Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA

RICHARD R SINDEN†, VLADIMIR N POTAMAN, ELENA A OUSSATCHEVA, CHRISTOPHER E PEARSON*, YURI L LYUBCHENKO** and LUDA S SHLYAKHTENKO**

Laboratory of DNA Structure and Mutagenesis, Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University System Health Sciences Center, 2121 West Holcombe Blvd., Houston, TX 77030-3303, USA

**Hospital for Sick Children, Department of Genetics, 555 University Avenue, Elm Wing, Rm. 11-135, Toronto, Ontario, Canada M5G 1X8*

***Departments of Biology and Microbiology, Arizona State University, Tempe, AZ 85287-2701, USA*

†*Corresponding author (Fax, 713-677-7689; Email, Rsinden@IBT.TAMU.edu).*

Fourteen genetic neurodegenerative diseases and three fragile sites have been associated with the expansion of $(CTG)_n \bullet (CAG)_n$, $(CGG)_n \bullet (CCG)_n$, or $(GAA)_n \bullet (TTC)_n$ repeat tracts. Different models have been proposed for the expansion of triplet repeats, most of which presume the formation of alternative DNA structures in repeat tracts. One of the most likely structures, slipped strand DNA, may stably and reproducibly form within triplet repeat sequences. The propensity to form slipped strand DNA is proportional to the length and homogeneity of the repeat tract. The remarkable stability of slipped strand DNA may, in part, be due to loop-loop interactions facilitated by the sequence complementarity of the loops and the dynamic structure of three-way junctions formed at the loop-outs.

[Sinden R R, Potaman V N, Oussatcheva E A, Pearson C E, Lyubchenko Y L and Shlyakhtenko L S 2002 Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA; *J. Biosci*. (*Suppl. 1*) **27** 53–65]

1. Introduction

1.1 *Triplet repeats, human disease and alternative DNA structures*

Since 1991, fourteen genetic neurodegenerative diseases and three fragile sites have been associated with the expansion of $(CTG)_n \bullet (CAG)_n$, $(CGG)_n \bullet (CCG)_n$, or $(GAA)_n \bullet (TTC)_n$ repeat tracts. These diseases are listed in table 1 with information about repeat lengths and the possible molecular biochemical defects leading to disease. The occurrence of these "triplet repeat diseases" within populations ranges from fairly common [Fragile X syndrome and myotonic dystrophy type 1 (DM1)] to rare (Dentatorubral-pallidoluysian atrophy). Moreover, the frequency of occurrence varies among populations. For example, spinocerebellar ataxia type 2 (SCA2) is the most common of these diseases in India (Saleem *et al* 2000).

Among the triplet repeat diseases, the position of the repeats with respect to the gene coding region affected varies. The repeats are found either 5′ of the gene, within the coding region (in an exon), within an intron, or 3′ of the gene (figure 1). Given this difference in location, the role of the repeat in the molecular etiology of the disease is expected to be different in different diseases. The $(CGG)_n$ repeat-associated diseases appear to involve methylation of the DNA containing the expanded CGG repeats, leading to gene silencing. In DM, SCA8, and SCA12, the 3′ localization of the expanded repeat may lead to aberrant mRNA processing or the aberrant expres-

Keywords. DNA hairpins; DNA structure; slipped strand DNA; triplet repeat disease; triplet repeats

sion of flanking genes. The diseases involving genes containing $(CAG)_n$ tracts in coding regions form a group in which the triplet repeat encodes a polyglutamine tract. The molecular etiology involves production of a protein with an expanded glutamine tract, and in some mouse models a polyglutamine tract alone is neurotoxic (Heintz and Zoghbi 2000).

There are several remarkable features associated with this class of diseases. First, the expansion of the triplet repeat tract represents a novel type of mutation. Prior to its discovery in Fragile X syndrome (Fu *et al* 1991; Kremer *et al* 1991) and myotonic dystrophy (Brook *et al* 1992; Fu *et al* 1992; Mahadevan *et al* 1992), this mutation had not been observed in simple model systems traditionally used for the study of mutations such as the bacteriophage T4, *Escherichia coli*, *Drosophila* and yeast. Secondly, although their genetic instability has been now extensively examined in *E. coli* and yeast, these organisms fail to exhibit the propensity for expansion found in humans. Although expansions can be observed, deletions of the repeat tracts are the predominant events. Thus, to date, the expansion mutations associated with the disease, especially the very large intergenerational expansions, remain a human phenomenon. Recently, however, age-dependent large expansions have been reported in mice (Fortune *et al* 2000). Third, the nature of the mutation, which expands in successive generations, leads to non Mendelian genetics. These diseases exhibit *anticipation*, in which the penetrance and severity of the disease increase and the age of onset decreases in successive generations. There is no therapeutic strategy for preventing or slowing the expansion of triplet repeats, and thus, no approach to slowing or preventing the onset of these diseases. While the molecular biological or biochemical defects causing the disease manifestations vary within this class of 14 triplet repeat diseases, the same underlying genetic mutation (triplet repeat expansion) is ultimately responsible for all diseases.

Different models have been proposed for the expansion of triplet repeats (Pearson and Sinden 1998a; Sinden 1999). Expansions or deletions can occur by simple

Table 1. Trinucleotide repeats in human genetic disease.

Disease	Gene	Locus	Repeat[a]	Repeat length			Protein/possible biological effect of expansion
				Normal	Pre-mutation	Disease	
Fragile X syndrome	FMR1 (FRAXA)	Xq37·3	$(CGG)_n$	6–52	59–230	230–2000 (pure)	FMR1 protein (FMRP)/promoter methylation, gene silencing, loss of FMRP function
Fragile XE syndrome	FMR2 (FRAXE)	Xq28	$(CCG)_n$	4–39	? (31–61)	200–900	FMR2 protein/methylation
None, fragile site	Fragile XF (FRAXF)		$(CGG)_n$	7–40	?	306–1008	Methylation
None, fragile site	FRA16A		$(CCG)_n$	16–49	?	1000–1900	Methylation
None, Jacobsen syndrome	(FRA11B)		$(CGG)_n$	11	80	100–1000	Methylation
Spinobulbar muscular atrophy (SBMA) (Kennedy's disease)	AR	Xq13–21	$(CAG)_n$	14–32	?	40–55	Androgen receptor (AR)/polyglutamine tract expansion
Myotonic dystrophy type 1 (DM1)	DMPK	19q13	$(CTG)_n$	5–37	50–80	80–1000	Myotonic dystrophy protein kinase/altered mRNA processing and transport, altered gene expression due to chromatin changes
Huntington disease	HD	4p16·3	$(CAG)_n$	10–34	36–39	40–121	Huntingtin/polyglutamine expansion
Spinocerebellar ataxia 1	SCA1	6p23	$(CAG)_n$	6–44	–	39–82 (pure)	Ataxin-1/polyglutamine expansion
Spinocerebellar ataxia 2	SCA2	12q24·1	$(CAG)_n$	14–31	–	34–59 (pure)	Ataxin-2/polyglutamine expansion
Spinocerebellar ataxia 3 (Machado Joseph disease)	SCA3	14q32·1	$(CAG)_n$	13–44	NA	55–84	Ataxin-3/polyglutamine expansion
Spinocerebellar ataxia 6	SCA6	19p13	$(CAG)_n$	4–18	NA	21–33	a_{1A}-voltage-dependent calcium channel subunit/polyglutamine expansion
Spinocerebellar ataxia 7	SCA7	13p12–13	$(CAG)_n$	4–34	NA	37–306	Ataxin-7/polyglutamine expansion
Spinocerebellar ataxia 8	SCA8	13q21	$(CTG)_n$	15–27?	110–200		3′ to SCA8gene/alters gene expression of adjacent genes?
Spinocerebellar ataxia 12	SCA12	5q31–33	$(CAG)_n$	7–28	?	66–78	5′ to SCA12 gene/alters gene expression?
Dentatorubral-pallidoluysian atrophy (Haw River syndrome)	DRPLA	12p13·31	$(CAG)_n$	7–25	?	49–75	Polyglutamine expansion
Friedreich ataxia (FRDA)	X25	9q13–21·1	$(GAA)_n$	6–29	? (> 34–40)	200–900	Frataxin/repeat tract in an intron; altered mRNA production, altered replication

