

Slipped (CTG)·(CAG) Repeats of the Myotonic Dystrophy Locus: Surface Probing with Anti-DNA Antibodies

Mandy Tam¹, S. Erin Montgomery^{1,2}, Mariana Kekis^{1,2}, B. David Stollar³
Gerald B. Price⁴ and Christopher E. Pearson^{1,2*}

¹*Program of Genetics and Genomic Biology, The Hospital for Sick Children, Room 11-135 555 University Avenue, Elm Wing, Toronto, Ont., Canada M5G 1X8*

²*Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ont. Canada M5S 1A8*

³*Department of Biochemistry Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111 USA*

⁴*McGill Cancer Center, McGill University, Montréal, Que. Canada H3G 1Y6*

At least 15 human diseases have been associated with the length-dependent expansion of gene-specific (CTG)·(CAG) repeats, including myotonic dystrophy (DM1) and spinocerebellar ataxia type 1 (SCA1). Repeat expansion is likely to involve unusual DNA structures. We have structurally characterized such DNA, with (CTG)_n·(CAG)_n repeats of varying length ($n = 17-79$), by high-resolution gel electrophoresis, and have probed their surfaces with anti-DNA antibodies of known specificities. We prepared homoduplex S-DNAs, which are (CTG)_x·(CAG)_y where $x = y$, and heteroduplex SI-DNAs, which are hybrids where $x > y$ or $x < y$. S-DNAs formed many different species of slipped isomers, as indicated by its multiple electrophoretic species. In contrast, SI-DNAs formed distinct structures, as indicated by the limited electrophoretic species for all possible repeat length pairings. Sister SI-DNAs with an excess of CAG repeats always migrated slower than their sister SI-DNAs with an excess of CTG repeats. Strikingly, both the propensity to form slipped structures and the pattern of S-DNAs, but not SI-DNAs, varied for similar lengths of CTG/CAG repeats between the DM1 and SCA1 loci, highlighting a role for flanking *cis*-elements in S-DNA but not SI-DNA formation. Slipped structures bound structure and nucleotide-specific anti-DNA antibodies. Binding of anti-B-DNA antibodies was reduced for both S-DNAs and SI-DNAs relative to their linear forms. SI-DNAs bound anti-Z-DNA antibodies, while both S and SI-DNAs bound anti-cruciform antibodies, revealing shared characteristics between the corresponding DNA structures and slipped DNAs. Such features of the repeats may be recognized by cellular proteins known to bind such structures.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: myotonic dystrophy type 1; trinucleotide repeat; DNA slipped structures; anti-DNA antibodies; electrophoresis

*Corresponding author

Introduction

Expansion of (CTG)_n·(CAG)_n trinucleotide repeat sequences is associated with 15 human genetic diseases.¹ Among these are myotonic dystrophy (DM1), spinocerebellar ataxia type 1 (SCA1), and Huntington's disease (HD). The mechanism leading

to repeat instability is unknown, but alternative DNA structures are thought to be important mutagenic intermediates. Trinucleotide repeats have the propensity to form slipped strand DNA structures in which the complementary DNA strands display an out-of-register alignment. These structures may lead to aberrant DNA replication and/or error-prone DNA repair resulting in repeat amplification. An appreciation of the biophysical structure of slipped-strand DNAs will facilitate an understanding of their metabolism, which should shed light on the role of slipped DNA structures in instability.

We have created models of slipped-strand DNAs formed by disease-associated trinucleotide repeats. Homoduplex slipped structures S-DNAs are

S.E.M. and M.K. contributed equally to this study.

Abbreviations used: DM1, myotonic dystrophy 1; HD, Huntington's disease; SCA1, spinocerebellar ataxia type 1; S-DNAs, homoduplex slipped DNAs; SI-DNAs, heteroduplex slipped DNAs.

E-mail address of the corresponding author: cepearson@genet.sickkids.on.ca

induced by denaturation and renaturation (reduplexing) of DNA strands containing a trinucleotide repeat tract.² S-DNAs are (CTG)_x-(CAG)_y, where $x = y$. These structures involve an out-of-register alignment of the two strands, which are held together by Watson–Crick pairing in the tandem repeats. Hybridizing strands containing different repeat lengths creates heteroduplex slipped intermediates, SI-DNAs. SI-DNAs are (CTG)_x-(CAG)_y where $x \neq y$. S-DNA structures may provide an accurate model of instability intermediates in non-proliferating genomes,^{2–4} while SI-DNA structures may reflect expansion/deletion intermediates at replication forks or of nick/gap-repair.^{4–6}

Three lines of evidence support S-DNAs and SI-DNAs as good models for mutagenic intermediates of instability. The effects of (1) repeat tract length and (2) repeat tract purity on structure formation correlate with their effect on genetic stability in humans, and (3) biophysically S-DNAs and SI-DNAs are remarkably stable.² The formation of the structures was very sensitive to repeat length. Longer tracts could more easily form slipped structures with a drastic increase in the ability to form slipped structures at lengths of 30–50 repeats, the genetic stability threshold length.^{2,7} Non-repeat interruptions in the purity of trinucleotide tracts provide genetic stability, while pure tracts of similar length are genetically unstable and associated with disease. Slipped structure formation is abolished by the presence of repeat interruptions, suggesting a role for slipped structure formation in repeat instability.⁸ S-DNAs and SI-DNAs are unlike other unusual DNA structures (cruciforms, Z-DNA)⁸ as they do not require superhelical tension, suggesting that they may form *in vivo* and could occur at replication forks, sites of repair or recombination.

Here, we have further characterized DNAs containing DM1 (CTG)_n-(CAG)_n repeats of varying lengths (17, 30, 50, and 79) in the linear, S and SI-DNA forms. Analysis has been accomplished using gel electrophoresis and surface-probing with specific anti-DNA antibodies. In electrophoresis, both S-DNAs and SI-DNAs migrated slowly compared to their linear duplex forms, and SI-DNAs with an excess of CAG repeats in the hairpin always migrated slower than their sister SI-DNAs with an excess of CTG repeats. Probing with eight anti-DNA antibodies raised against cruciform DNA, Z-DNA, native B-DNA and specific bases of single-stranded DNA implied the presence of three-way or four-way DNA junctions, unbase-paired thymidine residues and non-B-form structures in slipped-strand DNAs.

Results

Electrophoretic migration of slipped-strand DNAs

To investigate the effect of repeat length on the

formation of slipped DNAs, structures were prepared from a series of DM1 genomic clones that differed only in the length of the (CTG)_n-(CAG)_n repeat tract. Tract lengths ranged from the normal genetically stable range ($n = 17$ or 30) to the expanded unstable lengths ($n = 50$ or 79). Slipped DNAs were formed by denaturation/renaturation (Materials and Methods). Homoduplex S-DNAs, containing two strands with the same number of repeats, were formed by reduplexing 17, 30, 50 and 79 CTG/CAG repeat tracts uniquely ³²P-labeled on one of the strands. Similarly, SI-DNAs, containing two strands with different numbers of repeats, were formed by heteroduplexing all paired permutations of 17, 30, 50, and 79 CTG/CAG repeat tracts uniquely ³²P-labeled on one of the strands. Each of the radiolabeled fully duplexed linear DNAs, S-DNAs and sister SI-DNAs was digested with a secondary restriction enzyme to release the repeat-containing fragment and individually resolved on polyacrylamide gels (Figure 1).

Slipped S-DNAs migrated more slowly than their linear forms as a series of closely spaced distinct bands that appear as a smear (Figure 1A–F, compare lanes 1 and 3; lanes 2 and 4). The amount of S-DNA and complexity of the electrophoretic species formed increased with repeat tract length (Figure 1 compare lanes 3 and 4 between A and F). The degree of retarded migration was expressed as an R_{bp} -value, which is the ratio of the apparent length in base-pairs of an S-DNA based on its electrophoretic migration (bp_{app}), to its actual length in bp (bp_{len}). Values for bp_{app} were determined from a plot of the \log_{10} of the length *versus* the distance migrated for the 123 bp ladder (not shown). R_{bp} -values for each S-DNA species are summarized in Figure 2. The S-DNAs had R_{bp} -values from 1.25 to 1.35 (Figure 2), migrating as though they were 25–35% longer than expected for their lengths in the fully duplexed form.

The fully duplexed linear DNAs for each repeat length migrated anomalously fast (Figure 1A–F, lanes 1 and 2), with R_{bp} -values from 1 to 0.94 (Figure 2) electrophoretically appearing to be as much as 6% shorter than their actual length in base-pairs, a result that agrees with previously reported findings that (CTG)_n-containing DNAs display increased mobility relative to control fragments.^{9,10} The faster migration of the fully duplexed repeat DNAs reveals how strikingly slow the S-DNAs were.

Each of the sister SI-DNAs migrated predominantly as a distinct product (Figure 1A–F, compare lanes 5–8). Unique radio-labeling of the CTG strand or the CAG strand revealed that the SI-DNA having an excess of CTG repeats was electrophoretically distinct from its sister SI-DNA, having an excess of CAG repeats. The SI-DNA with an excess of CAG repeats migrated slower than its sister with an excess of CTG repeats (Figure 1A–F, compare lane 5 with 7 and lane 6 with 8). R_{bp} -values for the various SI DNAs ranged from 1.61 to 3.41, i.e. they migrated from 61% to

