遗传学报 *Acta Genetica Sinica*, January 2006, 33 (1): 1-11 ISSN 0379-4172

Mechanism of Trinucleotide Repeats Instabilities: The Necessities of Repeat Non-B Secondary Structure Formation and the Roles of Cellular Trans-acting Factors

PAN Xue-Feng [1,2,①]

1. *Institute of Cell and Molecular Biology, Edinburgh University, Edinburgh EH9 3JR, UK*;
2. *School of Life Sciences &Technology, Beijing Institute of Technology, Beijing* 100081, *China*

Abstract: The mechanism underlying CAG·CTG, CGG·CCG and GAA·TTC trinucleotide repeats expansion and contraction instabilities has not been clearly understood. Investigations *in vitro* have demonstrated that the disease causing repeats are capable of adopting non-B secondary structures that mediate repeats expansion. However, *in vivo*, similar observations have not been easily made so far. Investigations on the non-B secondary structure formation using *E.coli*, yeast etc cannot simulate the suggested repeats expansion instability. These could leave a space to infer a disassociation of the suggested repeats non-B secondary structure formation and the repeats expansion *in vivo*. Although longer trinucleotide repeats may be theoretically easier to form non-B DNA secondary structures in replication or in post-replication, however such non-B secondary structures are likely to cause repeat fragility rather than repeat expansion. In fact, repeat expansion as seen in patients may not necessarily require trinucleotide repeats to form non-B secondary structures, instead the repeat expansions can be produced through a RNA transcription-stimulated local repeat DNA replication and a subsequent DNA rearrangement.

Key words: trinucleotide repeats; expansion and contraction instability; fragility; human neurological-muscular disease; trans-acting factor

Trinucleotide repeats array distributed exclusively in human genome, however the biological significance of these distributions has not been clearly understood. It is now known that expansion instabilities of 4 in 10 possible trinucleotide repeats, e.g. CAG·CTG, GAA·TTC, CGG·CCG and CGA·TCG, can cause more than 15 human genetic diseases and cancers. By using different model systems, many investigations have been carried out to understand the molecular mechanisms underlying these DNA repeats maintenance, mainly to test the paradigm that repeat DNA non-B secondary structures mediating repeats expansion. However, this expected correlation of repeats DNA non-B secondary structures and repeats expansion could not be clearly demonstrated after more than ten years worldwide investigation. This could suggest that the hypothesis itself might be misleading. With this consideration, the necessities of repeat non-B secondary structure in repeat expansions *in vivo* and the roles of cellular trans-acting factors on repeats expansion are discussed. In the same time, a model of RNA transcription-stimulated additional repeat DNA replication and subsequently a DNA rearrangement produce repeats expansion in the absence of non-B DNA secondary formation is proposed.

1 Trinucleotide Repeats Expansion Through Repeats Non-B Secondary Structure Formation: When and How?

Expansion and contraction instabilities associ-

Received: 2005-08-21; Accepted: 2005-09-09

This Work Was Supported by Medical Research Council of United Kingdom and Beijing Institute of Technology (No. BIT-UBF-200506B4201).

① Corresponding Author. E-mail: xuefengpancam@yahoo.com.cn; xpan@staffmail.ed.ac.uk

ated with the propagation of trinucleotide repeats CAG·CTG, GAA·TTC, CGG·CCG are responsible for more than 15 human neurological genetic diseases (Table 1) [1-4]. The underlying molecular mechanisms have been investigated *in vitro* and in the model systems. *In vitro* investigations suggested that these disease-causing trinucleotide repeats can form non-B DNA secondary structures, and which could mediate repeat expansion and contraction instabilities. It was recently found that repeat expansion and contraction instabilities are not solely associated with CAG·CTG, GAA·TTC, CGG·CCG repeats propagation, but also can be seen with the propagation of a short CGA·TCG repeat array, of which the repeats instability is responsible for pseudoachondroplasia and multiple epiphyseal dysplasia [5]; and a 12-mer "CCCCGCCCCGCG" repeat array instability, which is related to the progressive myoclonus epilepsy [6]. The biochemical and biophysical investigations revealed that disease-causing repeats (including the trinucleotide repeats and the 12-mer "CCCGCCCCGCG" repeats) own potentials to adopt different types of non-B DNA structures, such as pseudohairpins, H-DNA and G-quadruplex etc., which are thought to be the intermediates for repeats expansion [3, 7-9], similar investigations have been carried out *in vivo* condition as well. Two research groups have attempted to detect repeat DNA secondary structure formations and their effects in model systems. One group was to use an λ-*E.coli* system to evaluate the folding potential of short CGG and CAG repeats during λ phage-plating [10], and the other was to use the artificially created CAG, CGG loops and yeast cells to test the repair capacity for the pre-existing CAG and CGG repeat loops [11]. The first experiment was conducted by inserting short CGG and CAG repeats into a palindrome DNA centre [10]. As it was observed that the folding of a palindrome DNA sequence into hairpin or cruciform structures *in vivo* depends mainly on the context of the DNA sequence in the palindrome centre. Indeed, the CAG and CGG DNA sequences that fold in the palindrome centre region can be the determinative factors for the palindrome DNA sequence folding, which then can be reflected through the phage plaque variation in plating [10]. The second report showed that at a particular cellular stage (meiotic cells), the pre-existing DNA loops couldn't be entirely corrected by any of the pathways of DNA repair and recombination [11]. These put together argued that the disease-causing repeats can be foldable *in vivo*, and once they are folded can also bypass cellular corrections.