[a]Typically repeat tracts contain sequence interruptions. See Pearson and Sinden (1998a) for a discussion of the sequence interruptions. ?, Potential mutagenic intermediate length. Not all diseases are associated with a premutation length repeat tract or premutation disease condition. –, None; NA, Not applicable.

primer template misalignment, in which the 3′ end of the nascent primer unpairs from the template and then rehybridizes with repeats upstream or downstream of the site of unwinding. Slippage upstream will lead to an expansion, while slippage downstream will lead to a deletion. Genetic recombination or gene conversion can also lead to expansion. A single slippage or recombination event will only lead at the most to a doubling of the repeat length. In many diseases expansion occurs by a factor of 10 or more, suggesting that multiple slippage or recombinational events must occur. Reiterative DNA synthesis, involving repetitive slippage events (Kornberg *et al* 1964) can lead to expansions longer than the length of the original repeat tract (Schlotterer and Tautz 1992; Petruska *et al* 1998; Hartenstine *et al* 2000). Sinden and Wells (1992) suggested that massive expansion may result from reiterative DNA synthesis mediated by a physical block to replication. The physical block to replication may involve an alternative DNA structure formed from the triplet repeats (Brahmachari *et al* 1995; Pearson and Sinden 1998b; Sinden 1999). Triplet repeats present a block to replication in *E. coli* in an orientation dependent fashion (Samadashwily *et al* 1997).

Alternative DNA structures differ from the canonical B-form DNA structure first described in 1953 (Watson and Crick 1953). Alternative structures include bent DNA, flexible DNA, parallel DNA, triplex DNA, quadruplex DNA, hairpins, cruciform structures, left-handed Z-DNA, and slipped stranded DNA (figure 2). While the list of alternative structures is large (and this list is by no means complete), not all DNA sequences can form alternative DNA structures. Specific sequence symmetry elements are required for the formation of certain alternative structures (table 2). This paper will briefly review work on the myriad of DNA structures formed from triplet repeats associated with human diseases. It will then focus on recent work from our laboratory on the structure of slipped strand DNA formed from $(CAG)_n \bullet (CTG)_n$ and $(CGG)_n \bullet (CCG)_n$ repeats associated with many neurodegenerative diseases.

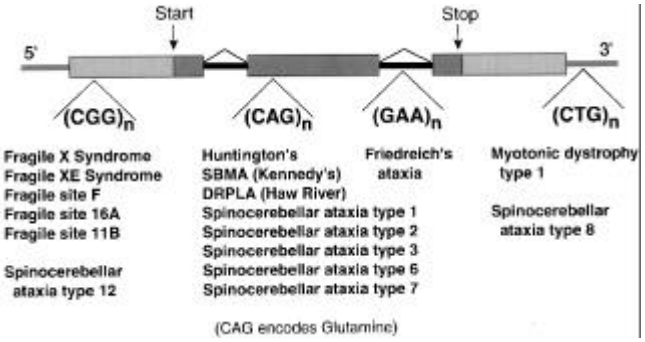

Figure 1. Location of repeat tract in triplet repeat diseases genes. An idealized gene is shown with the dark shaded area representing the coding region. The transcription start and stop are indicated. The light shaded area represents the 5′ upstream and 3′ downstream untranslated regions (UTR). Introns are shown as black lines and the flanking region as a shaded line. Both disease-associated $(CGG)_n$ triplet repeats and three fragile sites are located in the 5′ UTR. The $(CAG)_n$ repeat in SCA-12 is also located in the 5′ UTR. Many glutamine encoding CAG repeats are located within exons of the gene coding regions. The $(GAA)_n$ repeat in Friedreich's ataxia is located in an intron. The $(CTG)_n$ repeats in myotonic dystrophy type 1 and SCA-8 are located in the 3′ UTR of associated genes.

1.2 *Alternative DNA structures formed within triplet repeats*

Trinucleotide repeat sequences possess sequence motifs and symmetry elements that allow an increased flexibility of double-helix and the formation of several structures alternative to double-stranded DNA (table 2). These include single-stranded hairpins, triplex and quadruplex DNA, and slipped-strand DNA. Studies have also confirmed that the formation of alternative structures occurs in trinucleotide repeat sequences.

Linear DNA molecules containing triplet repeats have unusual helical parameters that become pronounced upon accumulation over a long stretch of repeats. The anomalously fast migration of trinucleotide repeat sequence-containing DNA fragments in polyacrylamide gel (Chastain *et al* 1995) was attributed to the increased flexibility of $(CAG)_n \bullet (CTG)_n$ and $(CGG)_n \bullet (CCG)_n$ double helices

Table 2. DNA repeats, symmetry elements and alternative DNA structures.

Repeat	Symmetry elements[a]	Alternative DNA structures[b]
$(GT)_n \bullet (AC)_n$	DR	Z-DNA
$(GC)_n \bullet (GC)_n$	DR, IR	Hairpins (ss), Z-DNA
$(AT)_n \bullet (AT)_n$	DR, IR	Hairpins (ss), cruciforms
$(CTG)_n \bullet (CAG)_n$	DR, QP	Flexible helix, mismatched hairpins (ss), slipped strand DNA
$(CGG)_n \bullet (CCG)_n$	DR, QP	Flexible helix, mismatched hairpins (ss), slipped strand DNA, quadruplex DNA
$(GAA)_n \bullet (TTC)_n$	DR, MR	Intramolecular triplex DNA, slipped strand DNA?
$(GGGGTT)_n$	DR	Quadruplex DNA (ss)

[a]DR, Direct Repeat; IR, Inverted Repeat; QP, Quasipalindrome (imperfect inverted repeat); MR, Mirror Repeat (see Sinden 1994) for definitions.
[b]Structures are those formed in double stranded DNA, unless indicated for single stranded DNA (ss).

