

## ARTICLE

# The competing mini-dumbbell mechanism: new insights into CCTG repeat expansion

Pei Guo and Sik Lok Lam

CCTG repeat expansions in intron 1 of the cellular nucleic acid-binding protein gene are associated with myotonic dystrophy type 2. Recently, we have reported a novel mini-dumbbell (MDB) structure formed by two CCTG or TTTA repeats, which potentially has a critical role in repeat expansions. Here we present a mechanism, called the competing MDB mechanism, to explain how the formation of MDB can lead to efficient mismatch repair (MMR) escape and thus CCTG repeat expansions during DNA replication. In a long tract of CCTG repeats, two competing MDBs can be formed in any segment of three repeats. Fast exchange between these MDBs will make the commonly occupied repeat behave like a mini-loop. Further participations of the 5'- or 3'-flanking repeat in forming competing MDBs will make the mini-loop shift in the 5'- or 3'-direction, thereby providing a pathway for the mini-loop to escape from MMR. To avoid the complications due to the formation of hairpin conformers in longer CCTG repeats, we made use of TTTA repeats as model sequences to demonstrate the formation of competing MDBs and shifting of mini-loop in a long tract of repeating sequence.

*Signal Transduction and Targeted Therapy* (2016) 1, e16028; doi:10.1038/sigtrans.2016.28; published online 2 December 2016

