

## DNA INVERTED REPEATS AND HUMAN DISEASE

John J. Bissler

*The Children's Hospital Research Foundation, Cincinnati, Ohio*

### TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Inverted Repeats
4. Cruciform Formation
5. *In Vivo* Formation of Cruciforms
  - 5.1. Chemical Modification of the Hairpin Loop
  - 5.2. Protein Recognition of the Extruded Hairpin
    - 5.2.1. Antibody Studies
    - 5.2.2. Endonuclease
  - 5.3. Chemical Cross-linking
6. Biological Roles of Inverted Repeats
  - 6.1. RNA Effects
  - 6.2. DNA Effects
7. Role of Inverted Repeats in Mutagenesis
  - 7.1. Palindrome Mediated Mutagenesis
  - 7.2. Quasipalindrome Mediated Mutagenesis
  - 7.3. Diseases Associated with Imperfect Inverted Repeats
    - 7.3.1. Hereditary Angioneurotic Edema
    - 7.3.2. Triplet Repeat Mediated Diseases
    - 7.3.3. Duchenne Muscular Dystrophy
    - 7.3.4. Osteogenesis Imperfecta
    - 7.3.5. Antithrombin Deficiency
    - 7.3.6. Silent Serum Cholinesterase
    - 7.3.7. Lesch-Nyhan Syndrome
    - 7.3.8. Kearn-Sayre Syndrome
    - 7.3.9. Biotinase Deficiency
    - 7.3.10. Familial Hypercholesterolemia
8. Perspectives
9. Acknowledgments
10. References

### 1. ABSTRACT

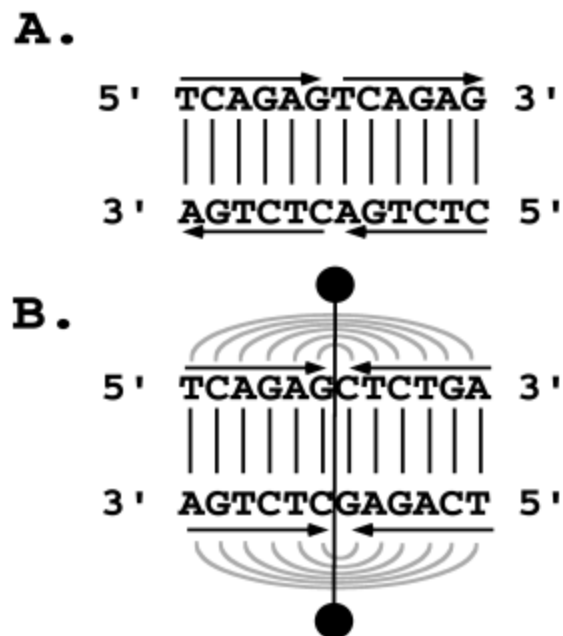
Inverted repeats are important elements in the human genome. Because of their nature, inverted repeats can engage in intra- and intermolecular basepairing. The ability to adopt hairpin and cruciform secondary structures is associated with frameshift mutations. These sequences also can be utilized by the polymerase allowing both intra- and interstrand switching events. Such mechanisms can involve imperfect inverted repeats and lead to additional mutations. Several human genetic diseases illustrate inverted repeat mediated mutagenesis.

### 2. INTRODUCTION

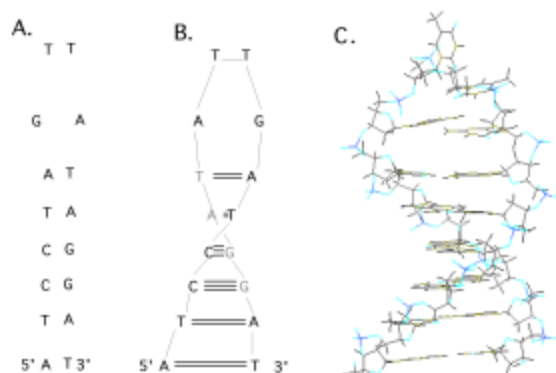
Mutations in the genome can result in inherited as well as malignant diseases. Studying the different mechanisms causing mutation will lead to strategies that promote health by reducing the mutagenic potential. Governmental regulations governing use and disposal of some genotoxic substances, such as radioactive material, are such a strategy. Like external mutagenic agents, innate mechanisms of mutation also lead to human disease. Two

examples of innate mechanisms leading to mutation are deficiencies in DNA repair systems (1), and mutations that evade repair (2). The combined impact of innate and environmental mutagenesis over millennia has been to cause genetic variation and, as a result of natural selection, facilitate evolution.

The fidelity of the genetic code is particularly vulnerable to mutagenic mechanisms during replication. During cell cycle phases other than S-phase, DNA is compared to the complementary strand and, if needed, repaired. During replication of the genetic code, point and frameshift mutations can occur in the newly synthesized strand. Replication is a rapid process. One estimate is that replication occurs at a rate of 1,000 bp per second (3). The DNA replication rate (4) and fidelity varies depending on the template sequence (5-7). Certain template sequences are more prone to replication errors. In this review, I will focus on the inverted repeat and its mechanistic role in contributing to genetic disease.



**Figure 1:** Direct and inverted DNA repeats. A. Direct repeats are illustrated with an arrow above the sequence that is repeated. Dark vertical bars represent interstrand binding between complementary strands. B. Inverted repeats are illustrated with the opposing arrows above the sequence. The vertical line perpendicular to the sequence denoted the center of the inverted repeat. Black vertical lines represents interstrand binding between complementary strands, while the gray concentric lines denote the intrastrand complementary nature of the bases in the inverted repeat.



**Figure 2:** Inverted repeats can form hairpin structures. A. An inverted repeat drawn as a hairpin structure. This can occur in both strands of complementary DNA forming a cruciform structure. B. A more realistic illustration of the hairpin structure illustrating the twist. Gray letters indicate the sequence is behind the overlying black sequence. C. Crystal structure of the inverted repeat plotted using RasMac Molecular Graphic Macintosh Version 2.6 and clearly shows the unpaired loop at the top of the hairpin structure.

### 3. INVERTED REPEATS

The human genome is far from random and has a number of repeat sequence types. The easiest to understand is the direct repeat in which the sequence is duplicated (figure 1A). The arrows over the sequence indicate the sequence that is repeated. Linguistically, an example of a direct repeat is "bye-bye". An inverted repeat, also called a palindrome, is similar except the second half of the repeat is in the complementary strand (figure 1B). A palindrome linguistically is a string of characters that reads the same in both directions, such as "noon". A consequence of these repeats being inverted is that they can base pair to the complementary strands through intermolecular base pairing, or to themselves through intramolecular base pairing. In Figure 1B, the black vertical lines represent the intermolecular base pairing and the gray concentric circles illustrate the potential intramolecular base pairing. Practically speaking, single-stranded DNA cannot bend sharply enough so that adjacent nucleotides on one strand can hydrogen bond as base pairs. There is a mandatory 3-5 bp spacer or loop between the 5' and 3' stems that allow the DNA to fold back on itself. In figure 2A, a single-stranded inverted repeat is written as a hairpin, however this is far from reality, and figure 2B illustrates the twist of the DNA. The crystallized hairpin structure is illustrated in figure 2C (8). These stem-loop structures can form in single-stranded DNA and are called "hairpins", or can form in both strands of the duplex DNA and are called "cruciforms". The terms inverted repeat and palindrome imply a perfect nature to the sequence. Sequences that are not perfect are called imperfect inverted repeats or quasipalindromes.

