We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

21

The Gratuitous Repair on Undamaged DNA Misfold

Xuefeng Pan[1,2], Peng Xiao[1], Hongqun Li[2], Dongxu Zhao[1] and Fei Duan[2]
[1]*School of Life Science, Beijing Institute of Technology, Beijing*
[2]*Health Science Center, Hebei University, Baoding*
China

A Paradox of Life
B-DNA is needed for maintenance of genetic stability, while it will convert into non-B DNA in replication, repair, transcription or recombination, leading to exposure of bases, single strands, and even introduction of distortions. All these could intrigue gratuitous repair on undamaged DNA using the conventional repair, recombination mechanisms.
Repair or not Repair, turns to be a question?

1. Introduction

In natural genomes, tens of DNA structure analogous to B-DNA conformation have been found to be formed through compiling weak interacting forces, including hydrophobic, Van der Waals and hydrogen-bond accepters and donors and inductions of certain agents (Rao et al., 2010). Of which, hairpins, cruciform junctions, Z-DNA, G-tetrads/quadruplexes, helices, loops and bulges are most studied so far.

Since the late 1950s, the roles of the non-B DNA structures in biological functions have begun to be enlightened (Watson & Crick, 1953; Wilkins et al., 1953a, 1953b; Svozil et al., 2008). Piling up results suggest that non-B conformations, such as cruciforms, triplexes, tetraplexes, can interact with proteins involving DNA metabolism, including replication, gene expression and recombination, or influence nucleosomes and other supramolecular structures formation (Wang & Griffith, 1996; Shimizu et al. 2000). However, non-B DNA secondary structures may also be treated as DNA mis-folds by DNA repair systems. Because of which the non-B DNA secondary structures can serve as end points for several types of genome rearrangements seen in some diseases (Wang & Vasquez, 2006; Wells, 2007; Bacolla & Wells, 2009; Chen et al., 2010).

2. DNA sequences which are susceptible to abnormal folding

The non-B DNA structure forming sequences are found to be rich in genomes from divergent organisms (Table 1) (Cox & Mirkin, 1997; Svozil et al., 2008; Cerz et al., 2011). For example, nearly half of the human genome consists of repetitive sequences, which can be arranged as inverted, direct tandem, and homopurine–homopyrimidine mirror repeats.

These repeat sequences are major contributors to forming non-B DNA structures, although the unusual structures can also be formed by various other sequences that are not repeating tracts (Svozil et al., 2008; Cerz et al., 2011). Repeat DNA sequences may adopt either orthodox right-handed B-DNA or non-B DNA conformations at specific sequence motifs as a function of negative supercoil density, created by transcription, protein binding, and other reasons. For example, inverted repeats can form B conformation in cells, while also forming hairpin structures, slipped structures with looped-out bases, four-stranded G-quartet structures, left-handed Z-DNA and intramolecular triplex DNA structures (H-DNA) depending on the base compositions and the arrangements.

Structural Feature	human	Chimpanzee	Macaque	Dog	Mouse
Cruciform	197910	190736	128334	172032	188532
Slipped Motif	347969	314516	305285	404750	695150
Triplex Motif	179623	105640	140580	303385	565479
Z-DNA Motif	294320	278928	280982	261012	690276
G-Tetraduplex	374545	314171	298142	492535	559280
Direct repeats	871045	787335	765798	968955	1593107
Inverted repeats	1044533	998249	843889	814080	801242
Mirror Repeats	1651723	1485135	1455025	1849897	1651723

Table 1. Non-B DNA motifs in different mammalian genomes (Cer et al., 2011)

2.1 Cruciform motif

DNA sequence that reads the same from 5' to 3' in either strand of a duplex is called as inverted repeat or palindrome DNA sequence. This subset of inverted repeat sequences may fold-back and form intramolecular, antiparallel, double helices stabilized by Watson–Crick hydrogen bonds (van Holde & Zlatanova, 1994; Courey, 1999; Smith, 2008).

As a whole, the interstrand hydrogen bonds in the inverted repeats must be broken, and intrastrand hydrogen bonds form between the complementary bases in each single strand, forming two hairpin-like arms with small (3-4 unpaired bases) loop at their tips. The structure looks similar to a four-way junction, of which the nucleobases in and around the junction are fully involved in base pairing.

2.2 Potential quadruplex sequences

Potential quadruplex sequences are usually G-rich, such as the DNA sequences in eukaryotic telomeres, and in non-telomeric genomic DNA, like the nuclease-hypersensitive promoter regions (Burge et al., 2006; Rawal et al., 2006; Qin & Hurley, 2008; Sannohe & Sugiyama, 2010). To form a quadruplex, the DNA sequences have to form overlapping four G-blocks. Each contains the same number (n) of G bases (n vary from 3 to 7), on each strand, and/ or separated by 1–7 nt (Burge et al., 2006). The potential unimolecular G-quadruplex forming sequences (i.e. intramolecular) can be expressed as follows (Burge et al., 2006):

GaXbGaXcGaXdGa

Where "a" is the number of G residues in each short G-tract, which are usually directly involved in G-tetrad. Xb, Xc and Xd can be any combination of residues, including G, forming the loops.
The potential quadruplex sequences were therefore restricted to:

$$\mathbf{G_{3\text{-}5}NLoop1G_{3\text{-}5}NLoop2G_{3\text{-}5}NLoop3G_{3\text{-}5}}$$

Where NLoop1-3 are loops of unknown length, within the limits 1<NLoop1-3 <7 nt.

2.3 Z-DNA motif

In 1979, DNA sequence of d (CpGpCpGpCpG) was crystallized and found to adopt a left-handed conformation (the Z-DNA conformation) with altered helical parameters relative to right-handed B-form (Rich et al., 1983; Mirkin, 2008). Later, it was realized that DNA sequences with alternating pyrimidines and purines, such as $(CA:TG)_n$ and $(CG:CG)_n$, may wind a double helix into a left-handed zigzag form (Z-DNA). Z-DNA is thinner (18 Å) than B-DNA (20 Å), due to its bases shifting to the outskirts of a double helix. It has only one deep, narrow groove equivalent to the minor groove in B-DNA.
In general, five or more tandem repeats, each comprising an alternating pyrimidine–purine dinucleotide motif, in which the pattern YG is preserved on at least one of the DNA strands can adopt Z-DNA.

2.4 Triplex motif

A subset of mirror repeat sequences comprise only purines (A and G, R) or pyrimidines (C and T, Y) on the same strand of a double stranded DNA, separated by few (0~8) nucleotides. These DNA motifs can adopt various intramolecular three-stranded analogous (triplex, H-DNA) stabilized by Hoogsteen hydrogen bonds (Casey & Glazer, 2001; Mukherjee & Vasquez 2011).
For a sequence requirement in forming triplex DNA is thought to be that only R · Y-containing mirror repeats can yield A: A*T and G: G * C triads. When the hydrogen bonds in the A · T and G · C base pairs are formed in canonical B-form DNA, several hydrogen bond forming groups in the bases can still be free unpaired. Each purine base has two hydrogen bond forming groups on the edges that are posed in the major groove. These unpaired bases can be used to form base triads that are unit blocks of triple-stranded DNA (see the following explanation for detail).
In theory, a homopurine-homopyrimidine duplex can form triplexes of either purine (Pu) motif (purine, antiparallel motif) or pyrimidine (Py) motif (pyrimidine, parallel motif). However, under physiological conditions, cytosine protonation is not favored, and C·G*G becomes therefore the most stable triad in a Pu motif. To form an intermolecular or intramolecular triplex, adjoining homopurine-homopyrimidine tracts of at least 10 base pairs are normally required for a duplex acceptor, since shorter than that the triplexes formed can be unstable under physiological conditions (Fox & Brown, 2011).
A triplex may be mutagenic *in vivo*, as double-strand breaks may occur in or near the triplex site, which if with DNA replication, recombinational repair may produce triplex mediated mutagenesis (Chan et al., 1999; Faruqi et al., 2000).
Triplex can also be formed in RNA transcription, although it is a kinetically unfavored compared to duplex annealing. However triplex RNA and DNA are stable, showing half-lives on the order of days, which may involve the molecular mechanism of Friedreich's ataxia (FRDA) (Pan et al., 2009).

3. The non-B DNA structures and non-B DNA structure-induced genetic instability

3.1 DNA loops/ bulges and slipped DNA

DNA loops and bulges are similar non-B DNA structures sharing common features of unpaired bases of different number (Fig. 1). They can be formed in anywhere by any DNA sequence in natural genome, therefore they may be the most frequent non-B DNA conformations in genomes. For example, $(CA \cdot TG)_n$ DNA sequences are found to exist everywhere in eukaryotic genomes as of 60 base pairs tracts. $(CA \cdot TG)_n$ forms both classical right-handed DNA double helix, and diverse alternative conformations including small DNA loops or bulges (Kladde et al., 1994; Ho, 1994).

Genomic instabilities can also be caused by DNA loops and bulges, which are often seen as slippage instabilities or insertion/deletion (I/ D) instabilities (Pan, 2004). Proteins that bind DNA loops and bulges are also found and mainly known to be mismatch repair proteins (Parker & Marinus, 1992; Carraway & Marinus, 1993; Fang et al., 2003; Kaliyaperumal et al., 2011).

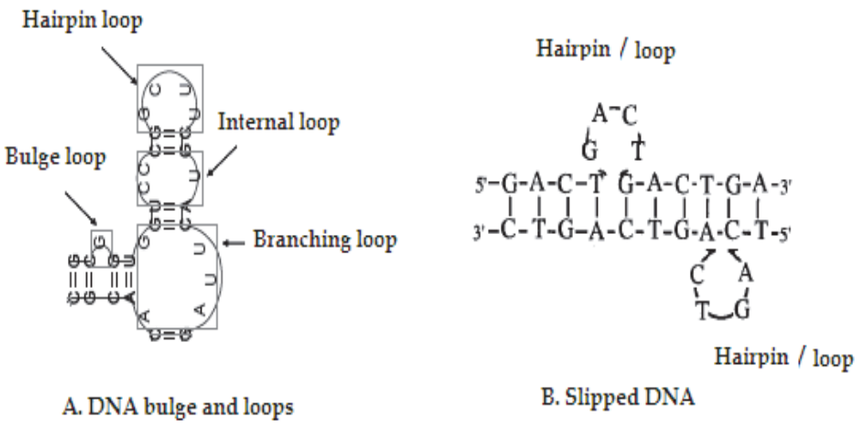


Fig. 1. DNA Loops and bulges

3.2 Branched structures

A branched DNA structure refers to a non-B DNA secondary structure with structured or unstructured "branch". For example, DNA intermediates appeared in homologous recombination as 3- and 4- way junctures are such branched DNA structures with differently oriented double helix arms. Similarly, flapped DNA structures appeared in processing Okazaki fragment in the lagging strand DNA replication also belong to branched DNA. Branched DNA migrates more slowly than their B-DNA conformation having same molecular weight and base composition. Importantly, branched DNA structures can also make genomic instability when in processing.

3.3 Hairpin/ cruciform and genetic instability

A hairpin can be formed at one strand of an inverted repeat, whereas a cruciform consists of two hairpin structures, both in each strand at the same position of the DNA (Fig.2)(Courey, 1999). Similarly some tandem arranged trinucleotide repeats such as CAG, CTG, CCG, CGG,

AAT, ATT etc. can also adopt hairpin structures with mismatched base pairs in the stem (McMurray, 1999; Trotta, et al., 2000).

To form a hairpin/cruciform, DNA duplex needs to be unwound in replication, transcription, and/or DNA repair processing; affording single-stranded repeat sequences the opportunity to base pair with itself in an intramolecular fashion. The term of "cruciform" originates from forming two duplex arms, which adopts either an "open" form, allowing strand migration or a "stacked" (locked) form, where the helices stack on each other (Courey, 1999; Khuu et al., 2006; Lilley, 2010). In both cases, the overall conformation and the intraduplex angles behave like the Holliday junction recombination intermediates (Fig.2A) (Courey, 1999; Khuu et al., 2006;; Lilley, 2010).