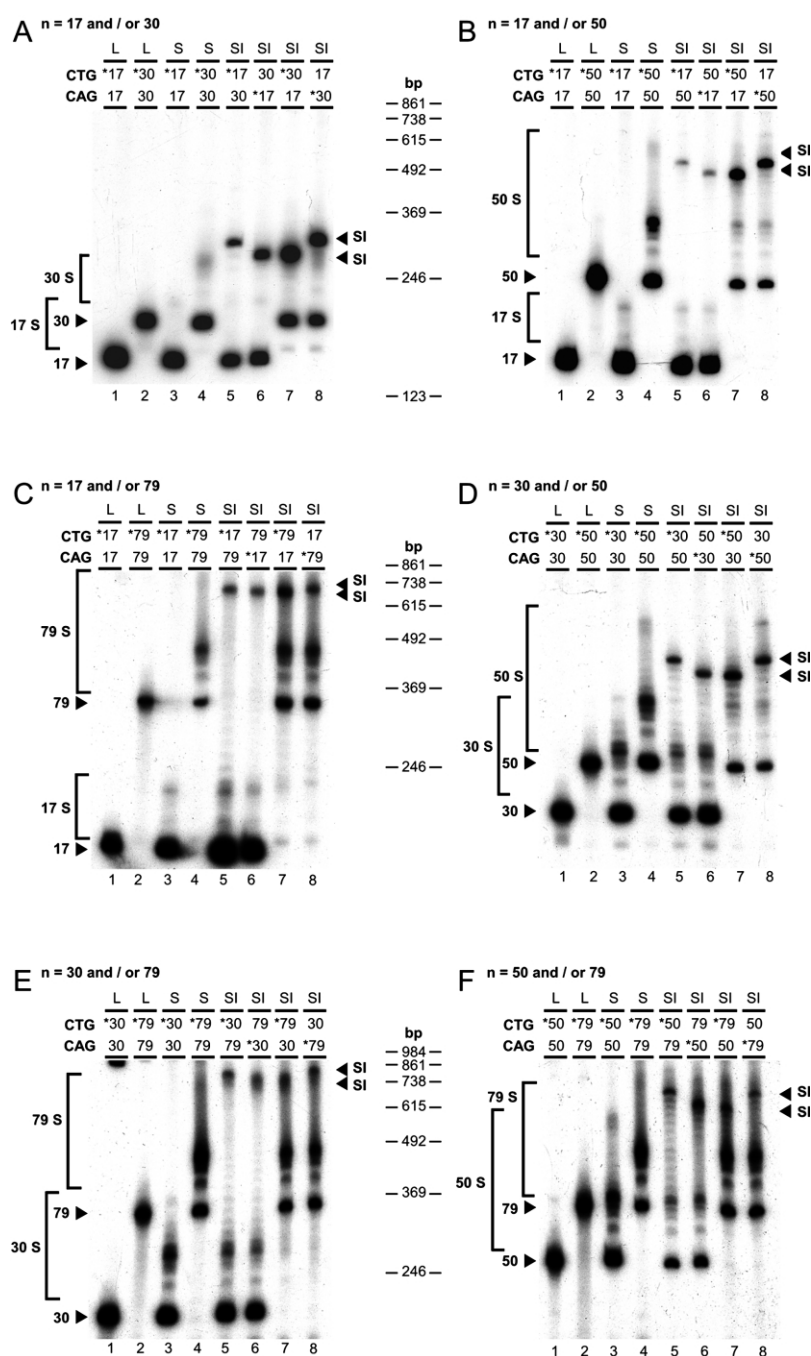


Figure 1. Sister SI-DNAs are not structurally equivalent. Various linear, S and SI-DNA samples were resolved on a polyacrylamide gel. Shown are the resulting autoradiograms. The ³²P-labeled strand is indicated by an asterisk. DNAs that have never been denatured are in the linear form (L) (A–F, lanes 1 and 3). Homoduplexed S-DNAs (S, square bracket) (A–F, lanes 2 and 4). The autoradiographic pattern for the linear and S-DNAs was the same regardless of which strand was radiolabeled (*). Heteroduplexed SI-DNAs (SI) ³²P-labeled on the *(CTG)_x·(CAG)_y (A–F, lane 5), (CTG)_y·*(CAG)_x (A–F, lane 6), *(CTG)_y·(CAG)_x (A–F, lane 7) and (CTG)_x·*(CAG)_y (A–F, lane 8). The distinct bands formed by the sister SI-DNAs are indicated (arrowheads). The locations and sizes (bp) of the 123 bp DNA marker fragments used for sizing, are indicated in the center and applies to gel panels on either side (they were the same gel). Only the lower portions of the gels are shown.


41% slower than expected (Figure 2). Importantly, for all permutations of repeat length pairings, the migration of the sister SI-DNA with an excess of CAG repeats, relative to that of its sister with an excess of CTG repeats, is greater than 1, ranging from 1.02 to 1.10 (Figure 2). Thus, the sister SI-DNA with an excess of CAG repeats always migrated 2–10% slower than its sister with an excess of CTG repeats regardless of the repeat lengths involved in the heteroduplex.

To determine whether the differential migration of the sister SI-DNAs was the result of their different molecular masses, we plotted their electrophoretic mobilities as a function of the log₁₀

of the actual molecular mass of the repeat-containing DNAs (Figure 3). The fully duplexed linear repeat DNAs migrated faster than expected for their molecular masses, indicating that the repeat sequence, rather than base composition is responsible for this effect (Figure 3, compare line of best-fit for linear repeat fragments to that of the 123 bp ladder). The degree of aberrant migration of each species was expressed as an *R*_{MW}-value (Figure 2), which is the ratio of the observed electrophoretic migration with the expected migration (calculated from the plot shown in Figure 3). The value for expected migration does not take secondary structure into account. All linear repeat DNAs had

Linear-DNA CTG	CAG	Actual Base Pairs	Apparent Base Pairs	$R_{bp} = bp_{app}/bp_{len}$	$R_{MW} = \frac{\text{Observed}}{\text{Expected}}$
17	17	168	174	1.04	1.01
30	30	207	205	0.99	1.04
50	50	267	257	0.96	1.04
79	79	354	331	0.94	1.06

S-DNA CTG	CAG	Actual Base Pairs	Apparent Base Pairs	$R_{bp} = bp_{app}/bp_{len}$	$R_{MW} = \frac{\text{Observed}}{\text{Expected}}$
17	17	168	227	1.35	0.87
30	30	207	271	1.31	0.83
50	50	267	333	1.25	0.86
79	79	354	448	1.27	0.85



SI-DNA CTG	CAG	Actual Base Pairs	Apparent Base Pairs	$R_{bp} = bp_{app}/bp_{len}$	$R_{MW} = \frac{\text{Observed}}{\text{Expected}}$
17	30	175	302	1.72	0.75
30	17	175	282	1.61	0.79
17	50	185	503	2.72	0.56
50	17	185	466	2.52	0.59
17	79	199	680	3.41	0.47
79	17	199	666	3.35	0.48
30	50	217	442	2.04	0.65
50	30	217	404	1.86	0.70
30	79	232	732	3.16	0.44
79	30	232	699	3.01	0.47
50	79	282	656	2.32	0.55
79	50	282	601	2.13	0.60

Figure 2. Molecular masses, migration data and calculated R_{bp} -values and R_{MW} -values. R_{bp} -values for each species represent the ratio of actual base-pairs (bp_{len}) to apparent base-pairs (bp_{app}). The actual number of base-pairs contained in each S and SI-DNA-containing fragments were obtained by totaling the base-pairs contained in duplex and the slip-out(s) as though they were extensions of the linear duplex (see schematic at top of table; black lines indicate duplex backbone, grey lines indicate hairpin base-pairs). Values for apparent base-pairs (bp_{app}) were determined from a plot of the \log_{10} of the length *versus* the distance migrated for the 123 bp ladder (not shown). For sister SI-DNAs, R_{bp} -values are compared as a ratio for when CAG is in excess to when CTG is in excess. The R_{MW} -values for each species represent the ratio of the observed electrophoretic migration with the expected migration. To determine the expected electrophoretic migration of each DNA species, the actual molecular mass values for the S and SI-DNA fragments were entered into the equation for the line of best-fit through the 123 bp fragment values on a plot of mobility *versus* \log_{10} of their molecular mass (shown in Figure 3). The value for expected migration does not take secondary structure into account. For sister SI-DNAs, R_{MW} -values are compared as a ratio for when CAG is in excess to when CTG is in excess. Actual mass were calculated from sequence contents, using adenine = 312.20 Da, cytosine = 288.18 Da, guanine = 328.2 Da, and thymine = 303.18 Da. The molecular mass of the non-repeating sequence flanking the repeat in the *EcoRI-HindIII* fragments was 68093.73 Da, and each (CTG)_n-(CAG)_n unit is 1848 Da. Values are the average of three to five independent experiments.

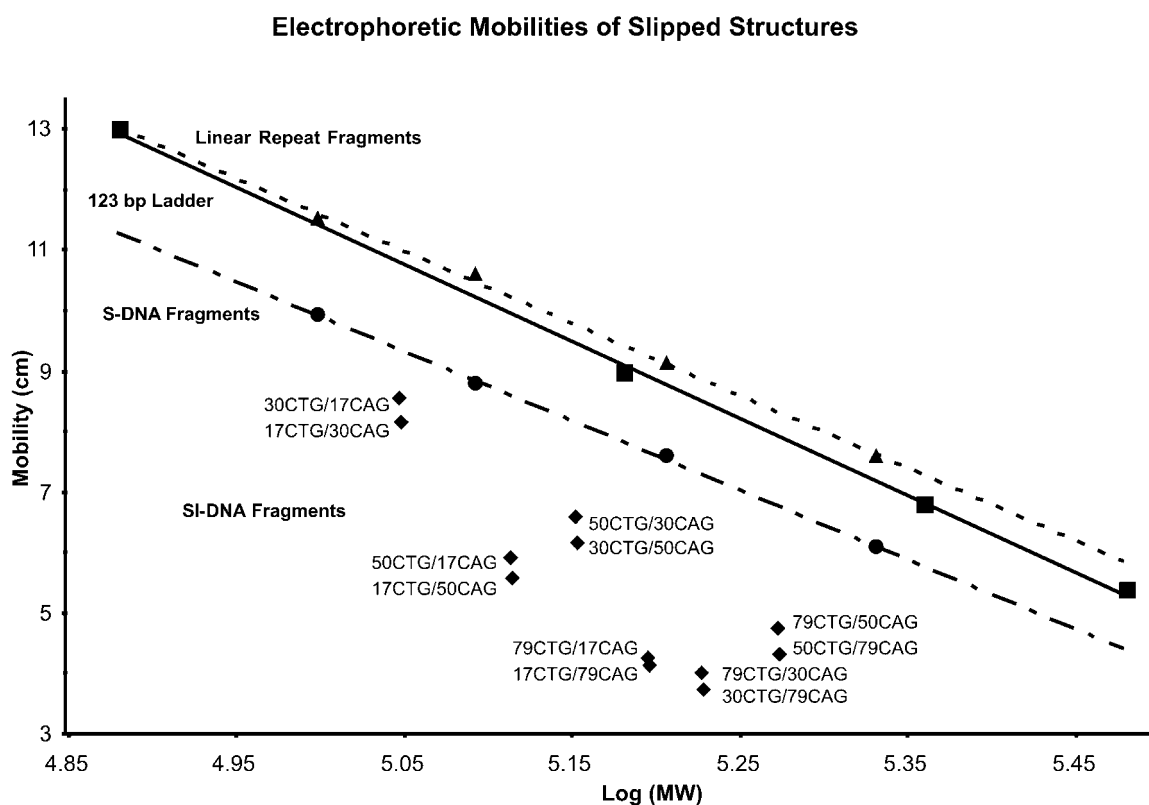


Figure 3. Relative electrophoretic mobilities of DNA slipped structures. The observed electrophoretic migration (cm) of each repeat-containing species was plotted relative to the \log_{10} of their actual molecular masses (Materials and Methods). The actual number of repeats contained in each species is indicated. For comparison the electrophoretic migration of the marker 123 bp fragments were also plotted. All species above or below the line of best-fit through the 123 bp marker migrated aberrantly fast or slow, respectively. To determine the expected electrophoretic migration of each DNA species (Figure 2) the actual molecular mass values for the S and SI-DNA fragments were entered into the equation for the line of best-fit through the 123 bp fragment values on a plot of mobility *versus* \log_{10} of their actual molecular masses. A plot of electrophoretic migration relative to length (base-pairs) yielded similar trends (not shown) and was used to calculate the apparent size (bp) in Figure 2.