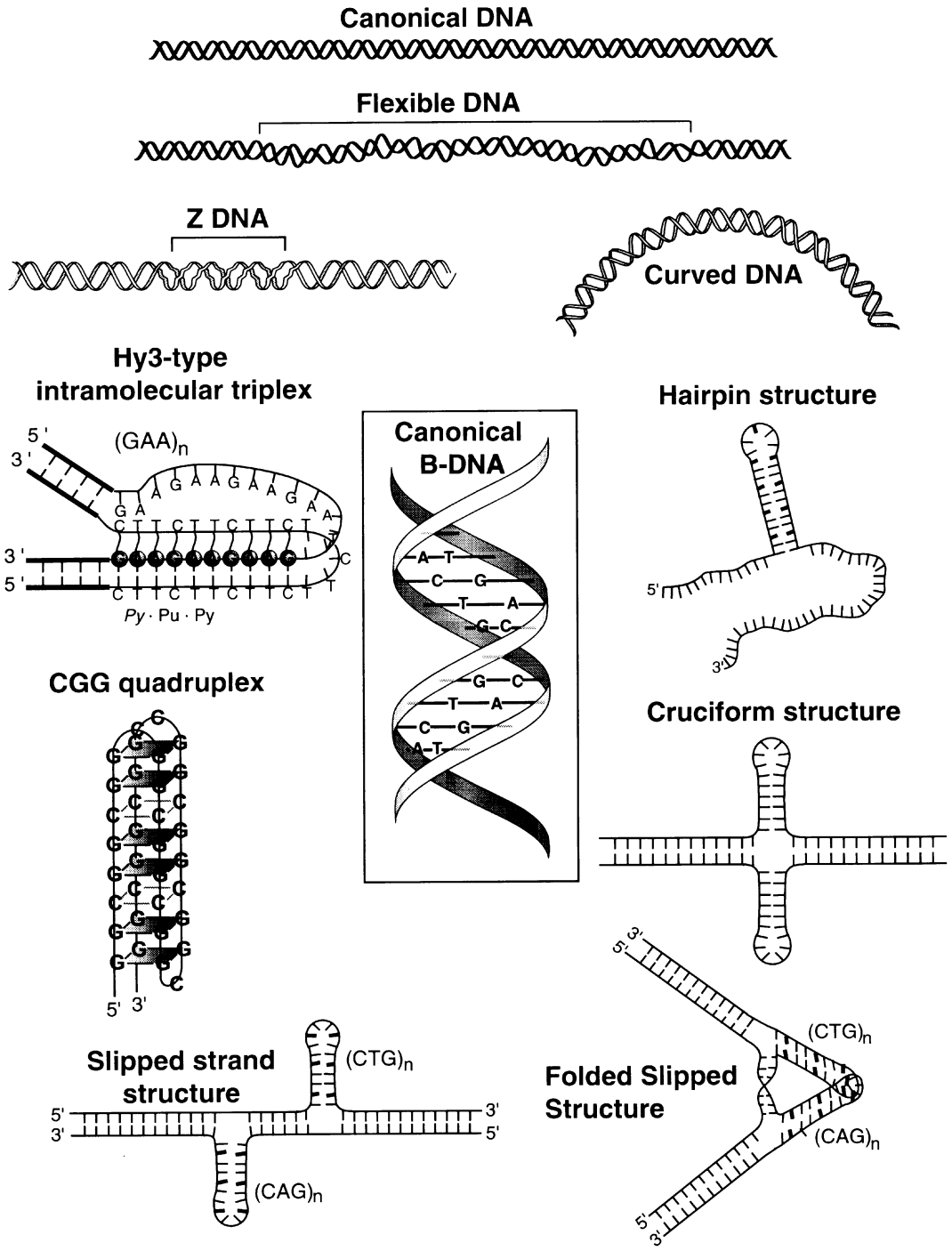

Figure 2. Alternative DNA structures formed in repeated DNAs. A canonical B-form helix is shown in a box in the center of the figure. A representation of this helix is also shown at the top of the figure. Although a straight molecule is shown, the DNA inherently possesses a certain degree of flexibility. $(CTG)_n$•$(CAG)_n$ and $(CGG)_n$•$(CCG)_n$ repeats, however, are extremely flexible compared to mixed sequence B-form DNA (shown below canonical DNA). The next two representations, Z-DNA and stable curved DNA, occur in $(CG)_n$•$(CG)_n$ and $(GT)_n$•$(AC)_n$ runs and DNA containing phased A_{4-5}•T_{4-5} tracts (and other sequences), respectively, but have not been identified in triplet repeats. Proceeding clockwise, hairpin structures form in most disease-relevant triplet repeat single strands (to some degree). Cruciform structures form in perfect inverted repeats and have not been identified in triplet repeat sequences. Folded slipped strand structures and simple slipped strand structures form in $(CTG)_n$•$(CAG)_n$ and $(CGG)_n$•$(CCG)_n$ repeats, as discussed in the text. Quadruplex structures have been identified in oligonucleotides containing CGG and TGG repeats. In these structures the guanine are held together by G•G Hoogsteen bonds. A Hy3-type intramolecular triplex (as well as bi-triplex structures) can form in the $(GAA)_n$•$(TTC)_n$ repeat associated with Friedreich's ataxia.

(Chastain and Sinden 1998). Based on the persistence length and torsional rigidity of DNA derived from the crystal structure database (Bhattacharyya *et al* 1999), an increased duplex flexibility may be explained by an accumulation of the larger flexibility of dinucleotide steps d(CA)•d(TG), d(GC)•d(GC) and d(CG)•d(CG) that are parts of triplet repeats. Experimental estimates show that the flexibility associated with as few as four CTG repeats [$(CTG)_4$] may be as large as that of a two-nucleotide (TT)•(TT) mismatch (Chastain and Sinden 1998). This flexibility is manifested in the more efficient cyclization of triplet repeat sequences compared to that of mixed sequence DNA (Bacolla *et al* 1997; Chastain and Sinden 1998). The increased flexibility of CTG repeats may explain the affinity of CTG repeats to wrap around histone octamers to form nucleosomes (Wang *et al* 1994; Godde and Wolffe 1996). However, CGG repeats behave differently, in that they are refractory to forming nucleosomes. Moreover, their assembly into nucleosomes occurs in a length and methylation state dependent manner (Godde *et al* 1996; Wang *et al* 1996; Wang and Griffith 1996). While relatively flexible, the inability of CGG repeats to easily form nucleosomes may be explained by their stable radius of curvature which is inappropriate for wrapping around a histone octamer (Chastain and Sinden 1998).

Upon strand separation, as in replication or transcription, single-stranded $(CTG)_n$, $(CAG)_n$, $(CGG)_n$ and $(CCG)_n$ repeat tracts can form hairpins. In such hairpins, every two out of the three bases of the triplet are involved in intrastrand hydrogen bonding with the complementary bases of another triplet (figure 2) (see for review Mitas 1997; Darlow and Leach 1998a; Pearson and Sinden 1998b). CGG and CCG strands can each fold into hairpin structures in two ways involving either a G•G or C•C mismatch. In contrast, CTG and CAG repeats can each form only one hairpin structure with a T•T or A•A mismatch (Mitas 1997; Darlow and Leach 1998a; Pearson and Sinden 1998b). Single-stranded GAA repeats may also fold into hairpins with G•A pairs and A•A mismatches (Suen *et al* 1999). A few other single-stranded triplet repeats such as the GTC and GAC repeats, whose relevance to human diseases has not yet been established, can also form hairpins (Mitas 1997).

$(CTG)_n$ hairpins are more stable than the $(CAG)_n$ ones because the smaller T•T mispairs are better stacked in DNA helix than bulky A•A mispairs (Petruska *et al* 1996; Mitas 1997; Gacy and McMurray 1998). If the hairpin folding patterns are not dictated by flanking sequences, loops with even numbers of unpaired bases may be somewhat more stable than those with an odd number of unpaired bases because they each have one more base pair stabilizing the hairpin bend (Darlow and Leach 1995; Petruska *et al* 1998; Hartenstine *et al* 2000). The pronounced preference of the even-numbered loops results in a more frequent polymerization slippage by two triplets than by one triplet (Petruska *et al* 1998).

Hairpins formed by $(CGG)_{15}$ are more stable than those formed by $(CCG)_{15}$ (Mitas 1997). In this case, the smaller C•C mismatch may be more destabilizing than the larger G•G mismatch because in a long CCG tract the duplex forms an e-motif (Gao *et al* 1995) in which C•C mispairs are not well stacked in the helix and cytosines become extrahelical. Similar to $(CTG)_n$ and $(CAG)_n$ hairpins, in long stretches of $(CGG)_n$ and $(CCG)_n$ even-numbered loops are preferred to the odd-numbered ones (Darlow and Leach 1995; Darlow and Leach 1998b).

Individuals with the $(GAA)_n$•$(TTC)_n$-related disease Friedreich's ataxia have a deficiency of mature mRNA transcribed from the *X25* gene coding for a protein, frataxin (Campuzano *et al* 1996). The low levels of mRNA may result from a block to transcription, such as a GAA hairpin (Suen *et al* 1999). An intramolecular Py•Pu•Py type triplex can form in a $(GAA)_n$•$(TTC)_n$ tract under the influence of negative supercoiling and low pH (figure 2) (Hanvey *et al* 1988). An alternative structure, possibly the Py•Pu•Pu triplex, can form at neutral pH in $(GAA)_n$•$(TTC)_n$ tracts with $n = 38$ (but not $n = 16$) (Ohshima *et al* 1996). This structure could also form a block to transcription (Bidichandani *et al* 1998). A triplex formation has been suggested between the (GAA)•(TTC) duplex and a GAA mRNA strand (Ohshima *et al* 1998). However, this DNA-RNA triplex formation is unlikely, as DNA does not form triplexes with a purine RNA strand (Semerad and Maher 1994). Finally, a bi-triplex structure, termed sticky DNA, may form by the association of the two Py•Pu•Pu triplexes in negatively supercoiled DNA molecules (Sakamoto *et al* 1999).

DNA sequences containing tracts of guanine nucleotides can form quadruplex DNA stabilized by Hoogsteen-type hydrogen bonds between guanine bases (figure 2) (see for review, Sinden 1994). Fry and Loeb (1994) showed that short $(CGG)_n$ oligonucleotides, where $n = 4–7$, associate to form intermolecular quadruplex structures. A polymerization block in $(CGG)_{20}$ was also interpreted as a consequence of quadruplex formation (Usdin and Woodford 1995). Similar quadruplex structures may block replication in $(CGG)_n$, $(AGG)_n$, and $(TGG)_n$ repeats, as well as pure poly(G) sequences (Usdin 1998). Although longer CGG repeats ($n = 8$, 11) may prefer the formation of a hairpin rather than the quadruplex formation (Nadel *et al* 1995), the association of two hairpins may lead to a quadruplex (Weisman-Shomer *et al* 2000). Such a quadruplex may be weakened by AGG triplets in spite of the latter being capable of quadruplex formation (Usdin 1998).