Unfortunately, these experiments cannot display the relation of the repeats expansion and the repeat secondary structures. It was therefore still unclear if the repeats structures are the real intermediates for the repeats expansion.

In principal, the folding of a repeated DNA sequence can be influenced by more than two parameters. Firstly, loss of the interactions with repeat-binding proteins, such as histone proteins in eukaryotic cells, or the HU protein in *E.coli*, etc. [12], can change the states of the DNA molecules *in vivo*, and which then might affect the repeats DNA fold ability. In fact, the assembly of nucleosome proteins onto short CTG or methylated CCG are actually being favored [13]. It has been reported that both short $(CTG)_n$ and $(CGG)_n$ repeats containing the methylated CpG dinucleotides can form very stable complexes with histone proteins [13], but long CGG tracts (>76 repeats) cannot. More interestingly, the long CGG tracts repel the histone binding when the CpG dinucleotides were methylated [14]. Secondly, negative supercoiling in DNA molecules can also be the determinative factors in repeats DNA folding. Negative supercoiling can be generated through either DNA replication or DNA transcription [15,16], which will promote certain types of DNA repeats (sequences) to form non-B secondary structures. For example, the negative supercoiling accumulated in DNA replication and in RNA transcription can drive AT-rich DNA repeats to form hairpin and/or cruciform structures *in vivo* [17]. A similar observation has been made during a $(AAT)_{24}$ trinucleotide repeat propagating in *E.coli*. In this case, transcription of $(AAT)_{24}$ repeat has

Table 1 Human genetic diseases associated with the trinucleotide repeats expansion and contraction instabilities

Diseases	Repeats	Location	Repeats variation	The affected gene
FragileX	CGG	Xq27.3		FMR1
Chronic lymphocytic leukaemia (CLL)	CCG	11q22-q23	CCG8 no expansion	Deletion
Oculopharyngeal muscular dystrophy	GCG	14q11	GCG6 GCG8-13	Pabp2
FragileXE	CGG	Xq28		FMR2
Jacobsen syndrome (FRA11B)	CGG		GCC7-35 GCC130-150GCC230-750	
Myotonic dystrophy	CTG	19q13	CTG5-37 CTG50-3000	3′-UTR
Friedreichi's ataxia	GAA	9q13-21.1(Intron1)	GAA6-34 GAA80 to GAA112-1700	Intron1 Frataxin
Spinobular muscular atrophy	CAG	Xq13-q21	CAG11-33 CAG38-66	Androgen receptor
Huntington's disease	CAG	4p16.3	CAG6-39 to CAG36-121	Huntingtin
Dentatorubral-pallidduysian atrophy	CAG	12p13.31	CAG6-35 to CAG51-88	Atrophin-1
Friedreich's ataxia	GAA	9q13-q21.1	GAA6-29 to GAA200-900	Frataxin
Multiplepiphyseal dysplasia	CGA	?	CGA5 to CGA4-6	?
Spinobulbar ataxia 1	CAG	6p23	CAG6-44 to CAG39-82	Ataxin-1
Spinobulbar ataxia 2	CAG	12q24.1	CAG14-31 to CAG34-59	Ataxin-2
Spinobulbar ataxia 3	CAG	14q32.1	CAG13-44 to CAG55-84	Ataxin-3
Spinobulbar ataxia 6	CAG	19p13	CAG4-18 to CAG21-33	Calcium channel
Spinobulbar ataxia 7	CAG	13p12-p13	CAG4-34 to 37-306	Ataxin-7
Spinobulbar ataxia 8	CTG	13q21	CTG15-27? to 110-200	?
Spinobulbar ataxia 12	CAG	5q31-q33	CAG7-28 to 66-78	?
Haw river syndrome	CAG	12q13-q21.1	CAG6-29 to 200-900	Polyglutamine

clearly shown as a reason for unusual DNA secondary structures formation, and DNA propagation instability generation, as well as repeats contraction instability (Pan and Leach, to be published results).

Despite these non-B DNA structure formation *in vivo*, subsequent genetic effects they might cause will still depend on the thermodynamic stability of the structure. Ongoing DNA polymerases can be stalled by a stable repeats DNA secondary structure, and further, the DNA replication fork can sometimes also be stalled, or to be converted into other DNA substrates, such as single-stranded gaps or double strand breaks [18]. In the last few years, as revealed from the investigations, cells can manage to resume the stalled DNA replication with the assistance of some of the homologous recombination proteins [19]; to repair or bypass the single-stranded gap by DNA translesion synthesis or homologous recombination [20], and to repair the double strand break damages by homologous recombination or non-homologous end joining [21]. Failure in any of these repair pathways may elevate the genomic instabilities, either lose the replicating DNA molecules, cause cell viability, or lead to the aberrant DNA rearrangements. The unviable cell may hide trinucleotide repeats instabilities, and the aberrant DNA rearrangements may cause repeats expansions. By contrast, non-B secondary structures formed through RNA transcription may not directly threaten DNA duplication and cell proliferation, but may still be a possible pathway causing repeats expansion (will be discussed in the following).