### 4. CRUCIFORM FORMATION

Energetically, it may seem very unlikely that a stem loop structure with less base pairing in the unpaired loop, would form in any abundance in DNA because the entire duplex structure would be favored on the basis of overall basepairing. Isolated linear DNA has a formidable energy barrier shielding the encoded information. The barrier consists of the energy of the hydrogen bonding and stacking forces that hold complementary strands of DNA together. The energy added to the molecule in living cells to make this information more readily available for transcription and replication is the unwinding of the DNA (9). The unwinding of DNA consists of twisting the two strands in the opposite direction of the helical coil. This energy strains the hydrogen bonds and stacking forces. The strain is relieved by the writhe of the DNA duplex upon itself, similar to twisting of a phone cord after it has been stretched. The relationship between twist and writhe is a constant known as the linking number. The negative superhelical tension can be restrained by wrapping DNA around proteins such as histones. The negative superhelical tension is dynamic, and is affected by both transcription and replication, as well as by changes in protein-DNA binding such as mediated histone acetylation (10,11). This superhelical tension is the driving energy for the formation of alternative DNA structures (12).

### 5. *IN VIVO* FORMATION OF CRUCIFORMS

The DNA features that cause cruciform extrusion *in vitro* have been demonstrated to also lead to alternative secondary structures *in vivo* (13). The basic approaches used to study cruciforms have been to demonstrate the extruded hairpin by chemically modifying the single-stranded loop, having protein recognition of the extrusion itself, or by chemically cross-linking bases in the cruciform stem.

#### 5.1. Chemical Modification of the Hairpin Loop

Chemical reactivity of the bases is influenced by base pairing. Single-strand bases, such as in hairpin loops, have different reactivity than in duplex DNA. Electrophiles like osmium tetroxide have been used to identify the extruded hairpin loop at basepair resolution (14-16). These sites can be detected by strategies employing either primer extension, because DNA polymerization arrests at the modified site, or by cleavage of the phosphodiester backbone at these sites by piperidine. Following modification of most bases, the glycosidic bond is more readily hydrolyzed resulting in an abasic site.

#### 5.2. Protein Recognition of the Extruded Hairpin

##### 5.2.1. Antibody Studies

An immunologic approach also has been employed using antibodies that recognize the sequence independent structure of the extruded cruciforms (17-20). This approach has identified antibody binding in eukaryotic cells. Ward *et al.* successfully used this approach to support the hypothesis that many inverted repeats exist as cruciform structures *in vivo* (18.)

##### 5.2.2. Endonuclease

Another approach to detect cruciform extrusion in living *Escherichia* used the enzyme T7 endonuclease VII controlled by an inducible promoter. Induction of the enzyme led to fragmentation of the DNA. The fragmentation implied that cruciforms in the bacterial genome were extruded and therefore were substrates for the enzymatic cleavage (21).

#### 5.3. Chemical Cross-linking

The last and most elegant technique illustrating cruciform formation *in vivo* employed the bi-functional cross-linking agent 4,5,8-trimethylpsoralen (22, 23). Trimethylpsoralen is freely permeable into cells and non-toxic unless activated by 360nm light. Therefore, the DNA structures within the cells are caught in a much more natural state. This approach has identified the extrusion of a well characterized inverted repeat in prokaryotic cells (13). The collective conclusion drawn from these studies is that cruciforms can and do form *in vivo*.

### 6. BIOLOGICAL ROLES OF INVERTED REPEATS

#### 6.1. RNA Effects

There is clear evidence that inverted repeats affect transcription. Virtually every molecular biology textbook has a discussion about rho independent termination (24- 27). In this mechanism the inverted repeat

from the 3' end of the mRNA facilitates termination of transcription. Furthermore, both ferritin and the transferrin receptor employ a iron response element (IRE) that regulates translation as a function of alternative secondary structure involving the IRE and polyribosome formation (28-30). The IRE in ferritin mRNA is 28 nucleotides long at the 5' end; whereas, in transferrin receptor mRNA the IRE is in the 3' untranslated region. When iron is in low abundance, the ferritin mRNA does not form polyribosomes whereas the transferrin receptor mRNA is translated. When iron is adequate, ferritin mRNA forms polyribosomes and transferrin receptor mRNA is degraded. This likely is regulated by protein binding competing with DNA secondary structure formation.

Likewise, transcription regulation of the human enkephalin gene is dependent on a 23 bp imperfect palindrome, or quasipalindrome, sequence (nucleotides -84 to -106). This sequence appears to be crucial for gene regulation (31, 32). Finally, mutations in the type XVII collagen gene result in the disease bullous pemphigoid (33-35). This gene also has an imperfect inverted repeat in the 5' end that has been postulated to regulate protein production (36).

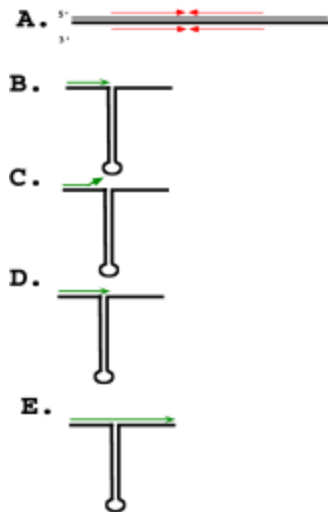
Besides regulating protein production through mRNA intermediates, perfect and imperfect inverted repeats have structural roles in tRNA, ribozymes, and some ribonuclear proteins.

#### 6.2. DNA effects

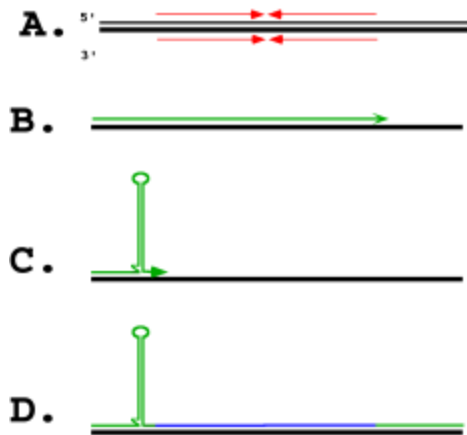
Inverted repeats have a number of biological roles. The easiest to understand is the DNA binding site for homodimeric proteins. Through interaction at their dimerization domains, the proteins become oriented in an antiparallel fashion. Their DNA binding domains also would be positioned in opposite directions, requiring an inverted repeat for binding. The classic example of this resides in restriction endonucleases. There is no evidence that these small inverted repeats extrude into secondary structure to achieve DNA cleavage.