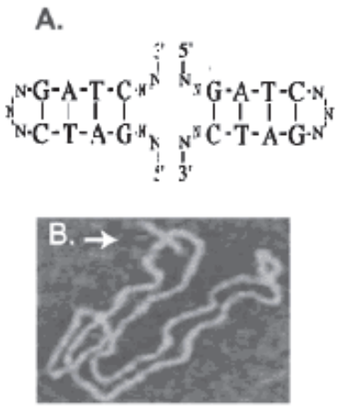


Fig. 2. Hairpin/cruciform of DNA

Both inverted repeats and tandem arranged trinucleotide repeats were found to be mutagenic, causing genomic instability. Inverted repeats were initially found to cause deletions in *E. coli* (Sinden et al., 1991), and then were seen in humans as (8; 22) (q24.13; q11.21), and many types of t (11; 22) translocations. The breakpoints of these translocation mutations were localized at the center of AT-rich palindromic sequences on 11q23 and 22q11, respectively. So far, t (11; 22) is the only known recurrent, non-Robertsonian translocation in humans, in some cases leads to male infertility and recurrent abortion (Kurahashi et al., 2000, 2006, 2010; Kurahashi & Emanuel, 2001). Furthermore, deletions stimulated by a poly (R.Y) sequence from intron 21 of the polycystic kidney disease 1 gene (PKD1) have also been characterized (Bacolla et al., 2001； Patel et al., 2004). And a long $(CCTG\text{-}CAGG)_n$ repeat in *E.coli* was also found to form cruciform (Pluciennik et al., 2002; Dere & Wells, 2006). Interestingly, cruciform-forming inverted repeats have mediated many of the microinversions in evolution that distinguish the human and chimpanzee genomes (Kolb et al., 2009).

In cells, DNA double strand breaks can be derived from cruciform, because hairpin/ cruciform are substrates for several structure-specific nucleases and/ or repair enzymes, such as SbcCD in *E.coli* and Mre11-Rad50 in eukaryotes. The actions of such enzymes make strand breaks, which may result in rearrangements or translocation of chromosomes (Smith, 2008).

In addition, proteins working in nucleotide excisonal repair (NER) can also recognize the helical distortions in hairpin, therefore NER may recognize DNA hairpin to resolve the hairpin in the DNA.

Besides, some other proteins were also found to bind the structural elements in cruciforms. For example, HMG proteins, replication initiation protein RepC, cruciform binding protein CBP, and four-way junction resolvases have all been indentified to bind cruciforms (Pearson et al., 1996; Jin et al., 1997; Novac et al., 2002; Lange et al., 2009; lilley, 2010).

3.4 Z-DNA and genetic instability

Z-DNA can be seen as the high-energy conformers of B-DNA that forms *in vivo* during transcription as a result of torsion strain generated by a moving polymerase (Wang, 1984; Casasnovas & Azorin, 1987; Johnston, 1988; Hebert & Rich, 1996). It has been thoroughly studied since 1957, how a right-handed B-DNA adopting a Z-DNA *in vitro* through ''flipping" the base pairs upside down, and rotating every other purine from *anti* to *syn* conformation (Johnston, 1988; Hebert & Rich, 1996). Compared to B-DNA, Z-DNA does not have a major groove, therefore could potentially impact transcription by physically blocking RNA polymerase, or by relaxing negative supercoiling turns, or by acting as an enhancer through recruiting transacting factors.

In Z-DNA, the guanosine nucleotides are in *syn* position where the bases are found over the sugar without protection, thus more accessible to DNA damaging factors, more resistant to processing by DNA repair enzymes. For example, alkylating damage such as N^7-methylguanine, which is typically removed by a DNA glycosylase in B-DNA is not efficiently repaired when present in Z-DNA (Pfohl-Leszkowicz et al., 1983; Boiteux et al., 1985).

Further, DNA sequences with the potential to adopt Z-DNA are associated with recombination hot spots in eukaryotic cells (Wang et al., 2006). A hot spot of 1000 bp in the major histocompatibility complex (MHC) in mice, containing several copies of long GT repeats, may account for up to 2% of the recombination events occurring on the chromosome (Crouau-Roy, 1999). In *E.coli*, the RecA molecules show a much higher binding affinity for Z-DNA than for normal B-DNA and single-stranded DNA, and show a Z-DNA structure-stimulated ATPase activity, implicating a recombination hot spot of Z-DNA in prokaryotes as well. Genetic recombination in Z-DNA can potentially induce deletion instability and/ or produce DNA double-strand breaks. For example, a CG (12) sequence forming Z-DNA induces high levels of genetic instability in both bacterial and mammalian cells (Casasnovas & Azorin, 1987).

Recently, proteins binding Z-DNA are found, including specific proteins, such as Zα domain-containing proteins ADAR1 and ESL, and fairly low specific proteins, such as HMG proteins (Suda et al., 1996; Lange et al., 2009).

3.5 H-DNA and H-DNA induced DSBs and genetic instability

H-DNA, alternatively known as triplex DNA can be classified into either pyrimidine motif or purine motif according to the orientation and composition of the third strand in a triple stranded DNA structure (Fig. 3). The third strand can form either Hoogsteen or reverse-Hoogsteen hydrogen bonds with the purine-rich strand of the duplex DNA. Therefore, the third strand can be both pyrimidine-rich and parallel to the complementary strand (Y* R: Y) or purine-rich and antiparallel to the complementary strand (R* R: Y), producing either pyrimidine motif or purine motif triple stranded DNA (as described previously).

Whereas (R* R: Y) triplexes form under conditions of physiological pH, triplex of the (Y* R: Y) composition form most readily under conditions of acidic pH. At physiological pH, triplex

may be stabilized by negative supercoiling, modified with phosphorothioate groups, or polyvalent cations such as spermine and spermidine. For the R* R: Y intramolecular triplexes and T: A* T and C^+: G* C triplets for the Y* R: Y intramolecular triplexes are included since these are considered the most stable triplet combinations.

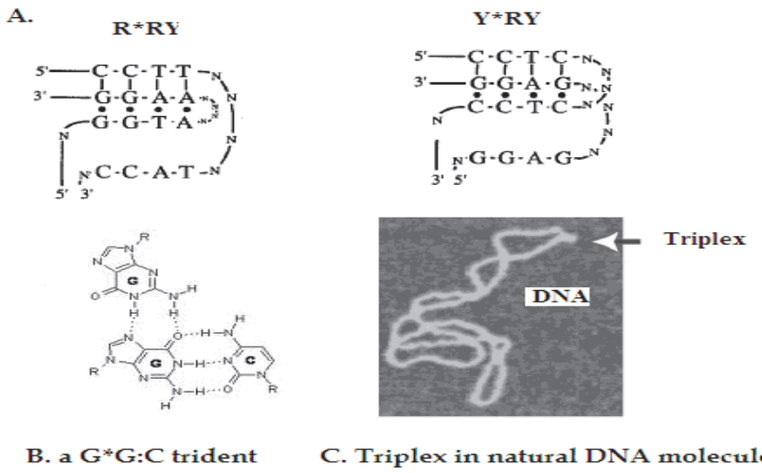


Fig. 3. H-DNA (Star/ Dot marks Hoogsteen hydrogen bonded bases; colon/ line shows Watson–Crick hydrogen bonded bases)

In general, formation of a triplex DNA was a role of sequence, topology (supercoil density), ionic conditions, protein binding, methylation, carcinogen binding, and other factors. Global negative supercoil density acts in concert with local transient waves of topological changes produced by replication or transcription, and both have a critical influence on forming and stabilizing triplex DNA *in vivo*. It has been reported that a higher negative supercoiling destabilized long CTG· CAG, CCG· CGG, and GAA· TTC repeats in *Escherichia coli*. Similarly a 2.5-kb poly (R · Y) tract from the human PKD1 gene lowered the viability of the host cells (Bacolla et al., 2001; Patel et al., 2004).

Several types of DNA damages induced by H-DNA have been reported, including single and/ or double strand breaks. For example, the endogenous H-DNA forming sequences from the human c-*myc* promoter was shown to be intrinsically mutagenic in mammalian cells because of the generation of either single or double strand breaks in the H-DNA, or near the H-DNA locus. Besides, the single-stranded area, or the triplex region is also a target of various nucleases, resulting in single or DSBs formation, and the increased mutagenesis or recombination (Wang & Vasquez, 2006).

Although triplex (H-DNA) DNA occurs mainly at poly (purine·pyrimidine) $((R \cdot Y)_n)$ tracts, it can also be induced to form with the sequence specific DNA recognition and binding of some synthetic triplex-forming oligonucleotides (TFOs) (Casey & Glazer, 2001; Mukherjee & Vasquez, 2011). TFOs bind to the major groove of homopurine-homopyrimidine stretches of double-stranded DNA to induce forming the triplex (Casey & Glazer, 2001; Mukherjee & Vasquez, 2011). During which the duplex DNA may have to undergo helical distortions on TFO binding and the distortions trigger endogenous recombination and repair mechanisms in the cell (Raghavan et al., 2004, 2005).

Indeed it has been reported that formation of TFO-induced triplex can induce sequence-specific DNA damages both in cells and in animals (Chan, et al., 1999; Kalish et al., 2005).

However, mismatch repair proteins are not involved in this TFO-induced mutagenesis. Several reports have now shown that cells that are deficient in the MutS and MutL homologues MSH2, MLH1, MSH3, or MSH6, do not show any change in TFO-induced mutagenesis. In contrasts, NER factors can recognize the intermolecular triplex at least in part. Therefore NER was involved in the triplex-induced mutagenesis and recombination in cells. For example, in *E.coli*, NER proteins, such as UvrB and UvrC, were necessary for H-DNA-induced cell growth retardation and cell lysis, similarly, recombination induced by TFOs depends also on the NER pathway (Faruqi et al., 2000).

3.6 G-tetraduplex and genetic instability

G-quadruplexes are higher-order DNA or RNA structures formed from G-rich DNA or RNA sequences that are built around tetrads of hydrogen-bonded guanine bases (Lipps & Rhodes, 2009; Sannohe & Sugiyama, 2010). Despite the wide prevalence of genomic sequences that have G-rich property and that can potentially fold into tetraplex / quadruplexes structures, a direct demonstration of their existence *in vivo* proved to be a difficult undertaking. Only recently has there evidence started to increase for their presence and role *in vivo* (Lipps & Rhodes, 2009), since most of the tetraplex/ quadruplexes forming sequences are fairly short and quadruplexes are likely to be transiently formed. G-quadruplexes (tetraduplex) may have several isomers which can be formed intramolecularly and intermolecularly (Fig. 4).

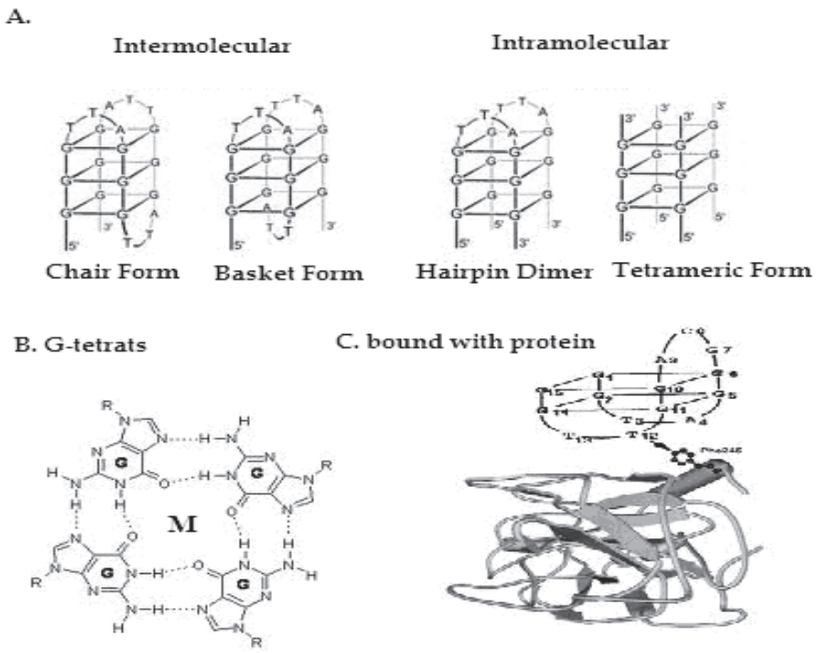


Fig. 4. Tetraplex DNA

Recent progress of the related studies revealed that G-quadruple could provide a nucleic acid based mechanism, such as regulating telomere maintenance, transcription, replication as well as translation. In the same time, various G-quadruplexes binding proteins, such as, a G4 quadruplex and purine Motif triplex nucleic acid-binding protein have also been characterized (Dyke et al.,2004), many others have been summarized in the reference (Fry, 2007).
The existence of cellular proteins that preferentially interact with tetraplex DNA provides a strong argument for the existence of quadruplex formations in genomic DNA.