2. Results and discussion

2.1 *Slipped strand DNA structures in $(CTG)_n \cdot (CAG)_n$ and $(CGG)_n \cdot (CCG)_n$ repeats*

Slipped strand DNA structures can form within repeated DNA sequences from an out-of-register alignment of complementary duplex strands within the repeat region (figure 2) (Sinden 1994). Such out-of-register misalignment will result in the formation of loops in the complementary strands. In the case of $(CTG)_n \bullet (CAG)_n$ and $(CGG)_n \bullet (CCG)_n$ repeats, the looped-out regions can form hairpins, as described above.

To form slipped strand DNA, the DNA duplex must unwind and the complementary strands must re-pair in an out-of-register alignment. *In vivo*, DNA helicase can unwind the DNA helix, as happens ahead of DNA polymerases in a replication fork. *In vitro*, following denaturation and renaturation of DNA molecules containing CTG or CGG repeats, a high proportion of the DNA population adopted an alternative conformation as indicated by the retarded mobility of molecules in polyacrylamide gels. Experimentally, linear plasmid DNA fragments containing expanded $(CTG)_n \bullet (CAG)_n$ repeats from the DM1 locus or the SCA1 locus show a complex, but reproducible patterns of slowly migrating products on a polyacrylamide gel following denaturation/renaturation protocols (figure 3). In addition, the $(CGG)_n \bullet (CCG)_n$ repeats from FRAXA genomic clones also show slowly migrating products. Slipped strand DNAs should contain looped-out $(CTG)_n$ and $(CAG)_n$ [or $(CGG)_n$ and $(CCG)_n$] repeat regions which are likely to form mispaired hairpins, creating a three-way DNA junction. Three- and four-way junctions introduce bends or kinks into DNA (Lilley and Clegg 1993; Shlyakhtenko *et al* 1994; Seeman and Kallenbach 1994; Lilley 1995; Stuhmeier *et al* 1997). A slipped strand DNA structure creates a molecule with at least two 3-way junctions which will introduce a considerable bend in DNA (Oussatcheva *et al* 1999) which is consistent with the low mobility observed following denaturation/renaturation.

The amount of slipped strand DNA structure formed is proportional to the repeat tract length and homogeneity (Pearson and Sinden 1996; Pearson *et al* 1998a,b). The amount of DNA migrating with reduced mobility increases sharply from about 3% for $(CTG)_{17} \bullet (CAG)_{17}$ to about 70–80% for $(CTG)_{50} \bullet (CAG)_{50}$. This increase in the amount of slipped strand DNA formed occurs between the repeat lengths typical of normal individuals and intermediate allele lengths found in affected individuals.

While loop-outs of different sizes can theoretically form anywhere within a triplet repeat tract, one might expect a continuum of products in the gel. Typically several major products appear over a background of minor products (Pearson and Sinden 1996; Pearson *et al* 1997, 1998a,b). The reproducible pattern of products with reduced mobility formed from a 878 bp DNA fragment containing $(CTG)_{255} \bullet (CAG)_{255}$ (with several sequence interruptions) is shown in figure 3. Note that because of the inherent flexibility of this DNA (see above), it migrates with an anomalous rapid mobility (at a position corresponding to about 790 bp). A range of products with reduced mobility are observed from about 800–1700 bp in length. Three major doublets are observed at about 875, 1000 and 1150 bp. Sequence interruptions within the repeat tract reduce the overall amount of the alternative DNA structure as well as reduce the heterogeneity of products formed. Moreover, sequence interruptions limit the range of stable slipped strand DNA isomers formed resulting in the distinct pattern of stable products formed.

Biochemical and electron microscopy experiments mapped the site of the unusual structures to within the triplet repeat region (Pearson and Sinden 1996; Pearson *et al* 1998b). Other experiments demonstrated that these slipped stranded structures were remarkably stable. The structures were stable through electroelution, multiple buffer changes, phenol extraction, and ethanol precipitation.

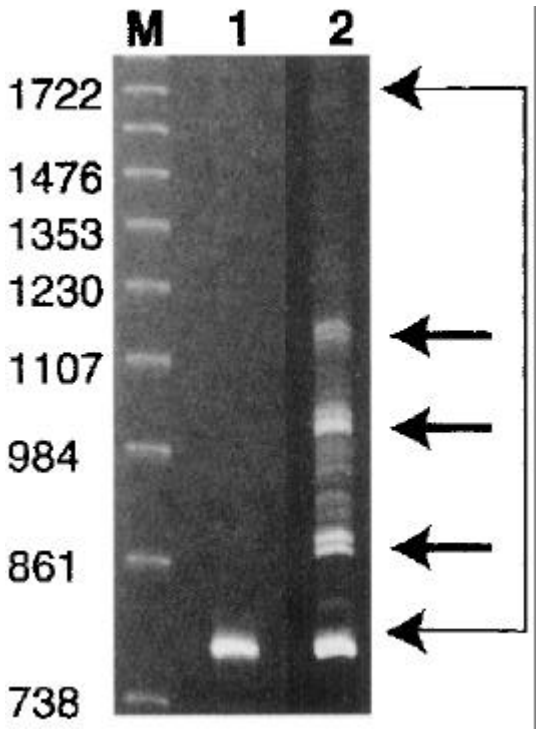

Figure 3. Slipped strand DNA structures migrate as heterogeneous products with reduced mobility on polyacrylamide gels. Lane 1, ladder of molecular weight markers. Lane 2, 878 bp DNA fragment containing $(CTG)_{255} \bullet (CAG)_{255}$ (with several sequence interruptions) derived from the *DM* locus. Following denaturation and renaturation many slow migrating bands (bracketed area), including three major pairs of doublets (arrows) are observed.

The structures were also stable when incubated overnight at 37°C or 55°C, but a 60 min incubation at 85°C did result in some interconversion of the isolated major slipped strand DNA isomer to other isomers, as well as some linear duplex (Pearson and Sinden 1996). The remarkable thermodynamic stability may result from the combination of base pairing within the hairpin loop and in the duplex DNA between the loop-outs, which would have to unpair completely for the structure to convert back to the linear form (Pearson and Sinden 1996). The stability may additionally result from loop-loop interactions between $(CTG)_n$ and $(CAG)_n$ hairpins (figure 2). Loss of slipped strand DNA necessarily must involve unpairing of bases in the loops as well as base pairs in the duplex between the loop-outs.

2.2 *Studies on model DNAs containing one or two 3-way junctions*

Important features of slipped structures including favourable sizes and numbers of slipped-out loops as a function of repeat tract length, their stability, and preferable positions, if any, in the repeat tract are not known. The characterization of slipped strand DNA has proven difficult due to the possibility of multiple isomer formation. In an approach to understand slipped strand DNA structure, we have started heteroduplex characterization using relatively simple structures made by hybridization of the *Pvu*II-*Pvu*II fragments of plasmid pUC8 and/or pUC8-based plasmids containing sequences that may form defined structural elements at defined positions. Heteroduplex models include only one hairpin or a loop that forms a 3-way junction. The two strands in such molecules are base paired everywhere except in the insert sequences that may form hairpins or single-stranded loops. Slipped-strand DNA models consist of two 3-way junctions that were formed by hybridization of two different insert-containing DNA fragments, in which the inserts are at different restriction sites (figure 4). Such defined constructs allow an analysis of relationships between their structural parameters and data from polyacrylamide gel mobility, from the shape and dynamics of molecules visualized using atomic force microscopy (AFM), and from chemical susceptibility of nucleobases at the loop and junction sites.