In some prokaryotic systems, inverted repeats appear to be crucial for DNA replication. *Escherichia coli* contains a 245 bp origin of replication that contains many inverted repeats. However, there is no evidence that these inverted repeats function in an alternative conformation (37). Other systems actually require the inverted repeat to engage in intrastrand base pairing for function. The single-stranded bacteriophage G4 contains inverted repeats ranging from 20 to 44 base pairs long that are required for proper function of the origin of replication (38). The protein DnaG binds to such an inverted repeat and this interaction is required for initiation of replication. The plasmid pT181 contains an inverted repeat that actually functions as an extruded cruciform structure to initiate replication (39,38). The bacteriophage N4 employs an extruded cruciform, that binds SSB on the single-strand loop to promote transcription of early genes (40, 41).

Eukaryotic DNA replication systems can also employ inverted repeats. The SV40 virus contains inverted



**Figure 3:** Deletion mutation of DNA mediated by an inverted repeat. A. Two complementary strands of DNA are depicted by two lines of differing thickness. The opposing arrows mark the site of the inverted repeat. At either end of the inverted repeat, a direct repeat sequence, must be present for slippage to occur. B. After the two strands are separated, the replication fork encounters the extruded stem-loop structure in the thick strand. C. The nascent strand and polymerase partially dissociate from the template strand. The nascent strand has copied the 5' direct repeat. D. The nascent strand slips forward over the stem-loop structure and comes in register with the 3' direct repeat. E. Continued polymerization excludes the sequence from the hairpin structure.



**Figure 4:** Insertion mutation of DNA mediated by an inverted repeat. A. Two complementary strands of DNA are depicted by two lines of differing thickness. The opposing arrows mark the site of the inverted repeat. At either end of the inverted repeat, a direct repeat sequence, must be present for slippage to occur. B. The replication passes through the inverted repeat and through the 3' direct repeat. C. The nascent strand and polymerase dissociate from the template strand and the hairpin structure forms in the nascent strand. The 3' direct repeat anneals to the complementary bases for the displaced 5' direct repeat. The template strand fails to form the stem-loop structure. D. Continued polymerization creates a duplication.

repeats that are critical for DNA replication; however, evidence thus far suggests their function is the linear form (42). The human pathogen HSV likewise employs inverted repeats. The genome contains inverted repeats  $oriL_1$ , a 144 bp inverted repeat and  $oriL_2$ , a 136 bp inverted repeat that is unstable when cloned into a bacterial system (43, 44). These examples raise the possibility that one strategy for replicating DNA employs inverted repeats, both in the duplex form with the complementary strand as well as in the extruded conformation.

Inverted repeats whether perfect or imperfect are imbedded in genomes and have critical functions. Yet, these structures pose a special impediment to DNA replication fidelity and are associated with several human disease related genes.

## 7. ROLE OF INVERTED REPEATS IN MUTAGENESIS

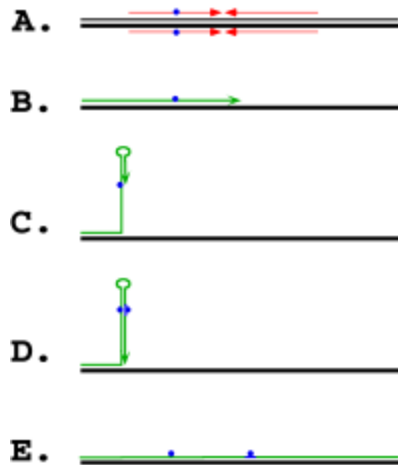
### 7.1. Palindrome mediated mutagenesis

Regions of DNA alternative secondary structure pose a barrier to replication fidelity. Work in *Escherichia coli* has demonstrated that the extruded cruciform facilitates frameshift mutations by bringing the DNA slippage sites (direct repeats) into close proximity (45,44). This can lead to deletion as well as insertion mutations. Figure 3A illustrates duplex DNA depicted by a wide and narrow line. The inverted repeat is illustrated by the opposing arrows. The replication of the strand is interrupted by the encounter of the replication fork and the extruded cruciform (figure 3B). The extrusion of the stem loop structure brings the distant DNA direct repeats into close proximity. In panels C and D the replication fork pauses, the 5' direct repeat sequence melts and the nascent strand reanneals on the direct repeat 3' to the extruded sequence. Replication continues resulting in a heteroduplex with one non-mutant strand and one deletion mutant. These can be made permanent by another round of DNA replication. In figure 4, one mechanism for an insertion mutation is illustrated where the inverted repeat in the copied strand folds into a cruciform structure and the replicated 3' direct repeat comes into register over the 5' complementary sequence allowing continued DNA synthesis resulting in a duplication.

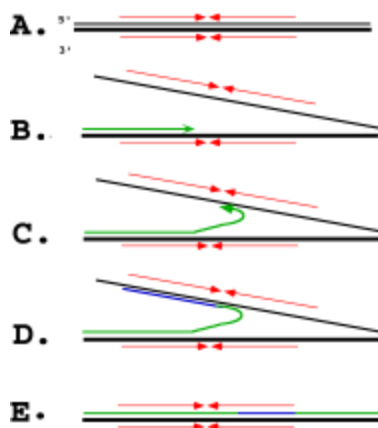
### 7.2. Quasipalindrome mediated mutagenesis

Imperfect palindromes theoretically also can mediate similar mutations as perfect palindromes and in addition engage in intra- and interstrand switching further increasing the potential mutation spectrum (46, 47).

Intrastrand switching requires the polymerase to replicate through the center of the palindrome and, with the nascent chain, disassociate from the template strand and engage in intrastrand annealing into a hairpin structure and then continue to polymerize (48, 49). This situation is transient and following additional polymerization, the nascent strand and polymerase switch back to the original template strand. If the template is a quasipalindrome,



**Figure 5:** Intrastrand switching. A. Two complementary strands of DNA are depicted by two lines of differing thickness. The opposing arrows mark the site of the imperfect inverted repeat. The blue dots illustrate the site of the base pairs difference between the inverted repeats. B. The DNA is polymerized through the center of the inverted repeat. C. The polymerization machinery and nascent strand dissociate from the template strand and the nascent strand forms a stem-loop structure creating an alternative template for DNA synthesis. D. Synthesis of the DNA occurs. The blue line represents the mutation arising from intrastrand templating. E. The nascent strand and polymerized DNA again dissociate and anneal back to the original template and synthesis continues. The mismatched in blue may be removed by DNA repair, or may be fixed in the genome by a second round of replication resulting in one original sequence and one mutant.