3.7 Unwound DNA

Unwound DNA is known to be formed by A+T -rich sequences (Fig. 5). Since A · T base pairs contain two hydrogen bonds and C · G base pairs contain three, A · T-rich tracts are less thermally stable than C · G -rich tracts in DNA. In the presence of superhelical energy, A+T -rich regions can unwind and remain unwound under conditions normally found in the cell. Such sites often provide places for DNA replication proteins to enter DNA to begin the chromosome duplication. Unwound DNA can therefore be alternatively called as DNA unwinding elements (DUEs) that have been identified in both prokaryotic and eukaryotic DNA sequences. DUEs are AT-rich sequences about 30-100 bp long. They share little sequence similarity except for being AT-rich. Under torsion stress, unwinding of the double helix occurs first in AT-rich sequences; therefore, DUEs can be maintained as unpaired DNA regions in the presence of negative supercoiling. The single-stranded area of the unwound structure may be target of nuclease activity resulting in single or DSBs, leading to enhanced mutagenesis or recombination.

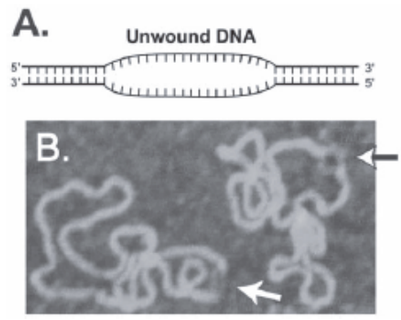


Fig. 5. Unwound DNA

3.8 Curved DNA

Normally, curved DNA is often seen in DNA segment containing runs of three or four bases of A in one strand and a similar run of T in the other and spaced at 10-base pair intervals. Interestingly we have recently found that trinucleotide repeats AAT can also adopt curved DNA in *E.coli*, which can be repressed by H-NS and its stimulated IS1E transposition (Pan et al., 2010)

4. Biological significance of DNA abnormal folding

Apart from the roles in DNA replication, transcription and gene regulation, non-B DNA may also lead to gene instability, including chromosomal translocation, deletion and amplification in cancer and other human diseases (Bacolla & Wells, 2009; Chen et al., 2010). Since non-B DNA abnormal folds have been addressed with generating DNA breaks, including both single and double strand DNA breaks. Non-B DNA structures recruit DNA repair machinery to the breaking sites, which then make gene mutations and chromosomal rearrangements during repair.

4.1 Effects of non-B DNA structures on DNA replication / transcription

Some regions of DNA forming non-B DNA structures in replication or transcription, which may turn to affect the DNA transactions (Van Holde & Zlatanova, 1994; Samadashwily et al., 1997; Krasilnikova et al., 2004; Lin et al., 2006; Mirkin & Mirkin, 2007)

One of the well-studied effects of the non-B structures on replication is a block to polymerases because of template folding, which was shown for cruciforms/ hairpins and H-DNA (Samadashwily et al., 1997; Krasilnikova et al., 2004; Voineaqu et al., 2009).
It has been found that triplex DNA can adversely affect DNA replication and potentially lead to replication fork collapse (Samadashwily et al., 1997; Krasilnikova et al., 2004; Voineaqu et al., 2009). The polypurine strand of a triplex forming duplex may not be a potential template, therefore giving increased chance of being single stranded, and forming intermolecular or intramolecular triplex (Hile & Eckert, 2004; Urban et al.,2010). Besides,a non-B DNA structure itself may also directly slow the progression of replication fork (Samadashwily et al., 1997; Mirkin & Mirkin, 2007; Trinh & Sinden, 1991). Such non-B DNA structures may be an obstacle to fork progression or a target for nucleolytic attack, thus allowing DNA breakage leading to deletion or recombination (Mirkin, 2006; Kim et al., 2006).
In contrast, the single-stranded parts in a cruciform or H-DNA may serve as the recognition elements for replication initiation proteins. For example, cruciform binding proteins (CBP), such as 14-3-3sigma in HeLa cells recruits replication proteins to a cruciform to start replication (Alvarez et al., 2002; Novac et al., 2002). Therefore, it is possible for a hairpin/cruciform DNA sequence behaves like a replication "origin", inducing an origin independent DNA replication. The similar way of DNA replication has been found in *E.coli* and named as stable DNA replication. More interestingly, the origin independent DNA replication has also been proposed as a mechanism for the production of expanded DNA repeats (Pan 2006).
In addition, certain non-B DNA structures can also interfere with RNA transcription and recombination (Van Holde & Zlatanova, 1994; Broxson et al., 2011). Similarly RNA transcription can also promote forming non-B DNA structures, including hairpin, triplexs and G_4DNA (Van Holde & Zlatanova, 1994; Broxson et al., 2011).

4.2 Modulation of supercoiling and promoting transcription

The extent of supercoiling in a DNA segment is known to affect transcription, recombination, and replication such that an ideal DNA topology may be critical for them. It has been found that formation of cruciforms, Z-DNA and H-DNA caused partial relaxation of excessive superhelicity in a topological domain. Specific cases of DNA replication and gene expression have also been described as superhelicity dependent events induced by formation of cruciforms, Z-DNA and H-DNA.

4.3 Accumulation of DNA Damages causing increased mutability within non-B DNA structure forming sequences or their flanking sequences

DNA sequences that are prone to adopting non-B DNA secondary structures are associated with hot spots of genomic instability, where repeat expansions, chromosomal fragility, or gross chromosomal rearrangements can be often seen. For example, long repeating tracts of CTG · CAG, CCTG · CAGG, and GAA · TTC are associated with the etiology of myotonic dystrophy type 1 (DM1), type 2(DM2), and Friedreich's ataxia (FRDA) (Wells, 2007). The repeating sequences involved have potentials to adopt a variety of non-B DNA secondary structures (McMurray, 1999; Pan, 2004, 2006, 2009). Studies in various model systems, including *Escherichia coli* and mammalian cell lines, such as COS-7, CV-1, and HEK-293, have revealed that conditions promoting formation of non-B DNA structures enhanced the repeats instabilities. Such instabilities can occur both within the repeat sequences and in the flanking sequences of up to ~4 kbp (Wojciechowska et al., 2006).

Indeed, it has been found that DNA double-stranded breaks (DSBs) can sometimes be accumulated at or around the repeating sequences, and error-prone repair pathways were also proposed to be involved in forming gross DNA rearrangements (Kurahashi et al., 2006). Moreover, DNA breaks may also happen in the single-stranded area, or the structured region when they serve as targets of nuclease activity, leading to enhanced mutagenesis or recombination. The breakpoints of the disease-causing translocation cluster within a 150-bp genomic region of the *bcl*-2 gene were seen potentially form a triplex DNA structure (Adachi & Tsujimoto, 1990; Raghavan et al., 2004, 2005).

It has long been found that, the efficacies of DNA replication in the leading and lagging strand templates were differently performed in *E. coli* chromosome. Replication errors and SOS mutator effects occurred preferentially in the lagging strand, while intermolecular strand switch events during DNA replication occurred preferentially in the leading strand (Iwaki et al., 1995; Trinh & Sinden, 1995; Iwaki et al., 1996; Fijalkowska et al., 1998; Sinden et al., 1999; Maliszewska-Tkaczyk, 2002; Gawel et al., 2002; Hashem & Sinden, 2005). Similarly, unequal fidelities have also been found with deletions between direct repeats in the leading strand template (Hashem & Sinden, 2005). This may attribute to potential of non-B DNA structure formation in the leading and lagging strand template in DNA replication. Similarly, the replication fidelities of various inverted repeats, direct repeats, including trinucleotide repeats can also be compromised if they adopt non-B DNA conformations, such as hairpin, cruciform, triplex, tetra-duplex DNA, leading potentially to mutations or rearrangements (Pan & Leach, 2000; Sinden et al., 2002).

4.4 Nucleosome exclusion

In eukaryotes, chromosomal DNA wrapping around histones in nucleosomes interferes with the protein binding to promoters and origins of replication. Nucleosome formations, on one hand, and formation of cruciform, Z-DNA and triplex DNA, on the other hand, are mutually exclusive. Thus, the alternative structure-forming DNA sequences may expose nucleosome-free DNA, making them accessible to transcription, replication, recombination proteins as well as nucleases, producing fragile sites in chromosome (chwartz et al., 2006; Lukusa & Fryns, 2008).

Fragile sites are specific loci that appear as constrictions, gaps, or breaks on chromosomes from cells exposed to partial inhibition of DNA replication (Schwartz et al., 2006; Lukusa & Fryns, 2008). In chromosomal level, fragile sites always lack nucleosomes, and sometimes can be associated with trinucleotide repeats (TNRs) of CGG·CCG, CAG·CTG, GAA·TTC and GCN·NGC, with specific G-rich tetra- to dodecanucleotide repeats or with long AT-rich repeats, such as the 33 or 42 minisatellites in the FRA16B and FRA10B common fragile sites (Wang & Griffith, 1996). In the same time, fragile sites can be classified as rare or common, depending on their frequency within the population and their specific mode of induction. So far, there are more than 89 common fragile sites listed in GDB (Gene Databases), which are considered to be an intrinsic part of the chromosomal structure presented in all individuals. Six common fragile sites have been cloned and characterized, including FRA3B (Huebner & Croce, 2001; Lettessier et al., 2011), FRA7G, FRA7H, FRA16D (Shah et al., 2010), FRAXB , and FRA6F. Common fragile site instability was attributed to the fact that they contain sequences prone to form secondary structures that may impair replication fork movement, possibly leading to fork collapse and resulting in DNA breaks.

Most rare fragile sites are induced by folate shortage, and others are induced by DNA minor groove binders. So far, seven folate sensitive (FRA10A, FRA11B, FRA12A, FRA16A, FRAXA,

FRAXE and FRAXF) and two nonfolate sensitive (FRA10B and FRA16B) fragile sites have been molecularly characterized. Interestingly, almost all these fragile sites are found to have expanded DNA repeats resulting from mutation involving the normally occurring polymorphic CCG/CGG trinucleotide repeats and AT-rich minisatellite repeats (Balakumaran et al., 2000; Voineagu et al., 2009).

The expanded repeats were also demonstrated to have the potentials, at least under certain circumstances, to form stable secondary non-B DNA structures, including intrastrand hairpins, slipped strand DNA or tetrahelical structures, or to present flexible repeat sequences. Both of which are expected to affect the replication. In addition, these DNA sequences are also found to decrease the efficiency of nucleosome assembly, resulting in decondensation defects seen as fragile sites (Wang & Griffith, 1996; Freudenreich, 2007).

5. Genes and gene products that are involved in abnormal folding

A numerous proteins that interact with non-B DNA secondary structures have been characterized recently. These proteins may also be called as DNA structure-specific proteins, such as Rad1, Rad2, Rad10, Msh2, Msh3, BLM, WRN and Sgs1 (Bhattacharyya & Lahue, 2004; Nag & Cavallo, 2007; Kantelinen et al., 2010; Pichierri et al., 2011). These DNA structure-specific proteins can be further classified by function into several distinct groups, depending on their possible effects on the formation/ stability of non-B DNA structure. Some of the binding proteins may increase the stability of the bound non-B DNA secondary structures; and some may promote forming non-B DNA secondary structures; or destabilize non-B DNA secondary structures. Indeed, the available data implicate various proteins participating in mismatch repair, nucleotide excision repair, base excision repair, homologous recombination, recognize non-B DNA secondary structures in trying to avoid "so called" structure-directed mutagenesis.