R_{MW} -values greater than 1. The S-DNAs and SI-DNAs migrated much more slowly than expected based on their molecular mass alone, indicating a structural basis for their retarded migrations (Figure 3). However, both S-DNAs and SI-DNAs were monomeric, as their slow migration was considerably faster than would be expected of multiple molecule aggregates.¹¹ Importantly, the ratio of the differential migration of the sister SI-DNAs was greater than would be expected of their differential molecular masses (Figure 2, note that the relative R_{MW} -values were all less than 1, ranging from 0.92 to 0.98). Thus, the differential migration of SI-DNAs with an excess of CAG repeats ($A = 312.2$ Da) and SI-DNAs with an excess of CTG repeats ($T = 303.18$ Da) cannot be explained by the small differences in molecular mass between the sisters. We conclude that the differential migration of sister SI-DNAs is the result of structural differences.

The resolution of each sister SI-DNA into predominantly one distinct electrophoretic product suggested that each formed a distinct preferred structure. Differences between sister SI-DNAs may

involve the geometries of the three-way slip-out junctions and/or the conformation of the slipped-out CAG compared to the slipped-out CTG (random coils *versus* hairpins and hairpin loops). The angles subtended by the arms of various three-way DNA junctions are sensitive to stacking interactions of the bases immediately adjacent to, the penultimate and possibly one more base at the junctions.^{29,30} While it is likely that both junction geometry and slip-out conformation contribute to the differential migration of the sister SI-DNAs, other explanations cannot be ruled out. For example, the location of the slip-out extrusion may vary between sister SI-DNAs; permuting this location relative to the complementary paired repeats would be expected to result in differences of electrophoretic migration. Another possible explanation for the differential migration of sister SI-DNAs is structural alteration, such as bent DNA in the non-repeating sequences flanking the slip-out junction, a phenomenon observed for sister three-way DNA junctions formed by cloned inverted repeats.¹² We have tested these possibilities below.

Effect of magnesium upon the differential migration of sister SI-DNAs

To determine if magnesium ions affected the geometry of the junction arms, slipped DNAs were electrophoresed in the presence or in the absence of magnesium ions. A concentration of 200 μ M magnesium is known to affect the relative geometries of some but not all four-way or three-way DNA junctions.^{13,14,43} In the presence of magnesium the migration pattern of the linear, and SI-DNA molecules was similar to that in its absence (Figure 4A). Magnesium did not affect the rate of anomalously fast migration of the linear repeat DNAs (compare lanes 1 and 4), yielding R_{bp} -values that were comparable to that in the absence of the metal. While magnesium did not affect the slower migration of SI-DNAs having an excess of CAG repeats over that of SI-DNAs with an excess of CTG repeats (compare lanes 3 and 6), it did reduce the degree of their slow migration, each yielding R_{bp} -values that were proportionately less in the presence of the metal. Importantly, the

sister SI-DNA molecules, which contained predominantly one slipped junction yielded relative R_{bp} -values that were greater than 1 in either the presence or in the absence of the metal. In contrast, S-DNA molecules, each of which harbor multiple junctions, displayed a broader electrophoretic pattern in the presence of the metal than in its absence (compare lanes 2 and 5) having R_{bp} -values ranging from 1.04 to 1.28. One interpretation is that the effect of Mg^{2+} on the geometric trajectory of molecules harboring one uniquely located slipped junction at varying locations (SI-DNAs) is limited compared to that of its effect on molecules with multiple junctions at varying locations (S-DNAs) in which it may be more complex. While the biophysical effect of metal ions on slipped DNAs is beyond the scope of the present study, these preliminary results argue against a simple geometric isomerization as an explanation for the differential migration of sister SI-DNAs.

Source of differential migration of sister SI-DNAs maps within the repeat tracts

To determine if the source of the differential migration of sister SI-DNAs was due to structural alterations (bends) in the non-repeating flanking sequences, these segments were removed, and the relative migration of the shortened sister SI-DNAs assayed. Heteroduplex mixtures of *Eco*RI/*Hind*III fragments uniquely radiolabeled on the *Eco*RI end of either the (CTG)30 or (CAG)30 strand were subjected to further restriction digestion using *Dde*I, which cuts 35 bp 3' of the CTG tract (Figure 5A). The relative degree of differential migration between sister SI-DNAs was unchanged following removal of this flanking sequence (Figure 5B, compare lanes 1 and 2 with 3 and 4). Thus, it is unlikely that the source of differential migration is due to a bend in the non-repeating sequence 3' of the CTG tract. A similar experiment was performed using heteroduplex mixtures of *Eco*RI/*Hind*III fragments uniquely radiolabeled on the *Hind*III end of either the (CTG)30 or (CAG)30 strand that were digested with *Msp*I, which cuts 8 bp 5' of the CTG tract, removing this non-repeating flank. Again, the differential migration between sister SI-DNAs was unchanged (not shown). We conclude that the source of differential migration of sister SI-DNAs maps to within the repeat tract and is not the result of bent DNAs in the flanking sequences.

Zippering direction does not affect the structure of either S-DNAs or SI-DNAs

The consistently slower migration of sister SI-DNAs with an excess of CAG *versus* an excess of CTG repeats may be the result of different locations of the excess repeats such that the location of slip-out extrusion would vary between the sister SI-DNAs. Changing the location of slip-out extrusion would be expected to change the electrophoretic migration. The location of slip-out

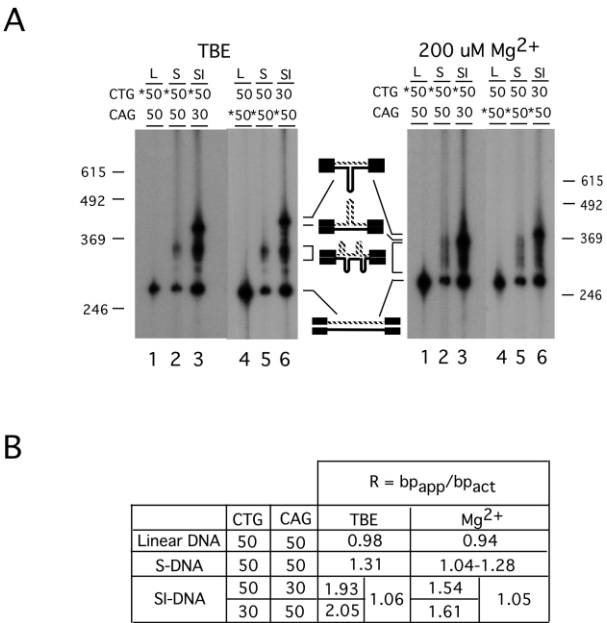


Figure 4. Effect of magnesium on electrophoretic migration. *Eco*RI/*Hind*III restriction digests of linear, S-DNA and SI-DNA mixtures of [(CTG)30·(CAG)30 × (CTG)50·(CAG)50] uniquely ³²P-labeled at the *Eco*RI end of either the (CTG)50 or (CAG)50 strand (indicated by an asterisk). Reactions were analyzed on a 4% polyacrylamide gel with Tris–borate EDTA (TBE) or with Tris–borate magnesium (200 μ M) and electrophoresed (Materials and Methods). Gels were dried, and exposed for autoradiography. R_{bp} -values were calculated as described for Figure 2. DNAs that have never been denatured are in the linear form; homoduplexed S-DNAs (square bracket); and the distinct bands formed by the sister SI-DNAs are indicated. The locations and sizes (bp) of the 123 bp DNA marker fragments used for sizing, are indicated on the left and right sides of each gel. Only the lower portions of the gels are shown.

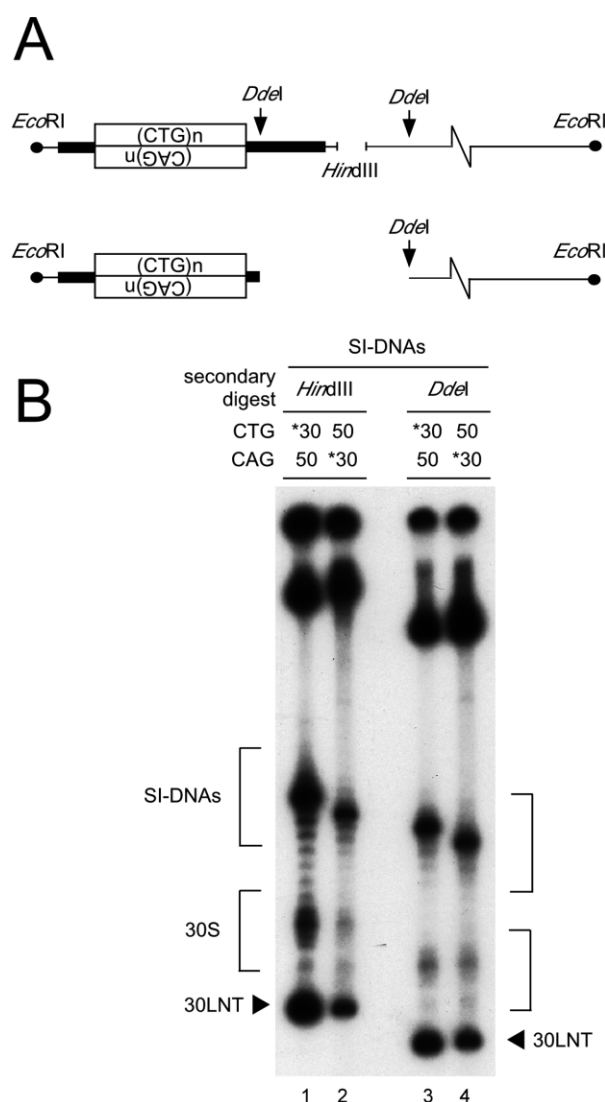


Figure 5. Source of the anomalous migration maps within the repeat tract. **A**, Restriction sites for *EcoRI*, *HindIII*, and *DdeI* are indicated. Human sequences are in bold, and the ^{32}P radiolabel is indicated by a dot. **B**, *EcoRI/HindIII*, and *EcoRI/DdeI* restriction digests of heteroduplex mixtures of [(CTG) $_{30}$ ·(CAG) $_{30}$ × (CTG) $_{50}$ ·(CAG) $_{50}$] uniquely ^{32}P -labeled at the *EcoRI* end of either the (CTG) $_{30}$ or (CAG) $_{30}$ strand (indicated by an asterisk). Reactions were analyzed on a 4% polyacrylamide gel, dried, and exposed for autoradiography.

extrusion may be affected by the mechanics of strand annealing: i.e. the isomers of slipped-stranded DNAs formed may depend upon the direction from which the two strands are “zippered” together (Figure 6).¹⁵ The annealing of complementary repeat-containing strands may occur in the 5'-C → T → G-3' direction or in the 5'-C ← T ← G-3' direction. The direction of strand zippering would be determined by the location of the primary inter-strand nucleation, from which zippering would emanate.¹⁵ Directional zippering of complementary strands containing different numbers of repeat units may force the extrusion of excess repeats from either of the extreme ends of

the repeat tract. The slip-out would form at the immediate end of the repeat tract on the opposite side from where nucleation occurred (Figure 6, Pathway I, compare SI-DNAs). If inter-strand nucleation were to occur on the opposite side, changing the direction of zippering might force the slip-out to occur on the opposite end of the repeat tract. Molecules with slip-outs extruding at opposite ends might be expected to display different electrophoretic migration patterns, much like the variable migration of molecules with bends occurring at different locations.¹⁶ Alternatively, intra-strand hairpins formed prior to inter-strand nucleation may resist “migration” by inter-strand zippering (Figure 6, Pathway II). In this case the location of the slip-out would be determined by the preferred location for intra-strand hairpin formation (Figure 6, Pathway II, compare SI-DNAs). If slipped structures formed predominantly by Pathway II over Pathway I, one would expect to observe the same electrophoretic pattern of slipped products regardless of zippering direction.