**Figure 6:** Interstrand switching. A. Two complementary strands of DNA are depicted by two lines of differing thickness. The opposing arrows mark the site of the inverted repeat. B. The DNA is polymerized into the inverted repeat. Unlike intramolecular switching, in intermolecular switching the DNA synthesis need not be through the center of symmetry. C. The nascent strand switches its template strand to the complementary strand. D. The nascent strand elongates by DNA polymerization on the complementary strand. E. The nascent strand switches back to the original template strand and continues synthesis. Similar mutations can arise from both intra- and interstrand switch mechanisms.

intrastrand switching will convert the quasipalindrome into a perfect inverted repeat. The mutation can then be fixed by DNA repair or a second round of replication. Figure 5 illustrates such a process. Figure 5C is the first strand switch. The mutation arising from intramolecular templating is illustrated in blue. The second strand switch occurs in figure 5D.

Interstrand switching is similar to intrastrand switching in that the nascent strand and polymerase separate from the template strand, however, in an interstrand switching event, the nascent strand switches to the other DNA strand being replicated (46). Figure 6 illustrates an interstrand switching event. Figure 6A again represents complementary strands including an inverted repeat. In figure 6B, replication occurs into one half of the inverted repeat. Figure 6C illustrates the first interstrand switch followed by polymerization in figure 6D, and then a second switch and continued polymerization in figure 6E. Both types of switching have been demonstrated to occur *in vivo* (46, 47).

### 7.3. Diseases Associated with Imperfect Inverted Repeats

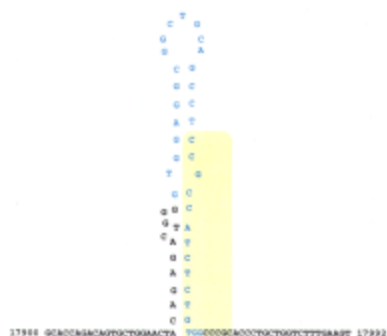
#### 7.3.1. Hereditary Angioneurotic Edema

There is strong evidence for imperfect inverted repeat mediated mutagenesis resulting in the human disease hereditary angioneurotic edema. This autosomal dominant disease is caused by mutations that reduce production of functional C1 inhibitor (50). C1 inhibitor is a pivotal regulatory protein in inflammation and regulates proteins involved in both complement, contact, and clotting cascades. When insufficient amounts of functional inhibitor are available for regulation, bradykinin (and possibly other kinins) are produced leading to tissue edema of the skin and mucosal surfaces. This causes symptoms ranging from temporary disfigurement to bowel obstructions or far worse, asphyxiation. The frequency of this disease is estimated at 1:50,000 to 1:100,000. During mutation screening, we identified two separate kindreds with mutations (51) that appeared to be mediated by a mechanism involving an imperfect inverted repeat (figure 7). This quasipalindrome was examined in a prokaryotic system and it has been found to cause mutations of similar types and sizes (Bissler, *et al.*, in preparation).

#### 7.3.2. Triplet repeat mediated diseases.

So far, twelve human genetic diseases have been associated with variations in the length of triplet repeats. These diseases include myotonic dystrophy, fragile X syndrome, and Huntington's chorea (52- 60). Clinically, these triplet repeat diseases are associated with a phenomenon called anticipation. Anticipation implies progressively earlier onset and worsening severity of the disease with successive generations (55, 58, 61). Anticipation is associated with expansion of the triplet repeat sequence in the disease associated gene. The mechanism of expansion is unclear, and many hypotheses involve abnormal replication or repair processes on a slipped strand structure within the triplet repeat region. These large triplet tracts are quasipalindromic, and Pearson *et al.* identified novel DNA secondary structures forming in

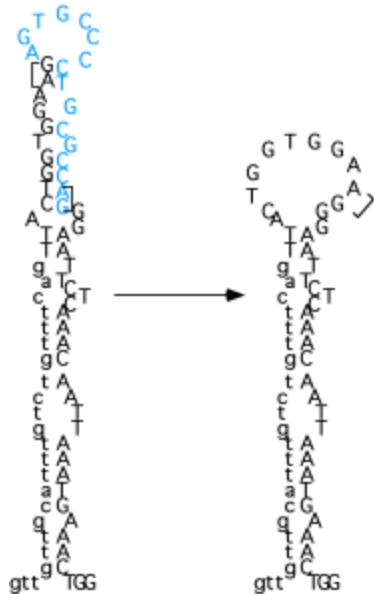
## Inverted repeats and disease



**Figure 7:** Quasipalindrome from the C1 Inhibitor gene. The blue sequence was deleted in one kindred and the sequence in the yellow background was tandemly duplicated in another kindred.



**Figure 8:** Sequence from d(CGG)n d(GCC)n repeat. A. A single strand of the triplet repeat. B. The same sequence in 8A configured into hairpin structure.



**Figure 9:** Genomic quasipalindrome from the Dystrophin gene. The lower case letters are intronic, the uppercase letters are exonic. The blue nucleotides are the 17 bases deleted. This could have happened by simple slipping on the AG direct repeat denoted by the brackets, or an interstrand switch mechanism.

(CTG) $_n$ (CAG) $_n$  tracts from the DM and FRAXA genes. They identified non-B-DNA structures, termed S-DNA (62). Such S-DNA can be remarkable thermostable, and this stability has been hypothesized to be due to hairpin formation in the loop-out regions (60). An example of hairpin formation in (CCG) $_n$  is illustrated in figure 8.

### 7.3.3. Duchenne Muscular Dystrophy

Duchenne muscular dystrophy is an X-linked recessive disease. Approximately one out of 3,500 male newborns are affected and will suffer a severe muscular wasting that will attenuate their lifespan to approximately 20 years of age. This disease has been mapped to the dystrophin gene at Xp21. Most cases (65%) reported have large (kilobase) intragenic deletions. The others are partial duplications, small frameshifts, or point mutations. One of these other mutants is a microdeletion between nucleotides 6982-6998 (63). There are several ways an imperfect inverted repeat contributes to mutagenesis in this gene (figure 9). For example, the polymerase machinery may simply slip on the AG direct repeat, or may resolve the hairpin structure up to the more G'C rich stem and then, during the resolution of the hairpin structure, slip-mispair on the direct repeats resulting in the deletion. With this 17 basepair deletion, the proposed hairpin has significantly less basepairing and presumably would be a weaker structure (figure 9).