As discussed previously, DNA structures can often induce DNA mutations. This DNA structure mediated mutagenesis may be because of the following reasons: the abnormal positioning of the bases and sugar in non-B DNA conformations, which impact the function of some DNA repair proteins on damaged DNA. For example, alkylating damage such as N^7-methylguanine or O^6-methylguanine is not repaired as efficiently in Z-DNA as it is in B-DNA. Alternatively, forming DNA secondary structures near DNA damage sites might influence the damage repair processing, depending on the types of damages, the environments, and the nature of the secondary structures (Pfohl-Leszkowicz et al., 1983; Boiteux et al., 1985).

5.1 MMR proteins

It has long been studied that MMR deficiency is associated with microsatellite sequence instability and human disease. For example, the instability of TNRs and AT-rich minisatellites is associated with their capacity of adopting unusual secondary structures, such as hairpins or DNA triplexes. This feature is common to different types of repeated DNA. Therefore, repeat instability is dependent on MMR in mice and yeast, consistent with the observation that sequences at repetitive DNA sites form short hairpins or small loops that are targets of the Msh2–Msh6 MMR (Modrich, 2006).

MMR proteins bind to non-B DNA secondary structures mainly through its capacity of recognizing mismatched base pairs. It has been found that MMR binds mismatches in a CNG triplet repeats hairpin stem. Although the MSH2–MSH3 complex of MMR also

binds perfect hairpin formed by inverted repeats (lacking mismatched regions), affinity is low, suggesting that mismatches are important for the MMR protein binding (Kantelinen et al., 2008). In addition, MutS has also been reported to bind parallel G4 DNA in humans (Fry, 2007).

5.2 NER and HR proteins

NER proteins, such as the UvrB and UvrC in *E.coli*, and the XPA, XPG, XPC in eukaryotes and homologous recombination proteins, such as RecA, HsRad51, were found to be involved in H-DNA mediated repair and recombination (Bacolla et al., 2001). UvrB and UvrC may preferentially recognize the helical distortions, while RecA recognizing single stranded DNA region in an H-DNA.

5.3 Helicases and junction resolvases

Proteins that preferentially catalyze the unwinding of DNA non-B DNA secondary structures are DNA helicases in ATP-hydrolysis dependent manner. Helicases are DNA unwinding enzymes that preferentially melt some of the non-B DNA structures. The selectivity of helicases on non-B DNA secondary structures has been identified in simian virus 40 (SV40), yeast and human cells. The most studied helicases are members of RecQ family, whose roles are found in a broad range of organisms from *E. coli* RecQ to humans WRN, BLM and RecQL4 (Mohaghegh et al., 2001; Bachrati & Hickson, 2003; Cobb & Bjergbaek, 2006; Masai, 2011). All the non-B DNA secondary structure unwinding helicases act catalytically and all require for their hydrolysis of nucleotide triphosphate, normally ATP, and the presence of Mg^{2+} ions. For example, G-quadruplex DNA substrates are unwound by RecQ helicase with a 3'→5' polarity and need the tetraplex to hold a short 3' single-stranded tail that serves as a "loading dock" for these enzymes (Jain et al., 2010). It should be emphasized, however, that none of the described helicases unwinds tetraplex DNA only and all the enzymes are also able to unfold, although at a lower efficiency, other DNA structures such as duplex DNA, Holliday junctions or triplex. Recently, DHX9 helicase from human cells was found to co-immunoprecipitate with triplex DNA, suggesting a role in maintaining genome stability (Jain et al., 2010). DHX9 displaced the third strand from a specific triplex DNA and catalyzed the unwinding with a 3′ to 5′ polarity for the displaced third strand ((Jain et al., 2010).

5.3.1 RecQ helicases BLM, WRN, RECQL4 and Sgs1

RecQ helicases are a group of DNA helicases that are conserved from bacteria to man (Bachrati & Hickson, 2003). *Rec*Q helicase is named after the *rec*Q gene of *Escherichia coli* and has the activity of unwinding DNA in the 3′–5′ direction in relation to the DNA strand in which the enzyme is bound (Mohaghegh et al., 2001). There are at least five homologues in humans, three of which are associated with genetic diseases. The yeast homologue of RecQ is Sgs1, whose function was found to be similar to most of the members in the RecQ family (Bachrati & Hickson, 2003; Cejka & Kowalczykowski, 2010; Masai, 2011).

It has been reported that, without a functional RecQ helicase, DNA replication does not advance normally. In humans, lacking of WRN or BLM protein accumulates aberrant replication intermediates (Harrigan et al., 2003; Cheok et al., 2005), this may allow for certain non-B DNA structure forming (Mohaghegh et al., 2001; Bacolla et al., 2011). Therefore, it is not surprising to see that more and more reports are going to be published

which specify the important roles of RecQ in resolving the non-B DNA structures, including those G4-DNA (Kamath-Loeb et al., 2001; Fry & loeb, 1999). Similarly the large T antigen and Dna2 helicase/ exonuclease have also been found to unwind the G-tetraduplex (Masuda-Sasa et al., 2008).

5.3.2 Junction resolvases

A cruciform is similar in appearance to a recombination intermediate, a four-way Holliday junction. Therefore, Holliday junction resolvases, RuvABC in prokaryotes, or Mus81, Sgs1 and Sgs2 in yeast might also have activity on cruciforms formed at inverted repeats (Cejka & Kowalczykowski, 2010; Lilley, 2010; Ashton et al., 2011; Mankouri et al., 2011),.

5.4 Topoisomerase

Non-B DNA structures can be substrates for DNA topoisomerase I and II (Howard et al., 1993; Froelich-Ammon et al., 1994). It has shown that DNA topoisomerase II binds and cleaves hairpins (e.g., hairpin formed at a negatively supercoiled 52-bp palindromic sequence in the human β-globin gene), but not cruciforms. DNA topoisomerase II cleavage sites near human immunodeficiency virus integration sites in the human genome consist of Z-DNA forming sequences and other repetitive sequence (Howard et al., 1993); in contrast, DNA topoisomerase I promotes forming parallel G4 DNA in humans. Similarly RAP1, Hop1 in yeast, and Thrombin in humans are also found to promote form of G4 DNA.

5.5 Single strand binding protein (SSB/RPA)

RPA–ssDNA serves as intermediate in many DNA repair processes. For example, ssDNA-RPA can be made through nuclease and helicase actions in repair of UV-induced thymine dimers by nucleotide excision repair, and in a replication fork where DNA polymerase is paused but without pausing DNA helicase accompanied. RPA may prevent or destabilize a non-B DNA structure formation. For example, RPA in humans has been found to destabilize a G'4 DNA (Fig. 1). As for a triplex, the polypyrimidine strands are preferred to bind with RPA, which will then form complex with XPA, XPC-hHR23B (Vasquez et al., 2002; Thomas et al., 2005). In mammalian cells, RPA binds 50-fold more strongly to pyrimidines than to purines, therefore, makes the polypyrimidine strand single-stranded in an intramolecular triplex structure at neutral pH. Moreover, persistent RPA binding may lead to RPA hyper-phosphorylation that triggers repair reactions (Thomas et al., 2005). In addition, RPA-ssDNA and an ssDNA–dsDNA junction can also act as initial signals for cells response to DNA damages, which activates the ATR pathway (Ball et al., 2004; Choi et al., 2010) .

5.6 DNA structure-specific nucleases

Proteins consist of nucleases that specifically cleave DNA next to or within a non-B DNA secondary structures have been well studied. The earliest protein having such functions was identified in *Saccharomyces cerevisiae*, the gene *KEM*1 (also called SEP1, DST2, XRN1 and RAR5) (Liu et al., 1994, 1995). *KEM*1 was initially characterized as a telomere binding protein, and later, it was found to cleave DNA that includes a four-stranded G4 domain but show low or no nucleolytic activity toward single- or double-stranded DNA substrates. Other well-known DNA structure specific nucleases are SbcCD (Connelly & Leach, 1992,

1996, 2004; Connelly et al., 1998, 1999) and its eukaryotic homologue of Mre11-Rad50 (Paull & Gellert, 1998, 2000; Sonoda et al., 2006; Carter et al., 2007; Delmas et al., 2009).

5.6.1 SbcCD

It is now known that influences of repetitive DNA sequences on genomic instabilities were often attributable to forming non-B DNA secondary structures *in vivo*. Once a non-B DNA structure is stable, which will interfere with DNA replication, repair and/ or transcription *in vivo*, resulting in unstable genome. These deleterious non-B DNA secondary structures have already been found to form in *E.coli*, such as the large hairpin formed by the long palindrome DNA sequences (Leach, 1994). The stable hairpin can be cleaved by SbcCD, leading to forming DNA double strand breaks, and then be repaired by using homologous recombination (Connelly & Leach, 1996,; Connelly et al., 1992, 1998, 1999).

Long palindrome sequences are significantly more stable in nuclease-deficient (SbcCD) strains of *E. coli* than in wild-type strains. The SbcCD protein complex is a member of the structural maintenance of chromosomes (SMCs) family found in bacteriophage, bacteria, yeast, *Drosophila*, mouse, and human. SbcCD has both 3′–5′ exonuclease activity on double-stranded DNA and endonuclease activity on single-stranded DNA (Connelly et al., 1999). *In vitro*, it can recognize and bind hairpin structures and cleave at the loop, 5′ immediately next to the loop/ stem junction.

Further degradation of the hairpin cleavage products can occur by the ATP-dependent double-stranded DNA exonuclease activity of the SbcCD protein complex. This structure-specific endonuclease activity does not need a 3′ or 5′ terminus (Connelly & Leach, 1992, 1996; Connelly et al., 1998, 1999).

5.6.2 Mre11-Rad50-Nbs1 (MRN) / Mre11-Rad50-Xrs2 (MRX)

Rad50 and Mre11 are the eukaryotic homologues of SbcCD that have not been shown to bind hairpin/cruciform directly. Mre11 and Rad50, forming complex with Nbs1 (in human cells) or Xrs2 (in yeast), show a hairpin structure cleaving activity i*n vitro*. And which participate in processing double strand breaks *in vivo* by homologous recombination or non-homologous end-joining (Paull & Gellert, 1998, 2000; Sonoda et al., 2006; Delmas et al., 2009). In hairpin cleavage, MRN/ MRX interacts with BRCA1 which preferentially binds four-way branched DNA, similar to cruciforms. Mre11 shows an incision activity at hairpin/ cruciform, and acts as a selective endonuclease in yeast to bind to G4 DNA or to G'2 quadruplex DNA and cleaves the G4 DNA.

5.6.3 other nucleases

Besides the DNA structure specific nucleases such as SbcCD and its eukaryotic homologue Mre11-Rad50-Nbs1 (Xrs1), many other DNA structure-specific DNA nucleases have also been determined. These nucleases recognize and cleave the non-B DNA structures or even the DNA sequences that have non-B DNA secondary structures adopted, playing important roles in various DNA transactions including DNA replication, repair and recombination. For example, Rad1-Rad10 (XPF or ERCC1) has shown to cleave branched intermediates/ Flapped DNA in repair (Li et al., 2008; Muñoz et al., 2009). And Rad2 family of nucleases, such as human XPG (Class I), FEN1 (Class II), and HEX1/ hEXO1 (Class III), have shown both substrate specific 5' to 3' exonuclease activity and endonuclease activity in repair, recombination, and/ or replication. Among them, Rad2 domain of human exonuclease 1

(HEX1-N2) has high activity on single- and double-stranded DNA substrates as well as a flap structure-specific endonuclease activity but does not have specific endonuclease activity at 10-base pair bubble-like structures, G:T mismatches, or uracil residues (Lee & Wilson, 1999). FEN-1, a structure-specific endonuclease is essential for DNA replication and repair, removes RNA and DNA 5' flaps (Tsutakawa et al., 2011). FEN-1 was thought to be involved in hairpin structure processing, and was found to be involved in CNG triplet repeat stability in the lagging strand template (Spiro et al., 1999; Singh et al., 2007). Similarly, Deletions in PCNA, RPA, and the Bloom protein (BLM), a 3′-5′ helicase can also increase CNG repeat expansion or deletion, which reportedly interacts with FEN-1 in cleaving flaps. Recently NucS from *Pyrococcus abyssi* was found to be the equivalent of FEN-1 that cleaves the flapped DNA in Okazaki frangment processing in the lagging strand DNA replication (Ren et al., 2009; Creze et al., 2011).