2 Dose trinucleotide repeats expansion really need repeat non-B secondary structure to be intermediates?

Recent investigations found that the repeats instabilities in mammals can happen in the non-proliferation cells, which seemed do not involve any genomic DNA replication [22], and cells isolated from the fragile X patients did not show any increased sensitivity to DNA damaging reagents, indicating that DNA damage repair functions in patients are normal [23]. Moreover, the disease-causing CTG, CGG and GAA duplex DNA sequences appear to be no reason to adopt non-B secondary structures unless that they are in DNA replication or transcription. Taken these together, it is suggesting that the repeat expansion and contraction instabilities are unlikely relevant to cell proliferation, genomic DNA replication and DNA damage repair pathways as well as repeats non-B DNA secondary structure formation.

As being demonstrated *in vitro* and *in vivo*, gene locus carrying disease causing repeats are normally the active chromosomal sites, where in most situations need to be transcribed for gene expression. Interestingly, most of the expanding repeats were likely to stall DNA polymerase and RNA polymerase in replication and transcription [4, 24]. These may be suggestive for a reasonable repeat expansion mechanism, which in this paper was called as RNA transcription-stimulated local repeats DNA replication (TSLDR) and an unequal crossing-over recombination (Fig.1A). In such a process, RNA transcription passing through the trinucleotide repeats is suggested to be interrupted by the repeat DNA conformation, however, which may not be necessarily any typical non-B DNA secondary structures. The interrupted RNA transcription in turn leaves an RNA moiety for DNA polymerase(s) to initiate a local DNA replication. The repeats expansion can be made through the newly synthesized repeat fragment as if it was "being inserted" into the non-template repeat strand (Fig.1 A, A_1 and A_1'). Given this happens, the non-template repeat strand would need to be broken by at least one of many single-stranded endonucleases, such as Mre11-Rad50 etc (Fig.1 A_0). Alternatively, repeats expansion can also be made through a double strand breaks rearrangement pathway. This might happen once if the second DNA replication was also being initiated on the non-template strand of transcription bulb (Fig.1A, B_0). In these situations, Mre11-Rad50 and many other DNA nucleases that are capable of processing the double strand breaks might become essential for generating and repairing the repeats double strand breaks (Fig.1A). As can be seen in Fig.1B, these repeats expansion pathways do not necessarily need the trinucleotide repeats to form any non-B secondary structures for the expansion, and do not necessarily need the whole genomic DNA replication and cell proliferation. However, on such repeat expansion mechanism, cells do need the functional DNA repairs and recombination. All of these are consistent with the present findings as described in the references [22, 23] (Fig.1A). In addition to the TSLDR repeat expansion mechanism, other possible routes in repeats expansion while do not necessarily need non-B secondary structures intermediates may be associate with the actions of certain types of transposable elements, the transposition process can also generate tandem repeats at each ends of a transposon. Interestingly we have found that IS1E inserting elements could recognize AAT repeats for target (Pan and leach unpublished results), and recently Mu phage was also found to target the CGG repeats [25].

3 What Roles of Cellular Trans-acting Factors Can Play in Repeats Expansion?

The less involvement of the repeats expansion to cell proliferation, genomic DNA replication and DNA repair and recombination pathways in patients can be nicely interpreted by using a TSLDR mechanism. Therefore, it becomes interesting to discuss the roles of the cellular trans-acting factors in repeats expansion combining to the TSLDR mechanism.