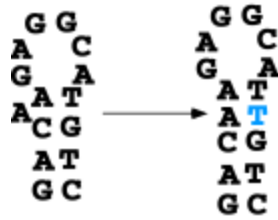
#### 7.3.4. Osteogenesis Imperfecta

The autosomal recessive disease Osteogenesis Imperfecta is manifest by remarkably brittle bones. This disease is associated with mutations in the type I procollagen gene. These mutations include deletions, insertions and point mutations often involving the triple helical coding region. The point mutations often disrupt the GXY repeat motif resulting in destabilization and delayed formation of the collagen triple helix. The kinetic alterations allow excessive post-translational hydroxylation and glycosylation of lysine. Bateman et al. report a mutation that results in a structural abnormality in the C-propeptide of the pro- $\alpha 1$  (I) chain (64-66). Carboxyl terminal propeptides of pro- $\alpha 1$  (I) and pro- $\alpha 2$  (I) are thought to be critical in initiating the process of chain assembly and helical propagation in the assembly of collagen. Mutations in this site, therefore, also would result in a disease such as Osteogenesis Imperfecta. The mutation identified was a T-insertion that can easily be explained by an inter- or intrastrand switch phenomenon resulting in the conversion of an imperfect inverted repeat to a palindromic sequence (figure 10).

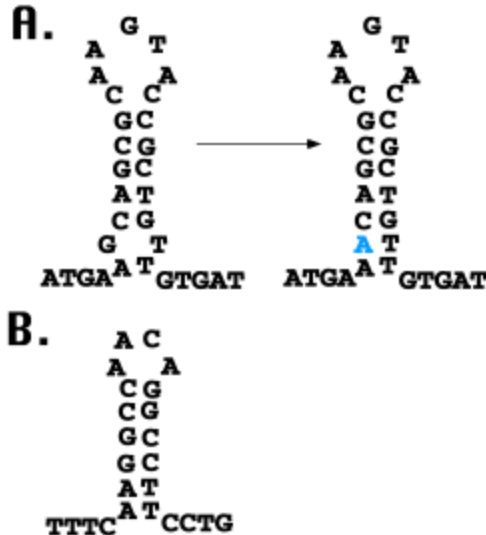
### 7.3.5. Antithrombin Deficiency

Mutations in the antithrombin gene can result in the development of venous thromboembolism at a young age (reviewed in 67). Mutation analysis of this inhibitor is consistent with the hypothesis that the reactive center coding region exhibits sequence directed mutagenesis involving inverted repeats. The P1 and P1' residues, on either side of the reactive center bond (involved in the inhibition reaction), are CpG dinucleotides. Twelve mutations have been reported involving these two codons. These mutations may best be explained by the methylation

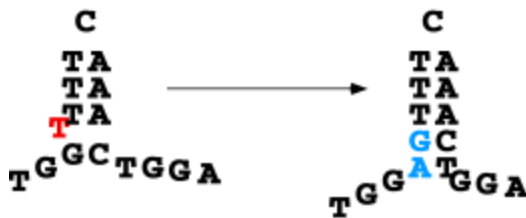




**Figure 10:** Sequence from the collagen I gene configured in a stem-loop structure. Using a strand switch mechanism (either intra- or interstrand) the stems of the inverted repeat are made perfectly complementary.



**Figure 11:** Quasipalindrome and palindrome from the Antithrombin gene. A. Nucleotides 13,800 to 13,831 are quasipalindromic and are illustrated as a stem-loop structure. The arrow shows the strand switch product that converts the stems of the quasipalindrome to a perfectly complementary sequence. The blue 'A' could result from synthesis templating from the 3' 'T' that would be mispaired with the 'G' residue. B. The quasipalindrome from 13,855 to 13,879. Only the loop, which does not basepair in the hairpin configuration is not palindromic. The sequence in bold identify the codons mutated.



**Figure 12:** Sequence from the cholinesterase gene. This structure likely could not form without unpairing of the intrastrand A/T on either side or the 'C' residue at the top. The arrow indicates that through a strand switch reaction, the sequence is made more palindromic as seen by the loss of the 'T' residue on the left, and its replacement with the 'AG' illustrated in blue, leading to a frameshift.

of the cytosine base followed by the spontaneous deamination resulting in a thymidine base substitution. In addition, mutations resulting in deficiency occur in two regions containing five codons on either side of the reactive center bond.

The coding region of the antithrombin III gene contains two mutable codons and is imperfectly palindromic. The sequence 5' to the reactive center coding region (nucleotides 13,800 to 13,831, accession number X68793) can be drawn into an imperfect cruciform structure containing the two codons in this region that mutate (figure 11A). This stem loop structure also may explain many of the point mutations in which the codon GCA is converted to ACA. A plausible explanation may be that during replication, a strand switch event occurred followed by a second round of replication that fixed this substitution in the genome and lead to disease. A carboxyl region of the reactive center coding region (nucleotides 13,855 to 13,879) can be configured into a stem loop structure as well (figure 11B). This region contains three codons that mutate and cause disease. It is tempting to suggest that these five codons on either side of the reactive center bond cluster mutations because these sequences can adopt a peculiar alternative secondary structure. No mutation has been found that actually deletes one of these potential stem loop structures. Perhaps no deletions have been identified because these are reasonably weak cruciform structures, or the sequence effects of contiguous nucleotides reduce the likelihood of deletion. The remnant effect may be on replication fidelity as the holoenzyme encounters the geometry and strength of such a structure.

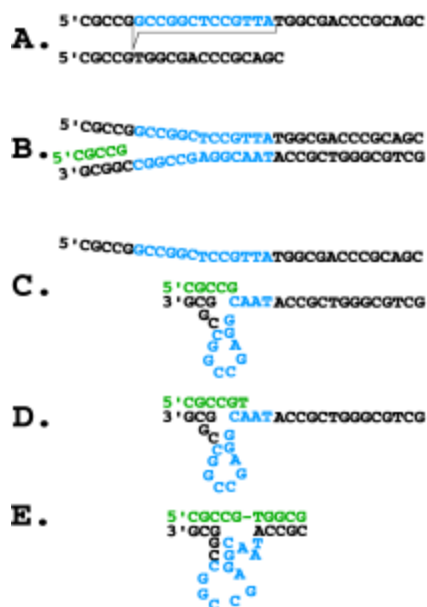
### 7.3.6. Silent Serum Cholinesterase

Deficiency of human serum cholinesterase is characterized by absence of enzymatic activity. The autosomal recessive phenotypic manifestation of this disorder is of concern for anesthesiologists and intensivists alike because this enzyme degrades the paralyzing agent succinyl choline. This drug is used in operating room and intensive care settings. The homozygous frequency of this disorder is estimated to be approximately 1:100,000 in the Caucasian population, but it is as high as 1:100 in western Alaskan Eskimos and 1:50 in the population of Andra Pradesh, India. Nogueira *et al.* found a patient that had the conversion of a quasipalindrome into a palindromic sequence that could be best explained by an intrastrand switch mechanism (68) (figure 12).