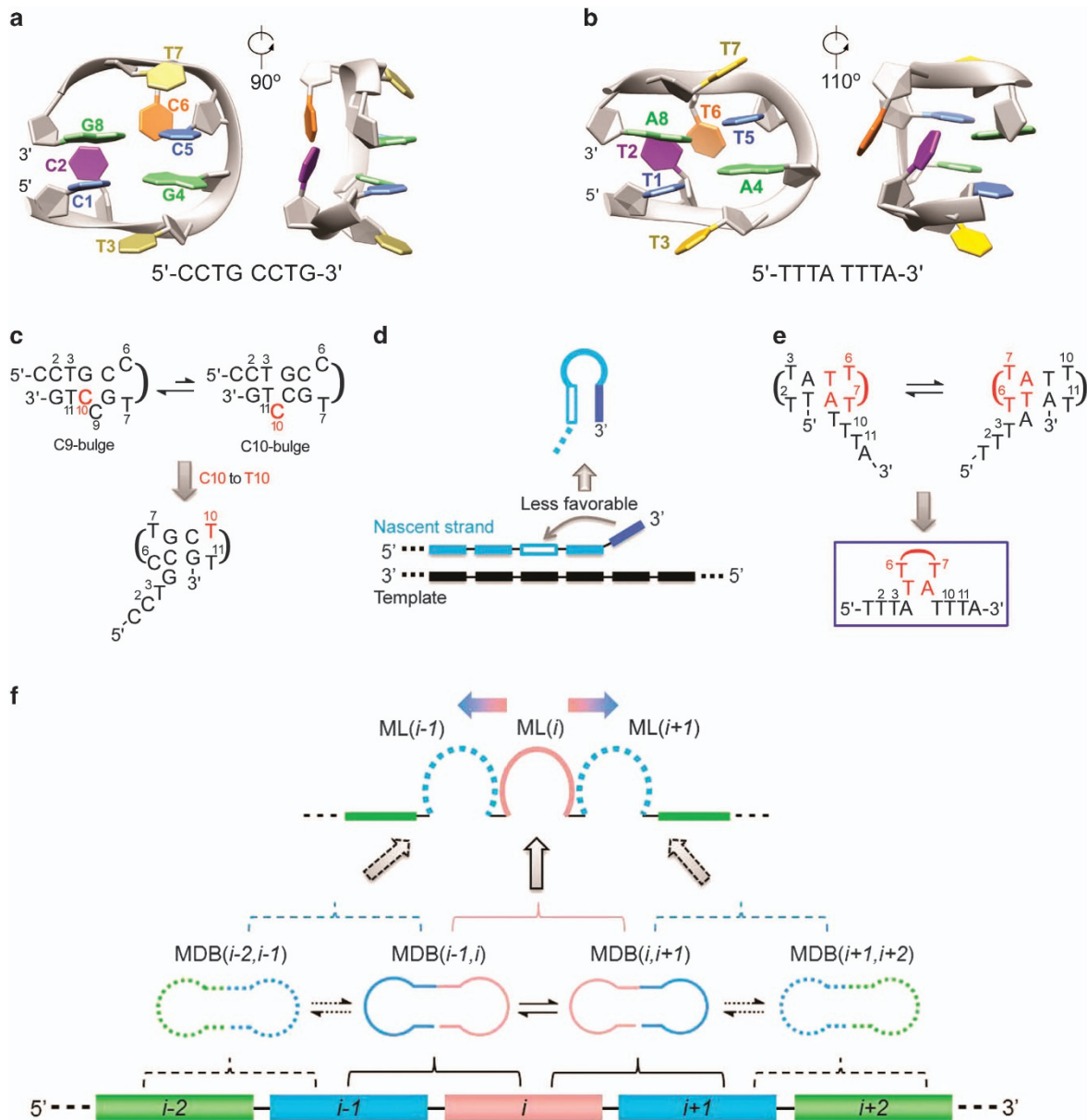
## INTRODUCTION

In the human genome, expansions of CCTG repeats found in intron 1 of the cellular nucleic acid-binding protein (*CNBP*) gene on chromosome 3q21 are associated with a complex multisystem disorder called myotonic dystrophy type 2 (DM2).<sup>1</sup> These repeats are located in a part of the complex motif (TG)<sub>14–25</sub>(TCTG)<sub>4–10</sub>(CCTG)<sub>*n*</sub> (where *n* is the repeat length).<sup>1–3</sup> In normal individuals, *n* is usually below 30 and the CCTG repeat tract is interrupted by one or more A/G/TCTG motif(s). In DM2 patients, *n* can vary between 55 and ~11 000 and the expansions of CCTG repeats are extremely variable.<sup>4</sup> At present, the mechanism of CCTG repeat expansions remains elusive.

For repeat expansions to occur, one commonly accepted pathway involves the formation of an unusual structure in the nascent strand during DNA replication.<sup>5–7</sup> Meanwhile, the unusual structure can also form during DNA repair or recombination.<sup>8–13</sup> In general, the unusual structure will be recognized and removed by mismatch repair (MMR), which is a post-replication repair system to maintain the fidelity of DNA replication. MMR proteins will search for structural artefacts such as non-Watson–Crick base pairs or unpaired nucleotides and then excise them. As a result, both the formation of unusual structure and repair escape must occur to bring about repeat expansions. For CCTG repeats, it has been shown that they can adopt a variety of different unusual structures including hairpin, dumbbell and mini-dumbbell (MDB).<sup>14–16</sup> Yet it remains unclear how these unusual structures escape from MMR. In this study, we present a novel mechanism, called the competing MDB mechanism, to explain how efficient MMR escape can occur via the formation of MDBs.

For the reported MDB formed by two CCTG repeats, it comprises of a 5'- and a 3'-type II loops.<sup>15,16</sup> A type II loop is defined by its distinctive folding geometry in which the first and fourth loop residues form the loop-closing base pair, the second

loop residue is positioned in the minor groove, whereas the third loop residue stacks on the loop-closing base pair.<sup>17</sup> In the 5'- and 3'-loops of this MDB, C1 and G4, and C5 and G8 form the two loop-closing base pairs (Figure 1a). C2 and C6 fold into the minor groove and form a mispair, whereas T3 and T7 stack on C1–G4 and C5–G8, respectively. Apart from the pairing geometry of the minor groove residues, the structural features of CCTG MDB are similar to those of another MDB formed by two TTTA repeats (Figure 1b).<sup>15,18</sup> Instead of forming a mispair, the minor groove residues T2 and T6 of TTTA MDB stack with each other. In these two MDBs, there are multiple stabilizing loop-loop interactions which are absent in larger dumbbell structures.<sup>14,19,20</sup> To understand how potentially CCTG MDB participates in MMR escape, we first consider the dynamic processes that have been observed in a DNA tract containing three CCTG repeats.<sup>14</sup> Instead of forming any MDB, two hairpin conformers were formed due to the stabilizing interactions between the first and third repeats (Figure 1c, top). During DNA replication, these interactions are expected to be less favorable in the nascent strand containing CCTG repeats because the 5'-end is hybridized with the template and end-fray occurs mainly at the 3'-end (Figure 1d). Through single-site substitution experiment to weaken these interactions between the first and third repeats, we demonstrated that an MDB could be formed (Figure 1c, bottom).<sup>16</sup> Thereby, in a segment of three repeats in a longer CCTG tract, there can be the formation of two possible types of MDBs, one containing a 5'-overhanging repeat and the other containing a 3'-overhanging repeat. Interestingly, these MDBs have been found in a DNA sequence containing three TTTA repeats and fast exchange between them makes the sequence behave like a mini-loop as evidenced by the unusually shifted <sup>31</sup>P signals of the second repeat at 25 °C (Figure 1e).<sup>18</sup> At lower temperatures, all <sup>1</sup>H and <sup>31</sup>P peaks were broadened, suggesting the presence of conformational exchange between two



**Figure 1.** The averaged structures of (a) CCTG MDB (PDB ID: 5GWL) and (b) TTTA MDB (PDB ID: 5GWQ). The two minor groove residues C2 and C6 form a mispair in CCTG MDB whereas T2 and T6 stack with each other in TTTA MDB. (c) Hairpins were formed in a sequence containing three CCTG repeats. Through single-site substitution experiment to weaken the stabilizing interactions between the first and third repeats, an MDB with a 5'-overhang was formed. (d) During DNA replication, end-fray occurs mainly at the 3'-end of the nascent strand and the 5'-end is hybridized with the template. Therefore, the 3'-terminal repeat cannot interact with its preceding repeats feasibly, thus hindering the formation of hairpin. (e) Two types of MDBs were formed in a sequence containing three TTTA repeats. Fast exchange between them made the sequence behave like a mini-loop. (f) A schematic diagram of the competing MDB mechanism. The mini-loop can shift in both the 5'- and 3'-directions.  $ML(i)$  cannot coexist with either  $ML(i-1)$  or  $ML(i+1)$ .

competing MDBs as this process involves the unfolding and refolding of the first and third repeats. If an hairpin was formed in this sequence, these signals would remain sharp due to a further stabilization of the hairpin at lower temperatures. Because there is no complication due to the formation of hairpin in TTTA repeats, here we employed longer sequences of TTTA repeats to serve as models to rationalize how CCTG repeat expansions occur via the competing MDB mechanism.

To elaborate the competing MDB mechanism, we first explain the concept of competing MDBs with a longer tract of repeats. In this longer tract, every two adjacent repeats are capable of forming an MDB (Figure 1f). For instances, the folding of the  $(i-1)$ th

and  $i$ th repeats results in the formation of  $MDB(i-1, i)$ , and the folding of the  $i$ th and  $(i+1)$ th repeats results in the formation of  $MDB(i, i+1)$ . If the formation of two MDBs requires the use of a common repeat, for example, both  $MDB(i-1, i)$  and  $MDB(i, i+1)$  require the use of  $i$ th repeat, these two MDBs will compete with each other and we call them the competing MDBs. Fast exchange between these competing MDBs will make the  $i$ th repeat behave like a mini-loop,  $ML(i)$ . If the 5'- or 3'-flanking repeat of these competing MDBs also participates in the formation of  $MDB(i-2, i-1)$  or  $MDB(i+1, i+2)$ , then fast exchange between (i)  $MDB(i-2, i-1)$  and  $MDB(i-1, i)$ , and (ii)  $MDB(i, i+1)$  and  $MDB(i+1, i+2)$  will result in the formation of  $ML(i-1)$  and  $ML(i+1)$ , respectively.

Thereby, a shifting of the mini-loop from the  $i$ th to  $(i-1)$ th or  $(i+1)$ th repeat will occur and this bi-directional shifting of the mini-loop forms the basis of the competing MDB mechanism.

## MATERIALS AND METHODS

### DNA samples

The sequences containing four to eight TTTA repeats were investigated in this study, and they were named as '(TTTA)<sub>4-8</sub>'. To investigate the effects of 5'- and 3'-flanking residues on the formation of MDB, we also designed two other sequences by adding an adenine residue to the 5'-terminal and a thymine residue to the 3'-terminal of two TTTA repeats, respectively. These sequences were named as 'A(TTTA)<sub>2</sub>' and '(TTTA)<sub>2</sub>T'. All DNA samples were synthesized using an Applied Biosystems model 394 DNA synthesizer (Applied Biosystems, Foster City, CA, USA). They were purified by denaturing polyacrylamide gel electrophoresis and diethylaminoethyl Sephacel anion exchange column chromatography, and finally desalted using Amicon Ultra-4 centrifugal filtering devices (Millipore Corporation, Billerica, MA, USA). Nuclear magnetic resonance (NMR) samples were prepared by dissolving 0.3 μmol purified DNA into 500 μl buffer solutions containing 10 mM sodium phosphate (pH 7.0), and 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid.

### NMR study

All NMR experiments were performed using Bruker AV-500 and/or AV-700 spectrometers (Bruker BioSpin AG, Faellanden, Switzerland). For studying the labile protons, the samples were prepared in a 90% H<sub>2</sub>O/10% D<sub>2</sub>O buffer solution. One-dimensional imino and two-dimensional (2D) nuclear Overhauser effect spectroscopy experiments were performed using excitation sculpting to suppress the water signal.<sup>21</sup> For studying the non-labile protons, the solvent was exchanged with a 99.96% D<sub>2</sub>O solution and a 2-s presaturation pulse was used to suppress the residual water signal. 2D nuclear Overhauser effect spectroscopy spectra were acquired with a data size of 4096 × 512 and a mixing time of 300 ms unless otherwise specified. The acquired data sets were zero-filled to give 4096 × 4096 spectra with a cosine window function applied to both dimensions. Backbone <sup>31</sup>P signals were assigned using 2D total correlation spectroscopy with a mixing time of 75 ms and <sup>1</sup>H-<sup>31</sup>P heteronuclear single-quantum coherence experiments. The <sup>31</sup>P spectral width was set to 6 p.p.m. and a data size of 4096 × 200 was collected. The <sup>31</sup>P chemical shifts were indirectly referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid using the derived nucleus-specific ratio of 0.404808636.<sup>22</sup>

### Native gel assay

Native gels composed of 25% polyacrylamide were prepared to investigate the oligomeric states of (TTTA)<sub>4-8</sub>. The gel loading samples were prepared in the same NMR buffer solution. For the reference lane, a DNA ladder composed of 10–100-nt single-strand oligomers was used. The polyacrylamide gel electrophoresis experiments were conducted at both ~25 and ~5 °C. To maintain the pH at 7.0, an electrophoretic buffer containing 9 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid), 20 mM bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane and 1 mM ethylenediamine-tetraacetic acid was used. DNA bands were visualized by post-staining the gels with stains-all solution.

## RESULTS AND DISCUSSION

Sequential assignments of (TTTA)<sub>4-8</sub>, A(TTTA)<sub>2</sub> and (TTTA)<sub>2</sub>T were made from the 2D nuclear Overhauser effect spectroscopy H6/H8-H1' fingerprint regions using standard methods<sup>23,24</sup> and the results are shown in Supplementary Figures S1–S4. Based on the H3' assignments from total correlation spectroscopy spectra, the <sup>31</sup>P assignments of (TTTA)<sub>4-8</sub>, A(TTTA)<sub>2</sub> and (TTTA)<sub>2</sub>T were made using <sup>1</sup>H-<sup>31</sup>P heteronuclear single-quantum coherence spectra and the results are shown in Supplementary Figures S5–S8.

(TTTA)<sub>4</sub> behaves like a mini-loop at the third repeat

In a DNA strand containing four TTTA repeats, three types of MDBs can be formed using the first two, the middle two and the last two repeats, respectively. These result in MDB(1,2) with a two-repeat

3'-overhang, MDB(2,3) with one repeat in each of the overhangs, and MDB(3,4) with a two-repeat 5'-overhang (Figure 2a). Based on the competing MDB mechanism, fast exchange between MDB(1,2) and MDB(2,3) makes the sequence behave like a mini-loop at the second repeat, ML(2). Similarly, fast exchange between MDB(2,3) and MDB(3,4) results in a mini-loop at the third repeat, ML(3). If ML(2) and ML(3) underwent slow exchange, the NMR spectral features of these two mini-loops would appear. If there were fast exchange between ML(2) and ML(3), there would be a loss of the mini-loop spectral features. Surprisingly, we observed only the NMR spectral features of ML(3) at 25 °C, for example, the unusual <sup>31</sup>P chemical shifts of T10 and T11 (Figure 2b), which are the NMR characteristics of type II TTTA loop.<sup>18,25</sup> As supported by the fact that most of the proton signals from the last three repeats were relatively more broadened than those from the first repeat (Figure 2c), competing MDBs seem to form more preferably in the last three repeats.

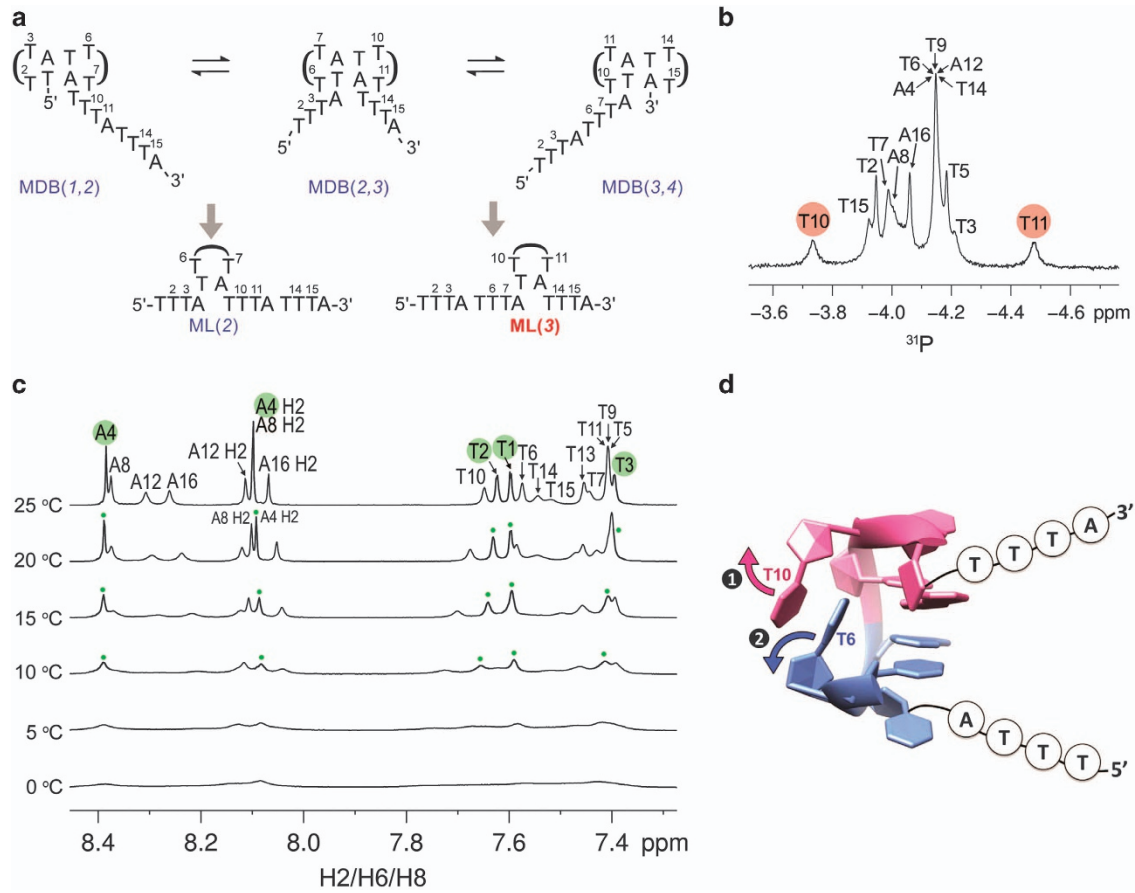
To rationalize the appearance of only ML(3) at 25 °C, we made use of the competing MDB mechanism. In (TTTA)<sub>4</sub>, ML(2) comes from the fast exchange between MDB(1,2) and MDB(2,3), whereas ML(3) comes from the fast exchange between MDB(2,3) and MDB(3,4). As the formation of both mini-loops involves the exchange with MDB(2,3), therefore the unfolding of MDB(2,3) has a critical role in these exchange processes. In the detailed TTTA MDB structure (Figure 1b), the minor groove residue T2 is sandwiched between the loop-closing base pairs and T6, making T2 being better stabilized in the minor groove than T6. Therefore, the unfolding of the 3'-loop (third repeat) is expected to occur before that of the 5'-loop (second repeat) in MDB(2,3) (Figure 2d). This will make the unfolded third repeat interact with the last repeat to form MDB(3,4) before the second repeat interacts with the first repeat to form MDB(1,2). As a result, MDB(2,3) will exchange more efficiently with MDB(3,4) than MDB(1,2), leading to the appearance of only ML(3). Upon lowering the temperature, conformational exchange between MDB(2,3) and MDB(3,4) became slower, promoting also the exchange between MDB(2,3) and MDB(1,2) and thus the formation of ML(2). As evidenced by the peak broadenings of all <sup>1</sup>H signals at 10 °C and below (Figure 2c), all four repeats were found to participate in the two MDB exchange processes, resulting in both ML(2) and ML(3). The presence of these two mini-loops suggests that the mini-loop can shift in both the 5'- and 3'-directions.

The presence of flanking repeats drives the unfolding of MDB

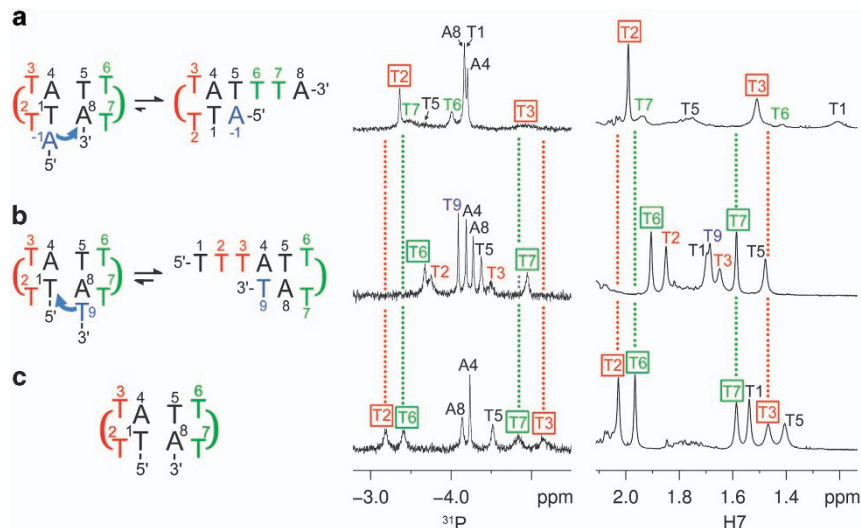
From the above analysis, the appearance of ML(3) in (TTTA)<sub>4</sub> at 25 °C was originated from the prior unfolding of the 3'-loop in MDB(2,3), which then interacted with the 3'-flanking repeat to form MDB(3,4). Upon slowing down the exchange at lower temperatures, the 5'-loop in MDB(2,3) became also available to interact with the 5'-flanking repeat to form MDB(1,2). To better understand the effects from the flanking repeats, we also determined the influences from the 5'- and 3'-nearest neighboring residues of TTTA MDB in this study.

For studying the 5'-nearest neighbor effect, we analyzed the NMR spectroscopic results of A(TTTA)<sub>2</sub> and found that the <sup>31</sup>P/H7 chemical shifts of T2 and T3 were close to those of TTTA MDB (Figure 3a). However, their T6 and T7 chemical shifts were quite different. These results suggest that the presence of the 5'-flanking A-1 does not affect much on the 5'-loop of TTTA MDB but it leads to the unfolding of the 3'-loop, probably due to A-1 competes with A8 to form a Watson–Crick base pair with T5. For (TTTA)<sub>2</sub>T, which shows the 3'-nearest neighbor effect, the <sup>31</sup>P/H7 chemical shifts of T6 and T7 remain close to those of TTTA MDB (Figure 3b). However, their T2 and T3 chemical shifts were very different. These suggest that the 3'-flanking T9 has no significant effect on the 3'-loop of TTTA MDB but it probably competes with T1 to form a Watson–Crick base pair with A4, thereby leading to the unfolding of the 5'-loop. The results of A(TTTA)<sub>2</sub> and (TTTA)<sub>2</sub>T suggest that the presence of 5'- or 3'-flanking residues facilitates the unfolding of MDB.





**Figure 2.** (a) Three types of MDBs can be formed in  $(TTTA)_4$  (top). Owing to steric clashes, the formation of MDB(1,2) disfavors the formation of MDB(3,4) and vice versa. Fast exchange between MDB(1,2) and MDB(2,3), and MDB(2,3) and MDB(3,4) results in ML(2) and ML(3), respectively. (b) At 25 °C, the unusually downfield T10 and upfield T11  $^{31}\text{P}$  signals suggest  $(TTTA)_4$  behaves like ML(3) with a rapid exchange between MDB(2,3) and MDB(3,4). (c) Peak broadenings were observed mainly from the proton signals belonging to the last three repeats. For easy identification, the sharper peaks from the first repeat were labeled in green. At lower temperatures, all proton signals were broadened, suggesting all four repeats were involved in conformational exchange processes. (d) Owing to T6 being sandwiched between T10 and the loop-closing base pairs, the 3'-loop unfolds before the 5'-loop in MDB(2,3), making the formation of MDB(3,4) occur before that of MDB(1,2).



**Figure 3.**  $1\text{D } ^{31}\text{P}$  and  $^1\text{H}$  NMR spectra of (a)  $A(TTTA)_2$ , (b)  $(TTTA)_2T$  and (c) TTTA MDB reference. The  $^{31}\text{P}/\text{H7}$  chemical shifts of T6 and T7 in  $A(TTTA)_2$ , and T2 and T3 in  $(TTTA)_2T$  are very different from those of TTTA MDB reference, suggesting the 3'-loop of  $A(TTTA)_2$  and the 5'-loop of  $(TTTA)_2T$  were unfolded, respectively. The 5'-flanking A-1 in  $A(TTTA)_2$  and the 3'-flanking T9 in  $(TTTA)_2T$  probably compete with A8 and T1 to form base pairs with T5 and A4, respectively. All these spectra were acquired at 10 °C.

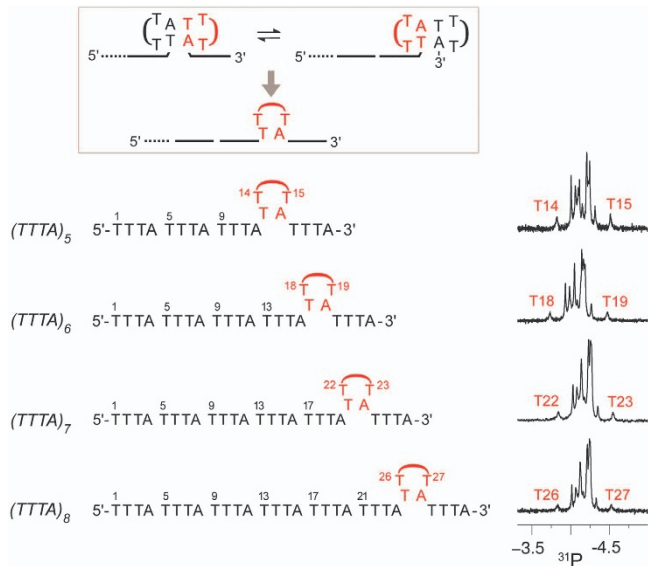
Preferential appearance of mini-loop at the 3'-penultimate repeat  
 To further demonstrate that the competing MDB mechanism provides a pathway for the unusual structure to shift to a different position along a tract of TTTA repeats, we also prepared longer

sequences containing five to eight TTTA repeats. Interestingly, all these sequences show a mini-loop at the 3'-penultimate repeat at 25 °C as supported by the unusually downfield or upfield signals of T14 and T15 in (TTTA)<sub>5</sub>, T18 and T19 in (TTTA)<sub>6</sub>, T22 and T23 in (TTTA)<sub>7</sub>, and T26 and T27 in (TTTA)<sub>8</sub> (Figure 4). The preferential 3'-penultimate repeat position can be rationalized by the sequential unfolding of the 3'- and 5'-loops that we encountered in (TTTA)<sub>4</sub>, further consolidating the presence of competing MDBs in these sequences.

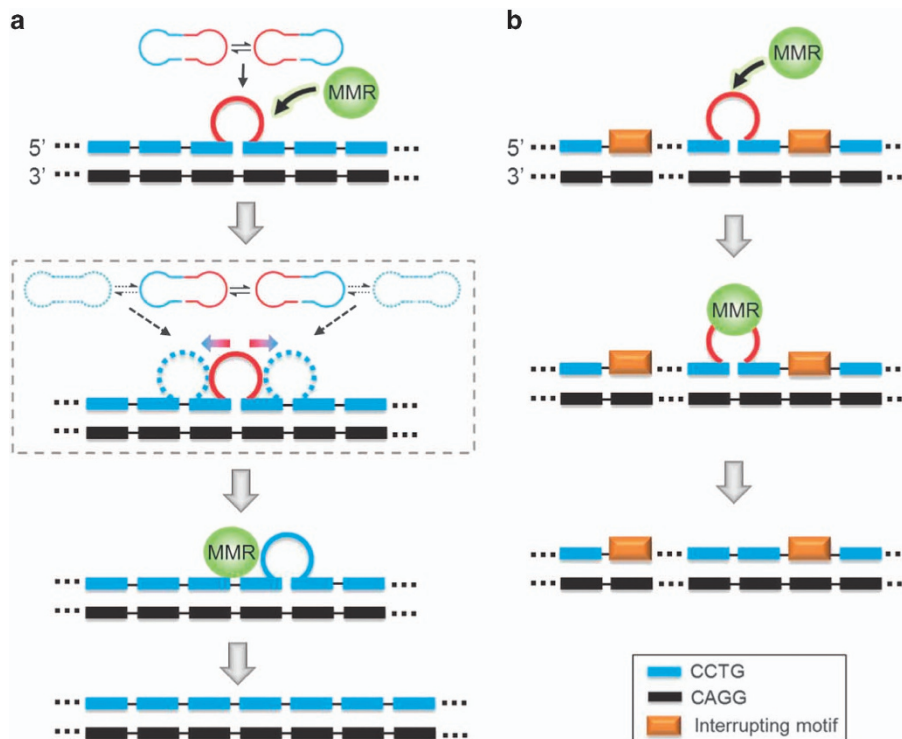
Upon lowering the temperature, conformational exchange between the two competing MDBs in the last three repeats became slower, making the formation of competing MDBs towards the 5'-direction more feasible. As supported by the broadenings of all <sup>31</sup>P signals at lower temperatures (Supplementary Figure S9), all repeats were found to participate in MDB exchange processes. Thereby, the resulting mini-loops are capable to shift in both the 5'- and 3'-directions. To verify that these peak broadenings were not resulting from the formation of multimeric conformers, we also performed native gel analysis. As the mobilities of these repeating sequences were found to be similar to those of single-strand references, these results suggest that all these sequences adopt monomeric conformations at both ~25 and ~5 °C (Supplementary Figure S10).

#### Competing MDBs contribute to efficient repair escape

The unusual structure formed in the nascent strand during DNA replication is usually recognized and removed by MMR proteins. Generally, MMR proteins make use of the local weakening due to mismatch/misalignment to locate the unusual structure.<sup>26</sup> As there is a conserved binding site in MMR proteins for the phosphate of unpaired or mismatched nucleotides,<sup>27-30</sup> it is expected that local structural changes of the mismatch/misalignment site will affect the recognition by MMR proteins, thus providing a possible pathway for MMR escape. In this study, our



**Figure 4.** The unusually downfield or upfield <sup>31</sup>P signals of T14 and T15 in (TTTA)<sub>5</sub>, T18 and T19 in (TTTA)<sub>6</sub>, T22 and T23 in (TTTA)<sub>7</sub>, and T26 and T27 in (TTTA)<sub>8</sub> suggest a mini-loop appears at the penultimate repeat of these sequences. The mini-loop is resulting from fast exchange between two competing MDBs formed in the last three repeats. The <sup>31</sup>P spectra shown here were acquired at 25 °C.



**Figure 5.** (a) In the uninterrupted CCTG repeat tract of DM2 patients, the mini-loop can shift in both directions, resulting in a more efficient escape from MMR. (b) In normal individuals, the presence of interrupting motifs in the CCTG repeat tract hinders the shifting of mini-loop. Thereby, the mini-loop can be recognized and removed more efficiently by MMR proteins.

results reveal that in a long tract of TTTA repeats, the mini-loop resulting from competing MDBs has a tendency to shift from the 5'- to 3'-direction at 25 °C because there is a sequential unfolding of the 3'- and 5'-loops in TTTA MDB. We believe the translational movement of the mini-loop along the repeat tract will provide a more efficient MMR escape pathway than that resulting from local structural changes. At lower temperatures, shifting of the mini-loop can happen in both the 5'- and 3'-directions due to slower exchange in competing MDBs. This bi-directional movement will further enhance the ability of the mini-loop to escape from MMR.

In CCTG MDB, there is no sequential unfolding of the 3'- and 5'-loops because the two minor groove residues do not stack with each other. Instead, they align in the same plane and form a mispair.<sup>15</sup> As a consequence, the mini-loop resulting from competing CCTG MDBs can shift in both the 5'- and 3'-directions. In the long and uninterrupted CCTG repeat tract of DM2 patients, intrinsically, the mini-loop formed in the nascent CCTG strand has an equal chance to shift towards the 5'- and 3'-directions. During DNA replication, the local environment of the nascent strand will keep changing due to changes in torsional stress resulting from supercoiling induced by replication activities such as the approaching and departure of different proteins.<sup>31</sup> Therefore, the shifting direction of the mini-loop will also be affected, making the mini-loop shift in an unpredictable direction along the CCTG repeat tract and thus enhancing the capability of the mini-loop to escape from MMR (Figure 5a). In normal individuals, intron 1 of the *CNBP* gene contains only a short CCTG repeat tract that is interrupted by A/G/TCTG motifs. As a consequence, the effect of bi-directional shifting of mini-loop will be limited and thereby the mini-loop can be recognized and removed by MMR more efficiently (Figure 5b).

## ACKNOWLEDGEMENTS

We thank Professor H.N.C. Wong for his continuing support on the research activities of our group. The work described in this paper was supported by General Research Fund (project no. CUHK14302114) from the Research Grants Council of the Hong Kong Special Administrative Region and a direct grant (project ID: 3132671) from the Faculty of Science of The Chinese University of Hong Kong.

## COMPETING INTERESTS

The authors declare no conflict of interest.

## REFERENCES

- Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL *et al.* Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 2001; **293**: 864–867.
- Bachinski LL, Czernuszewicz T, Ramagli LS, Suominen T, Shriver MD, Udd B *et al.* Premutation allele pool in myotonic dystrophy type 2. *Neurology* 2009; **72**: 490–497.
- Bachinski LL, Udd B, Meola G, Sansone V, Bassez G, Eymard B *et al.* Confirmation of the type 2 myotonic dystrophy (CCTG)<sub>n</sub> expansion mutation in patients with proximal myotonic myopathy/proximal myotonic dystrophy of different European origins: a single shared haplotype indicates an ancestral founder effect. *Am J Hum Genet* 2003; **73**: 835–848.
- Meola G, Cardani R. Myotonic dystrophies: an update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim Biophys Acta* 2015; **1852**: 594–606.
- Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 1993; **365**: 274–276.
- Richard GF, Paques F. Mini- and microsatellite expansions: the recombination connection. *EMBO Rep* 2000; **1**: 122–126.
- Mirkin SM. Expandable DNA repeats and human disease. *Nature* 2007; **447**: 932–940.
- Dere R, Wells RD. DM2 CCTG<sup>n</sup>CAGG repeats are crossover hotspots that are more prone to expansions than the DM1 CTG<sup>n</sup>CAG repeats in *Escherichia coli*. *J Mol Biol* 2006; **360**: 21–36.

- Savouret C, Brisson E, Essers J, Kanaar R, Pastink A, te Riele H *et al.* CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J* 2003; **22**: 2264–2273.
- van den Broek WJ, Nelen MR, Wansink DG, Coerwinkel MM, te Riele H, Groenen PJ *et al.* Somatic expansion behaviour of the (CTG)<sub>n</sub> repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet* 2002; **11**: 191–198.
- Manley K, Shirley TL, 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.
- Entezam A, Lokanga AR, Le W, Hoffman G, Usdin K. Potassium bromate, a potent DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X pre-mutation mouse model. *Hum Mutat* 2010; **31**: 611–616.
- Goula AV, Berquist BR, Wilson DM 3rd, Wheeler VC, Trottier Y, Merienne K. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet* 2009; **5**: e1000749.
- Lam SL, Wu F, Yang H, Chi LM. The origin of genetic instability in CCTG repeats. *Nucleic Acids Res* 2011; **39**: 6260–6268.
- Guo P, Lam SL. Minidumbbell: a new form of native DNA structure. *J Am Chem Soc* 2016; **138**: 12534–12540.
- Guo P, Lam SL. New insights into the genetic instability in CCTG repeats. *FEBS Lett* 2015; **589**: 3058–3063.
- van Dongen MJ, Mooren MM, Willems EF, van der Marel GA, van Boom JH, Wijmenga SS *et al.* Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A base pair in the loop. *Nucleic Acids Res* 1997; **25**: 1537–1547.
- Guo P, Lam SL. Unusual structures of TTTA repeats in *icaC* gene of *Staphylococcus aureus*. *FEBS Lett* 2015; **589**: 1296–1300.
- Ippel JH, Lanzotti V, Galeone A, Mayol L, Van den Boogaart JE, Pikkemaat JA *et al.* Thermodynamics of melting of the circular dumbbell d < pCGC-TT-GCG-TT >. *Biopolymers* 1995; **36**: 701–710.
- Erie D, Sinha N, Olson W, Jones R, Breslauer K. A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry* 1987; **26**: 7150–7159.
- Stott K, Stonehouse J, Keeler J, Hwang TL, Shaka AJ. Excitation sculpting in high-resolution nuclear magnetic resonance spectroscopy: application to selective NOE experiments. *J Am Chem Soc* 1995; **117**: 4199–4200.
- Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids—(IUPAC Recommendations 1998). *Pure Appl Chem* 1998; **70**: 117–142.
- Wijmenga SS, van Buuren BNM. The use of NMR methods for conformational studies of nucleic acids. *Prog Nucl Magn Reson Spectrosc* 1998; **32**: 287–387.
- Feigon J, Wright JM, Leupin W, Denny WA, Kearns DR. Use of two-dimensional NMR in the study of a double-stranded DNA decamer. *J Am Chem Soc* 1982; **104**: 5540–5541.
- Blommers MJ, van de Ven FJ, van der Marel GA, van Boom JH, Hilbers CW. The three-dimensional structure of a DNA hairpin in solution two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur J Biochem* 1991; **201**: 33–51.
- Natrajan G, Lamers MH, Enzlin JH, Winterwerp HH, Perrakis A, Sixma TK. Structures of *Escherichia coli* DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res* 2003; **31**: 4814–4821.
- Gupta S, Gellert M, Yang W. Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops. *Nat Struct Mol Biol* 2012; **19**: 72–78.
- Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. Structure of the human MutSalpha DNA lesion recognition complex. *Mol Cell* 2007; **26**: 579–592.
- Obmolova G, Ban C, Hsieh P, Yang W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 2000; **407**: 703–710.
- Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G×T mismatch. *Nature* 2000; **407**: 711–717.
- Levens D, Benham CJ. DNA stress and strain, *in silico*, *in vitro* and *in vivo*. *Phys Biol* 2011; **8**: 035011.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

Supplementary Information accompanies the paper on the *Signal Transduction and Targeted Therapy* website (<http://www.nature.com/sigtrans>)