Three-way junction formation results in DNA bending, and the heteroduplex migrates slower in a polyacrylamide gel than the fully double-stranded molecules of similar size. The migration of the DNA containing 3-way junctions is reduced with increasing size of the hairpin/loop protrusion (figure 5). Nearby intrinsic bends may add

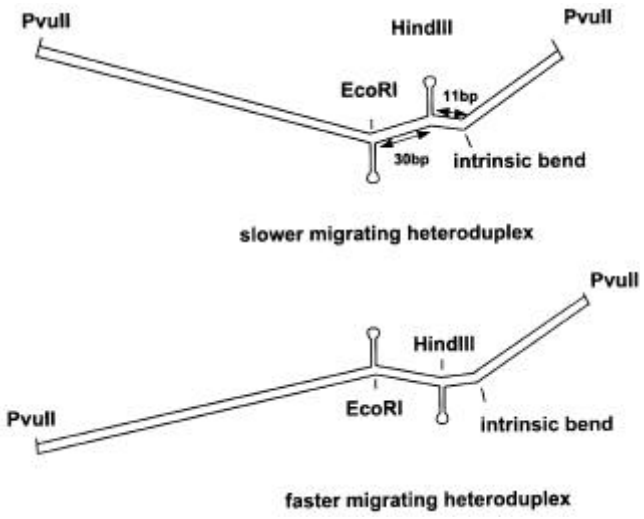

Figure 4. Differential bending in model slipped strand DNA structures. Relationship of hairpin isomer bends to intrinsic bends in DNA is shown. If an intrinsic bend exists in the DNA, then the formation of hairpins (e.g. at the *Eco*RI and *Hind*III sites) will introduce a second and third bend in the DNA. The direction of the hairpin-induced bends in the two sister heteroduplex isomers will be either in- or 180° out-of-phase with each other and with the intrinsic bend, and thus may act to introduce curvature in a concerted or antagonistic fashion in the overall molecule. The slow migrating heteroduplex should have the most distant bends oriented in the same direction (top) while the faster migrating heteroduplex should have these bends oriented in the opposite direction (bottom).

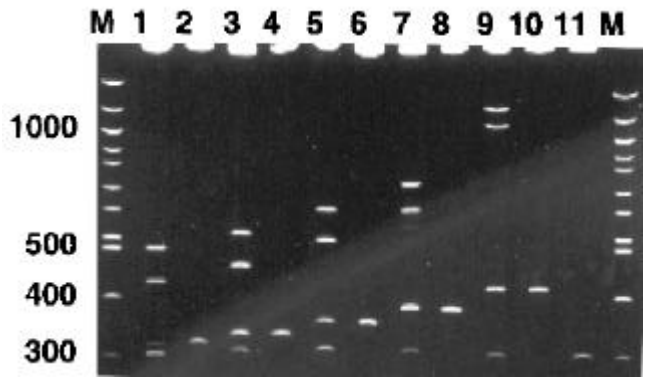

Figure 5. Electrophoretic separation in a polyacrylamide gel of 3-way junction heteroduplex isomers containing hairpins. This gel shows the separation of the hybridization products in a 5% native polyacrylamide gel. *Pvu*II-digested plasmids with inserts in the *Eco*RI site were hybridized with *Pvu*II-cut pUC8 and the products were separated in a polyacrylamide gel. The hybridization products contain 7 bp (lane 1), 12 bp (lane 3), 18 bp (lanes 5, 12), 27 bp (lane 7), and 50 bp (lane 9) hairpins. *Pvu*II-digested insert-containing plasmids were included to identify the parental duplex fragments for the 7 bp- (lane 2), 12 bp- (lane 4), 18 bp- (lane 6), 27 bp- (lane 8), 50 bp-hairpin-containing fragments (lane 10), and pUC8 (lane 11), which contains no insert. Lanes M, 100 bp ladder molecular size markers. (This figure is modified from figure 2 from Oussatcheva *et al* 1999. Copyright, Academic Press.)

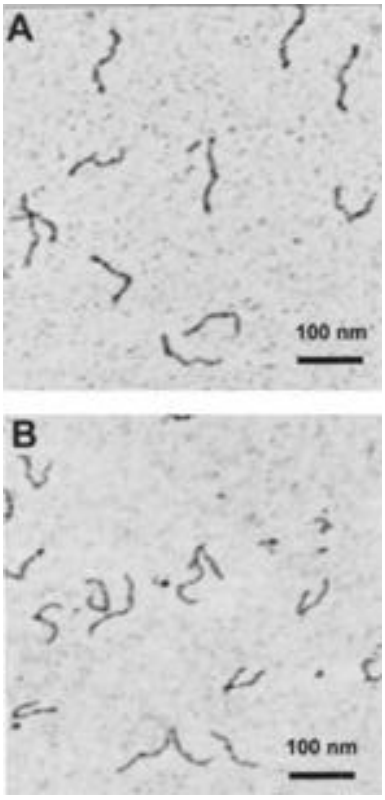

Figure 6. AFM images of sister heteroduplex isomers. These images show different bending angles at the 3-way junctions in the two isomers shown in figure 5 for the 27 bp hairpin (lane 8). (**A**) Faster migrating isomer. (**B**) Slower migrating isomer.

in-phase or out-of-phase bends at the 3-way junction, so that two possible heteroduplex isomers are separated in a gel (Oussatcheva *et al* 1999). AFM analysis of molecular shapes of the two, faster and slower migrating, heteroduplex species indicates that the slow migrating isomers have stronger overall bends than the faster migrating ones (figure 6). A very dynamic structure of 3-way junctions is inferred from the wide distributions of interarm angles, as well as from the fragment end-to-end distances determined from the images of dried samples. In addition, the junction mobility is observed directly when the samples are imaged in liquid. Time-resolved images of the same area show that both the long and the short heteroduplex arms change their orientations between consecutive scans (figure 7). Measurements of the shorter heteroduplex arm length produce a shorter apparent length than predicted from the known sequence. This is consistent with the pyramidal geometry of the junction so that the heteroduplex arm is elevated above the AFM surface and its projection on the plane seems shorter (Shlyakhtenko *et al* 2000). Other analyses have shown that the arms of 3-way junctions form dynamic trigonal pyramidal structures (Ma *et al* 1986; Shlyakhtenko *et al* 1994; Stuhmeier *et al* 1997; Oussatcheva *et al* 1999).

Heteroduplex constructs with two 3-way junctions have two bends at different sequence positions and also migrate slower than the fully double-stranded molecules of similar size. The spatial relationship of the two bends (and other possible intrinsic bends in flanking sequences) may strongly influence the three-dimensional shapes of hetero-

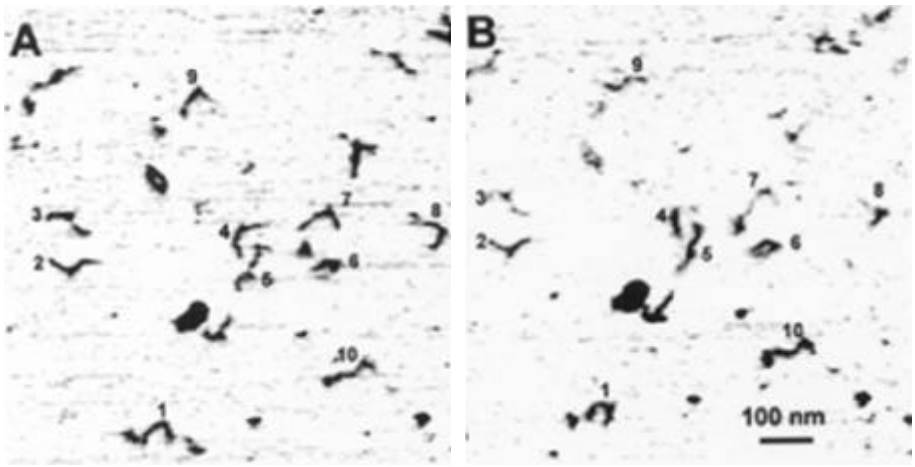

Figure 7. AFM analysis of 3-way junction dynamics in liquid. Two consecutive AFM scans of 301-bp DNA molecules containing a 7-bp hairpin. The DNA, in TE buffer with 200 mM NaCl, was injected into the fluid cell mounted on the top of APS mica. Ten molecules in the field are numbered. Most molecules change shape in the time between the two scans the shape. Molecules 7 and 9, in particularly, change shape dramatically. The time between successive scans was 2 min 40 s.

duplex isomers and their migration order. The AFM appearance of double heteroduplexes even with short inverted repeat arms shows a complex pattern of flat representations of the three-dimensional structures (Oussatcheva *et al* 1999).