### 7.3.7. Lesch-Nyhan Syndrome

Mutations that inactivate the human hypoxanthine phosphoribosyl transferase (HPRT) gene lead to the X-linked neurological disease Lesch-Nyhan Syndrome. This syndrome is associated with polyarthrosis, spasticity, mental retardation, self mutilation, and an elevated serum uric acid. Gibbs *et al.* identified a 13 bp deletion that could result from a slip 3' (figure 13). Following the first slip there is additional polymerization of a 'T' residue, then a second slip that would place the nascent 3' end in register with the complementary DNA. Both slips could be facilitated by a hairpin structure in the template strand thereby eliminating the inverted repeat

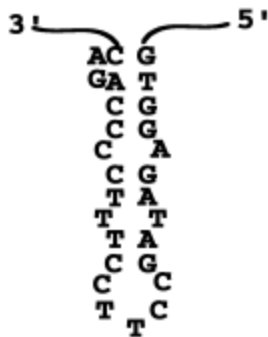
## Inverted repeats and disease



**Figure 13:** Sequence from the hypoxanthine-phosphoribosyl transferase gene. A. The top strand is the wild type coding strand from the hypoxanthine-phosphoribosyl transferase gene. The blue nucleotides depict the deleted sequence. The bottom sequence is the resulting deletion. B. The sequence is illustrated as duplex DNA. The melting of DNA is depicted by the 'Y' configuration and the green nucleotides represent the nascent strand. C. The nascent strand partially melts and the template strand forms a hairpin structure. The nascent strand comes into register so that the 3' can continue polymerization. D. The polymerase adds a 'T' residue to the nascent strand. E. A second realignment between the template and nascent strands occurs, making the hairpin structure stronger, and the nascent continues polymerization.



**Figure 14:** Mitochondrial DNA with the breakpoint identified by the different color sequence. The blue nucleotides are included in the deletion, and the black nucleotides are retained.



**Figure 15:** Quasipalindrome configured as a hairpin structure from the biotinase gene. Flawed resolution of this structure during DNA replication has been postulated to lead to deletion in the biotinase gene (72).

sequence. This mutation could then be fixed in the genome by another round of replication (69).

### 7.3.8. Kearns-Sayre Syndrome

Kearns-Sayre Syndrome is a constellation of physical findings including external ophthalmoplegia, retinal degeneration, and a cardiac conduction block. The genetic basis for this mitochondrial cytopathy resides in deletions in the mtDNA. Frequently, the deletions are flanked by a variable length direct repeat and is most easily explained by slip mispairing or errors in recombination. Remes *et al.* identified a 38 year old man who had deletion breakpoint flanked by an imperfect inverted repeat in a highly conserved sequence block (CSB II) that had been implicated in cleavage of RNA primers during initiation of DNA replication by MRP ribonuclease (70). The deletion includes nucleotides 10170 through 13406. The 3' breakpoint of this deletion resides in the middle of a small inverted repeat configured into a hairpin structure (figure 14). The blue nucleotides represent the portion of the inverted repeat that is included in the deletion, the black nucleotides depict those nucleotides remaining. The stem loop structure is small and of low energy, and will require confirmation of the role, if any, in mutagenesis.

### 7.3.9. Biotinase Deficiency

Biotinase Deficiency is an autosomal recessive disorder resulting in the inability to recycle biotin. This water-soluble vitamin is an essential cofactor for several carboxylases. Lack of functional biotinase can result in severe deficiency of biotin and manifest in clinical features including seizures, ataxia, developmental delay, acidosis, and coma (71). Pomponio *et al.* identified a complex deletion/insertion mutation that they postulated was mediated by a quasipalindrome that adopted a hairpin configuration in the template strand (72) (figure 15).

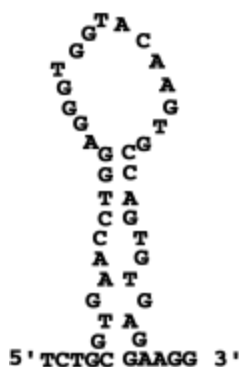
### 7.3.10. Familial Hypercholesterolemia

Familial hypercholesterolemia is an autosomal dominant disease affecting approximately 1 out of 500 people (73). While several different types of mutations can lead to this disease, Yamakawa-Kobayashi *et al.* identified a quasipalindrome that they postulated was responsible for a complex deletion/insertion mutation in exon 8 of the LDL receptor. The quasipalindrome is illustrated in figure 16. These authors postulated a model that involved a deletion and an intramolecular strand switch mechanism in two kindred they felt were unrelated (74) (figure 16).

## 8. PERSPECTIVES

Innate and environmentally mediated mutagenesis challenges the integrity of the human genome. On the grand scale, such mutation mechanisms facilitate the evolutionary process. This review focused on mutagenesis caused by imperfect inverted repeats. Understanding the basic biology of DNA replication and the challenges posed by alternative DNA structures can begin to lead to predictions and rules explaining such mutagenic mechanisms. This may lead to better strategies for mutation screening as well as for more stable DNA constructs for investigation and therapeutic applications.





**Figure 16:** Quasipalindrome sequence from the LDL receptor gene configured into a hairpin structure. Yamakawa-Kobayashi *et al.* postulated that this structure may be involved in the mutagenic mechanism leading to familial hypercholesterolemia (74).

## 9. ACKNOWLEDGMENT

I wish to thank Dr. Richard R. Sinden for his critical review and beneficial discussions of this manuscript. This work was supported by NIH Grants DK 02418-02 and NIAMS-96-1.

## 10. REFERENCES

1. Sarasin, A. and A. Stary: Human cancer and DNA repair-deficient diseases. *Cancer Detection & Prevention* 21: 406-11 (1997)
2. Minnick, D. T. and T. A. Kunkel: DNA synthesis errors, mutators and cancer. *Cancer Surveys* 28: 3-20 (1996)
3. J. E. Darnell, H. Lodish and D. Baltimore: Molecular Cell Biology. Scientific American Books, New York (1990)
4. Weaver, D. T. and M. L. DePamphilis: Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. *J Biol Chem* 257: 2075-86 (1982)
5. Cooper, D. N. and M. Krawczak: Mechanisms of insertional mutagenesis in human genes causing genetic disease. *Hum Genet* 87: 409-15 (1991)
6. Cooper, D. N. and M. Krawczak: The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum Genet* 85: 55-74 (1990)
7. Krawczak, M. and D. N. Cooper: Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum Genet* 86: 425-41 (1991)
8. van Dongen, M. J., M. M. Mooren, E. F. Willems, G. A. van der Marel, J. H. van Boom, S. S. Wijmenga and C. W. Hilbers: Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A base pair in the loop. *Nucleic Acids Res* 25: 1537-47 (1997)
9. Bauer, W. R., H. Ohtsubo, E. Ohtsubo and C. J. Benham: Energetics of coupled twist and writhe changes in closed circular pSM1 DNA. *J Mol Biol* 253: 438-52 (1995)