To determine whether directional zippering affected the structural characteristics of the slipped-strand DNAs formed, we reannealed DNAs that were linearized on either side of the repeat tract, such that 85–90% of the plasmid length was either 5' (*HindIII*) or 3' (*EcoRI*) of the CTG tract (Figure 6). The increased probability that the primary inter-strand nucleation occurred in the non-repeating adjacent sequence 5' or 3' of the repeat tract favored zippering in either the 5'-C → T → G-3' or in the 5'-C ← T ← G-3' direction, respectively. Following reannealing the repeat-containing fragments were released by a secondary restriction digestion using *EcoRI* or *HindIII* and products were separated by ultra-high-resolution gel electrophoresis. Figure 7 shows the zippering analyses for the formation of S-DNAs and SI-DNAs formed by repeat tracts of 30 and/or 50 units. The electrophoretic pattern of S-DNAs formed was unaltered by the direction of strand zippering (Figure 7, compare lane 1 with 2 and lane 3 with 4). Similarly, the pattern of SI-DNAs formed between tracts of 30 and 50 units was indistinguishable whether they were zippered from the 5' or the 3' directions (Figure 7, compare lane 5 with 6 and lane 7 with 8). This high-resolution gel resolved the “smear” into clearly distinct bands revealing the numerous distinct slipped S-DNA and SI-DNA isomers contained within each reaction. Regardless of the direction of inter-strand zippering, an identical pattern of S-DNAs and SI-DNAs was formed (Figure 7 compare lane 1 with 2 and lane 3 with 4; compare lane 5 with 6 and lane 7 with 8). The similar pattern of slipped products formed supports the predominance of an intra-strand process (Figure 6 Pathway II) for the formation of slipped DNAs. We conclude that S and SI-DNAs form distinct isomers and that the locations of the slipped-out repeats, and hence the relative amount of each different slipped isomer formed, were unaffected

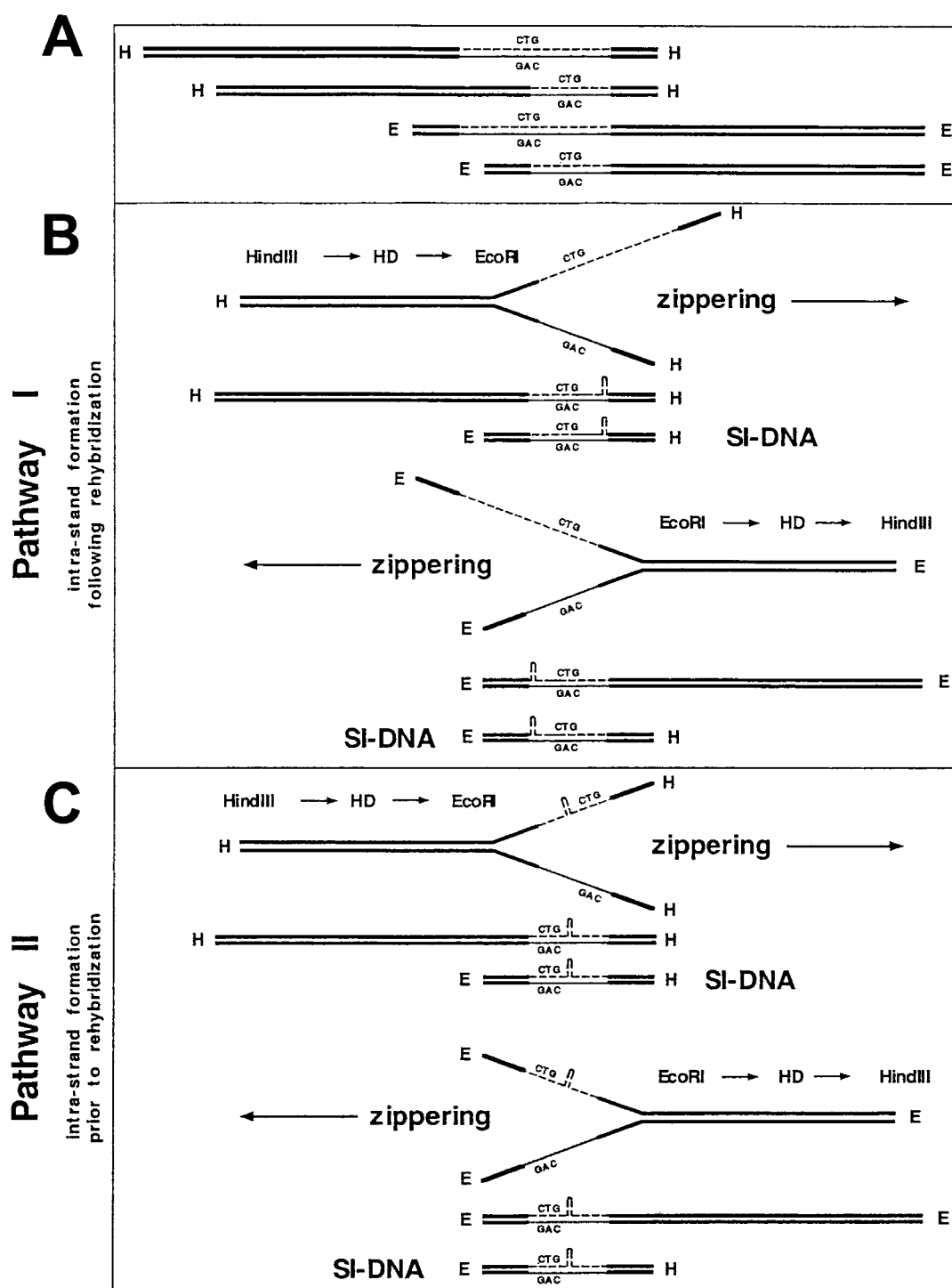


Figure 6. Two modes of intra-strand hairpin formation and inter-strand nucleation result in different locations for SI-DNAs. A, Schematic representation of two pairs of DM1 clones of different repeat lengths. The first set is linearized 3' (*Hind*III) of the CTG/CAG repeat tract; the second set is linearized 5' (*Eco*RI) of the CTG/CAG repeat tract. B, Pathway I, intra-strand structure formation following rehybridization. Directional zippering of complementary strands containing different numbers of repeats to form SI-DNA structures at opposite ends of the repeat tract. Top, zippering in the *Hind*III → *Eco*RI direction might be expected to result in SI-DNA structures toward the 3'-end of the repeat tract. Bottom, zippering in the *Eco*RI → *Hind*III direction might be expected to result in SI-DNA structures toward the 5'-end of the repeat tract. C, Pathway II, intra-strand structure formation prior to rehybridization. Directional zippering of complementary strands containing different numbers of repeats where one strand harbors an intra-strand hairpin. Zippering in either the *Hind*III → *Eco*RI or *Eco*RI → *Hind*III direction would result in SI-DNA structures located at the preferred site for intra-strand hairpin formation.

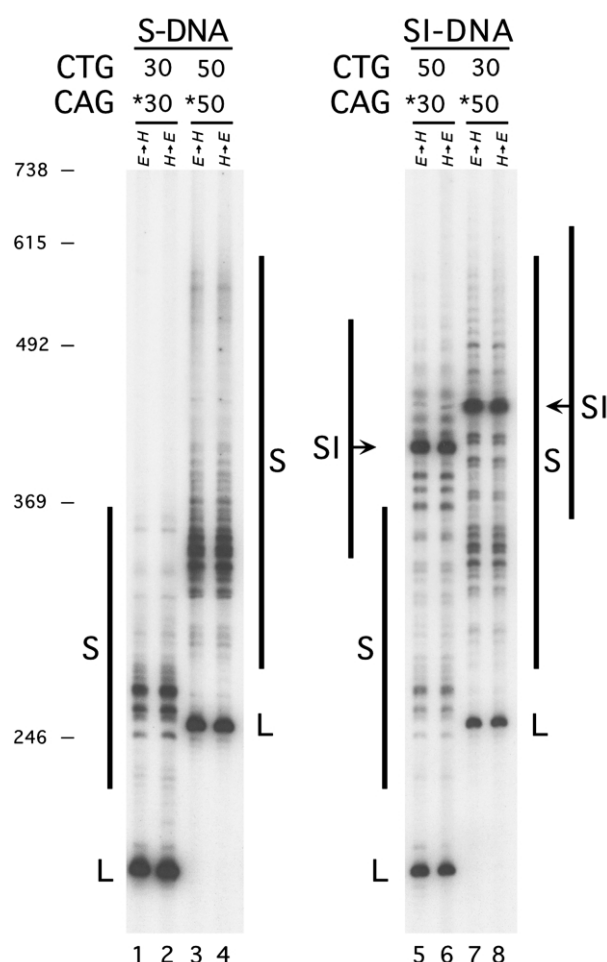


Figure 7. Zippering direction does not affect the structure of S or SI-DNAs. Zippering analysis for the formation of S (lanes 1–4) and SI-DNAs (lanes 5–8) formed by repeat tracts of 30 and/or 50 repeats. ^{32}P -labeled S and SI-DNA preparations differing in the direction of zippering ($\text{EcoRI} \rightarrow \text{HindIII}$ or $\text{HindIII} \rightarrow \text{EcoRI}$) were restriction digested to release the repeat-containing fragment and analyzed by ultra-high-resolution PAGE (Materials and Methods). Linear 30 repeat and 50 repeat fragments are indicated (L). Bold lines labeled S indicate S-DNAs. Bold lines labeled SI indicate SI-DNAs. Arrows indicate the preferred SI-DNA isomer. See the text and Figure 6 for details. Only the lower portion of the ultra-high-resolution gel is shown.

by the direction of complementary strand annealing.