SLX1 and SLX4 are other structure-specific endonucleases acting as heteromer that cleave branched DNA substrates, particularly simple-Y, 5'-flap, or replication fork structures. It also cleaves the strand bearing the 5' nonhomologous arm at the branch junction and generates ligatable nicked products from 5'-flap or replication fork substrates (Fricke & Brill, 2003).

RAGs is a complex consisting of RAG1, RAG2, and HMGB1 that cleaves 3′ overhangs in multiple locations at the duplex/ single-stranded transitions (Fugmann, 2001). RAGs complex is able to cleave different non-B DNA structures such as symmetric bubbles, heterologous loops and proposed triplex DNA. For example, RAGs complex cleaves the *bcl*-2 Mbr at 3′ overhang and non-B DNA structures under physiological buffer conditions (Adachi & Tsujimoto, 1990; Fugmann, 2001; Raghavan et al., 2004, 2005).

In addition, many single-strand specific nucleases, like S1, P1, and mung bean nucleases, are also efficient at cleaving single stranded DNA in the non-B DNA structures, though at low pH. Since some non-B DNA structures, e.g. H-DNA and G4 DNA disclose an unstructured single-stranded DNA region, which therefore serve as substrates for those single-strand specific nucleases. Recently, a more specific nuclease that cuts single-stranded DNA 5' to a G4 domain was isolated from human cells. This enzyme, initially named G quartet nuclease 1 (GQN1) is thought to be involved in immunoglobulin heavy chain class switch recombination in B cells, does not digest single- or double-stranded DNA, Holliday junctions or tetraplex RNA. It specifically cuts single-stranded DNA located few nucleotides 5' to either G'2 or G4 domains (Sun et al., 2001). However, GQN1 cannot incise tetraplex RNA, showing a significant difference from a mouse cytoplasmic exoribonuclease (mXRN1p) which cleaves G4 RNA (Bashkirov et al., 1997).

6. Gratuitous repair on undamaged DNA misfolds by multiple proteins

DNA damage and repair are always active in living cells regardless of the proliferation status of the cells. And unpaired bases and the helix distortions/ junctions in most of the non-B DNA secondary structures can therefore be targets for the structure specific proteins working in DNA repair, e.g. mismatch repair, nucleotide excision repair etc., launching DNA repairs or activating checkpoints repair (Voineagu et al., 2009).

6.1 Repair by singular pathway of DNA repair

Small DNA loops/ bulges, triplex DNA may be readily corrected by an individual repair, such as a mismatch repair or a nucleotide excision repair. For example, helix distortion and/ or mismatched base pairs in a hairpin, which sometime also occurs with imperfect hairpin

structures at CAG repeats, can be recognized by mismatch repair machinery (Yang, 2006). Msh2/ Msh3 complex in eukaryotic cells specifically binds CAG-hairpins, and the ATP-ase activity of the Msh2 / Msh3 complex can be altered by the binding. However, the repair is dependent on the number of loops/ bulges. A few of them may be repaired by MMR, but too many may not because of interfering MMR by multiple MutS binding, suggesting that repair on a particular non-B DNA conformation will be conditional, depending on locations and environments. Further, nucleotide excision repair (NER) proteins can bind intermolecular triplex, which are involved in the triplex mediated mutagenesis and recombination (Wang & Vasquez, 2006). In bacterial cells, NER proteins UvrB and UvrC were responsible for triplex-induced cell growth retardation. Given the likenesses of the intermolecular and intramolecular triplex, it is possible for NER contributing to the H-DNA-induced mutagenesis and recombination.

6.2 Competitions among multiple repair proteins

Apart from initiating an individual pathway of DNA repair, some non-B DNA structures can also be recognized by more than one repair proteins working in different repair pathways, resulting in competitions between proteins on same DNA structures.

Competition of repair proteins on a non-B DNA structure may be needed for a cooperative repair, setting up a cooperative new DNA repair to repair; in contrast, the competition may sometimes be internecine, failing in repair of either pathway. Under this circumstance, the repair on a non-B DNA structure by the compositing actions of the DNA structural recognition proteins would be compromised. For example, a stable hairpin may be needed for starting DNA replication, but such a stable hairpin would also be repaired by SbcCD or Mre11-Rad50, making a DNA break for homologous recombination to repair (Leach, 1994). Similarly, unwound DNA or small DNA loops may also be needed for DNA replication or for transcription. While they may also be recognized and bound by repair proteins, such as DNA mismatch and nucleotide-excision repair proteins, recombination proteins, instead of SSB/ RPA (Kirkpatrick & Petes, 1997).

A good demonstration for the internecine competition between multiple repair proteins was the foldings of TGG and AGG repeats in the lagging strand template in a replication fork (Pan & Leach, 2000; Pan et al., to be published results). TGG, AGG and CGG repeats are a group of NGG repeats which own significant potential of folding into non-B DNA secondary structures (Usdin, 1998; Pan & Leach, 2000). AGG repeats formed triplex (Suda et al., 1996; Mishima et al., 1996, 1997), homoduplex (Suda et al.,1995), tetra-duplex (Yang & Hurley, 2006), and a special G-quadruplex, known as tetrad:heptad:heptad:tetrad ((G:H:H:G) or (T:H:H:T)) (Matsugami et al., 2001a, 2001b, 2002, 2003), while CGG and TGG repeats formed pseudo-hairpin and tetra-duplex, respectively (Darlow & Leach, 1998; Usdin, 1998; Pan & Leach, 2000; Zemánek et al., 2005).

It was shown by Pan and Leach, that replication of TGG repeats in the lagging strand template experiences repeats misfolding, during which both MutS and SbcCD were found to affect the later processing by homologous recombination. Binding MutS to the non-B DNA structure formed by TGG repeats may stabilize the structure, while hindering SbcCD cleaving the structure. Interestingly, the roles of MutS and SbcCD in this case seemed complex, since TGG repeats can replicate either without MutS or SbcCD, suggesting that they also play same role in stabilizing the TGG repeat structure. In contrast, similar sized AGG repeats was found also to fold into non-B DNA structures in a similar lagging strand template of a replication fork.

However, the non-B DNA structure formed by AGG repeats was found to be incapable of binding with MutS protein, and being cleaved by SbcCD. This made consistence with the reports though AGG repeats belong to a same group of NGG trinucleotide repeats with TGG repeats, they form various G-rich DNA secondary structures, including quadruplex, triple helical, homoduplex and tetrad:heptad:heptad:tetrad ((G:H:H:G) or (T:H:H:T)). Obviously, some of these non-B DNA structures folded may not be recognized by MutS protein *in vivo*, making significant differences in DNA structure formation between AGG repeats and TGG repeats (Pan et al., unpublished results).

The examples of a coordinated repair by different repair proteins on the same non-B DNA structures are the repair of DNA loops by MMR and NER proteins (Kirkpatrick & Petes, 1997; Zhao et al., 2009, 2010). It has been found that both MSH2 and XPA proteins are involved in the instabilities of CAG repeats, possibly through some so far unidentified roles (Kirkpatrick & Petes, 1997; Lin & Wilson, 2009; Zhao etal., 2009, 2010). Knocking down both MSH2 and XPA proteins did not further reduce CAG repeat contraction, suggesting a new role for these proteins in the same pathway. Similarly, it has also been reported the MSH2 and XPA are also involved in H-DNA metabolism but once again the DNA structure may not be processed via canonical MMR or NER mechanisms (Zhao et al., 2009, 2010).

6.3 Repair proteins can be defeated by DNA secondary structure

It may be feasible by postulating that more non-B DNA structures might be formed by DNA sequences in the genomes. However the repair machinery in the cells may only be limited to a few types, such as those MMR, NER single / double strand breaks etc. It therefore raises a question as if all non-B DNA structures possibly form could be recognized and processed by those repair proteins? The answer to this question is presently unknown; however some of the known secondary structures cannot easily be repaired, including large DNA loops and the flapped DNA etc.

6.3.1 Large loops

Stable base pairing prevents recognition by repair enzymes of bases or junctions requiring repair. For example, in *E.coli*, small loops (or secondary structure) may allow mispairing of bases that are corrected by MMR enzymes, leading to loss of base interruption (Parker & Marinus, 1992; Carraway & Marinus, 1993). However, DNA loops made up of less than four unpaired bases are efficiently corrected by methyl-directed mismatch repair (MMR), but loops larger than that cannot be repaired effectively (Parker & Marinus, 1992; Carraway & Marinus, 1993; Fang et al., 2003). The reason for this inefficacy was found to be due to the failure in loop recognition using MutS proteins, leaving the large looped DNA unrepaired by MMR.

6.3.2 Flapped DNA

Flap endonuclease (RAD27 in *Saccharomyces cerevesiae*; FEN-1 in humans) can destabilize simple tandem repeat loci. The 5′ to 3′ flap endonuclease FEN-1/ RAD27 is a structure-specific nuclease required for Okazaki fragment processing in the lagging strand DNA replication. FEN-1, a structure-specific endonuclease is also thought to be involved in CNG triplet repeat stability. It has been reported that a stable hairpin formed by CTG or CAG repeats at the flap region can block the activity of FEN-1. Which then join the upstream Okazaki fragment, resulting in repeats expansion during the next cycle of replication, marking the activity of FEN-1 can be defeated by stable DNA structure (Spiro et al., 1999; Singh et al., 2007).

6.4 Cellular response to non-B DNA structures by activating checkpoints

The existence of cellular proteins that interact with non-B DNA structures provides both strong argument for the existence of non-B DNA structure formations in genomic DNA, and suggestion for cell having intrinsic response to the formation of non-B DNA structures. However, it seems that not all non-B DNA secondary structures, unless they make severe troublesome to DNA metabolism such as making DNA double strand breaks, or generating long single stranded region, were recognized as "DNA damage". Even if cruciforms / hairpins, triplexes, slipped conformations, quadruplexes, and left-handed Z-DNA have all been reported to be chromosomal targets for DNA repair, recombination, and aberrant DNA synthesis, leading to repeat expansion or genomic rearrangements associated with neurodegenerative and genomic disorders. Some of them may also raise more severe response by cells (Voineagu et al., 2009).

The situations for a non-B DNA secondary structure intriguing a cellular response may be addressed at the competing recognition and processing by multiple repair proteins, resulting in incomplete / partial / opposing processing of the non-B DNA structure. Such intermediates may be recognized by proteins capable of activating a cellular response. Alternatively non-B DNA structure bears components that can be recognized by proteins capable of activating a cellular response (Voineagu et al., 2009). In support of this idea, DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 were found to play opposite roles in yeast gene targeting, a triple stranded DNA mediated process. During which Rad1, Msh2, and Msh3 facilitated forming triplex DNA, while Sgs1 prevented forming triplex DNA (Langston & Symington, 2005), therefore should a cellular response be intrigued in gene targeting may have to wait for processing the structure-specific proteins.

The ssDNA region in a non-B DNA structure may likely be coated by single-stranded DNA-binding protein (RPA) directly, or RPA coats the ssDNA after the non-B DNA structure is processed. Either way makes a common intermediate of ssDNA-RPA that activates ATR signaling in response to all of the genotoxic lesions (Krejci et al., 2003; Hu et al., 2007). Indeed, the ssDNA-RPA complex has been found to be a common intermediate in the processing of many types of damaged DNA, including DSBs, UV-induced thymidine dimers, intrastrand cross-links, and mismatches in base-pairing (Ball et al., 2005; Choi et al., 2010). The RPA–ssDNA complex will promote the loading of the 9–1–1 and ATR–ATRIP complexes (Dore et al., 2009). The juxtaposition of these complexes allows ATR to phosphorylate Chk1, which then promotes cell cycle arrest, causing a cellular response to non-B DNA structure formation. Alternatively, ssDNA-RPA complex can recruit Cut5, by which ATR (ATR-ATRIP) (Mec1-Ddc2 in yeast), DNA polymerase α, Rad50-Mre11-Nbs1 (MRN) and clamp loader Rad24 (Rad17 in mammals) can all be recruited to the ssDNA-RPA (Cortez et al., 2001; Zou & Elledge, 2003; Robison et al., 2004).