2.3 *Studies on model DNAs containing $(CTG)_n$ and $(CAG)_n$ loop-outs*

In a second series of experiments, molecules containing looped-out CTG or CAG repeats in one strand, were constructed in heteroduplex reactions between pUC8 and a pUC8 derivative carrying $(CTG)_{23}$•$(CAG)_{23}$. The gel migration properties and AFM images of the 3-way junctions formed by triplet repeat sequences $(CTG)_{23}$ and $(CAG)_{23}$ (figure 8A) were very similar to the hairpin-containing model systems (Oussatcheva *et al* 1999). The distributions of interarm angles for 3-way junction molecules containing a 27 bp hairpin (Oussatcheva *et al* 1999) and $(CTG)_{23}$ were also very similar (data not shown) suggesting that the shape of both types of DNA molecules are similar.

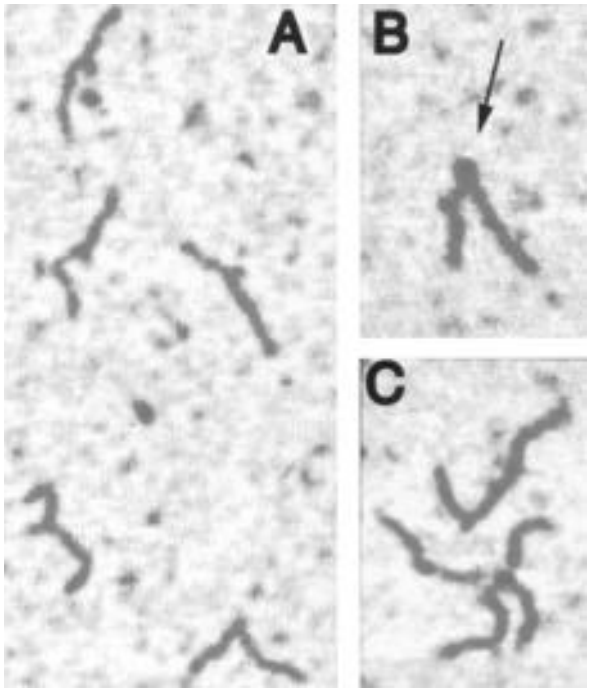

Figure 8. AFM images of $(CTG)_{23}$ and $(CAG)_{23}$ loop-outs. (**A**) Molecules with a single $(CTG)_{23}$ loop-out. (**B**) A molecules with both a $(CTG)_{23}$ and $(CAG)_{23}$ loop-out in opposite strands with 30 bp spacing. The two complementary loop-outs likely interact forming a bulky protrusion indicated by the arrow. (**C**) Representative dimer molecule, where interactions of complementary loops of the two double loop-out fragments stabilize the association of two molecules.

Next, the slipped-strand DNA models with two looped-out CTG and CAG repeat tracts separated by 30 bp of non-repeating sequence to prevent reannealing of these heteroduplexes into perfect duplexes were analysed. If the two incompletely base-paired triplet repeat hairpins in the heteroduplex are small (8 repeats) the isomers migrate in a gel in a manner similar to those containing pairs of perfect inverted repeat hairpins and/or unstructured loops. Figure 9A (top) shows the assignment of the faster and slower migrating isomers for the $(CAG)_8$•$(CTG)_8$ model of slipped-strand DNA. When the two looped-out triplet repeats are large enough to contact each other across the intervening sequence [i.e. $(CTG)_{23}$ and $(CAG)_{23}$], a reversal in the gel mobility of isomers is observed (figure 9A, bottom). This is consistent with another experiment (figure 9B), where a certain order of migration of heteroduplex isomers containing 75 nucleotide loops of random non-complementary sequence in two DNA strands (two left lanes) is reversed when heteroduplex loops contain complementary sequences (two right lanes).

To further substantiate the hypothesis that loop-loop interactions exist in slipped strand DNA, we isolated heteroduplex isomers containing presumed interacting hairpins (loops) and used chemical probe assays to test for the protection of certain nucleobases from modification as expected if these bases are involved in inter-hairpin (loop) base pairing. The preliminary chemical

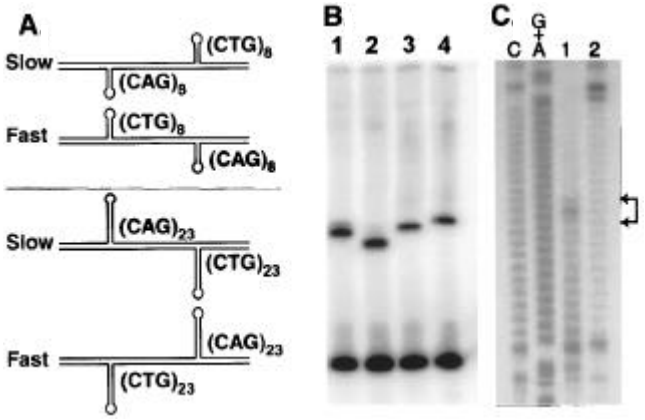

Figure 9. Properties of model slipped DNA constructs: effects of potential loop-loop interactions. (**A**) Different migration orders for heteroduplex isomers with short (non-interacting) and long (possibly interacting) triplet repeat loops. (**B**) The migration order for the long loops when the random sequences in two DNA strands cannot interact due to non-complementarity (lanes 1, 2) and when they are complementary (lanes 3, 4). Both pairs show 3′ and 5′ loop-outs in the top strand, respectively, as for the top pair of structures in **A**. (**C**) Chloroacetaldehyde reactivity of unpaired adenines and cytosines when the heteroduplex contains a single $(CAG)_{23}$ loop (lane 1), and when it contains a complementary $(CTG)_{23}$ loop 30 bp apart making an inter-loop interaction possible (lane 2).

probing data (figure 9C) show that central bases in a single 23 triplet repeat loop are susceptible to chloroacetaldehyde, whereas they become protected from reaction when a complementary loop is present at a 30 bp distance suitable for the inter-loop base pairing.

AFM images of heteroduplex fragments with a $(CTG)_{23}$ loop-out and $(CAG)_{23}$ loop-out separated by 30 bp are shown in figure 8B,C and provide additional evidence for loop–loop interaction. Double loop-out "slipped strand model structures" often show a much larger protrusion than single loop-out structures. For example, the molecule, shown in figure 8B, has a very bright protrusion that is not seen in images of one-loop structures. Interestingly, this molecule is about 10% shorter than one-loop structures suggesting a zigzag path of the double-stranded DNA molecule in the vicinity of the two loops that interact with each other. Two 3-way junctions may provide the necessary flexibility for two strands to form a zigzag. The most striking observation with two-loop structures is the formation of dimers through the loop-loop interaction. An example of such complexes is shown in figure 8C. Two heteroduplex molecules touching each other produce a bright interaction point whose position is consistent with the positions of looped structures in each molecule. Note that such dimers have never been observed in similar two-hairpin heteroduplexes, strongly suggesting more extensive interaction between the triplet repeat loops with less internal base pairs than the hairpins. This intermolecular interaction is consistent with our previous observation that a large amount of triplet repeat-containing DNA migrates as aggregated products on a polyacrylamide gel following denaturation and renaturation (Pearson and Sinden 1996). This type of intermolecular interaction may depend on ionic conditions, temperature, etc.

Altogether, these preliminary observations suggest a certain degree of tertiary interaction between spatially separated triplet repeat loops. This is possible if the complementary triplet repeat loops are close enough for a direct contact so as to form inter-loop base pairs. The sufficient mobility (flexibility, dynamics) of the 3-way junction arms may provide the favourable orientation necessary for tertiary interaction. Thus, the inter-loop base pairing may be an important factor that stabilizes slipped-strand DNA and prevents its reannealing into a fully double-stranded structure.

