would form Lk3 so efficiently (>90%), as only an extra 4-bp-long helix (TTGT/ACAA) is present besides the 17-bp-long LD. A2 looks like a more plausible structure because it has a longer helix (Fig. 5 *A*). However, this structure is not sufficiently stable to be present at 37°C (Fig. 3 *A*, lane 6). Here, the thermostability of the helix containing an internal loop, which is an interesting issue for studying secondary structure of ssDNA, was simply investigated by changing the temperature of ligation. Thus, by using our approach to DNA catenane formation, we obtained important secondary

TABLE 2 Contributions of secondary structures containing noncomplementary regions to the linking number of formed catenanes

Combination of ODNs	Main catenane	Structure figure	Sequence of secondary structure of noncomplementary regions contributing to Lk*
60A + 70B	Lk2	Fig. 4 *B*	None
60A + 70C	Lk3	Fig. 5 *A2*	**ATTGTTTTGT/ACAGTGACAAT** (+1)
60A + 70D	Lk3, Lk4	Fig. 5 *B*	**ATTGTTTTGT***ATGTTTCAAGTAAGCGCAGGATTTGTTT***GTA/TACACAGTGACAAT** (+2)
60A + 70E	Lk5	Fig. 5 *C*	GTATAA/TTTTTT or GGAAG/TTTTT (+1)
60B + 70B	Lk2	Fig. 6 *A*	None
60B + 70C	Lk3	Fig. 6 *B*	**CTC***TAG***ATTGTTTTGT/ACAGTGACAATGAG** (+1)
60C + 70D	Lk5	Fig. 6 *C1*	**CTC***TAG***ATTGTTTTGT***ATGTTTCAA***GTAAGCGCAGGATTTGTTTGTATAT/ATATTATA CTTACACAGTG ACAATGAG** (+3)

*Sequences of Watson-Crick bps are shown in bold, bulge structures are shown in italics, and internal loop structures are underlined. Numbers in parentheses indicate the linking number contributions of these secondary structures.

structure information, which is difficult to obtain by other experimental methods.

For the DNA catenane formed of 70D and 60A, 55% of Lk4 and 45% of Lk3 were obtained (Fig. 2, *A–C*, lanes 5), although the structure obtained from *mfold* has only an additional 3-bp-long helix (GTA/TAC, Fig. 5 *B*) as compared with the one from 70C and 60A, which resulted in almost no Lk4. Note that the total domain is only 13/14 nt long, involving an internal loop of TT/GTG. Under ligation conditions, the two strands in the short internal loop (TT/GTG) were probably braided with each other, generating a continuous helix consisting of 13/14 nucleotides. This structure is not stable at 37°C because more than 80% Lk2 formed when the ligation was carried out at 37°C (data not shown). The two small hairpin structures in 70D do not contribute to the linking number value.

In the case of 70E and 60A (Fig. 5 *C*), Lk5 mostly formed after ligation (>75%, Supplemental Fig. 3 *B*, lane 7). The Lk value of the *mfold* structure for the catenane, which contains two 17-bp-long duplexes is expected to be 4. So braiding also happened to some extent in either of the two internal loops (Fig. 5 *C*). Possible short duplexes formed could be GTATAA/TTTTTT (in loop I) and GGAAG/TTTTT (in loop II). Note that these structures cannot be obtained by *mfold*, which determines that it is not stable enough because more than 50% of mismatches are present. Again, only Lk4 was observed at 37°C, indicating that these putative structures are not stable. We conclude that our approach may provide useful information about the structure of internal loops, whose free energy values are only determined by loop size during the *mfold* calculations (48) without taking into consideration the possibility of sequence-dependent secondary structures.

For 60B and 70C, the *mfold* results show that the secondary structure domain has a similar total helix length as that of 60A with 70D (Figs. 5 *B* and 6 *B*). Both of them have an 8-bp-long helix containing a bulge loop, a 3-bp-long helix, and an internal loop TT/GTG between the two helices. However, the obtained catenanes have different linking numbers: 60B and 70C gave only Lk3, whereas 60A and 70D gave 55% Lk4 and 45% Lk3. This difference is most likely caused by the different location of the specific secondary structure domains. For 60B and 70C, it is close to the 15-bp-long LD and the total helix is only 30 bp long (15+2+8+2+3). With the assumption that the helix periodicity in the internal loop is not shorter than 10.5 bp/turn, a 30 bp helix should result in Lk3. On the other hand, for 60A and 70D, the total helix length is 36 bp if the 6×15 internal loop (CTCTAG/GATTTTTTTTTTTTT) is included.