10. Esposito, F. and R. R. Sinden: DNA supercoiling and eukaryotic gene expression. [Review]. *Oxford Surveys on Eukaryotic Genes* 5: 1-50 (1988)
11. Wolffe, A. P., J. Wong and D. Pruss: Activators and repressors: making use of chromatin to regulate transcription. *Genes to Cell* 2: 291-302 (1997)
12. Sinden, R. R. : Supercoiled DNA: Biological Significance. *J Chem Ed* 64: 294-301 (1987)
13. Zheng, G., D. W. Ussery and R. R. Sinden: Estimation of superhelical density *in vivo* from analysis of the level of cruciforms existing in living cells. *J Mol Biol* 221: 122-9 (1991)
14. Palecek, E., P. Boublikova and K. Nejedly: Probing of DNA structure with osmium tetroxide. Effect of ligands. *Biophys Chem* 34: 63-8 (1989)
15. Greaves, D. R., R. K. Patient and D. M. Lilley : Facile cruciform formation by an (A-T)<sub>34</sub> sequence from a Xenopus globin gene. *J Mol Biol* 185: 461-78 (1985)
16. Lilley, D. M. and B. Kemper: Cruciform-resolvase interactions in supercoiled DNA. *Cell* 36: 413-22 (1984)
17. Ward, G. K., A. Shihab-el-Deen, M. Zannis-Hadjopoulos and G. B. Price: DNA cruciforms and the nuclear supporting structure. *Exp Cell Res* 195: 92-8 (1991)
18. Ward, G. K., R. McKenzie, M. Zannis-Hadjopoulos and G. B. Price: The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. *Exper Cell Res* 188: 235-46 (1990)
19. Frappier, L., G. B. Price, R. G. Martin and M. Zannis-Hadjopoulos: Monoclonal antibodies to cruciform DNA structures. *J Mol Biol* 193: 751-8 (1987)
20. Steinmetzer, K., M. Zannis-Hadjopoulos and G. B. Price: Anti-cruciform monoclonal antibody and cruciform DNA interaction. *J Mol Biol* 254: 29-37 (1995)
21. Panayotatos, N. and A. Fontaine: A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. *J Biol Chem* 262: 11364-8 (1987)
22. Kramer, P. R. and R. R. Sinden: Measurement of unrestrained negative supercoiling and topological domain size in living human cells. *Biochemistry* 36: 3151-8 (1997)
23. Ussery, D. W., R. W. Hoepfner and R. R. Sinden: Probing DNA structure with psoralen *in vitro*. *Methods in Enzymol* 212: 242-62 (1992)
24. Chan, C. L., D. Wang and R. Landick: Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J Mol Biol* 268: 54-68 (1997)
25. von Hippel, P. H., W. A. Rees and K. S. Wilson: RNA interactions in the regulation of transcription. *Nucleic Acids*

*Symposium Series* 1-4 (1995)

26. Wilson, K. S. and P. H. von Hippel: Stability of *Escherichia coli* transcription complexes near an intrinsic terminator. *J Mol Biol* 244: 36-51 (1994)

27. Farnham, P. J. and T. Platt: Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Res* 9: 563-77 (1981)

28. Laing, L. G. and K. B. Hall: A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry* 35: 13586-96 (1996)

29. Sierzputowska-Gracz, H., R. A. McKenzie and E. C. Theil: The importance of a single G in the hairpin loop of the iron responsive element (IRE) in ferritin mRNA for structure: an NMR spectroscopy study. *Nucleic Acids Res* 23: 146-53 (1995)

30. Jaffrey, S. R., D. J. Haile, R. D. Klausner and J. B. Harford: The interaction between the iron-responsive element binding protein and its cognate RNA is highly dependent upon both RNA sequence and structure. *Nucleic Acids Res* 21: 4627-31 (1993)

31. McMurray, C. T.: Mechanisms of DNA expansion. *Chromosoma* 104: 2-13 (1995)

32. McMurray, C. T., N. Juranic, S. Chandrasekaran, S. Macura, Y. Li, R. L. Jones and W. D. Wilson : Hairpin formation within the human enkephalin enhancer region. 2. Structural studies. *Biochemistry* 33: 11960-70(1994)

33. Darling, T. N., J. A. McGrath, C. Yee, B. Gatalica, R. Hametner, J. W. Bauer, G. Pohla-Gubo, A. M. Christiano, J. Uitto, H. Hintner and K. B. Yancey: Premature termination codons are present on both alleles of the bullous pemphigoid antigen 2/type XVII collagen gene in five Austrian families with generalized atrophic benign epidermolysis bullosa. *J Invest Derm* 108: 463-8 (1997)

34. Chavanas, S., Y. Gache, G. Tadini, L. Pulkkinen, J. Uitto, J. P. Ortonne and G. Meneguzzi: A homozygous in-frame deletion in the collagenous domain of bullous pemphigoid antigen BP180 (type XVII collagen) causes generalized atrophic benign epidermolysis bullosa. *J Invest Derm* 109: 74-8 (1997)

35. Gatalica, B., L. Pulkkinen, K. Li, K. Kuokkanen, M. Ryyanen, J. A. McGrath and J. Uitto: Cloning of the human type XVII collagen gene (COL17A1), and detection of novel mutations in generalized atrophic benign epidermolysis bullosa. *Am J Hum Gene* 60: 352-65 (1997)

36. Li, K., K. Tamai, E. M. Tan and J. Uitto: Cloning of type XVII collagen. Complementary and genomic DNA sequences of mouse 180-kilodalton bullous pemphigoid antigen (BPAG2) predict an interrupted collagenous domain, a transmembrane segment, and unusual features in the 5'-end of the gene and the 3'-untranslated region of the mRNA. *J Biol Chem* 268: 8825-34 (1993)

37. Sinden, R. R.: DNA Structure and Function. Academic Press, New York (1994)

38. Kornberg, A., T. A. Baker: DNA Replication. Freeman Press, New York (1992)

39. Jin, R., M. E. Fernandez-Beros and R. P. Novick: Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform? *EMBO J* 16: 4456-66 (1997)

40. Markiewicz, P., C. Malone, J. W. Chase and L. B. Rothman-Denes: *Escherichia coli* single-stranded DNA-binding protein is a supercoiled template-dependent transcriptional activator of N4 virion RNA polymerase. *Genes & Devel* 6: 2010-9 (1992)

41. Glucksmann, M. A., P. Markiewicz, C. Malone and L. B. Rothman-Denes: Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. *Cell* 70: 491-500 (1992)

42. Toose, J.: Molecular Biology of Tumor Viruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1982)

43. Lockshon, D. and D. A. Galloway: Cloning and characterization of oriL2, a large palindromic DNA replication origin of herpes simplex virus type 2. *J Virol* 58: 513-21 (1986)

44. Schaaper, R. M., B. N. Danforth and B. W. Glickman: Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli* lacI gene. *J Mol Biol* 189: 273-84 (1986)

45. Glickman, B. W. and L. S. Ripley: Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc Nat Acad Sci USA* 81: 512-6 (1984)

46. Rosche, W. A., T. Q. Trinh and R. R. Sinden: Leading strand specific spontaneous mutation corrects a quasipalindrome by an intermolecular strand switch mechanism. *J Mol Biol* 269: 176-87 (1997)