3.1 Effects of DNA replication and DNA replication components

Investigations with the model systems, such as

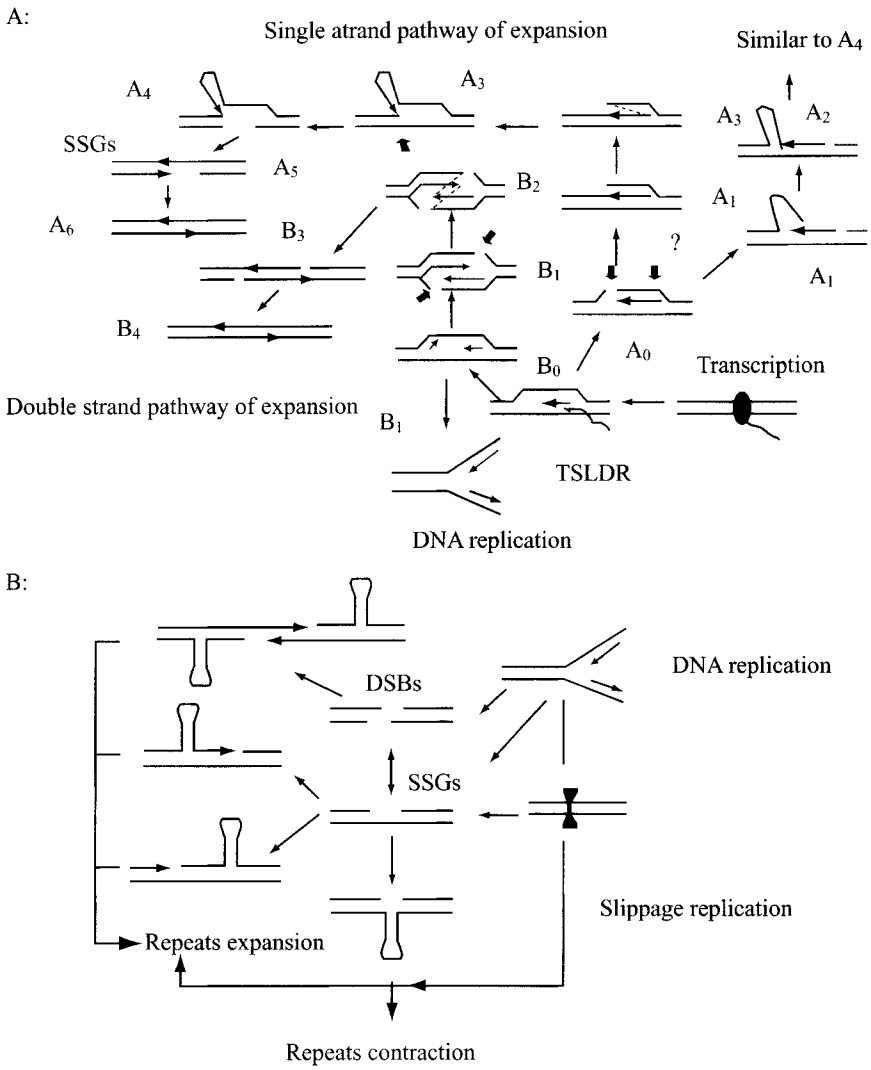

Fig. 1 Pathways of trinucleotide repeat expansion

A: Transcription-promoted local DNA replication (TSLDR) leads to repeat expansion. RNA transcription through the repeats array generates a RNA moiety after being interrupted that can be recognized as a primer RNA for local repeats DNA replication. Abolishment of this repeats DNA replication by DNA nucleases through making single strand breaks leads to the extra repeats array being "inserted" into the non-template repeats strand and produce single strand expansion (A_0 to A_5). Alternatively, double strand breaks can also be made (B_1 to B_4). B: Pathways of repeats expansion through the repair of single strand and double strand breakage. Single-stranded gaps (SSGs) and double strand breaks (DSBs) may be generated during repeats propagation (either by a stalled DNA replication fork or by the repair of the repeat non-B secondary structure etc.). Repair for these SSGs and DSBs leads to repeats expansion.

E.coli, yeast etc., have already suggested a correlation between the generations of trinucleotide repeat expansion and contraction instabilities and the unfaithful DNA replication. Indeed several reports with these model systems have suggested that mutations in *dnaQ (mutD)*, the proof-reading activity, *dna E,* the polymerase activity of *E.coli pol* Ⅱ complex, *pol* Ⅰ in Haemophilus influenza and *polδ* in eukaryotic organisms affected some repeats instability [26-30]. Similarly, mutations in helicases, which was known to cause replication fork stalling, such as *dnaB, helD, rep* etc., also provoke repeats instabilities. The remaining question will be if these mutants also affect trinucleotide repeats instability [19]. Recently, several new DNA polymerases have been characterised in *E.coli* and in other eukaryotic cells, e.g. DNA polymerase Ⅳ and Ⅴ in *E.coli*; Y family DNA polymerases in yeast and mammalian cells [31,32]. It is be-

lieved that some DNA polymerases worked in the different types of DNA replications, such as the DNA *pol*Ⅴ and DNA *pol*Ⅲ in *E.coli* DNA translesion synthesis etc [31]. These alternative DNA replications, though have not been fully understood either in prokaryotic or in eukaryotic cells [31,32], may start to suggest a possibility that one of the DNA replication modes might be critical in trinucleotide repeat expansion and contraction instability, such as in the proposed TSLDR pathway.

3.2 Okazaki fragment and DNA flap processing

In theory, unfaithful DNA replication also includes the abnormal processing of the Okazaki fragments in the lagging strand DNA replication. It has long been clear that the nascent DNA strand on a conventional lagging strand DNA replication is replicated through making pieces of Okazaki fragments. Maturation of these discontinues Okazaki DNA fragments then require several gene products to process, such as the RAD2 and Rth1/Rad 27 in yeast, the Fen1 in humans, and the RnhA/B/ DNA polymerase Ⅰ in *E.coli* [33-35]. In yeast, deficient in *rth*1/*rad* 27 gene increases trinucleotide repeats instability has been demonstrated [36]. Similarly, the yeast DNA ligase mutants that known to ligate the Okazaki fragments, e.g. *cdc*9-1 and *cdc*9-2 were also found to increase the CAG repeat instability [37]. These findings infer that the DNA replicative intermediates, like unprocessed Okazaki fragment can mediate the trinucleotide repeats instability in repeats replication, either directly [38], or by the homologous recombination repair of the single strand gaps and/or double strand breaks that subsequently converted [39]. Otto *et al.*[40] have examined the *fen*1 gene in Huntington disease patients as that is as normal as the wildtype gene. However, several demonstrations have argued that the repeat Okazaki fragment was capable of forming Fen1- resisting hairpin structure [40].

3.3 Mismatch repair function and its proteins

Mismatch repair functions are conserved in both prokaryotic and eukaryotic organisms. Mismatch repair functions correct either mismatched base pairs or small DNA loops that appeared in DNA replication, DNA damage, repair and recombination [41,42]. In humans, it has been known that loss of mismatch repair function caused hereditary nonpolyposis colon cancer, which was believed due to an increased risk of microsatellite instability in genes playing caretaker and gatekeeper roles [43]. The effects of mismatch repair on trinucleotide repeat expansion and contraction instabilities have been examined. To date, there were two roles of the mismatch repair: the binding activities of the mismatch repair proteins, MutS and Msh2, as well as the repair functions as the whole versus repeat secondary structure formation and the final output of repeat instability have been found. Sinden's lab found that Msh2, the common subunit of the MutSα and MutSβ recognition protein complex in the yeast and human mismatch repair pathway recognised the mismatched base pairs in a CAG repeats pseudo-hairpin structure [44]. This finding made consistency with the Jarwarse *et al.*'s observation that the increased repeats instabilities in plasmid in *E.coli* required the presence of methyl-directed mismatch repair components, MutS/L/H, in a repeats size dependent manner[45,46]. Mismatch repair proteins seemed to be capable of stimulating repeats larger deletions by binding onto the repeats secondary structure while decrease the small slippage instability by the repair function [44-47]. This two opposing roles have further observed as that MMR influenced the CAG·CTG larger deletion instability in an orientation dependent manner, while preventing its small slippage instability in an orientation-independent manner [48]. More recently, it was also found that the Msh2 in mouse increased the repeats instability in somatic and post-meiotic sperm cells, otherwise caused repeat deletions when absent [49,50]. The significance of these findings is that the mismatch repair proteins, such as MutS and Msh2 may directly contribute to the repeat larger instabilities presumably by acting with the repeat secondary structure in the way of doing the repair. Interestingly, the MutS-binding activity was found to be separated from its repair function. We found that the binding activity of MutS, plus the nu-

clease activity of SbcCD upon the TGG trinucleotide repeat secondary structure during the lagging strand replication in plasmid in *E.coli* was detrimental. In our hands, the involvement of MutS, SbcCD (the structural-specific nuclease activity) and RecA (the recombination protein) are found to reside in a synthesised event, which caused TGG repeat propagation instability but not the expansion and contraction instabilities [51]. Since the homologous proteins of MutS, SbcCD and RecA proteins are conserved in mammalian cells, and the expanded CGG trinucleotide repeats in human cause fragile X syndrome, it therefore becomes interesting to suggest a similar synthesised repair process may also be implicated in human fragile disease.

3.4 Recombination and recombination proteins

The homologous recombination in prokaryotic and eukaryotic cells plays major roles in DNA damage tolerance, repair and replication [18]. Homologous recombination can repair both the DNA single-stranded gaps and the double strand breaks [18, 19, 52]. Moreover, some homologous recombination proteins are also being required for the salvage for the stalled replication forks while not doing the conventional recombination [20, 53]. Apparently, there are connections between DNA homologous recombination and DNA replication *in vivo* [54]. The roles of recombination and recombination-related proteins in trinucleotide repeat instabilities have been thoroughly examined in *E.coli*, yeast and mammalian cells [55-61]. In 1998, Freudenreich *et al.* reported that a CTG repeat propagation in yeast caused double strand breakage in the absence of Rad27[56], and short after, several groups have found that such double strand breaks produced repeat expansion and contraction instabilities, especially when the Mre11-Rad50-Xrs2 and Rad52 proteins were absent[57-61]. These findings turned to be that the aberrant homologous recombination for the repeats double strand breaks repair causes repeats expansion and contraction instabilities. However, the similar effect of the mammalian homologous recombination on myotonic dystrophy CTG repeats instabilities has failed to be seen [61].

3.5 DNA secondary structure processing proteins

As discussed above, once the speculated non-B DNA secondary structures have been formed *in vivo*, cells cannot always correct them. In *E.coli,* the structural-specific protein SbcCD has been found to cleave the palindrome DNA hairpin structures [52]. SbcCD introduced double strand breaks by cleaving the palindromic DNA hairpin structure, which in turn left to the functional homologous recombination to repair. It is known that the SbcCD is the functional homologues of Rad50-Mre11-Xrs2 (yeast) or Rad50-Mre11-Nbs1 (mammalian). However, these Rad50-Mre11-Xrs2 (Nbs1) proteins play rather complex roles in DNA replication, repair, and cell cycle control [62]. Considering the Rad50-Mre11-Xrs2 mutants generating significant CAG repeat instabilities [57-59], it should be cautious to correlate the non-B secondary structure formation of the trinucleotide repeats sequences.

3.6 Repeats location

Interestingly, even in the patients, it is not all their CAG, CGG or GAA trinucleotide repeats in chromosomes went to expand and contract. In fact, the expandable trinucleotide repeat arrays of CAG, CGG and GAA need to reside in the particular gene loci (Table 1). This implied that the location of the disease-causing repeat in patients could also be important in affecting the repeat expansion and contraction instabilities. In 1999, Brock *et al.* reported a positive correlation between the CAG·CTG repeat expandability and its flanking sequences. They found that the most expandable repeats appeared to be located within CpG islands [63]. Recently Cleary *et al.* reported that a trinucleotide repeat location versus its replication origin might be responsible for this "site-specific" effect [64]. Conceivably, the flanking DNA sequences might be capable of influencing the repeat conformation in DNA replication [65].

3.7 Repeat- binding proteins

Non-B DNA structure-binding proteins have been reported in literature, including Z-DNA binding

proteins, triplex-DNA binding proteins, NSEP-1, cruciform-DNA binding proteins, HMG1 etc [66]. The CGG repeat-binding proteins have also been isolated from human cell lines. These CGG repeat-binding proteins can bind to the unmethylated CGG repeated array specifically. In particular, the expanded long CGG repeats can be poorly methylated in their CpG dinucleotides (unmethylation or hypomethylation problem), and can be bound by CGG repeat-binding proteins (not the histone proteins). However, the significance of this protein-binding is not clear [67-69].

4 Concluding Remarks

Molecular mechanisms underlying the CAG, CGG and GAA trinucleotide repeats expansion and contraction instabilities in certain chromosome loci (Table 1) have not been fully understood after more than 10 years investigations. To date, lots of efforts have been made to investigate whether or not these disease-causing trinucleotide repeats can form non-B secondary structures *in vivo*, and whether or not the non-B secondary structures can lead to repeat expansion and contraction instabilities. As discussed here, it appears to be that multiple factors and pathways relating to DNA replication repair and recombination may influence repeat instabilities. An expansion model has thus been proposed in Fig.1A [70]. As can be seen in Fig.1B, trinucleotide repeats expansion may not be necessarily involving any significant non-B secondary structures, cell proliferation, genomic DNA replication and the loss of any DNA repair and recombination functions [22,23], an RNA transcription stimulated local repeat DNA replication may be capable of making repeats expansion. As a comparison, the present model for repeats expansion was also outlined in Fig.1B, which relay on the cell proliferation, genomic DNA replication and the loss of any DNA repair and recombination functions. As Fig.1B indicated that DNA single strand gaps and/or the double strand breaks could be generated in trinucleotide repeats replication or in repeats DNA damage repair and recombination. Trinucleotide repeats may be capable of forming non-B secondary structures through loss of interaction with repeat-binding proteins or through the drive of supercoiling. Repair for these DNA structures may cause repeat expansion and contraction instabilities as well as fragility (Fig.1).

Acknowledgements: The author is grateful to Professor D Leach in the University of Edinburgh for discussion and suggestion.

References:

[1] Richards R I, Sutherland G R. Dynamic mutations: a new class of mutations causing human disease. *Cell*, 1992, 70 : 709-712.

[2] Richards R I, Sutherland G R. Dynamic mutation: possible mechanisms and significance in human disease. *Trends Biochem Sci*, 1997, 22 : 432-436.

[3] Wells R D, Dere R, Hebert M L, Napierala M, Son L S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acid Res*, 2005, 33(12) : 3785-3798.

[4] Usdin K, Grabczyk E. DNA repeat expansions and human disease. *Cell Mol Life Sci*, 2000, 57 : 914-931.

[5] Delot E, King L M, Briggs M D, Wilcox W R, Cohn D H. Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene. *Hum Mol Genet*, 1999, 8 : 123-128.

[6] Lalioti M D, Mirotson M, Buresi C, Peitsch M C, Rossier C, Ouazzani R, Baldy-Moulinier M, Bottani A, Malafosse A, Antonarakis S E. Identification of mutations in cystatin B, the gene responsible for the Unverricht-Lundborg type of progressive myoclonus epilepsy (EPM1). *Am J Hum Genet*, 1997, 60 : 342-351.

[7] Sinden R R, Potaman V N, Oussatcheva E A, Person C E, Lyubchenko Y, Shlyahtenko L S. Triplet repeat DNA structures and human genetic disease: Dynamic mutations from dynamic DNA. *J Biosci*, 2002, 271 : 53-65.

[8] Mitas M. Trinucleotide repeats associated with human disease. *Nuclei Acids Res*, 1997, 259 : 2245-2253.

[9] Pearson C E, Ewel A, Sinden R R. In: Wells and Warren eds. Genetic instabilities and hereditary neurological diseases. Academic Press, 1998, 585-621.

[10] Darlow J M, Leach D R F. The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in *Escherichia coli* suggest hairpin folding preferences *in vivo*. *Genetics*, 1995, 141 : 825-832.

[11] Moor H, Greemwell P W, Liu C P, Arnheim N, Petes T D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc Natl Acad Sci USA*, 1999, 96 : 1504-1509.

[12] Rouviere-Yaniv J, Germond J E, Yaniv M. *E. coli* DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. *Cell*, 1979, 17 : 265-274.

[13] Godde J S, Kass S U, Hirst M C, Wolffe A P. Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. *J Biol Chem*, 1996, 271 : 243525-243588.

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

[15] Champoux J J. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, 2001, 70 : 369-472.

[16] Postow L, Crisona N J, Peter B J, Hardy C D, Cozzarelli N R. Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci USA*, 2001, 98 : 8219-8226.

[17] Courey A J, Wang J C. Influence of DNA sequence and supercoiling on the process of cruciform formation. *J Mol Biol*, 1988, 202 : 35-43.

[18] Marians K J. Replication and recombination intersect. *Curr Opin Genet Dev*, 2000, 10 : 151-156.

[19] Michel B. Replication fork arrest and DNA recombination. *Trends Biochem Sci*, 2000, 25 : 173-178.

[20] Cox M M, Goodman M F, Kreuzer K N, Sherratt D J, Sandler S J, Marians K J. The importance of repairing stalled replication forks. *Nature*, 2000, 404 : 37-41.

[21] Van Gent D C, Hoeijmakers J H J, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nature Rev Genet*, 2001, 2 : 196-206.

[22] McMurray C T, Kortun I V. Repair in haploid male germ cells occurs late in differentiation as chromatin is condensing. *Chromosoma*, 2003, 111 : 505-508.

[23] Wang T S, Hsieh L J, Hsu T Y, Chung C H, Li S Y. DNA damage and repair in lymphoblastoid cell lines from normal donors and fragile X syndrome patients. *Arch Med Res*, 2002, 33 : 128-135.

[24] Chandler S P, Kansagra P, Hirst M C. Fragile X (CGG)n repeats induce a transcriptional repression in cis upon a linked promoter: evidence for a chromatin mediated effect. *BMC Mol Biol*, 2003, 4 : 1-12.

[25] Manna D, Deng S, Breier A M, Higgins N P. Bacteriophage Mu targets the trinucleotide sequence CGG. *J Bacteriol*, 2005, 187(10) : 3586-3588.

[26] Bzymek M, Lovett S T. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc Natl Acad Sci USA*, 2001, 98 : 8319-8325.

[27] Saveson C J, Lovett S T. Enhanced deletion formation by aberrant DNA replication in *Escherichia coli.Genetics*, 1997, 146 : 457-470.

[28] Jin Y H, Obert R, Burgers P M, Kunkel T A, Resnick M A, Gordenin D A. The 3′→5′ exonuclease of DNA polymerase delta can substitute for the 5′ flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proc Natl Acad Sci USA*, 2001, 98 : 5122-5127.

[29] Bayliss C D, deVen T V, Moxon E R. Mutations in pol I but not mutSLH destabilize Haemophilus influenzae tetranucleotide repeats. *EMBO J*, 2002, 21 : 1465-1476.

[30] Iyer R R, Pluciennik A, Rosche W A, Sinden R R, Wells R D. DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in Escherichia coli. *J Biol Chem*, 2000, 275 : 2174-2184.

[31] Friedberg E C, Feaver W J, Geriach V L. The many faces of DNA polymerases: Strategies for mutagenesis and for mutational avoidance. *Proc Natl Acad Sci USA*, 2000, 97 : 5681-5683.

[32] Goodman M F, Tippin B. Sloppier copier DNA polymerases involved in genome repair. *Curr Opin Genet Dev*, 2000, 10 : 162-168.

[33] Schweitzer J K, Livingston D M. Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum Mol Genet*, 1998, 7 : 69-74.

[34] Henricksen L A, Veeraraghavan J, Chafin D R, Bambara R A. DNA ligase I competes with FEN1 to expand repetitive DNA sequences *in vitro*. *J Biol Chem*, 2002, 277 : 22361-22369.

[35] White P J, Borts R H, Hirst M C. Stability of the human fragile X (CGG)(n) triplet repeat array in Saccharomyces cerevisiae deficient in aspects of DNA metabolism. *Mol Cell Biol*, 1999, 19 : 5675-5684.

[36] Nagata Y, Mashimo K, Kawata M, Yamamoto K. The roles of Klenow processing and flap processing activities of DNA polymerase I in chromosome instability in *Escherichia coli* K12 strains. *Genetics*, 2002, 160 : 13-23.

[37] Irelanda M J, Reinkea S S, Livingston D M. The impact of lagging strand replication mutations on the stability of CAG repeat tracts in yeast. *Genetics*, 2000, 155 : 1657-1665.

[38] Gordenin D A, Kunkel T A, Resnick M A. Repeat expansion-all in a flap? *Nat Genet*, 1997, 16:116-118.

[39] Symington L S. Homologous recombination is required for the viability of *rad27* mutants. *Nucleic Acids Res*, 1998, 26 : 5589-5595.

[40] Otto C J, Almqvist E, Hayden M R, Andrew S E. The "flap" endonuclease gene FEN1 is excluded as a candidate gene implicated in the CAG repeat expansion underlying Huntington disease. *Clin Genet*, 2001, 59 : 122-127.

[41] Jiricny J. Replication errors: cha(lle)nging the genome. *MEBO J*, 1998, 17 : 6427-6436.

[42] Marra G, Schar P. Recognition of DNA alterations by the mismatch repair system. *Biochem J*, 1999, 338 : 1-13

[43] El-Rifai W, Powell S M. Molecular biology of gastric cancer. *Semin Radiat Oncol*, 2002, 12 : 128-140.

[44] Boyer J C, Umar A, Risinger J I, Lipford J R, Kane M, Yin S, Barrett J C, Kolodner R D, Kunkel T A. Microsatellite instability, mismatch repair deficiency, and genetic defects in human cancer cell lines. *Cancer Res*, 1995, 55 : 6063-6070.

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

[46] Jaworski A, Rosche W A, Gellibolian R, Kang S, Shimizu M, Bowater R P, Sinden R R, Wells R D. Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. *Proc Natl Acd Sci USA*, 1995, 92 : 11019-11023.