When random sequences in 60C and 70D instead of poly(dT) were used in both strands, larger than expected linking numbers were obtained, even at 37°C (Fig. 3 *B*). By *mfold* modeling, more than three secondary structures with similar free energy values were obtained. Plausible structures that can yield Lk5 at 10°C or 37°C are shown in Fig. 6 *C*. Again, some of the internal loops need to braid around each other to give Lk5. Other structures cannot yield Lk5, even if most of the internal loops form helices (Supplemental Fig. 5). The preference for helix formation by the two strands in the internal loop may be caused by the stacking tendency of bases for which a helical conformation of backbones is necessary. A structure similar to the one we described above, a noncomplementary DNA helical structure containing mispaired nucleotides and interwound backbones, was reported by Vologodskii et al. (49). Our results clearly indicate that ssDNA forms secondary structures much more easily than is normally expected, especially at lower temperature (Table 2). It is difficult, if not impossible, to avoid the formation of a secondary structure once a long ssDNA with a random sequence is present. It is noteworthy that the sequences we used are mostly AT rich. It can be expected that much more stable secondary structures can form for GC-rich sequences.

Validation of mismatched DNA duplex formation

We have demonstrated that DNA secondary structure formation can be monitored by our approach. Our data imply that this approach may be used as an experimental method for checking whether the duplex with noncanonical bps can form, especially when such potential duplex regions are juxtaposed to regular B-DNA helices. In such a case, the observations of larger than expected Lk values are indicative of

formation of a mismatched right-handed antiparallel double helix. The NMR structure of a mismatched RNA duplex part containing three consecutive sheared G·A pairs (UGGA/GAAG) has recently been reported (39). Our approach may also allow the determination of the pitch of the mismatched helix. For the M1 sequence, the total duplex length is 36 bp (15-bp-long normal B-DNA and 21-bp-long mismatched duplex). Thus, a higher yield of Lk4 (>50%) could be obtained if the mismatched duplex were formed with the pitch of 10.5 bp/turn. However, the main product is Lk3 (70%), indicating that the pitch of this mismatched duplex is >10.5 bp. Note that if the pitch were >14 bp, only Lk3 but not Lk4 would form (15-bp-long normal B-DNA and 8-bp-long mismatched duplex give two turns, and the remaining 13-bp-long mismatched duplex gives another turn). Because Lk4 (>15%) was also observed, the pitch most probably lies between 11 and 13 bp. For the M3 sequence (85% Lk3 and 15% Lk2), the pitch of mismatched duplex is probably >12 bp because no Lk4 was observed. The total duplex length is 35 bp (15-bp-long normal B-DNA and 20-bp-long mismatched duplex $(GGAGA)_4/(GAGGA)_4$). If the pitch were <12 bp, a certain amount of Lk4 should be observed (15-bp-long normal B-DNA and 7-bp-long mismatched duplex give two turns, and the length of remaining mismatched duplex is 13 bp). However, further studies are needed for determining precisely the pitch of these helices by investigating, for example, the change of linking number with the length of mismatched duplex. For M2, one reason for the nearly complete absence of mismatched duplex might be the formation of hairpin structures ($(AG)_n/(TT)_n$) within one strand.

CONCLUSIONS

By forming DNA catenanes and determining their topology, we have monitored how the two strands in the internal loop between two duplex segments twist with respect to one another. It provided us with information on ssDNA secondary structure formation, which is difficult to obtain by other experimental approaches. Combined with *mfold* modeling, these data make it possible to predict more precisely the secondary structure of ssDNA. We found that ssDNA can much more easily fold and form secondary structures than expected. The two strands of the ''internal loop'' in the folded ssDNA structure prefer to braid around each other, even in the presence of several consecutive mismatches. Our approach is promising for checking the stability of DNA secondary structures under various conditions such as temperature and ionic strength. We have also demonstrated that our approach is applicable to experimentally check the formation of double helices comprised of only noncanonical bps.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

This work was supported by National Institutes of Health grant GM59173 to M.D.F.-K.