47. de, B. J. and L. S. Ripley: Demonstration of the production of frameshift and base-substitution mutations by quasipalindromic DNA sequences. *Proc Nat Acad Sci USA* 81: 5528-31 (1984)

48. Papanicolaou, C. and L. S. Ripley: An in vitro approach to identifying specificity determinants of mutagenesis mediated by DNA misalignments. *J Mol Biol* 221: 805-21 (1991)

49. Papanicolaou, C. and L. S. Ripley: Polymerase-specific differences in the DNA intermediates of frameshift mutagenesis. *In vitro* synthesis errors of *Escherichia coli* DNA polymerase I and its large fragment derivative. *J Mol Biol* 207: 335-53 (1989)

50. Bissler, J. J., K. S. Aulak, V. H. Donaldson, F. S. Rosen, M. Cicardi, R. A. Harrison and A. E. Davis III : Molecular defects in hereditary angioneurotic edema. *Proc Asso. Am Physicians* 109: 164-173 (1997)

51. Bissler, J. J., V. H. Donaldson and A. E. Davis III: Contiguous deletion and duplication mutations resulting in type I hereditary angioneurotic edema. *Hum Genet* 93: 265-9 (1994)
52. Zoghbi, H. Y.: Molecular genetics and neurobiology of neurodegenerative and neurodevelopmental disorders. *Pediatr Res* 41: 722-6 (1997)
53. Mitas, M.: Trinucleotide repeats associated with human disease. *Nucleic Acids Res* 25: 2245-54 (1997)
54. Reddy, P. S. and D. E. Housman: The complex pathology of trinucleotide repeats. *Curr Opin Cell Biol* 9: 364-72 (1997)
55. La Spada, A. R.: Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol* 7: 943-63 (1997)
56. Nance, M. A.: Clinical aspects of CAG repeat diseases. *Brain Pathol* 7: 881-900 (1997)
57. Tsuji, S.: Molecular genetics of triplet repeats: unstable expansion of triplet repeats as a new mechanism for neurodegenerative diseases. *Intern Med* 36: 3-8 (1997)
58. Ashley, C. T., Jr. and S. T. Warren: Trinucleotide repeat expansion and human disease. *Ann Rev Genet* 29: 703-28 (1995)
59. Pearson, C. E. and R. R. Sinden: "Slipped Strand DNA, Dynamic Mutations, and Human Disease." In: Genetic Instabilities and Hereditary Neurological Diseases. Eds: Wells and Warren. Academic Press, New York (1998:in press)
60. Pearson, C. E. and R. R. Sinden: Trinucleotide Repeat DNA Structures: Dynamic Mutations from Dynamic DNA. *Curr Opin Struct Biol* (1998:in press)
61. Monros, E., M. D. Molto, F. Martinez, J. Canizares, J. Blanca, J. J. Vilchez, F. Prieto, R. de Frutos and F. Palau: Phenotype correlation and intergenerational dynamics of the Friedreich ataxia GAA trinucleotide repeat. *Am J Hum Genet* 61: 101-10 (1997)
62. Pearson, C. E. and R. R. Sinden: Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* 35: 5041-53 (1996)
63. Robinson, D. O., D. J. Bunyan, H. A. Gabb, I. K. Temple and S. C. Yau: A small intraexonic deletion within the dystrophin gene suggests a possible mechanism of mutagenesis. *Hum Genet* 99: 658-62 (1997)
64. Bateman, J. F., S. R. Lamande, H. H. Dahl, D. Chan, T. Mascara and W. G. Cole: A frameshift mutation results in a truncated nonfunctional carboxyl-terminal pro alpha 1(I) propeptide of type I collagen in osteogenesis imperfecta. *J Biol Chem* 264: 10960-4 (1989)
65. Lamande, S. R., S. D. Chessler, S. B. Golub, P. H. Byers, D. Chan, W. G. Cole, D. O. Sillence and J. F. Bateman: Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro alpha 1 (I) chain carboxyl-terminal propeptide which impair subunit assembly. *J Biol Chem* 270: 8642-9 (1995)
66. Chessler, S. D., G. A. Wallis and P. H. Byers: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of type I collagen result in defective chain association and produce lethal osteogenesis imperfecta. *J Biol Chem* 268: 18218-25 (1993)
67. Davis III, A. E., J. J. Bissler and K. S. Aulak: Genetic defects in the C1 inhibitor gene. *Complement Profiles* 1: 133-150 (1993)
68. Nogueira, C. P., M. C. McGuire, C. Graeser, C. F. Bartels, M. Arpagaus, A. F. Van der Spek, H. Lightstone, O. Lockridge and B. N. La Du: Identification of a frameshift mutation responsible for the silent phenotype of human serum cholinesterase, Gly 117 (GGT----GGAG). *Amer J Hum Genet* 46: 934-42 (1990)
69. Gibbs, R. A., P. N. Nguyen, A. Edwards, A. B. Civitello and C. T. Caskey: Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families. *Genomics* 7: 235-44 (1990)
70. Remes, A. M., K. J. Peuhkurinen, R. Herva, K. Majamaa and I. E. Hassinen: Kearns-Sayre syndrome case presenting a mitochondrial DNA deletion with unusual direct repeats and a rudimentary RNAase mitochondrial ribonucleotide processing target sequence. *Genomics* 16: 256-8 (1993)
71. Wolf, B., R. E. Grier, R. J. Allen, S. I. Goodman, C. L. Kien, W. D. Parker, D. M. Howell and D. L. Hurst: Phenotypic variation in biotinidase deficiency. *J Pediatr* 103: 233-7 (1983)
72. Pomponio, R. J., V. Narasimhan, T. R. Reynolds, G. A. Buck, L. F. Povirk and B. Wolf: Deletion/insertion mutation that causes biotinidase deficiency may result from the formation of a quasipalindromic structure. *Hum Mol Genet* 5: 1657-61 (1996)
73. Goldstein, J. L. and M. S. Brown: Familial hypercholesterolemia. Eds: C. C. Scriver, A.L. Beaudet, W.S. Sly and D. Valle. McGraw-Hill, New York (1989)
74. Yamakawa-Kobayashi, K., T. Kobayashi, H. Yanagi, Y. Shimakura, J. Satoh and H. Hamaguchi: A novel complex mutation in the LDL receptor gene probably caused by the simultaneous occurrence of deletion and insertion in the same region. *Hum Genet* 93: 625-8 (1994)

**Key words:** Molecular Biology, DNA Structure, Inverted Repeat, Cruciform, Quasipalindrome, Mutagenesis, Hairpin

Send correspondence to: John J. Bissler, M.D., Children's Hospital Research Foundation #5, 3333 Burnet Ave., Cincinnati, Ohio 45229-3039, Tel:513-636-4531 Fax: 513-636-7407, E-mail: john.bissler@chmcc.org

**Note:** The figures in this manuscript can be accessed in color at <http://www.bioscience.org/1998/v3/d/bissler/list.htm>