The purpose of activating DNA damage checkpoint in response to the formation of non-B DNA secondary structure is to regulate cell cycle events, for mediating appropriate repair and fork restart processes. While non-B DNA structure forming sequences per se are probably an infrequent trigger of DNA damage checkpoint responses, and, thus, should not be regarded as a real DNA damage by cells. There has extensive evidence suggesting that non-B DNA structure forming sequences can only induce checkpoint-triggering events when stable non-B DNA structures are adopted. The stable DNA structures may affect normal DNA metabolism, making DSBs or causing more severe effects on DNA metabolism, such as replication fork stalling, formation of nucleosome free sites (Chromosomal Fragile Sites) etc.

Consisting with that, mutations in checkpoint genes, such as Mec1, Ddc2, Rad9, Rad17, Rad24, or Rad53, produce repeat instabilities by a $CAG_{\sim 70}$, including both expansion and contraction instabilities. These suggested that DNA structure formed by long CAG repeats activated checkpoints in eukaryotes (Lahiri et al., 2004; Sundararajan & Freudenreich, 2011). Similarly, a CAG_{175} repeat on plasmids can also be recognized as ''DNA damage'' in *E. coli*, as witnessed by inducing SOS response (Majchrzak et al., 2006).

Surprisingly, it was found that even those shorter CAG repeats (containing 13–20 triplets) can also intrigue DNA damage checkpoint. By which repeats expansion can be prevented when the repeats formed non-B structures, suggesting that cells have endowed the checkpoint mechanism of responding to non-B DNA structure formation (Razidlo & Lahue, 2008).

Another example as intriguing cellular response for non-B DNA structure formation by derived structure processing is also found with human PKD1 gene. The 2.5-kb polypurine–polypyrimidine tract in intron 21 in human PKD1 gene potentially forms H-DNA structure, contributing to the high mutation rate of the PKD1 gene (Bacolla et al., 2001； Patel et al., 2004). A plasmid carrying this polypurine–polypyrimidine tract induced a stronge SOS response and severely delayed the host cell growth, resulting in a dramatic decrease in colony formation (Patel et al., 2004). However, the effect was largely reduced without UvrA (100-fold decrease in colony formation), and nearly vanished without UvrB or UvrC. These suggested the polypurine–polypyrimidine repeat sequence or the structure formed by the repeats per se was not involved in the effects, while the NER processing was essential (Bacolla et al., 2001).

6.5 Mre11-Rad50-Nbs1 (MRN)/ Mre11-Rad50-Xrs2 (MRX)

Apart from the nucleolytic activity, MRN / MRX can also play roles in activating the checkpoints as mentioned above (van den Bosch, et al., 2003; Sundararajan & Freudenreich, 2011). It was believed that a single stranded region in a non-B DNA structure forms ssDNA-RPA to the amount of triggering a checkpoint response (normally exceeds 300 bp). One way of Rad50-Mre11-Nbs1 (MRN) contributing to checkpoint response might be through Cut5 recruitment. Rad50-Mre11-Nbs1 (MRN) can be recruited to the single stranded region in the non-B DNA structure, and then participates in ATR checkpoint. Alternatively Rad50-Mre11-Nbs1 (MRN) can also secure DNA replication as implicated by its ortholog SbcCD in *E.coli* (Darmon et al., 2007; Zahra et al., 2007). Indeed, the MRN / MRX complex has been co localized in the replication machinery. In this context, the resection role of MRN / MRX on DSB initiated recombination repair may be no more necessary as long as the checkpoints mechanism prevented the DSB formation by checkpoint proteins (Mimitou & Symington, 2008; Zhu et al., 2008).

Non-B DNA structure forming sequences are potential triggers of DNA damage checkpoint responses mainly by inducing replication fork stalling and chromosomal breaks. Since the non-B DNA structures have specific DNA conformations at the damaged site, which may influence the checkpoint signaling, and the dynamics of checkpoint activation are likely to differ at different types of non-B DNA structure forming sequences.

7. Future perspectives

Many lines of evidence suggest that unusual DNA structures can form *in vivo* and play significant roles in DNA metabolism, while they may also serve as a source for the

generation of genomic instability. Strikingly, unusual DNA structures were often found to trigger some kinds of repair actions or avoidance responses that promote their removal of the structures once formed. Under this later circumstance, it becomes obvious that formation of non-B DNA structures *in vivo* was somehow similar to the appearances of some real DNA damages as induced by environmental DNA damaging agents. Certain unusual DNA structures have unpaired bases and regions with helix distortions/junctions etc., which may experience unprovoked repair in cells. Therefore triggering cellular responses of a non-B DNA structure is subject to its morphological/ topological properties, which could attract recognizing repair proteins. In fact, a non-B DNA structure is often recognized by more than one repair proteins, such as the proteins working in MMR, NER and recombination. Questions rose therefore as if individual pathways of DNA repair accounts enough for the repair of the non-B DNA structures? Or does it need multiple proteins working in different repair pathways reconstitute synthesized pathway(s) to repair? Nevertheless, progress in this field seems support an idea that enzymes/ proteins that recognize and/ or process the possible non-B DNA structures may be different because of the non-B DNA structures formed. Proteins that have been found to associate with non-B DNA instability might take part in an unexpected way in processing the non-B DNA structures. Therefore studies in the coming future may have to focus on the identifications of the types of non-B DNA structures that elicit certain kinds of mutations and the enzyme systems involved. It could be expected that more diseases will be recognized as because of mutations at non-B DNA structures. Also, strategies will have to make toward developing therapeutics to appease the devastating effects of the syndromes.

8. References

Adachi, M. & Tsujimoto. Y. (1990). Potential Z-DNA elements surround the breakpoints of chromosome translocation within the 5′ flanking region of *bcl*-2 gene. *Oncogene,* 5:1653–1657.

Alvarez, D., Novac, O., Callejo, M., Ruiz, M. T., Price, G. B. & Zannis-Hadjopoulos, M. (2002).14-3-3sigma is a cruciform DNA binding protein and associates in vivo with origins of DNA replication. *J Cell Biochem.* 87(2):194-207.

Antony, S., Arimondo, P. B., Sun, J-S. & Pommier, Y. (2004). Position- and orientation-specific enhancement of topoisomerase I cleavage complexes by triplex DNA structures. *Nucleic Acids Res.*, 32 (17): 5163-5173.

Arimondo, P. B., Riou, J-F., Mergny, J-L., Tazi, J., Sun, J-S., Garestier, T. & Hélène, C. (2000). Interaction of human DNA topoisomerase I with G–quartet structures. *Nucleic Acids Res.*, 28 (24): 4832-4838.

Ashton, T. M., Mankouri, H. W., Heidenblut, A., McHugh, P. J. & Hickson, I. D. (2011). Pathways for Holliday Junction Processing during Homologous Recombination in *Saccharomyces cerevisiae*. *Mol Cell Biol.*, 31(9):1921-1933.

Bachrati, C. Z. & Hickson, I. D. (2003). RecQ helicases: suppressors of tumorigenesis and premature aging. *Biochem J.*, 374(Pt 3):577-606.

Bacolla, A., Jaworski, A., Connors, T. D. & Wells, R. D. (2001). Pkd1 unusual DNA conformations are recognized by nucleotide excision repair. *J Biol Chem.*, 276:18597–18604.

Bacolla, A. & Wells, R. D. (2009). Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog.*, 48:273-285.

Bacolla, A., Wang, G., Jain, A., Chuzhanova, N. A., Cer, R. Z., Collins, J. 1 R., Cooper, D. N., Bohr, V. A. & Vasquez, K. M. (2011). Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of base substitution in human cells. *J Biol Chem.,* 286(12):10017-10026.

Balakumaran, B. S., Freudenreich, C. H. & Zakian, V. A. (2000). CGG / CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Hum Mol Genet.,* 9(1):93-100.

Ball, H. L, Myers, J. S. & Cortez, D. (2005). ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. *Mol Biol Cell.,* 16(5):2372-2381.

Bashkirov, V. I., Scherthan, H., Solinger, J. A., Buerstedde, J. M. & Heyer, W. D. (1997). A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J Cell Biol.,* 136(4):761-773.

Bhattacharyya, S. & Lahue, R. S. (2004). *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol Cell Biol.,* 24(17):7324-7330.

Boiteux, S., Costa de Oliveira, R. & Laval, J. (1985). The *Escherichia coli* O6-methylguanine-DNA methyltransferase does not repair promutagenic O6-methylguanine residues when present in Z-DNA. *J Biol Chem.,* 260(15):8711-8715.

Boiteux, S. & Laval, F. (1985). Repair of O6-methylguanine, by mammalian cell extracts, in alkylated DNA and poly (dG-m5dC)·(polydG-m5dC) in B and Z forms. *Carcinogenesis,* 6:805–807.

Broxson, C., Beckett, J. & Tornaletti, S. (2011). Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human *c-myb* gene. *Biochemistry.,* [Epub ahead of print]

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.,* 34(19):5402-5415..

Carraway, M. & Marinus, M. G. (1993). Repair of heteroduplex DNA molecules with multibase loops in *Escherichia coli*. *J Bacteriol.,* 175(13):3972-3980.

Casasnovas, J. M. & Azorin, F. (1987). Supercoiled induced transition to the Z-DNA conformation affects the ability of a $d(CG/GC)_{12}$ sequence to be organized into nucleosome-cores. *Nucleic Acids Res.,* 15:8899–8918.

Casey, B. P. & Glazer, P. M. (2001). Gene targeting via triple-helix formation. *Prog Nucleic Acid Res Mol Biol.,* 67:163-192.

Cejka, P. & Kowalczykowski, S. C. (2010). The full-length *Saccharomyces cerevisiae* Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. *J Biol Chem.,* 285(11):8290-82301.

Cer, R. Z., Bruce, K. H., Mudunuri, U. S., Yi, M., Volfovsky, N., Luke, B. T., Bacolla, A., Collins, J. R. & Stephens, R. M. (2011). Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res.,* 39 (Database issue):D383-391.

Chan, P. P., Lin, M., Faruqi, A. F., Powell, J., Seidman, M. M. & Glazer, P. M. (1999). Targeted correction of an episomal gene in mammalian cells by a short DNA fragment tethered to a triplex-forming oligonucleotide. *J Biol Chem.*, 274, 11541-11548.

Chen, J. M., Cooper, D. N., Ferec, C., Kehrer-Sawatzki, H. & Patrinos, G. P. （2010）. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol.*, 20(4):222-233.

Cheok, C. F., Bachrati, C. Z., Chan, K. L., Ralf, C., Wu, L. & Hickson, I. D. (2005). Roles of the Bloom's syndrome helicase in the maintenance of genome stability. *Biochem Soc Trans.*, 33(Pt 6):1456-1459.

Choi, J. H., Lindsey-Boltz, L. A., Kemp, M., Mason, A. C., Wold, M. S. & Sancar, A. (2010). Reconstitution of RPA-covered single-stranded DNA-activated ATR-Chk1 signaling. *Proc Natl Acad Sci USA.*, 107(31):13660-13665.

Connelly, J., de Leau, E. S. & Leach, D. (1992). DNA cleavage and degradation by the SbcCD complex from *Escherichia coli*. *Nucleic Acid Res.*, 27, 1039-1046

Connelly, J. & Leach, D. (1996). The *sbc*C and *sbc*D genes of *Escherichia coli* encode a nuclease involved in palindromic inviability and genetic recombination. *Genes to Cell.*, 1,285-291.

Connelly, J., Kirkham, L. A. & Leach, D. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci U S A.*, 95, 7969-7974.

Connelly, J., delau, E.S. & Leach, D. R. F. (1999). DNA cleavage and degradation by the SbcCD complex from *Escherichia coli*, *Nucleic Acid Res.,* 27:1039–1046.

Courey, A. J. (1999). Analysis of altered DNA structures: cruciform DNA. *Methods Mol Biol.* 94:29-40.