REFERENCES

1. Zou, L., and S. J. Elledge. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science.* 300:1542–1548.
2. Lee, G. S., M. B. Neiditch, R. R. Sinden, and D. B. Roth. 2002. Targeted transposition by the V(D)J recombinase. *Mol. Cell. Biol.* 22: 2068–2077.
3. Cuenoud, B., and J. W. Szostak. 1995. A DNA metalloenzyme with DNA ligase activity. *Nature.* 375:611–614.
4. Breaker, R. R. 1997. DNA aptamers and DNA enzymes. *Curr. Opin. Chem. Biol.* 1:26–31.
5. Li, Y., and R. R. Breaker. 1999. Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. USA.* 96:2746–2751.
6. Sreedhara, A., Y. Li, and R. R. Breaker. 2004. Ligating DNA with DNA. *J. Am. Chem. Soc.* 126:3454–3460.
7. Gan, H. H., S. Pasquali, and T. Schlick. 2003. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res.* 31:2926–2943.
8. Kim, N., N. Shiffeldrim, H. H. Gan, and T. Schlick. 2004. Candidates for novel RNA topologies. *J. Mol. Biol.* 341:1129–1144.
9. Holbrook, S. R. 2005. RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol.* 15:302–308.
10. Smith, S. B., Y. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science.* 271:795–799.
11. Zhang, Y., H. Zhou, and Z. C. Ou-Yang. 2001. Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. *Biophys. J.* 81:1133–1143.
12. Rivetti, C., C. Walker, and C. Bustamante. 1998. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* 280:41–59.
13. Mills, J. B., E. Vacano, and P. J. Hagerman. 1999. Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). *J. Mol. Biol.* 285:245–257.
14. Kuznetsov, S. V., Y. Shen, A. S. Benight, and A. Ansari. 2001. A semi-flexible polymer model applied to loop formation in DNA hairpins. *Biophys. J.* 81:2864–2875.
15. Murphy, M. C., I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha. 2004. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* 86:2530–2537.
16. Buhot, A., and A. Halperin. 2004. Effects of stacking on the configurations and elasticity of single-stranded nucleic acids. *Phys. Rev. E.* 70:020902.
17. Cantor, C. R., and P. R. Schimmel. 1980. Biophysical Chemistry. Freeman, New York.
18. Bloomfield, V. A., D. M. Crothers, and I. Tinoco. 2000. Nucleic Acids: Structure, Properties, and Functions. University Science Books, Sausalito, CA.
19. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.
20. Dong, F., H. T. Allawi, T. Anderson, B. P. Neri, and V. I. Lyamichev. 2001. Secondary structure prediction and structure-specific sequence analysis of single-stranded DNA. *Nucleic Acids Res.* 29:3248–3257.
21. Gaspin, C., and E. Westhof. 1995. An interactive framework for RNA secondary structure prediction with a dynamical treatment of constraints. *J. Mol. Biol.* 254:163–174.
22. Wilson, W. D., M.-H. Dotrong, E. T. Zuo, and G. Zon. 1988. Unusual duplex formation in purine rich oligodeoxyribonucleotides. *Nucleic Acids Res.* 16:5137–5151.

23. Rippe, K., V. Fritsch, E. Westhof, and T. M. Jovin. 1992. Alternating $d(GA)_n$ sequences form a parallel-stranded DNA homoduplex. *EMBO J.* 11:3777–3786.

24. Suda, T., Y. Mishima, H. Asakura, and R. Kominami. 1995. Formation of a parallel-stranded DNA homoduplex by d(GGA) repeat oligonucleotides. *Nucleic Acids Res.* 23:3771–3777.

25. Huertas, D., and F. Azorin. 1996. Structural polymorphism of homopurine DNA sequences. $d(GGA)_n$ and $d(GGGA)_n$ repeats form intramolecular hairpins stabilized by different base-pairing interactions. *Biochemistry.* 35:13125–13135.

26. Vorlickova, M., I. Kejnovska, J. Kovanda, and J. Kypr. 1999. Dimerization of the guanine-adenine repeat strands of DNA. *Nucleic Acids Res.* 27:581–586.

27. Sakanoto, N., K. Ohshima, L. Montermini, M. Pandolfo, and R. D. Wells. 2001. Sticky DNA, a self-associated complex formed at long GAA·TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J. Biol. Chem.* 276:27171–27177.