[47] Parniewski P, Jaworski A, Wells R D, Bowater R P. Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. *J Mol Biol*, 2000, 299 : 865-874.

[48] Schmidt K H, Abbott C M, Leach D R. Two opposing effects of mismatch repair on CTG repeat instability in *Escherichia coli*. *Mol Microbiol*, 2000, 35 : 463-471.

[49] Manley K, Shirley T L, Flaherty L, Messer A. Msh2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet*, 1999, 23 : 471-473.

[50] Kovtun I V, McMurray C T. Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet*, 2001, 27 : 407-411.

[51] Pan X, Leach D R. The roles of *mutS*, *sbcCD* and *recA* in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res*, 2000, 28 : 3178-3184.

[52] Leach D R, Okely E A, Pinder D J. Repair by recombination of DNA containing a palindromic sequence. *Mol Microbiol*, 1997, 26 : 597-606.

[53] Courcelle J, Donaldson J R, Chow K H, Courcelle C T. DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science*, 2003, 299 : 1064-1067.

[54] Haber J E. DNA recombination: the replication connection. *Trends Biochem Sci*, 1999, 24 : 271-275.

[55] Jekupciak J P, Wells R D. Gene conversion (recombination) mediates expansions of CTG·CAG repeats. *J Biol Chem*, 2000, 275 : 40003-40013.

[56] Freudenreich C H, Kantrow S M, Zakian V A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, 1998, 279 : 853-856.

[57] Paques F, Leung W Y, Haber J. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol Cell Biol*, 1998, 18 : 2045-2054.

[58] Richard G F, Dujio B, Haber J E. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol Gen Genet*, 1999, 261 : 871-882.

[59] Richard G F, Goellner G M, McMurray C T, Haber J E. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. *EMBO J*, 2000, 19 : 2381-2390.

[60] Jankowski C, Nasar F, Nag D K. Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc Natl Acad Sci USA*, 2000, 97 : 2134-2139.

[61] Savouret C, Brisson E, Essers J, Kanaar R, Pastink A, teRiele H, Junien C. CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J*, 2003, 22 : 2264-2273.

[62] D'Amours D, Jackson S P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Rev Mol Cell Biol*, 2002, 3 : 317-327.

[63] Brock G J, Anderson N H, Monckton D G. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Gene*, 1999, 8 : 1061-1067.

[64] Cleary J D, Nichol K, Wang Y H, Pearson C E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat Gene*, 2002, 31 : 37-46.

[65] Pan X. Two-triplet CGA repeats impede DNA replication in bacteriophage M13 in *Escherichia coli*. *Microbiol Res*, 2004, 159 : 97-102.

[66] Holde K V, Zlatanova J. Unusual DNA structures, chromatin and transcription. *BioEssays*, 1994, 16 : 59-68.

[67] Deissler H, Behn-krappa A, Doerfler W. Purification of nuclear proteins from human HeLa cells that bind specifically to the unstable tandem repeat (CGG)n in the human *FMR1* gene. *J Biol Chem*, 1996, 271 : 4327-4334.

[68] Goddle J S, Kass S U, Hirst M C, Wolffe A P. Nucleosome assembly on methylated CGG triplet repeats in

the fragile X mental retardation gene 1 promoter. *J Biol Chem*, 1996, 271 : 24325-24328.

[69] Saveliev A, Everett C, Sharpe T, Webster Z, Festenstein R. DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*, 2003, 422 : 909-913.

[70] Pan X. Self-maintenance of gene and occurrenance of genetic diseases. Beijing: Science Press, 2004, 1-403.

三核苷酸重复序列失稳定机制：重复序列形成 non-B 二级结构的必要性和细胞反式作用因子的作用

潘学峰[1,2]

1. 英国爱丁堡大学细胞与分子生物学研究所，爱丁堡 EH9 3JR，英国；

2. 中国北京理工大学生命科学与技术学院，北京 100081，中国

摘 要：与三核苷酸重复序列 CAG·CTG、CGG·CCG 和 GAA·TTC 扩增和缺失有关的分子机制尚不能得到清楚的阐释。体外研究表明，上述疾病相关的重复序列可以在体外形成 non-B 二级结构，并介导重复序列扩增。然而，迄今为止，类似的观察尚未在体内研究过程中得以实现。利用模型生物大肠杆菌和酵母等进行的有关研究并不能模拟三核苷酸重复序列的扩增，这暗示三核苷酸重复序列的体内扩增可能与重复序列形成 non-B 二级结构关联性并不大。尽管理论上较长的三核苷酸重复序列可以在复制和后复制过程中较易形成 non-B DNA 二级结构，但这样的二级结构倾向于导致重复序列出现“脆性”，而不是扩增。事实上，患者所具有的三核苷酸重复序列扩增并非一定需要通过 non-B 二级结构的介导，这些重复序列的扩增是可以通过一种 RNA 转录诱导的局部 DNA 重复序列的复制和其后的 DNA 重排得以发生。

关键词：三核苷酸重复序列；扩增和缺失；脆性 DNA；神经-肌肉疾病；反式作用因子

作者简介：潘学峰，男，中国国籍，英国永久居民，微生物学与生物化学学士、遗传学硕士、生物化学博士，研究生导师，研究方向：重大疾病的分子基础