Cox, R. & Mirkin, S. M. (1997). Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci USA.*, 94: 5237–5242.

Creze, C., Lestini, R., Kühn, J., Ligabue, A., Becker, H. F., Czjzek, M., Flament, D. & Myllykallio H. (2011). Structure and function of a novel endonuclease acting on branched DNA substrates. *Biochem Soc Trans.*, 39(1):145-9.

Crouau-Roy, B. (1999). Trans-speciation maintenance in the MHC region of a polymorphism which includes a polymorphic dinucleotide locus, and the *de novo* arisal of a polymorphic tetranucleotide microsatellite. *Tissue Antigens.*, 54(6):560-4.

Darmon, E., Lopez-Vernaza, M. A., Helness, A. C., Borking, A., Wilson, E., Thacker, Z., Wardrope, L. & Leach, D. R. F., (2007). SbcCD regulation and localization in *Escherichia coli*. *J Bacteriol.,* 189(18), 6686-6694.

Dere, R. & Wells, R. D. (2006). DM2 CCTG*CAGG repeats are crossover hotspots that are more prone to expansions than the DM1 CTG*CAG repeats in *Escherichia coli*. *J Mol Biol.,* 360(1):21-36.

Dore´, A. S., Kilkenny, M. L., Rzechorzek, N. J. & Pearl, L. H. (2009). Crystal structure of the Rad9-Rad1-Hus1 DNA damage checkpoint complex— implications for clamp loading and regulation. *Mol Cell,* 34, 735–745.

Dyke, M. W. V., Nelson, L. D., Weilbaecher, R. G. & Mehta, D. V. (2004). Stm1p, a G4 Quadruplex and purine motif triplex nucleic acid-binding protein, interacts with ribosomes and subtelomeric Y′ DNA in *Saccharomyces cerevisiae*. *J Biol Chem.*, 279, 24323-24333.

Fang, W. H., Wang, B. J., Wang, C. H., Lee, S. J., Chang, Y. T., Chuang, Y. K. & Lee, C. N. (2003). DNA loop repair by *Escherichia coli* cell extracts. *J Biol Chem.,* 278(25):22446-22452.

Faruqi, A. F., Datta, H. J., Carroll, D., Seidman, M. M. & Glazer, P. M. (2000). Triple-helix formation induces recombination in mammalian cells via a nucleotide excision repairdependent pathway. *Mol Cell Biol.*, 20, 990-1000.

Fox, K. R. & Brown, T. (2011). Formation of stable DNA triplexes. *Biochem Soc Trans.,* 39(2):629-34.

Freudenreich, C. H. (2007). Chromosome fragility: molecular mechanisms and cellular consequences. *Front. Biosci.,* 12:4911-24.

Froelich-Ammon, S. J., Gale, K. C. & Osheroff, N. (1994). Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage. *J Biol Chem.,* 269:7719–7725.

Fry, M. & Loeb, L. A. (1999). Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence $d(CGG)_n$. *J Biol Chem.,* 274(18):12797-802.

Fry, M.（2007）. Tetraplex DNA and its interacting proteins. *Frontiers in Bioscience,* 12, 4336-4351.

Fugmann, S. D. (2001). RAG1 and RAG2 in V (D) J recombination and transposition. *Immunol Res.,* 23(1):23-39.

Gur-Arie, R., Cohen, C. J., Eitan, Y., Shelef, L., Hallerman, E. M. & Kashi, Y. (2000). Simple Sequence Repeats in *Escherichia coli*: Abundance, Distribution, Composition, and Polymorphism. *Genome Res.,* 10: 62-71.

Harrigan, J. A., Opresko, P. L., von Kobbe, C., Kedar, P. S., Prasad, R., Wilson, S. H. & Bohr, V. A. (2003). The Werner syndrome protein stimulates DNA polymerase beta strand displacement synthesis via its helicase activity. *J Biol Chem.,* 278(25):22686-22695.

Hebert, A. & Rich, A. (1996). The Biology of Left-handed Z-DNA. *J Biol Chem.,* 271(20): 11595–11598.

Ho, P. S. (1994). The non-B DNA structure of $d(CA/TG)_n$ does not differ from that of Z-DNA. *Proc Natl Acad Sci U S A.,* 91(20):9549-9553.

Howard, M.T. & Griffith, J. D. (1993). A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. *J Mol Biol.,* 232:1060–1068.

Inagaki, H., Ohye, T., Kogo, H., Kato, T., Bolor, H., Taniguchi, M., Shaikh, T. H., Emanuel, B. S. & Kurahashi, H. (2009). Chromosomal instability mediated by non-B DNA:Cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res.,* 19:191–198.

Jain, A., Bacolla, A., Chakraborty, P., Grosse, F., Vasquez, K. M. (2010). Human DHX9 helicase unwinds triple-helical DNA structures. *Biochemistry.,* 49(33):6992-6999.

Jin, R., Fernandez-Beros, M. E. & Novick, R. P. (1997). Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform? *EMBO J.,* 16(14):4456-66.

Johnston, B. H. (1988). Chemical probing of the B-Z transition in negatively supercoiled DNA. *J Biomol Struct Dyn.,* 6(1):153-166.

Kalish, J. M., Seidman, M. M., Weeks, D. L. & Glazer, P. M. (2005). Triplex-induced recombination and repair in the pyrimidine motif. *Nucleic Acids Res.,* 33(11):3492-502.

Kaliyaperumal, S., Patrick, S. M. & Williams, K. J. (2011). Phosphorylated hMSH6: DNA mismatch versus DNA damage recognition. *Mutat Res.,* 706(1-2):36-45.

Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M. & Fry, M. (2001). Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the $d(CGG)_n$ trinucleotide repeat sequence. *J Biol Chem.,* 276(19):16439-16446.

Kantelinen, J., Kansikas, M., Korhonen, M. K., Ollila, S., Heinimann, K., Kariola, R., Nystrm, M. (2010). MutSbeta exceeds MutSalpha in dinucleotide loop repair. *Br J Cancer.,* 102(6):1068-1073.

Khuu, P. A., Voth, A. R., Hays, F. A. & Ho, P. S. (2006). The stacked-X DNA Holliday junction and protein recognition. *J Mol Recognit.,* 19(3):234-242.

Kirkpatrick, D. T. & Petes T. D. (1997). Repair of DNA loops involves DNA mismatch and nucleotide-excision repair proteins. *Nature,* 387:929–931.

Kladde, M. P., Kohwi, Y., Kohwi-Shigematsu, T. & Gorski J. (1994). The non-B DNA structure of $d(CA/TG)_n$ differs from that of Z-DNA. *Proc Natl Acad Sci U S A.,* 91(5):1898-902.

Kolb. (2009). Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. *Chromosome Res.,* 17:469–483

Kurahashi, H., Shaikh, T. H., Hu, P., Roe, B. A., Emanuel, B. S. & Budarf, M. L. (2000). Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). *Hum Mol Genet,* 9:1665–1670.

Kurahashi, H. & Emanuel, B. S. (2001). Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum Mol Genet.,* 10(23):2605-17.

Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Kato, T. & Emanuel, B. S. (2006). Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst).,* 5(9-10):1136-45.

Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M. & Emanuel, B. S. (2010). The constitutional t(11;22): implications for a novel mechanism responsible for gross chromosomal rearrangements. *Clin Genet.,* 78(4):299-309.

Lahiri, M., Gustafson, T. L., Majors, E. R. & Freudenreich, C. H. (2004). Expanded CAG repeats activate the DNA damage checkpoint pathway. *Mol Cell.,* 15(2):287-93.

Lange, S. S., Reddy, M. C. & Vasquez, K. M. (2009). Human HMGB1 directly facilitates interactions between nucleotide excision repair proteins on triplex-directed psoralen interstrand crosslinks. *DNA Repair (Amst).,* 8(7):865-72.

Langston, L. G. & Symington, L. S. (2005). Opposing roles for DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 in yeast gene targeting. *EMBO J.,* 24, 2214–2223

Lee, B. I. & Wilson, D. M. 3rd. (1999). The RAD2 domain of human exonuclease 1 exhibits 5' to 3' exonuclease and flap structure-specific endonuclease activities. *J Biol Chem.,* 274(53):37763-9.

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachagès, A. M., Vogt, N., Hansen, R. S., Malfoy, B., Brison, O. & Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature.,* 470(7332):120-3.

Li, F., Dong, J., Pan, X., Oum, J. H., Boeke, J. D. & Lee, S. E. (2008). Microarray-based genetic screen defines SAW1, a gene required for Rad1 / Rad10-dependent processing of recombination intermediates. *Mol Cell.,* 30(3):325-335.

Lilley, D. M. (2010). The interaction of four-way DNA junctions with resolving enzymes. *Biochem Soc Trans.,* 38(2):399-403.

Lin, Y. & Wilson, J. H. (2009). Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. *DNA Repair (Amst).,* 8(8):878-885.

Lipps, H. J. & Rhodes, D. (2009). G-quadruplex structures: *in vivo* evidence and function. Trends *Cell Biol.* 19(8):414-22.

Liu, Z. & Gilbert, W. (1994). The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure. *Cell*, 77(7):1083-1092.

Liu, Z., Lee, A. & Gilbert, W. (1995). Gene disruption of a G4-DNA-dependent nuclease in yeast leads to cellular senescence and telomere shortening. *Proc Natl Acad Sci USA.,* 92(13):6002-6002.

Lukusa, T. & Fryns, J. P. (2008). Human chromosome fragility. *Biochim Biophys Acta.,* 1779(1):3-16.

Majchrzak, M., Bowater, R. P., Staczek, P. & Parniewski, P. (2006). SOS repair and DNA supercoiling influence the genetic stability of DNA triplet repeats in *Escherichia coli. J Mol Biol.,* 364(4):612-24.

Mankouri, H. W., Ashton, T. M., Hickson, I. D. (2011). Holliday junction-containing DNA structures persist in cells lacking Sgs1 or Top3 following exposure to DNA damage. *Proc Natl Acad Sci U S A.,* 108(12):4944-4949.

Masai, H. (2011). RecQL4: a helicase linking formation and maintenance of a replication fork. *J Biochem.,* 2011:149(6):629–631.

Masuda-Sasa, T., Polaczek, P., Peng, X. P., Chen, L. & Campbell, J. L. (2008). Processing of G4 DNA by Dna2 helicase / nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2 / RPA substrate recognition. *J Biol Chem.,* 283(36):24359-24373.

Mäueler, W., Bassili, G., Epplen, C., Key, H. G. & Epplen, J. T. (1999). Protein binding to simple repetitive sequences depends on DNA secondary structure(s). *Chromosome Res.,* 7(3):163-166.

McMurray, C. T. (1999). DNA secondary structure: A common and causative factor for expansion in human disease. *Proc Natl Acad Sci U S A.,* 96: 1823–1825.

Mimitou, E. P. & Symington, L. S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature,* 455(7214):770-774.

Mirkin, S. M. (2008). Discovery of alternative DNA structures: a heroic decade (1979-1989). *Front Biosci.* , 13:1064-1071.

Mishima, Y., Kaizu, H. & Kominami, R. (1997). Pairing of DNA fragments containing (GGA: $TCC)_n$ repeats and promotion by high mobility group protein 1 and histone H1. *J Biol Chem*, 272, 26578–26584

Mukherjee, A. & Vasquez, K. M. (2011). Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie,* 2011:93(8):1197-1208.

Muñoz, I. M., Hain, K., Déclais, A. C., Gardiner, M., Toh, G. W., Sanchez-Pulido, L., Heuckmann, J. M., Toth, R., Macartney, T., Eppink, B., Kanaar, R., Ponting, C. P.,

Lilley, D. M. & Rouse, J. (2009). Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol Cell,* 35(1):116-127.

Nag, D. K. & Cavallo, S. J. (2007). Effects of mutations in SGS1 and in genes functionally related to SGS1 on inverted repeat-stimulated spontaneous unequal sister-chromatid exchange in yeast. *BMC Mol Biol.,* 8:120.

Novac, O., Alvarez, D., Pearson, C. E., Price, G. B., Zannis-Hadjopoulos, M. (2002). The human cruciform-binding protein, CBP, is involved in DNA replication and associates *in vivo* with mammalian replication origins. *J Biol Chem.,* 277(13):11174-11183.