28. Manor, H., R. B. Sridhara, and R. G. Martin. 1988. Abundance and degree of dispersion of genomic $d(GA)_n{\cdot}d(TC)_n$ sequences. *J. Mol. Evol.* 27:96–101.

29. Gao, Y. G., H. Robinson, R. Sanishvili, A. Joachimiak, and A. H.-J. Wang. 1999. Structure and recognition of sheared tandem GA base pairs associated with human centromere DNA sequence at atomic resolution. *Biochemistry.* 38:16452–16460.

30. Chou, S.-H., and K.-H. Chin. 2001. Solution structure of a DNA double helix incorporating four consecutive non-canonical base pairs. *J. Mol. Biol.* 312:769–781.

31. Chen, J., and N. C. Seeman. 1991. The synthesis from DNA of a molecule with the connectivity of a cube. *Nature.* 350:631–633.

32. Seeman, N. C., J. Chen, S. M. Du, J. E. Mueller, Y. Zhang, T. J. Fu, Y. Wang, H. Wang, and S. Zhang. 1993. Synthetic DNA knots and catenanes. *New J. Chem.* 17:39–55.

33. Fu, T. J., Y. C. Tse-Dinh, and N. C. Seeman. 1994. Holiday junction crossover topology. *J. Mol. Biol.* 236:91–105.

34. Bucka, A., and A. Stasiak. 2002. Construction and electrophoretic migration of single-stranded DNA knots and catenanes. *Nucleic Acids Res.* 30:e24.

35. Kuhn, H., V. V. Demidov, and M. D. Frank-Kamenetskii. 2002. Rolling-circle amplification under topological constraints. *Nucleic Acids Res.* 30:574–580.

36. Libchaber, A., and R. Bar-Ziv. 2001. Effects of DNA sequence and structure on binding of RecA to single-stranded DNA. *Proc. Natl. Acad. Sci. USA.* 98:9068–9073.

37. Shaw, S. Y., and J. C. Wang. 1993. Knotting of a DNA chain during ring closure. *Science.* 260:533–536.

38. Rybenkov, V. V., N. R. Cozzarelli, and A. V. Vologodskii. 1993. Probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl. Acad. Sci. USA.* 90:5307–5311.

39. Chen, G., B. M. Znosko, S. D. Kennedy, T. R. Krugh, and D. H. Turner. 2005. Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry.* 44:2845–2856.

40. Laurie, B., V. Katritch, J. Sogo, T. Koller, J. Dubochet, and A. Stasiak. 1998. Geometry and physics of catenanes applied to the study of DNA replication. *Biophys. J.* 74:2815–2822.

41. Stasiak, A., V. Katritch, J. Bednar, D. Michoud, and J. Dubochet. 1996. Electrophoretic mobility of DNA knots. *Nature.* 384:122.

42. Vologodskii, A. V., N. J. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet, and A. Stasiak. 1998. Sedimentation and electrophoretic migration of DNA knots and catenanes. *J. Mol. Biol.* 278:1–3.

43. Pluen, A., B. Tinland, J. Sturm, and G. Weill. 1998. Migration of single-stranded DNA in polyacrylamide gels during electrophoresis. *Electrophoresis.* 19:1548–1559.

44. Hsu, T.-P., and C. Cohen. 1984. Observations on the structure of a polyacrylamide gel from electron micrographs. *Polymer.* 25:1419–1423.

45. Viovy, J.-L. 2000. Electrophoresis of DNA and other polyelectrolytes: physical mechanisms. *Rev. Mod. Phys.* 72:813–872.

46. Grosberg, A. Y., and A. R. Khokhlov. 1994. Statistical Physics of Macromolecules. AIP, New York.

47. Ng, P.-S., and D. E. Bergstrom. 2004. Protein-DNA footprinting by endcapped duplex oligodeoxyribonucleotides. *Nucleic Acids Res.* 32:e107.

48. SantaLucia, J., and D. Hicks. 2004. The thermodynamics of DNA structure motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33:415–440.

49. Vologodskii, A. V., X. Yang, and N. C. Seeman. 1998. Non-complementary DNA helical structure induced by positive torsional stress. *Nucleic Acids Res.* 126:1503–1508.