2.4 *AFM images of presumed slipped strand DNA*

We have obtained unique images of a 263 bp DNA containing presumed $(CTG)_{50}$•$(CAG)_{50}$ slipped strand structures (from the slower migrating minor product that constitutes about 5% of the DNA population). In contrast to the heterogeneity observed with EM (see figure 7 in Pearson *et al* 1998b), the AFM images were remarkably similar. About 80% of the images show pyramidal structures at the junction of a DNA bend (figure 10A–C). One interpretation is that the CTG and CAG arms of a slipped strand structure interact with each other to form hydrogen bonds between the loops of the hairpins (model shown in figure 2). In doing this they fold the DNA and this may bend the intervening flanking $(CTG)_n$•$(CAG)_n$ duplex. Remarkable structures have also been observed with DNA containing a $(CGG)_{54}$•$(CCG)_{54}$ repeat tract. Following denaturation and hybridization a very high

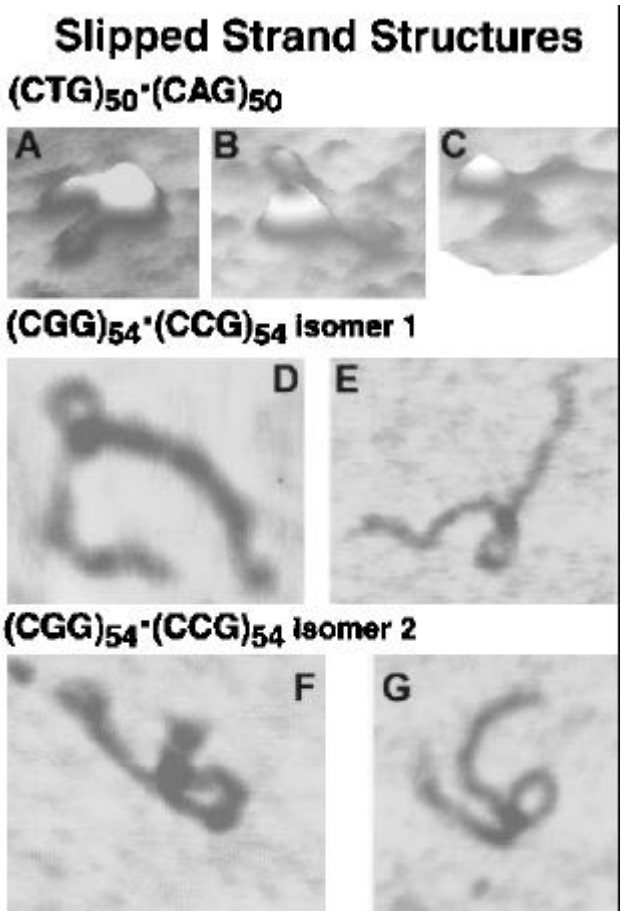

Figure 10. AFM images of representative slipped strand DNA structures. AFM images **A–C** show three-dimensional projections of three slipped strand DNA molecules [the most slowly migrating product as described (Pearson *et al* 1998b)] from $(CTG)_{50}$•$(CAG)_{50}$. A pyramidal structure is observed at the apex of a bent molecule. One interpretation is that this pyramidal mass represents two looped out strands interacting above the plane of the DNA molecule. AFM images D–E and F–G are the two major slipped strand isomers from $(CGG)_{54}$•$(CCG)_{54}$. A clear loop is seen in AFM images from both of these structural isomers. Again this may be the interaction of two long loop out structures or the looping of the very flexible (CGG)•(CCG) linear helix held together by interactions between short loops.

proportion of molecules with clear loops were observed (figure 10D–G). This is also consistent with our hypothesis that a folded loop forms via interaction of looped-out CGG and CCG strands as has been suggested to occur for other types of repeats (Coggins and O'Prey 1989; Coggins *et al* 1992; Ulyanov *et al* 1994, 1998). This DNA loop could either be the result of interaction between the loops of two long arms (as shown in figure 2), or the interaction of two short loops that are separated by a long distance. In the latter, the interaction between distant short loops would have the effect of forming a loop in the DNA, which may be what is observed in these structures. DNA structures of this type have been observed previously, although with much longer repeated DNA tracts (Coggins and O'Prey 1989; Coggins *et al* 1992).

3. Conclusion

To summarize, it may be said that slipped strand DNA structures are formed stably and reproducibly within the trinucleotide repeating sequences. Their propensity to form is proportional to the length and homogeneity of the repeat tract. A very dynamic structure of 3-way junctions formed at the loop-outs and the sequence complementarity of the loops may favour the formation of folded slipped strand DNA structures stabilized by loop-loop interactions.

References

Bacolla A, Gellibolian R, Shimizu M, Amirhaeri S, Kang S, Ohshima K, Larson J E, Harvey S C, Stollar B D and Wells R D 1997 Flexible DNA – genetically unstable CTG•CAG and CGG•CCG from human hereditary neuromuscular disease genes; *J. Biol. Chem.* **272** 16783–16792

Bhattacharyya D, Kundu S, Thakur A R and Majumdar R 1999 Sequence directed flexibility of DNA and the role of cross-strand hydrogen bonds; *J. Biomol. Struct. Dyn.* **17** 289–300

Bidichandani S I, Ashizawa T and Patel P I 1998 The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure; *Am. J. Hum. Genet.* **62** 111–121

Brahmachari S K, Meera G, Sarkar P S, Balagurumoorthy P, Tripathi J, Raghavan S, Shaligram U and Pataskar S 1995 Simple repetitive sequences in the genome: Structure and functional significance; *Electrophoresis* **16** 1705–1714

Brook J D, McCurrach M E, Harley H G, Buckler A J, Church D, Aburatani H, Hunter K, Stanton V P, Thirion J P, Hudson T, Sohn R, Zemelman B, Snell R G, Rundle S A, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper P S, Shaw D J and Housman D E 1992 Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member; *Cell* **68** 799–808

Campuzano V, Montermini L, Molto M D, Pianese L, Cossee M, Cavalcanti F, Montos E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani S I, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel P I, Di Donato S, Mandel J L, Cocozza S, Koenig M and Pandolfo M 1996 Friedreich's Ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion; *Science* **271** 1423–1427

Chastain P D, Eichler E E, Kang S, Nelson D L, Levene S D and Sinden R R 1995 Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes; *Biochemistry* **34** 16125–16131

Chastain P D and Sinden R R 1998 CTG repeats associated with human genetic disease are inherently flexible; *J. Mol. Biol.* **275** 405–411

Coggins L W and O'Prey M 1989 DNA tertiary structures formed in vitro by misaligned hybridization of multiple tandem repeat sequences; *Nucleic Acids Res.* **17** 7417–7426

Coggins L W, O'Prey M and Akhter S 1992 Intrahelical pseudoknots and interhelical associations medicated by mispaired human minisatelite DNA sequences in vitro; *J. Mol. Biol.* **121** 279–285

Darlow J M and Leach D R F 1995 The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in *Escherichia coli* suggest hairpin folding preferences in vivo; *Genetics* **141** 825–832

Darlow J M and Leach D R F 1998a Secondary structures in d(CGG) and d(CCG) repeat tracts; *J. Mol. Biol.* **275** 3–16

Darlow J M and Leach D F 1998b Evidence for two preferred hairpin folding patterns in d(CGG)•d(CCG) repeat tracts *in vivo*; *J. Mol. Biol.* **275** 17–23

Fortune M T, Vassilopoulos C, Coolbaugh M I, Siciliano M J and Monckton D G 2000 Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability; *Hum. Mol. Genet.* **9** 439–445

Fry M and Loeb L A 1994 The fragile X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure; *Proc. Natl. Acad. Sci. USA* **91** 4950–4954

Fu Y-H, Kuhl D P A, Pizzuti A, Pieretti M, Sutcliffe J S, Richards S, Verkerk A J M H, Holden J J A, Fenwick J R G, Warren S T, Oostra B A, Nelson D L and Caskey C T 1991 Variation of the CGG repeat at the Fragile X site results in genetic instability: resolution of the Sherman paradox; *Cell* **67** 1047–1058

Fu Y H, Pizzuti A, Fenwick Jr R G, King J, Rajnarayan S, Dunne P W, Dubel J, Nasser G A, Ashizawa T, DeJong P, Wieringa B, Korneluk R, Perryman M B, Epstein H F and Caskey C T 1992 An unstable triplet repeat in a gene related to myotonic muscular dystrophy; *Science* **255** 1256–1258

Gacy A M and McMurray C T 1998 Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA; *Biochemistry* **37** 9426–9434

Gao X, Huang X, Smith G K, Zheng M and Liu H 1995 A new antiparallel duplex motif of DNA CCG repeats that is stabilized by extrahelical bases symetrically localized in the minor groove; *J. Am. Chem. Soc.* **117** 8883–8884

Godde J S, Kass S U, Hirst M C and Wolffe A P 1996 Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter; *J. Biol. Chem.* **271** 24325–24328

Godde J S and Wolffe A P 1996 Nucleosome assembly on CTG triplet repeats; *J. Biol. Chem.* **271** 15222–15229

Hanvey J C, Shimizu M and Wells R D 1988 Intramolecular DNA triplexes in supercoiled plasmids; *Proc. Natl. Acad. Sci. USA* **85** 6292–6296

Hartenstine M J, Goodman M F and Petruska J 2000 Base stacking and even/odd behaviour of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase; *J. Biol. Chem.* **275** 18382–18390

Heintz N and Zoghbi H Y 2000 Insights from mouse models into the molecular basis of neurodegeneration; *Annu. Rev. Physiol.* **62** 779–802

Kornberg A, Bertsch L L, Jackson J F and Khorana H G 1964 Enzymatic synthesis of deoxyribonucleic acid, XVI. Oligonucleotides as templates and the mechanism of their replication; *Proc. Natl. Acad. Sci. USA* **51** 315–323

Kremer E J, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren S T, Schlessinger D, Sutherland G R and Richards R I 1991 Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$; *Science* **252** 1711–1714

Lilley D M J 1995 Kinking of DNA and RNA by base bulges; *Proc. Natl. Acad. Sci. USA* **92** 7140–7142

Lilley D M J and Clegg R M 1993 The structure of branched DNA species; *Q. Rev. Biophys.* **26** 131–175

Ma R I, Kallenbach N R, Sheardy R D, Petrillo M L and Seeman N C 1986 Three-arm nucleic acid junctions are flexible; *Nucleic Acids Res.* **14** 9745–9753

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-McDonald J, de Jons P J, Wieringa B and Korneluk R G 1992 Myotonic dystrophy mutation: An unstable CTG repeat in the 3′ untranslated region of the gene; *Science* **255** 1253–1255

Mitas M 1997 Trinucleotide repeats associated with human disease; *Nucleic Acids Res.* **25** 2245–2253

Nadel Y, Weisman-Shomer P and Fry M 1995 The fragile X syndrome single strand $d(CGG)_n$ nucleotide repeats readily fold back to form unimolecular hairpin structures; *J. Biol. Chem.* **270** 28970–28977

Ohshima K, Kang S, Larson J E and Wells R D 1996 Cloning, characterization, and properties of seven triplet repeat DNA sequences; *J. Biol. Chem.* **271** 16773–16783

Ohshima K, Montermini L, Wells R D and Pandolfo M 1998 Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo; *J. Biol. Chem.* **273** 14588–14595

Oussatcheva E A, Shlyakhtenko L S, Glass R, Sinden R R, Lyubchenko Y L and Potaman V N 1999 Structure of Branched DNA Molecules: Gel Retardation and Atomic Force Microscopy Studies; *J. Mol. Biol.* **292** 75–86

Pearson C E, Eichler E E, Lorenzetti D, Kramer S F, Zoghbi H Y, Nelson D L and Sinden R R 1998a 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

Pearson C E, Wang Y H, Griffith J D and Sinden R R 1998b Structural analysis of slipped-strand DNA (S-DNA) formed in $(CTG)_n \bullet (CAG)_n$ repeats from the myotonic dystrophy locus; *Nucleic Acids Res.* **26** 816–823

Pearson C E, Ewel A, Acharya S, Fishel R A and Sinden R R 1997 Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases; *Hum. Mol. Genet.* **6** 1117–1123

Pearson C E and Sinden R R 1996 Alternative DNA structures within the trinucleotide repeats of the myotonic dystrophy and fragile X locus; *Biochemistry* **35** 5041–5053

Pearson C E and Sinden R R 1998a Slipped strand DNA, dynamic mutations, and human disease; in *Genetic instabilities and hereditary neurological disorders* (eds) R D Wells and S T Warren (San Diego: Academic Press) pp 585–621

Pearson C E and Sinden R R 1998b Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA; *Curr. Opin. Struct. Biol.* **8** 321–330

Petruska J, Arnheim N and Goodman M F 1996 Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases; *Nucleic Acids Res.* **24** 1992–1998

Petruska J, Hartenstine M J and Goodman M F 1998 Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease; *J. Biol. Chem.* **273** 5204–5210

Sakamoto N, Chastain P D, Parniewski P, Ohshima K, Pandolfo M, Griffith J D and Wells R D 1999 Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia; *Mol. Cell* **3** 465–475

Saleem Q, Choudhry S, Mukerji M, Bashyam L, Padma M V, Chakravarthy A, Maheshwari M C, Jain S and Brahmachari S K 2000 Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation; *Hum. Genet.* **106** 179–187

Samadashwily G M, Raca G and Mirkin S M 1997 Trinucleotide repeats affect DNA replication *in vivo*; *Nature Genet.* **17** 298–304

Schlotterer C and Tautz D 1992 Slippage synthesis of simple sequence DNA; *Nucleic Acids Res.* **20** 211–215

Seeman N C and Kallenbach N R 1994 DNA branched junctions; *Annu. Rev. Biophys. Biomol. Struct.* **23** 53–86

Semerad C L and Maher L J III 1994 Exclusion of RNA strands from a purine motif triple helix; *Nucleic Acids Res.* **22** 5321–5325

Shlyakhtenko L S, Potaman V N, Sinden R R, Gall A A and Lyubchenko Y L 2000 Structure and dynamics of three-way DNA junctions: Atomic force microscopy studies; *Nucleic Acids Res.* **28** 3472–3477

Shlyakhtenko L S, Rekesh D, Lindsay S M, Kutyavin I, Appella E, Harrington R E and Lyubchenko Y L 1994 Structure of three-way DNA junctions. 1. Non-planar DNA geometry; *J. Biomol. Struct. Dyn.* **11** 1175–1189

Sinden R R 1994 *DNA structure and function* (San Diego: Academic Press)

Sinden R R 1999 Biological implications of the DNA structures associated with disease-causing triplet repeats; *Am. J. Hum. Genet.* **64** 346–353

Sinden R R and Wells R D 1992 DNA structure, mutations, and human genetic disease; *Curr. Opin. Biotech.* **3** 612–622

Stuhmeier F, Welch J B, Murchie A I, Lilley D M and Clegg R M 1997 Global structure of three-way DNA junctions with and without additional unpaired bases: a fluorescence resonance energy transfer analysis; *Biochemistry* **36** 13530–13538

Suen I S, Rhodes J N, Christy M, McEwen B, Gray D M and Mitas M 1999 Structural properties of Friedreich's ataxia d(GAA) repeats; *Biochim. Biophys. Acta* **1444** 14–24

Ulyanov N B, Bishop K D, Ivanov V I and James T L 1994 Tertiary base pair interactions in slipped loop-DNA: an NMR and model building study; *Nucleic Acids Res.* **22** 4242–4249

Ulyanov N B, Ivanov V I, Minyat E E, Khomyakova E B, Petrova M V, Lesiak K and James T L 1998 Immobile Slipped-loop Structures of DNA in Solution; in *Structure, motion, interaction and expression of biological macromolecules* (eds) R H Sarma and M H Sarma (Schenectady: Adenine Press) vol. 1, pp 75–88

Usdin K 1998 NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases; *Nucleic Acids Res.* **26** 4078–4085

Usdin K and Woodford K J 1995 CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*; *Nucleic Acids Res.* **23** 4202–4209

Wang Y-H, Amirhaeri S, Kang S, Wells R D and Griffith J D 1994 Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene; *Science* **265** 669–671

Wang Y-H, Gellibolian R, Shimizu M, Wells R D and Griffith J 1996 Long CCG triplet repeat blocks exclude nucleosomes – a possible mechanism for the nature of fragile sites in chromosomes; *J. Mol. Biol.* **263** 511–516

Wang Y-H and Griffith J 1996 Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion; *J. Biol. Chem.* **271** 22937–22940

Watson J D and Crick F H C 1953 Molecular structure of nucleic acids – a structure for deoxyribose nucleic acid; *Nature* (*London*) **171** 737–738

Weisman-Shomer P, Cohen E and Fry M 2000 Interruption of the fragile X syndrome expanded sequence d(CGG)(n) by interspersed d(AGG) trinucleotides diminishes the formation and stability of d(CGG)(n) tetrahelical structures; *Nucleic Acids Res.* **28** 1535–1541