Slipped DNAs in SCA1 repeats: the effect of flanking sequences

We next studied the possible contribution of non-repeat flanking sequence on structure and formation of S-DNA and SI-DNA by analyzing slipped structure formation in CTG/CAG repeats from the SCA1 locus whose flanking sequences are completely different from those flanking the repeats at the DM1 locus. The pattern and amount

of homoduplex S-DNA formed by either 30 repeats or 49 repeats in the SCA locus was considerably different from that formed by 30 repeats or 50 repeats in the context of the DM1 locus (compare lanes 2 and 6 of Figure 8 with lanes 3 and 4 of Figure 1D). Differences were also observed between S-DNAs formed by 60 SCA1 repeats and S-DNAs formed by 79 DM1 repeats (compare lane 10 of Figure 8 with lane 4 of Figure 1E). Thus, for S-DNAs the pattern and amount of structures formed by a given length of repeats can vary depending upon the sequence of the non-repeating DNAs flanking the repeat. However, unlike S-DNAs, SI-DNAs formed by SCA1 clones formed distinct electrophoretic species much like the sister SI-DNAs formed by DM1 clones (compare lanes 3, 4, 7, and 8 of Figure 8 with lanes 5, 6, 7, and 8 of Figure 1D). Thus, non-repeating sequences flanking repeat tracts do not seem to affect the formation or structure of SI-DNAs. Furthermore, for SCA1 DNAs, sister SI-DNAs having an excess of CAG repeat migrated slower than those with an excess of CTG repeats. These results provide further support that the differential migration of sister SI-DNAs was due to structural differences within the repeat tract and not to alterations within the non-repeats flanks as has been suggested for sister three-way junctions.¹²

Probing the surface of S-DNAs and SI-DNAs with specific anti-DNA antibodies

Specific anti-DNA antibodies were used to probe the surfaces of linear DNAs, S-DNAs and SI-DNAs formed by (CTG)-(CAG) repeat-containing DNAs. Anti-DNA antibodies with characterized specificities can be used as tools to probe the surface of DNA molecules to determine whether they contain corresponding structural characteristics.¹⁷ This is particularly useful when the antibodies used are specific for DNA structures that have been extensively characterized such as Z-DNA, three-way or four-way DNA junctions, which may be minor components in complex materials. Binding of anti-DNA antibodies to S-DNAs or SI-DNAs would therefore indicate a shared structural characteristic with the DNA that was used to elicit the antibody. Using eight different characterized anti-DNA antibodies, we probed the surface of gel-purified slipped strand DNAs. Antibodies included those raised against cruciform DNA, Z-DNA, single-stranded DNA and those specific for individual bases and native B-DNA (Figure 9E). Binding was determined using electrophoretic mobility-shift assays.^{17–19} We note that, since the repeat tracts range from 30 to 50 CTG/CAG repeats (90–150 bp) and structural epitopes may be repeated along this length, each DNA may bind single or multiple-Fab regions from individual or multiple antibodies. Such interactions would not be expected to result in distinct shifted complexes.

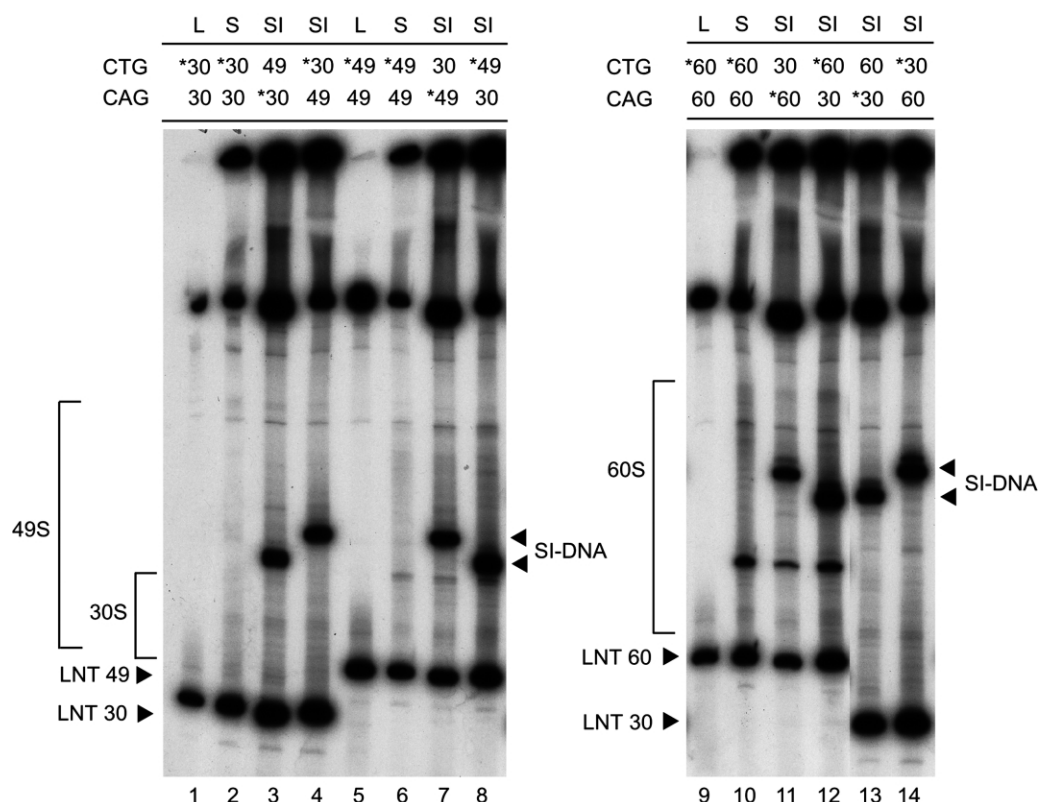


Figure 8. Similarity of SI-DNAs but not S-DNAs in SCA1 context. Both S-DNAs and SI-DNAs were prepared as for the DM1 clones. *Eco*NI/*Stu*I restriction digests of linear, S-DNA and SI-DNA mixtures of [(CTG)₃₀·(CAG)₃₀ × (CTG)₄₉·(CAG)₄₉] (lanes 1–8) and [(CTG)₃₀·(CAG)₃₀ × (CTG)₆₀·(CAG)₆₀] (lanes 9–14) uniquely ³²P-labeled at the *Eco*NI end of either the (CTG)_n or (CAG)_n strand (indicated by an asterisk). Reactions were analyzed on a 4% polyacrylamide gel, dried, and exposed for autoradiography.

Anti-cruciform antibody

Slipped-strand DNAs containing slip-outs that are imperfectly paired hairpin duplexes may contain defined three-way DNA junctions. To determine whether the major S-DNA and SI-DNA structures contained DNA junctions we used the anti-cruciform DNA antibody.^{18–20} The monoclonal 2D3 antibody has been extensively characterized to recognize three-way or four-way DNA junctions in a sequence-independent fashion.^{18,19} Both *n* = 30 and 50 S-DNAs were bound by 2D3, while the linear forms were not (Figure 9A, lanes 4 and 2; data not shown). The recognition of S-DNA by 2D3 was specific, since binding persisted in the presence of competitor plasmid DNA. Both sister SI-DNAs were efficiently bound by the anti-cruciform antibody (Figure 9A, lanes 6 and 8). These results support the presence of DNA junctions within these slipped repeat conformations. Furthermore, both S-DNAs and SI-DNAs yielded band-shifts that were relatively broader than the distinct shifts obtained from binding to molecules containing a unique three-way or four-way DNA junction.^{18,19} The broader distribution of the antibody–DNA complexes may reflect multiple antibodies bound at various sites and/or DNAs complexed with individual antibodies bound at

various locations, interpretations that would suggest that a single molecule could harbor multiple slip-outs and/or that slip-outs might occur throughout the repeat tract.

Base-specific anti-nucleoside antibodies

Slipped-strand DNAs may contain single-stranded regions involving runs of numerous bases (encompassing multiple repeat units in a random-coil conformation), a few nucleotides (as would occur in hairpin loops), or individual nucleotides such as T:T or A:A mismatches contained in hairpins formed by CTG or CAG repeats, respectively. To detect possible single-stranded regions or unpaired bases in S-DNAs and SI-DNAs we used several anti-nucleoside antibodies. Antibodies directed to individual bases can react with single-stranded DNAs in which the corresponding bases are exposed to solvent²¹ and detect small amounts of unpaired bases that may be present in a locally denatured region of otherwise helical DNA.²²

Each of the purified linear, S-DNA and SI-DNA repeat structures was probed with antibodies directed to specific bases. A polyclonal anti-thymidine serum, Ra492b²³ bound with higher affinity to S-DNA with 50 repeats than to its

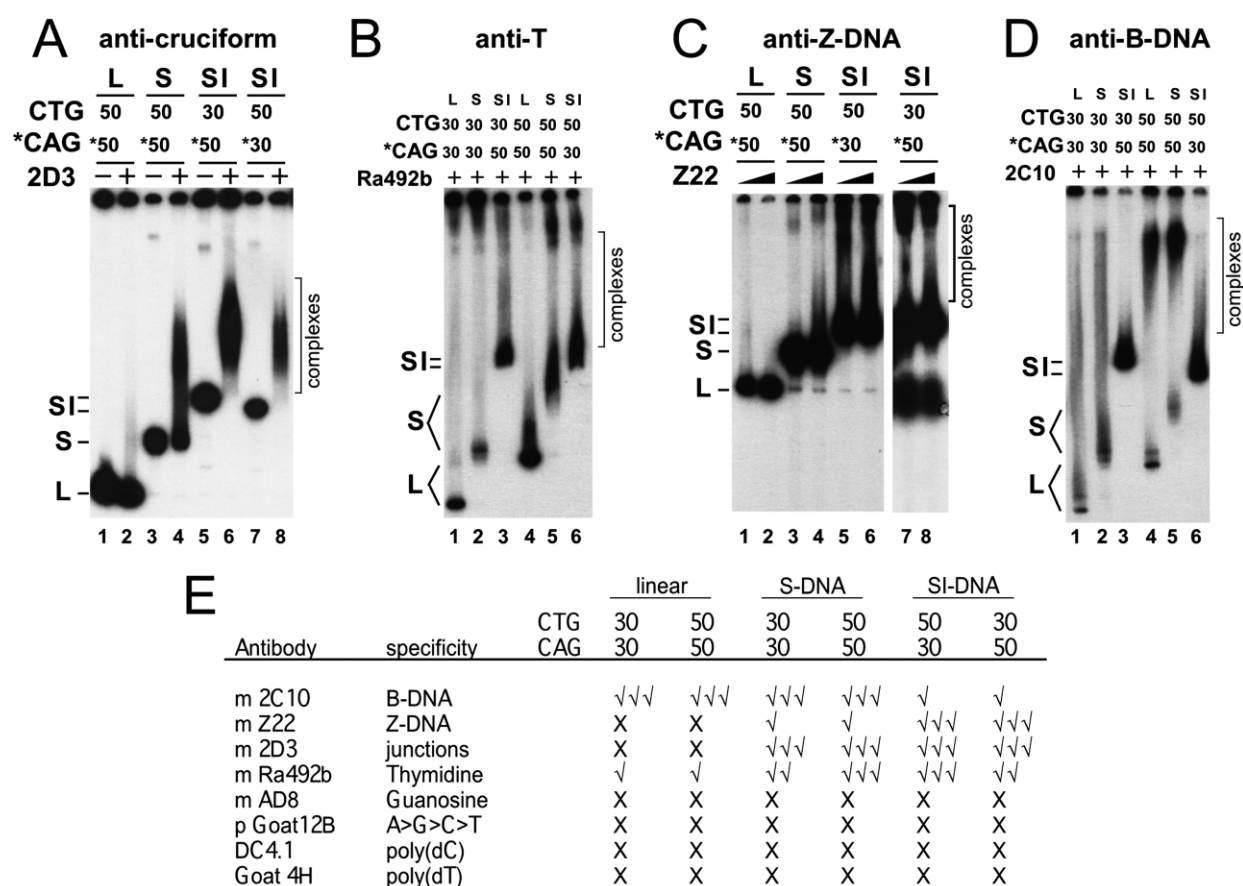


Figure 9. Binding of anti-DNA antibodies by mobility shift. A, Slipped-strand DNAs contain three-way DNA junctions. 32 P-labeled Linear (L), S-DNA (S) and SI-DNA (SI) substrates where $n = 30$ and/or 50 were incubated in the absence (-) or presence (+) of anti-cruciform DNA antibody 2D3 and analyzed by PAGE. High molecular mass DNA-2D3 complexes are indicated (square bracket). B, Slipped-strand DNAs are bound by the anti-thymidine antibody Ra492b. 32 P-labeled Linear (L), S-DNA (S) and SI-DNA (SI) substrates where $n = 30$ and/or 50 were incubated in the presence (+) of anti-thymidine antibody Ra492b and analyzed by PAGE. High molecular mass DNA-Ra492b complexes are indicated. C, Slipped-strand DNAs have Z-DNA character. 32 P-labeled Linear (L), S-DNA (S) and SI-DNA (SI) substrates where $n = 50$ and 30 were incubated with increasing amounts of anti-Z-DNA antibody Z22 and analyzed by PAGE. High molecular mass DNA-Z22 complexes are indicated. D, Anti-B-DNA antibody 2C10 binds strongly to both linear and S-DNAs but binds weakly to SI-DNA. 32 P-labeled Linear (L), S-DNA (S) and SI-DNA (SI) substrates where $n = 30$ and/or 50 were incubated with the anti-B-DNA antibody 2C10 and analyzed by PAGE. High molecular mass DNA-2C10 complexes are indicated. E, Summary of anti-DNA antibody binding to slipped DNAs. The antibodies used included those raised against native B-DNA (m2C10), Z-DNA (mZ22), cruciform DNA (m2D3) and individual bases (mRa492b, mAD8, pGoat12B, DC4.1 and Goat4H). Relative binding was determined using electrophoretic mobility-shift assays. (✓✓✓✓) very strong binding; (✓✓✓) strong binding; (✓✓) moderate binding; (✓) weak binding; (X) no binding; ns, data not shown.

corresponding linear form (Figure 9B, lanes 4 and 5). The S-DNA with 30 repeats bound with a slight preference over its corresponding linear form (Figure 9B, lanes 1 and 2). In addition, both sister SI-DNAs appeared to be preferentially recognized by the anti-thymidine serum (Figure 9B, lanes 3 and 6). However, a difference in binding affinity could not be detected between sister SI-DNAs with an excess of CTG or CAG repeats. We expected that the anti-thymidine activity would have detected differences between slipped-out CTG and slipped-out CAG repeats, as the latter would not contain any thymidine residues. It may be that unpaired T residues at the three-way slip-out junction of (CTG) $_{30}$ ·(CAG) $_{50}$ are bound with

similar affinity as are the unpaired T-residues contained in the hairpin loop of (CTG) $_{50}$ ·(CAG) $_{30}$.

The polyclonal serum Goat12B,²³ which has strong anti-adenosine and weaker anti-guanosine, anti-cytosine, and anti-thymidine activities did not bind to any of the DNAs (Figure 9E). Furthermore, the monoclonal anti-guanosine antibody AD8²³ also failed to bind to any of the DNA substrates. The lack of specific binding by either Goat12B or AD8 suggested that either the slipped-structures did not contain sufficient amounts of single-stranded regions or that in the single-stranded conformation the antigenic epitopes of the adenosine and guanosine were masked. We favor the latter interpretation, since mung bean nuclease analyses

of S-DNAs and SI-DNAs⁴ revealed a considerable amount of single-stranded regions.

Antibodies specific for poly(dC) or poly(dT) (dC4.1 and Goat4H, respectively)²³ failed to bind any of the (CTG)_n-(CAG)_n DNA conformations (Figure 9E). These antibodies were used as specificity controls; neither S nor SI-DNAs contain runs of C or T and hence would not be expected to be bound. We conclude that the (CTG)_n-(CAG)_n DNAs do not contain the tilted and dramatically twisted base-pairs, affecting depth and width of both major and minor grooves, that are typical of poly(dT-dA) and poly(dC-dG) sequences. Structural homology of these epitopes has not been observed in (CTG)_n-(CAG)_n repeat-containing DNA substrates.

Anti-Z-DNA antibodies

The structure of Z-DNA is well-characterized.²⁴ To test whether any of its structural features were shared by slipped-strand DNAs, we probed their surfaces with a well-characterized monoclonal antibody, Z22, which recognizes Z-DNA independent of nucleotide sequence.^{25,26} The fully base-paired linear forms of CTG/CAG repeats were not bound by Z22 and only weak binding occurred to the S-DNAs (Figure 9C, lanes 1 and 2; 3 and 4). In contrast, both sister SI-DNAs were bound by Z22 (Figure 9C, lanes 5 and 6 and lanes 7 and 8). These results support the presence of Z-DNA-like conformations in slipped DNAs, which might include the deformed sugar-phosphate backbone, the G8 residue or similarities to B-Z junctions, each of which are recognized by Z22.^{25,26}

Anti-B-DNA antibodies

As both a positive control and to assess the relative amounts of B-like character in each (CTG)_n-(CAG)_n DNA conformation, we probed their surfaces with monoclonal antibody 2C10, which preferentially recognizes native B-DNA.²⁷ Binding of 2C10 to the linear forms was observed as protein-DNA complexes that migrated more slowly in a gel-shift assay (Figure 9D, lanes 1 and 4). 2C10 bound with similar affinity to a DNA fragment of the PUC19 vector (data not shown), confirming that the recognition of the antibody is sequence-independent. Antibody 2C10 bound to S-DNAs composed of 30 repeats or 50 repeats and to their linear forms with similar affinity (Figure 9D, lanes 2 and 5). Much weaker binding was observed with both SI-DNA forms (Figure 9D, lanes 3 and 6). Since the non-repeating regions flanking the repeats in both S-DNAs and SI-DNAs are in a B-like configuration^{2,28} it is likely that the differential binding of 2C10 reflected the reduced B-like characteristics of the SI-DNAs. Both sister SI-DNAs may involve altered Watson-Crick base-pairing, as reflected by the poor binding of 2C10 and their relatively greater affinity for the anti-Z-DNA Z22 antibody.

Except for the anti-B-DNA autoantibody 2C10, the specific antibodies we tested were induced by immunization and recognized only the differences between the immunogen and native DNA; they do not react with naturally occurring native B-DNA. We note that a negative control using an anti-β-galactosidase antibody was performed to eliminate the possibility that the DNA fragments were recognized non-specifically by any antibody (data not shown).

Discussion

Structural differences between S-DNAs and SI-DNAs have been identified in this report: (1) for a given length of repeats, S-DNAs form numerous species of slipped isomers, while SI-DNAs form predominantly one distinct structure; (2) as determined by the reduced binding by anti-B antibodies, SI-DNAs contain an altered B-form double helix relative to either the S-DNA or linear forms; (3) SI-DNAs, unlike S-DNAs, may share more characteristics with left-handed Z-DNA; (4) differences between the sister SI-DNAs could not be detected by antibody binding, although their unique electrophoretic mobility and the known biophysical stability difference between intra-strand CTG and CAG hairpins suggest structural differences between the conformers.

The slower migration of SI-DNAs having an excess of CAG repeats indicated the existence of structural differences between sister SI-DNAs. There are four possible, but not mutually exclusive, explanations for the differential migration of sister SI-DNAs: first, the geometry and/or base-pairing scheme at the putative three-way slip-out junction may be significantly different between slipped-out CTG *versus* slipped-out CAG repeats; such differences can affect the angles subtended by the different arms of the slip-out junctions. Secondly, the conformation of the slipped-out CAG repeats may differ significantly from that of slipped-out CTG repeats. Third, the location of the slip-out-the point along the Watson-Crick duplex at which the slip-out extrudes, may differ between sister SI-DNAs. Lastly, the conformation of the flanking sequences in conjunction with the slip-out junction may affect the overall structure of the molecule. The evidence presented herein and elsewhere favors the first two explanations.

The conformation of the slipped-out CAG repeats has recently been shown to be different from that of the slipped-out CTG repeats.⁴ The slipped out CAG repeats can assume both random-coil and hairpin conformations, while slipped-out CTG repeats assume only the intra-strand hairpin conformations. These differences are likely to contribute to the slower migration of SI-DNAs having an excess of CAG repeats relative to those with an excess of CTG repeats.

The amount and pattern of S-DNA formed by a given length of (CTG)_n-(CAG)_n repeats varied

depending upon the sequence of the non-repeating DNAs flanking the repeat. The amount of S-DNA formed by repeats from the DM1 locus was considerably greater than that formed by similar lengths of repeats derived from the SCA1 locus. It has been reported that sequences flanking cruciforms³¹ and triplex forming tracts³² can affect the ability to extrude these structures. The manner through which flanking sequence may affect S-DNA formation is not clear but such differences may correlate with the levels of genetic instability displayed by CAG/CTG repeats at different disease loci.¹ It may be of interest that more S-DNA was formed by DM1 repeats which display high levels of genetic instability in humans, while less S-DNA was formed by SCA1 repeats, which, in humans, is not as genetically unstable as DM1 repeats.¹ Such an association may be related both to the propensity of S-DNA formation and/or the pattern or type of structures formed. Interestingly, flanking sequence (sequence context) is known to affect the biophysical structure of base-base mismatches and such structural differences correlate with differential abilities to be recognized and processed by bacterial mismatch repair.^{33–35} The ability of slipped-DNAs to be recognized and/or processed by human repair proteins is not known.

In contrast to S-DNAs, the formation of SI-DNAs and the slower migration of the sister with an excess of CAG repeats was unaffected by changing the flanking non-repeating sequences, thereby arguing against a role for flanking sequences in the number and size of slip-outs in SI-DNAs. These results provide further support for the conclusion that the differential migration of sister SI-DNAs was due to structural differences within the repeat tract and not to alterations within the non-repeats flanks as has been suggested for some sister three-way junctions.¹²

Anti-DNA antibody binding results suggest that slipped DNAs have structural features shared by DNA junctions, Z-DNA and un-paired thymidine residues. Anti-cruciform-DNA, anti-thymidine and anti-Z-DNA antibodies preferentially recognize slipped-strand DNA over linear duplex CTG/CAG DNAs. Interestingly, an anti-Z-DNA antibody bound preferentially to SI-DNAs and, to a lesser degree, to S-DNAs, indicating shared properties between Z-form DNA and slipped heteroduplexes. Each slipped DNA molecule probably contained multiple structural epitopes distributed along the repeat tract, as indicated by the “smeared” electrophoretic appearance of the antibody-DNA complexes, binding single or multiple-Fab regions from individual or multiple antibodies. While the stoichiometry of antibody to slipped DNAs is not known, our results clearly suggest that each molecule contains multiple sites and that these are distributed throughout the length of the repeat tract.

Recent nucleotide-resolution analysis of slipped structures using sequence-specific and structure-specific enzymatic probes⁴ revealed the presence

of three-way DNA junctions. Cruciform-binding proteins might recognize the stem, the loop, or the junction. It is known that binding by an anti-cruciform monoclonal antibody 2D3, requires junctions with fully base-paired stems.¹⁸ Three-way or four-way junctions are sufficient for its binding, making it a good probe for similar structures within slipped-strand DNA. Its recognition of S-DNA is therefore consistent with these models. Taken together, the results presented here are consistent with models suggested for S-DNA and SI-DNA, and provide further confirmation of some of the deduced features.

The lack of specific binding by either antibodies directed to adenosine, cytosine or guanine residues suggested that either the slipped-structures did not contain sufficient amounts of single-stranded regions or that the antigenic epitopes of these residues were masked or inaccessible to the antibodies. We favor the latter interpretation, since anti-nucleoside antibodies usually recognize bases in denatured DNAs very efficiently. Unpaired residues extruded from a hairpin or duplex may not extend far enough to be accessible to the anti-nucleoside antibodies.²³ The inability of the anti-A antibodies to detect the mispaired A:A bases in intra-strand hairpins formed by slipped-out CAG repeats may be similar to the inability of these mismatches to be cleaved by the single-strand nucleases mung bean and P₁ nucleases.^{4,8} The inaccessibility of the A:A residues may be due to the intra-helical stacking of the mismatched bases. We were surprised that the anti-A antibody did not bind either S or SI-DNAs, since we have previously shown that slipped-out CAG repeats have a more single-stranded character than slipped-out CTG repeats, which assume intra-strand hairpin. The CAG slip-outs were cut by mung bean nuclease at most ApG linkages but not at CpA linkages, possibly due to stacking of CpA.⁴ It is possible that the binding of A residues by anti-A antibodies is blocked in a similar fashion as is the CpA linkage from mung bean nuclease. We believe that the inability of the anti-A, C, and G antibodies to bind slipped-DNAs reflects inaccessibility of the bases by these antibodies.

Binding of an anti-Z-DNA antibody to slipped-DNAs suggests that they share structural features with Z-DNA. With guanosine residues in the uncommon *syn* conformation, Z-DNA contains sugar-phosphate backbones rotated 90° with respect to the native *B*-form and thus forms a zig-zag-like left-handed helix, which results in offset stacking of base-pairs and a change in tilt angles.²⁴ Z-22 protects four nucleotides in continuous Z-DNA²⁵ and makes minimal contacts with the bases of the helix, contacting the G8 residue, which is very near the sugar-phosphate backbone.¹⁷ Anti-Z binds better to SI than S, may be due to the longer lengths of the slip-outs in SI-DNAs, suggesting that Z22 interacts with the slipped-out DNA, rather than the slip-out junction.

The anti-B-DNA antibody, 2C10, bound linear

and S-form (CTG)_n-(CAG)_n DNAs with similar affinity, indicating a high proportion of B-duplex in the S-DNAs. 2C10 recognizes the double-helical structure of DNA and preferentially binds dA:dT over dG:dC base-pairs.²⁷ Since competition assays with synthetic oligonucleotides indicated that 2C10 could bind well to alternating dA:dT, to stretches of (dA)_n-(dT)_n, or to (dA:dG)_n-(dC:dT)_n sequences³⁶ the reduced affinity of sister (CTG)_n-(CAG)_n SI-DNAs for the 2C10 antibody further supported the suggestion that SI-DNAs involve altered Watson–Crick base-pairing.

Implications

S-DNAs and SI-DNAs formed by (CTG)_n-(CAG)_n repeats are alternative DNA structures stable enough to exist under physiological conditions. The formation of these structures correlates with the length of trinucleotide repeats, a factor that determines their genetic stability. Their involvement in the disease-associated dynamic mutagenesis is therefore possible. Here, the molecular features of S-DNA and SI-DNA have been inferred by probing with anti-DNA antibodies. Knowing the specific recognizable epitopes in such structures may facilitate the search for cellular proteins with binding capability. Certain cellular proteins are known to recognize specific-DNA conformations including cruciforms³⁷ Z-DNA²³ and specific bases in denatured DNA.²³ Proteins that can recognize DNA junctions include members of the ubiquitous HMG protein family, including HMG1, the testis-determining factor SRY³⁸ and the transcriptional activator NHP6A/B.³⁹ Z-DNA-binding proteins include the RNA-editing enzyme ADAR1²⁴ the *Ustilago rec1* protein⁴⁰ and HPP-1, a human protein involved in homologous pairing.⁴¹ It is possible that structural elements in S-DNA and SI-DNA are recognized by cellular proteins which may directly or indirectly be involved in metabolism of the trinucleotide repeats. In fact, a human mismatch repair protein, hMSH2, has been shown to bind preferentially to these structures prepared *in vitro*⁴², suggesting a possible role for the DNA mismatch repair pathway in CTG/CAG repeat instability. Having a detailed structural picture of S-DNA and SI-DNA may facilitate screening for more of such potentially important cellular components that are involved in trinucleotide repeat instability.

Materials and Methods

Plasmids/DNA

Plasmids containing human DM1 genomic (CTG)_n-(CAG)_n repeats ($n = 17, 30, 50$ or 79) and human non-repeating sequences flanking the repeat (sites 417–436 and 451–494 from accession number S86455) have been described.^{2,4,28} The repeat tracts are pure. The repeat tract is flanked by unique *Eco*RI and *Hind*III restriction sites 59 bp and 54 bp 3' and 5' of the CTG tract, respectively.

Plasmids harboring the human SCA1 genomic (CAG)_n-(CTG)_n repeats ($n = 30, 49, 60$ or 74) with human non-repeating sequences flanking the repeat (sites 936–1524 and 1614–3387 from accession number X79204) have been described.³ The repeat tracts are pure. The repeat tracts are flanked by unique *Eco*NI and *Stu*I restriction sites 358 bp and 62 bp 3' and 5' of the CAG tract, respectively. Plasmids were prepared from *Escherichia coli* cells, as described.²

DNA treatments

DM1 plasmids were linearized with *Hind*III then radiolabeled on the 5' or 3' ends with phage T4 PNK (USB) and [γ -³²P]ATP (NEN) or AMV reverse transcriptase (USB) and [α -³²P]dATP, respectively. Samples were reduplexed or heteroduplexed, as described (see below), and the repeat-containing fragment was then liberated by a secondary restriction digestion with *Eco*RI. For some experiments plasmids were linearized with *Eco*RI, radiolabeled, and released by *Hind*III or *Dde*I digestion. Reaction products were resolved on preparative 4% (w/v) polyacrylamide gels. Similarly, for the SCA1 plasmids both S-DNAs and SI-DNAs were prepared as above, except plasmids were linearized with *Eco*NI, ends radiolabeled and repeat-containing fragments released by *Stu*I digestion.

Structure formation

Homoduplex slipped S-DNAs of repeat-containing DNAs with a given length of repeats uniquely ³²P-labeled on the CTG or CAG strand, were formed by alkaline denaturation/renaturation as described in detail.^{4,7,28,42} Linearized plasmids uniquely ³²P-labeled on the CTG or CAG strand were dried in a speed vacuum, then denatured in a solution of 500 mM NaOH (yielding pH 13) and 0.5 M NaCl at room temperature for five minutes. Samples were then neutralized by the addition of a 50-fold volume of 50 mM Tris–HCl (pH 8), 5 mM EDTA, resulting in a final solution with 0.01 M NaOH and 0.03 M NaCl (pH 8). These conditions favor full renaturation. Samples were then incubated at 68 °C for three hours, followed by two precipitations in ethanol. Longer incubations did not change the pattern of products formed. Caution was taken to avoid sample dehydration. Heteroduplex slipped intermediates (SI-DNAs) were prepared as follows: DNAs uniquely ³²P-labeled on the (CTG)_x or (CAG)_x strand were mixed with an equimolar amount of unlabeled (CTG)_y-(CAG)_y and heteroduplexed by denaturation/renaturation (as above). Similarly, DNAs uniquely ³²P-labeled on the (CTG)_y or (CAG)_y strand, were mixed with an equimolar amount of unlabeled (CTG)_x-(CAG)_x and heteroduplexed by denaturation/renaturation (as above).

Gel electrophoresis and migration analysis

Native 4% polyacrylamide gels (acrylamide to bis-acrylamide, 40:2, w/w) were cast in TBE buffer and run at a constant voltage of 10–12 V/cm. All gels, except those in Figure 7C, were 13 cm × 9 cm × 0.15 cm. The ultra-high-resolution gels of Figure 7C, were identical having the dimensions 39 cm × 30.5 cm × 0.15 cm. Gels were dried on Whatman paper and exposed to radiographic film (Kodak). Loaded onto all gels was a DNA size marker, the 123 bp ladder (Gibco BRL). Relative electrophoretic migration was analyzed relative to this

marker. Details of assessment are indicated in the legends to Figures 1–3.

The effect of magnesium upon the differential migration of sister SI-DNAs was tested as follows: slipped DNAs were prepared as above and resuspended in Tris (pH 8). Magnesium was added to a final concentration of 200 μ M. Samples with or without the metal were incubated at room temperature for at least 15 minutes, then electrophoresed on Tris–borate gels perfused with 200 μ M magnesium or on Tris–borate EDTA gels, respectively. Gels were electrophoresed at room temperature with recirculation from the top to the bottom at a rate of 100 ml/minute.

Antibody-binding assays

The binding specificities of various anti-DNA antibodies were determined by gel mobility-shift assays. For the antibody-binding experiments the bands corresponding to the main DM1 *Eco*RI/*Hind*III linear, S-DNA, and SI-DNA forms were excised and purified by electroelution as described.²⁸ Linear forms were gel-purified from samples that had never been denatured, the S-DNA forms were purified from individual DNAs that had been reduplexed and only the SI-DNAs were purified from heteroduplexed DNA samples. Approximately 1 ng of radiolabelled DNA substrate was incubated with various amounts of antibodies in borate buffer (pH 7.6) or phosphate buffer (pH 7.5) for 15–30 minutes on ice. The assay mixture was electrophoresed in 1 \times TBE buffer on a 4% (w/v) polyacrylamide gel (acrylamide to bis-acrylamide, 19:1, w/w), at a constant 150 V. Following electrophoresis gels were dried and autoradiographed. Appropriate negative controls included use of anti- β -galactosidase antibody (a kind gift from Dr J. Rommens), and *Hind*III linearized PUC19 plasmid DNA were included in some assays to test for specificity of binding. The relative binding efficiency of each repeat structural conformation by different anti-DNA antibodies was determined by densitometric analysis of individual gel-shifted regions relative to the un-bound material within a lane. Densitometric readings of multiple gels were converted to a number of $\sqrt{\text{marks}}$ shown in Figure 9E. Relative sensitivities between structures to each antibody were determined by comparing the densitometric analyses of individual gel bands relative to the un-bound material. In this fashion, the relative bar lengths between strands were normalized. Such comparisons were possible because reaction conditions (DNA and antibody concentrations and reaction times) were essentially identical.

Acknowledgements

We thank our lab members for their discussions. We thank Dr J. Rommens for control anti- β -galactosidase antibodies. The authors note with regret the death of our co-author and colleague Gerald B. Price, during the course of this work and dedicate this work to him. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) and the Muscular Dystrophy Association USA (C.E.P.). C.E.P. is a CIHR Scholar and a Canadian Genetic Disease Network Scholar.

References

1. Cleary, J. D. & Pearson, C. E. (2003). The contribution of *cis*-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.* In the press..
2. Pearson, C. E. & Sinden, R. R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
3. Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L. & Sinden, R. R. (1998). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry*, **37**, 2701–2708.
4. Pearson, C. E., Tam, M., Wang, Y.-H., Montgomery, S. E., Dar, A., Cleary, J. D. & Nichol, K. (2002). Slipped-strand DNAs formed by long (CAG)(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucl. Acids Res.* **30**.
5. Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. (2002). Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet.* **31**, 37–46.
6. Panigrahi, G. B., Cleary, J. D. & Pearson, C. E. (2002). *In vitro* (CTG) \cdot (CAG) expansions and deletions by human cell extracts. *J. Biol. Chem.* **277**, 13926–13934.
7. Pearson, C. E., Wang, Y. H., Griffith, J. D. & Sinden, R. R. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG) n -(CAG) n repeats from the myotonic dystrophy locus. *Nucl. Acids Res.* **26**, 816–823.
8. Pearson, C. E. & Sinden, R. R. (1998). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr. Opin. Struct. Biol.* **8**, 321–330.
9. Chastain, P. D., II, Eichler, E. E., Kang, S., Nelson, D. L., Levene, S. D. & Sinden, R. R. (1995). Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry*, **34**, 16125–16131.
10. Chastain, P. D. & Sinden, R. R. (1998). CTG repeats associated with human genetic disease are inherently flexible. *J. Mol. Biol.* **275**, 405–411.
11. Gaillard, C. & Strauss, F. (1994). Association of poly-(CA)-poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2. *Science*, **264**, 433–436.
12. Oussatcheva, E. A., Shlyakhtenko, L. S., Glass, R., Sinden, R. R., Lyubchenko, Y. L. & Potaman, V. N. (1999). Structure of branched DNA molecules: gel retardation and atomic force microscopy studies. *J. Mol. Biol.* **292**, 75–86.
13. Duckett, D. R., Murchie, A. I. & Lilley, D. M. (1990). The role of metal ions in the conformation of the four-way DNA junction. *EMBO J.* **9**, 583–590.
14. Welch, J. B., Duckett, D. R. & Lilley, D. M. (1993). Structures of bulged three-way DNA junctions. *Nucl. Acids Res.* **21**, 4548–4555.
15. Wetmur, J. G. & Davidson, N. (1968). Kinetics of renaturation of DNA. *J. Mol. Biol.* **31**, 349–370.
16. Gough, G. W. & Lilley, D. M. (1985). DNA bending induced by cruciform formation. *Nature*, **313**, 154–156.
17. Sanford, D. G. & Stollar, B. D. (1992). Assay of anti-DNA antibodies. *Methods Enzymol.* **212**, 355–371.
18. Frappier, L., Price, G. B., Martin, R. G. & Zannis-Hadjopoulos, M. (1987). Monoclonal antibodies to cruciform DNA structures. *J. Mol. Biol.* **193**, 751–758.

19. Frappier, L., Price, G. B., Martin, R. G. & Zannis-Hadjopoulos, M. (1989). Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. *J. Biol. Chem.* **264**, 334–341.
20. Steinmetzer, K., Zannis-Hadjopoulos, M. & Price, G. B. (1995). Anti-cruciform monoclonal antibody and cruciform DNA interaction. *J. Mol. Biol.* **254**, 29–37.
21. Stollar, B. D. (1986). Antibodies to DNA. *CRC Crit. Rev. Biochem.* **20**, 1–36.
22. Sundquist, W. I., Lippard, S. J. & Stollar, B. D. (1986). Binding of *cis*- and *trans*-diaminedichloroplatinum(II) to deoxyribonucleic acid exposes nucleosides as measured immunochemically with anti-nucleoside antibodies. *Biochemistry*, **25**, 1520–1524.
23. Stollar, B. D. (1988). Immunological approaches to the definition of unusual nucleic acid structures. In *Unusual DNA Structures* (Wells, R. A., ed.), pp. 253–265, Springer, New York.
24. Herbert, A. & Rich, A. (1999). Left-handed Z-DNA: structure and function. *Genetica*, **106**, 37–47.
25. Runkel, L. & Nordheim, A. (1986). Conformational DNA transition in the *in vitro* torsionally strained chicken beta-globin 5' region. *Nucl. Acids Res.* **14**, 7143–7158.
26. Sanford, D. G. & Stollar, B. D. (1990). Characterization of anti-Z-DNA antibody binding sites on Z-DNA by nuclear magnetic resonance spectroscopy. *J. Biol. Chem.* **265**, 18608–18614.
27. Kubota, T., Akatsuka, T. & Kanai, Y. (1986). A monoclonal anti-double stranded DNA antibody from an autoimmune MRL/Mp-lpr/lpr mouse: specificity and idiotype in serum immunoglobulins. *Immunol. Letters*, **14**, 53–58.
28. Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. & Sinden, R. R. (1997). Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117–1123.
29. Lu, M., Guo, Q. & Kallenbach, N. R. (1991). Effect of sequence on the structure of three-arm DNA junctions. *Biochemistry*, **30**, 5815–5820.
30. van Buuren, B. N., Overmars, F. J., Ippel, J. H., Altona, C. & Wijmenga, S. S. (2000). Solution structure of a DNA three-way junction containing two unpaired thymidine bases. Identification of sequence features that decide conformer selection. *J. Mol. Biol.* **304**, 371–383.
31. Sullivan, K. M. & Lilley, D. M. (1986). A dominant influence of flanking sequences on a local structural transition in DNA. *Cell*, **47**, 817–827.
32. Kang, S., Wohlrab, F. & Wells, R. D. (1992). GC-rich flanking tracts decrease the kinetics of intramolecular DNA triplex formation. *J. Biol. Chem.* **267**, 19435–19442.
33. Werntges, H., Steger, G., Riesner, D. & Fritz, H. J. (1986). Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies. *Nucl. Acids Res.* **14**, 3773–3790.
34. Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H. *et al.* (1986). Structures of mismatched base pairs in DNA and their recognition by the *Escherichia coli* mismatch repair system. *EMBO J.* **5**, 3697–3703.
35. Hunter, W. N., Brown, T. & Kennard, O. (1986). Structural features and hydration of d(C-G-C-G-A-A-T-T-A-G-C-G); a double helix containing two G-A mispairs. *J. Biomol. Struct. Dynam.* **4**, 173–191.
36. Kubota, T., Watanabe, N., Kanai, Y. & Stollar, B. D. (1996). Enhancement of oxidative cleavage of DNA by the binding sites of two anti-double-stranded DNA antibodies. *J. Biol. Chem.* **271**, 6555–6561.
37. Pearson, C. E., Ruiz, M. T., Price, G. B. & Zannis-Hadjopoulos, M. (1994). Cruciform DNA binding protein in HeLa cell extracts. *Biochemistry*, **33**, 14185–14196.
38. Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. & Bianchi, M. E. (1994). Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 6115–6124.
39. Paull, T. T., Carey, M. & Johnson, R. C. (1996). Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation *in vivo* and assembly of preinitiation complexes *in vitro*. *Genes Dev.* **10**, 2769–2781.
40. Kmiec, E. B., Angelides, K. J. & Holloman, W. K. (1985). Left-handed DNA and the synaptic pairing reaction promoted by *Ustilago* rec1 protein. *Cell*, **40**, 139–145.
41. Fishel, R. A., Detmer, K. & Rich, A. (1988). Identification of homologous pairing and strand-exchange activity from a human tumor cell line based on Z-DNA affinity chromatography. *Proc. Natl Acad. Sci. USA*, **85**, 36–40.
42. Pearson, C. E., Zorbas, H., Price, G. B. & Zannis-Hadjopoulos, M. (1996). Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J. Cell. Biochem.* **63**, 1–22.
43. Assenberg, R., Weston, A., Cardy, D. L. N. & Fox, K. R. (2002). Sequence-dependent folding of DNA three-way junctions. *Nucl. Acids Res.* **30**, 5142–5150.

Edited by P. J. Hagerman

(Received 1 April 2003; received in revised form 7 July 2003; accepted 8 July 2003)