Novara, F., Beri, S., Bernardo, M. E., Bellazzi, R., Malovini, A., Ciccone, R., Cometa, A. M., Locatelli, F., Giorda, R. And Zuffardi, O. (2009). Different molecular mechanisms causing 9p21 deletions in acute lymphoblastic leukemia of childhood. *Hum Genet.,* 126:511-520.

Pan, X. & Leach, D. (2000). The roles of *mut*S, *sbc*CD and *rec*A in the propagation of TGG repeat in *Escherichia coli*. *Nucleic Acids Res,* 28: 3178–3184.

Pan, X. (2004). Self-maintenance of gene and occurrence of genetic diseases, *Science Press Beijing*, pp.1–403.

Pan, X. (2006). Mechanism of trinucleotide repeats instabilities: the necessities of repeat non-B secondary structure formation and the roles of cellular trans-acting factors. *J Genet Genomics,* 33: 1–11.

Pan, X., Ding, Y. F. & Shi, L. F. (2009). The roles of SbcCD and RNaseE in the transcription of GAA · TTC repeats in Escherichia coli. *DNA Repair* (Amst) , 8:1321–1327.

Pan, X., Liao, Y., Liu, Y., Chang, P., Liao, L., Yang, L. & Li, H. Q. (2010). Transcription of AAT · ATT Triplet Repeats in *Escherichia coli* Is Silenced by H-NS and IS1E Transposition. *PLoS ONE,* 5(12): e14271.

Pan, X., Chang, P., Long, J., Xiao, P., Li, H. Q. & Duan, F. (2011). The molecular maintenance of trinucleotide repeats AGG in *Escherichia coli* relies on both homologous recombination and UmuDC –catalyzed DNA translesion synthesis. *DNA Res.,* (submitted).

Parker, B. O. & Marinus, M. G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli. Proc Natl Acad Sci U S A.,* 89(5):1730-1734.

Patel, H. P., Lu, L., Blaszak, R. T. & Bissler, J. J. (2004). PKD1 intron 21: triplex DNA formation and effect on replication. *Nucleic Acids Res.,* 32(4):1460-1468.

Pearson, C. E., Zorbas, H., Price, G. B. & Zannis-Hadjopoulos, M. (1996). Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J Cell Biochem.,* 63(1):1-22.

Pfohl-Leszkowicz A, Boiteux S, Laval J, Keith G, Dirheimer G. (1983). Enzymatic methylation of chemically alkylated DNA and poly(dG-dC) X poly(dG-dC) in B and Z forms. *Biochem Biophys Res Commun.,* 116(2):682-688.

Qin, Y. & Hurley, L. H. (2008). Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie.,* 90(8):1149-1171.

Qiu, J. F. & Pan, X. (2008). The linkage of DNA replication, recombination and repair, *Prog. Biochem. Biophys.,* 35: 751–756.

Raghavan, S. C., Swanson, P. C., Wu, X., Hsieh, C. L. & Lieber, M. R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature*, 428, 88-93.

Raghavan, S. C., Chastain, P., Lee, J. S., Hegde, B. G., Houston, S., Langen, R., Hsieh, C. L., Haworth, I. S. & Lieber, M. R.(2005). Evidence for a triplex DNA conformation at the *bcl*-2 major breakpoint region of the t(14;18) translocation. *J Biol Chem.*, 280, 22749-22760.

Rao,S. R., Trivedi, S., Emmanuel, D., Merita, K., & Hynniewta, M. (2010). DNA repetitive sequences-types, distribution and function: A review. *J Cell and Mol Biol.,* 7(2) & 8(1): 1-11.

Rawal, P., Kummarasetti, V. B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S. K. & Chowdhury, S. (2006). Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.,* 16(5):644-55.

Razidlo, D. F. & Lahue, R. S. (2008). Mrc1, Tof1 and Csm3 inhibit CAG·CTG repeat instability by at least two mechanisms. *DNA Repair* (Amst), 7(4):633-640.

Ren, B., Kühn, J., Meslet-Cladiere, L., Briffotaux, J., Norais, C., Lavigne, R., Flament, D., Ladenstein, R. & Myllykallio, H. (2009). Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J.,* 28(16):2479-2489.

Rich, A., Nordheim, A. & Azorin, F. (1983). Stabilization and detection of natural left-handed Z-DNA. *J Biomol Struct Dyn.,* 1(1):1-19.

Sannohe, Y. & Sugiyama, H. (2010). Overview of formation of G-quadruplex structures. *Curr Protoc Nucleic Acid Chem.,* Chapter 17:Unit 17.2.1-17.

Schwartz, M., Zlotorynski, E. & Kerem, B. (2006). The molecular basis of common and rare fragile sites. *Cancer Lett.,* 232(1):13-26.

Shah, S. N., Opresko, P. L., Meng, X., Lee, M. Y. & Eckert, K. A. (2010). DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D. *Nucleic Acids Res.,* 38(4):1149-1162.

Shimizu, M., Mori, T., Sakuri, T. & Shindo, H. (2000).Destablization of nucleosomes by an unusual DNA conformation adopted by poly(dA)poly(dT) tracts *in vivo*. *EMBO J.*,19(13):3358-3365.

Sinden, R. R., Zheng, G. X., Brankamp R. G. & Allen, K. N. (1991). On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation *in vivo*. *Genetics*, 129, 991–1005.

Sinden, R. R., Hashem, V. I. & Rosche, W. A. (1999). DNA-directed mutations. Leading and lagging strand specificity. *Ann N Y Acad Sci.*, 870,173-189

Singh, P., Zheng, L., Chavez, V., Qiu, J. & Shen, B. (2007). Concerted action of exonuclease and Gap-dependent endonuclease activities of FEN-1 contributes to the resolution of triplet repeat sequences $(CTG)_n$- and $(GAA)_n$-derived secondary structures formed during maturation of Okazaki fragments. *J Biol Chem.,* 282(6):3465-3477.

Smith, R. (2008). Meeting DNA palindromes head-to-head. *Genes Dev.,* 22: 2612-2620.

Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X, Mariappan, S. V. & McMurray, C. T. (1999). Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol Cell,* 4(6):1079-1085.

Stros, M., Muselíková-Polanská, E., Pospísilová, S. & Strauss, F. (2004). High-affinity binding of tumor-suppressor protein p53 and HMGB1 to hemicatenated DNA loops. *Biochemistry.,* 43(22):7215-7225.

Suda,T., Mishma,Y., Takayanagi, K., et al. (1996). A novel activity of HMG domains promotion of the triple-stranded complex formation between DNA containing $(GGA/TCC)_{11}$ and $d(GGA)_{11}$ Oligonucleotides. *Nucleic Acids Res.,* 24, 4733-4740

Sun, H., Yabuki, A. & Maizels, N. (2001). A human nuclease specific for G4 DNA. *Proc Natl Acad Sci U S A.,* 98(22):12444-12449.

Sundararajan, R. & Freudenreich, C. H. (2011). Expanded CAG / CTG Repeat DNA Induces a checkpoint response that impacts cell Proliferation in Saccharomyces cerevisiae. *PLoS Genet.,* 7(3):e1001339.

Svozil, D., Kalina, J., Omelka, M. & Schneider, B. (2008). DNA conformations and their sequence preferences. *Nucleic Acids Res.,* 36(11):3690-3706.

Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., Tomlinson, C. G., Thompson, P., Sarker, A. H., Shen, B., Cooper, P. K., Grasby, J. A. & Tainer, J. A. (2011). Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell.*, 145(2):198-211.

Urban, M., Joubert, N., Purse, B. W., Hocek, M. & Kuchta, R. D. (2010). Mechanisms by which human DNA primase chooses to polymerize a nucleoside triphosphate. *Biochemistry.,* 49(4):727-735.

Van Holde, K. & Zlatanova, J. (1994). Unusual DNA structures, chromatin and transcription. *Bioessays.,* 16(1):59-68.

Vasquez, K. M., Christensen, J., Li, L., Finch, R. A. & Glazer, P. M. (2002). Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. *Proc Natl Acad Sci U S A.,* 99(9):5848-5853.

Voineagu, I., Freudenreich, C. H. & Mirkin, S. M. (2009).Checkpoint responses to unusualstructures formed by DNA repeats. *Mol. Carcinogenesis,* 48:309 – 318.

Voineagu, I., Surka, C. F., Shishkin, A. A., Krasilnikova, M. M. & Mirkin, S. M. (2009). Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat Struct Mol Biol.,* 16(2):226-228.

Wang, G. & Vasquez, K. M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A.,* 101(37):13448-13453.

Wang, G. & Vasquez, K. M. (2006). Non-B DNA structure induced genetic instability. *Mutat Res.*, 598, 103-119

Wang, J. C. (1984). DNA supercoiling and its effects on the structure of DNA. *J Cell Sci.* Suppl., 1:21-29.

Wang, Y. W. & Griffith, J. D. (1996). Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. *J Biol Chem.,* 271, 22937-22940.

Watanabe, Y. & Maekawa, M. (2010). Spatiotemporal regulation of DNA replication in the human genome and its association with genomic instability and disease. *Curr Med Chem.,* 17:222-233.

Watson, J. D. & Crick, F. H. C. (1953). Structure for deoxyribose nucleic acid, *Nature*, 171:737 – 738.

Weber, J. L. & Wong, C. (1993). Mutation of human short tandem repeats. *Hum Mol Genet.,* 2(8):1123-1128.

Wells, R. D. (2007). Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci.,* 32(6): 271-278.

Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. (1953a). Molecular structure of DNA, *Nature*,171:738-740.

Wilkins, M. H. F., Seeds, W. E., Stokes, A. R. & Wilson, H. R. (1953b). Helical structure of crystalline DNA, *Nature*, 172: 759 - 762.

Wojciechowska, M., Napierala, M., Larson, J. E. & Wells, R. D. (2006). Non-B DNA conformations formed by long repeating tracts of Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, and Friedreich's Ataxia genes, not the sequences per se, promote mutagenesis in flanking regions. *J Biol Chem.,* 281(34):24531-24543.

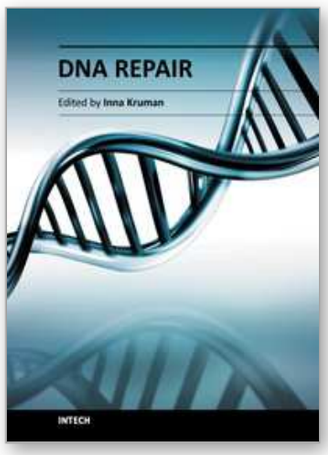
Yang, W. (2006). Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair* (Amst), 5(6):654-666.

Zahra, R., Blackwood, J. K., Sales, J. & Leach, D. R. F. (2007). Proofreading and secondary structure processing determine the orientation dependence of CAG • CTG trinucleotide repeat instability in *Escherichia coli*. *Genetics.,* 276:27-41.

Zhao, J., Jain, A., Iyer, R. R., Modrich, P. L. & Vasquez, K. M. (2009). Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res.,* 37(13):4420-4429.

Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci.,* 67:43-62.

IntechOpen



DNA Repair

Edited by Dr. Inna Kruman

ISBN 978-953-307-697-3
Hard cover, 636 pages
Publisher InTech
Published online 07, November, 2011
Published in print edition November, 2011

The book consists of 31 chapters, divided into six parts. Each chapter is written by one or several experts in the corresponding area. The scope of the book varies from the DNA damage response and DNA repair mechanisms to evolutionary aspects of DNA repair, providing a snapshot of current understanding of the DNA repair processes. A collection of articles presented by active and laboratory-based investigators provides a clear understanding of the recent advances in the field of DNA repair.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Xuefeng Pan, Peng Xiao, Hongqun Li, Dongxu Zhao and Fei Duan (2011). The Gratuitous Repair on Undamaged DNA Misfold, DNA Repair, Dr. Inna Kruman (Ed.), ISBN: 978-953-307-697-3, InTech, Available from: http://www.intechopen.com/books/dna-repair/the-gratuitous-repair-on-undamaged-dna-misfold

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen