Conformational distortions induced by periodically recurring A...A in d(CAG),d(CAG) provide stereochemical rationale for the trapping of MSH2.MSH3 in polyQ disorders

Yogeeshwar Ajjugal, Thenmalarchelvi Rathinavelan *

Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana State 502285, India

1. Introduction

Mismatch in the DNA occurs when two non-complementary bases erroneously align together and form a base pair (also known as non-canonical or non-Watson-Crick base pair) during the biological processes like DNA replication, recombination and spontaneous deamination etc. [1–3]. To maintain the genome integrity, the eukaryotic cells are equipped with sophisticated mismatch repair (MMR) proteins which recognize and correct the mismatched base pairs in the DNA [4]. MSH2.MSH6 (MutSβ) and MSH2.MSH3 (MutSβ) are the two heterodimeric complexes that play the prime role in the eukaryotic mismatch repair process [5]. While the former recognizes a single base mispair or 1–2 unpaired bases [6], the latter recognizes the insertion/deletion of 1–15 nucleotides (loops) as well as single base pair mismatches [7–9].

Polyglutamine diseases such as Huntington’s, several spinocerebellar ataxia etc. arise due to the expansion of a CAG repeat tract that encodes for a glutamine tract (polyQ) in the protein. The CAG repeat number lies in the range of 6–35 in the Huntingtonin (HTT) gene of the normal individuals. However, when the CAG repeat number expands beyond 35 in HTT gene, it leads to Huntington’s disease [10,11]. The mismatch repair MSH2.MSH3 protein complex is shown to have a major role in the expansion of CAG repeats [9]. The earlier recombination studies in yeast have shown that CAG/CTG repeats which tend to form stable hairpin structure have escaped from the repair pathway [12,13]. Indeed, it has been shown that the presence of A...A mismatch in the stem of the CAG repeat hairpin facilitates the binding of MSH2.MSH3 to

CAG repeat instability causes a number of neurodegenerative disorders. The unusual hairpin stem structure formed by the CAG repeats in DNA traps the human mismatch repair MSH2.MSH3 (MutSβ) complex. To understand the mechanism behind the abnormal binding of MutSβ with the imperfect hairpin stem structure formed by CAG repeats, molecular dynamics simulations have been carried out for Mutsβ-d(CAG),d(CAG),d(CTG),d(CTG) (1 A...A mismatch) and Mutsβ-d(CAG),d(CAG),d(CAG) (5 mismatches, wherein, A...A occurs periodically) complexes. The interaction of MSH3 residue Tyr245 at the minor groove side of A...A, an essential interaction responsible for the recognition by MutSβ, are retained in both the cases. Nevertheless, the periodic unwinding caused by the nonisostericity of A...A with the flanking canonical base pairs in d(CAG),d(CAG) distorts the regular B-form geometry. Such an unwinding exposes one of the A...A mismatches (that interacts with Tyr245) at the major groove side and also facilitates the on and off hydrogen bonding interaction with Lys546 sidechain (MSH2-domain-IV). In contrast, kinking of the DNA towards the major groove in Mutsβ-d(CAG),d(CAG),d(CTG),d(CTG) doesn’t facilitate such an exposure of the bases at the major groove. Further, the unwinding of the helix in d(CAG),d(CAG) enhances the tighter binding between MSH2-domain-I and d(CAG),d(CAG) at the major groove side as well as between MSH3-domain-I and MSH3-domain-IV. Markedly, such enhanced interactions are absent in Mutsβ-d(CAG),d(CAG),d(CTG),d(CTG) that has a single A...A mismatch. Thus, the above-mentioned enhancement in intra- and inter-molecular interactions in Mutsβ-d(CAG),d(CAG),d(CAG) provide the stereochemical rationale for the trapping of Mutsβ in CAG repeat expansion disorders.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
the hairpin and leads to CAG repeat expansion rather than performing the mismatch repair activity [9]. It has also been shown that more than one MSH2.MSH3 binds to expanded CAG hairpin indicating that the periodic occurrence of A..A mismatch acts as a multiple trapping point [9]. Thus, these suggest that the hairpin stem structure formed by expanded CAG repeat (with a periodic occurrence of A..A mismatch in the hairpin stem) acts as a key factor in misguiding the MSH2.MSH3 complex through the establishment of a strong binding between them [14]. However, the underlying mechanism behind such a tight binding between CAG repeat hairpin and the MSH2.MSH3 complex is unknown.

To derive the atomistic insights about the aforementioned tighter binding between the expanded CAG repeat and MSH2.MSH3 complex, molecular dynamics (MD) simulations have been carried out for MSH2.MSH3-d(CAG)2(CAG)2.d(CTG)2(CTG)2 (1 mismatch, Mutsb-CAG-1AA) and MSH2.MSH3-d(CAG)5.d(CAG)5 (5 mismatches, wherein, A..A occurs periodically, Mutsb-CAG-5AA). MD simulations indicate that Tyr245 (MSH3) interacts at the minor groove of the mismatch site, the essential interactions for the recognition, as also seen in the crystal structures (PDB ID: 3THX, 3THY, 3THZ and 3THW). Interestingly, the local distortions induced by the A..A mismatch due to its nonisostericity with the flanking canonical C..G and G..C base pairs facilitate such interactions and lead to bending in the DNA duplex. To our surprise, the periodic unwinding of the helix at the A..A mismatch in d(CAG)5.d(CAG)5 leads to an enhancement in the interaction within Mutsb complex as well as with the DNA substrate. Such enhanced interactions are not found in the case of d(CAG)2(CAG)2.d(CTG)2(CTG)2 with a single A..A mismatch. Thus, the tighter binding seen in MSH2.MSH3-d(CAG)5.d(CAG)5 complex, perhaps, is the reason behind the trapping of MSH2.MSH3 in the polyQ disorders.

2. Methods

2.1. Molecular dynamics simulation protocol

The MSH2.MSH3 (MutSb) complex in the crystal structure (PDB ID: 3THX, Fig. 1) was used to dock with 3 different 15-mer DNA substrates (Schemes (Table 1)) used in current investigation. The 15-mer DNA CAG duplex models, namely, CAG-1AA (has a single A..A) and CAG-5AA (has five A..A mismatches) obtained from the previous molecular dynamics (MD) simulations [15] were used as the starting models. However, CAG-WC (has only the canonical base pairs) was modeled using 3D-NuS web tool [16]. It is noteworthy that the native DNA duplex in MutSb-DNA crystal structure was replaced with the above-mentioned DNA duplexes in the respective simulation systems. Since some of the residues of MSH2 and MSH3 subunits were missing in the crystal structure, they were modeled using ModLoop web server [17]: 108–111, 137–144, 315–323, 518–519, 646–647, 820–836 residues of MSH2 and, 135–136, 160–168, 262–275, 315–323, 518–519, 724–733 residues of MSH3. Subsequently, MutSb-CAG-1AA, MutSb-CAG-5AA and MutSb-CAG-WC complexes were generated manually. In all the schemes, adenosine di phosphate (ADP) was retained in the ATPase domain of MSH2 as found in the crystal structure. Subsequently, these models were subjected to molecular dynamics simulations using pmemd.cuda module of AMBER16 suit [18]. The OL15 and ff14SB force field were used for the DNA [19] and the protein [20] respectively. The force field for ADP was taken from the AMBER parameter database (http://amber.manchester.ac.uk/). All the systems were explicitly solvated with TIP3P water box and Na+ counter ions were added to neutralize the system and a 10 Å cut-off was used for the non-bonded interactions. The long range electrostatic interactions were taken into account by Particle Mesh Ewald method [21] and the SHAKE algorithm was applied to

Table 1

| S.No | Scheme         | Protein | DNA                                                                 |
|------|----------------|---------|----------------------------------------------------------------------|
| 1    | MutSb-CAG-1AA  | MSH2,MSH3 | 5’ C1 A2 G3 C4 A4 G5 C6 A6 G7 A8 C9 A10 G11 A12 C13 G14 3’   |
| 2    | MutSb-CAG-5AA  | MSH2,MSH3 | 5’ C1 A2 G3 C4 A4 G5 C6 A6 G7 A8 C9 A10 G11 A12 C13 G14 3’   |
| 3    | MutSb-CAG-WC   | MSH2,MSH3 | 5’ C1 A2 G3 C4 A4 G5 C6 A6 G7 A8 C9 A10 G11 A12 C13 G14 3’   |

Fig. 1. Cartoon representation of the crystal structure of MSH2.MSH3 and a DNA (colored in cyan) having a bulge (PDB ID: 3THX). Note that the different domains of MSH2 and MSH3 are colored differently.
constrain bonds involving hydrogen atoms. A 2 fs time step was used during the simulation. All the systems were equilibrated for 50 ps (using a NVT ensemble) followed by a 500 ns production run with a NPT ensemble, wherein P was kept at 1 atm. During the equilibration run, the solute and the solvent were slowly relaxed in several steps as described in earlier studies [15,22–28].

2.2. Trajectory analysis

The root mean square deviation (RMSD) and protein...DNA interaction analysis of the MD trajectories were calculated using cpptraj module [29] of AMBER suite. GNUPLOT [30] software was used for plotting the data. The Pymol [31] and VMD [32] tools were used for the visualization of the trajectories.

2.3. Binding energy estimation

The gas phase binding energies of MSH2 and MSH3 interaction as well as MutSβ-CAG-1AA and MutSβ-CAG-5AA complexes (of schemes MutSβ-CAG-5AA and MutSβ-CAG-1AA) were calculated using the last 50 ns MD trajectories with a frame size of 25 ps. Note that the terminal 2 residues on both the sides of the DNA duplexes were ignored due to end fraying effect. AMBER suite was employed for the calculation [18]. The end-point binding energies (ΔE_{BE}) between the DNA substrate and MSH2,MSH3 as well as between MSH2 and MSH3 were independently extracted through post-processing the MD trajectories of schemes MutSβ-CAG-5AA and MutSβ-CAG-1AA using the following equations:

$$\Delta E_{BE} = \Delta E_{complex} - (\Delta E_{receptor} + \Delta E_{ligand})$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW}$$

Note that the energy (ΔE_{MM}) of the complex (ΔE_{complex}), receptor (ΔE_{receptor}) and ligand (ΔE_{ligand}) were estimated using the bond distance, bond angle and dihedral energy terms (ΔE_{int}) as well as using van der Waals (ΔE_{vdw}) and electrostatic (ΔE_{ele}) energy components obtained from the respective gas phase energy minimized trajectories. However, ΔE_{BE} is mainly contributed by ΔE_{vdw} and ΔE_{ele} as ΔE_{int} component becomes zero.

3. Results

The MD simulations of MutSβ-CAG-1AA (DNA having a single A...A mismatch) and MutSβ-CAG-5AA (DNA having five A...A mismatches) indicate that the complex attains a root mean square deviation (RMSD) of 4–5 Å quite early during the simulation (less than 10 ns) (Supplementary Fig. S1). Since the MutSβ amino acids surrounding the DNA are rich in arginine and lysine, they are involved in salt-bridge/hydrogen bonding interactions with the DNA backbone (Supplementary Fig. S2). These are non-specific interactions and are seen both in MutSβ-CAG-1AA and MutSβ-CAG-5AA, but with a difference in their interaction patterns due to the difference in the conformation of the substrates. Similarly, several nonspecific interactions are observed between the protein and the substrate DNA backbone. Intriguingly, several base specific interactions are observed in MutSβ-CAG-1AA and MutSβ-CAG-5AA which lead to differences in their interaction patterns as discussed below.

3.1. Tyr245 and Lys246 interactions at the A...A mismatch site lead to a kink in CAG-1AA

Detailed analysis of the CAG-1AA duplex of the MutSβ-CAG-1AA complex indicates that A8 and A23 disengage themselves from the hydrogen bonding interaction quite early during the simulation and continues in the same fashion till the end of the simulation (Fig. 2A). These adenines move out of plane with respect to each other and facilitate the interaction with the MSH3 through the formation of A23(N7)...Tyr245(O) and A23(N6)...Tyr245(O) as well as A8(O4)...Tyr245(N) (Fig. 2B) hydrogen bonds (Fig. 2C). The -syn glycosyl conformation of A23 exposes N6 and N7 to the minor groove side and facilitates this interaction. A previous mutagenesis study

![Fig. 2. MutSβ interaction with the CAG-1AA at the mismatch site. (A) Time vs hydrogen bond distance plot showing the complete loss of hydrogen bond between the mismatched A8 and A23 in the CAG-1AA substrate of MutSβ-CAG-1AA complex. (B) Time vs hydrogen bond distance plot showing the formation of A23(N7)...Tyr245(O), A23(N6)...Tyr245(O), A8(O4)...Tyr245(N), A23(N7)...Lys246(NZ) and A8(N3)...Lys246(NZ) hydrogen bonds. (C, D) Snapshot showing (C) the interaction of Tyr245 with A8 and A23 and, (D) the kink at the mismatch site of the DNA substrate at 500 ns. (E) Snapshot illustrating the interaction of Arg313(MSH3) to a base of the substrate (500 ns).](https://example.com/fig2.png)
has also shown the importance of Tyr245 (equivalent to Tyr157) in MSH2.MSH3 mediated mismatch repair activity in *Saccharomyces cerevisiae* [33]. Although Lys246 of MSH3 interacts with the mismatch through the formation of A23(N7)…Ly s246(NZ) and A6(N3)…Ly s246(NZ) hydrogen bonds, the interactions are transient in nature (Fig. 2B). The conformational flexibility seen at the mismatch site and the associated interactions with the protein molecule lead to a kink in the DNA duplex (Fig. 2D). Besides these, a few other amino acids are also found to interact with the DNA bases of CAG-1AA substrate. For instance, at the major groove side, Arg313 (MSH3) is involved in hydrogen bonding with the substrate base (Fig. 2E).

### 3.2. Periodic A…A mismatch in CAG-5AA tightens the interaction between MutSb and CAG-5AA

In line with the above, Tyr245 interacts (which is crucial for the mismatch recognition) with the central A8…A23 mismatch in CAG-5AA albeit the nature of interaction is different from CAG-1AA. In the first place, A8…A23 hydrogen bond is retained majority of the time during the simulation through N6(A8)…N7(A23) hydrogen bond (Fig. 3A) unlike in the previous case (Fig. 2A). Further, N7(A8) is also engaged in intermittent hydrogen bond formation with Ly s546(MSH2) side chain during the simulation (Fig. 3B, C (Right)). Such interactions are facilitated through the movement of A23 (-syn glycosyl conformation) towards the major groove. Further, Tyr245 (MSH3) is also engaged in N3(A8)…Tyr245(O) hydrogen bonding interaction (Fig. 3B, C (Left)). Among the other 2 A…A mismatches (A5…A26 and A11…A20) present in the helix (Note that the remaining two are ignored due to the end fraying effect, Table 1), A5…A26 retains the N6…N7 hydrogen bond (Fig. 3D). Nonetheless, A11…A20 hardly retains the hydrogen bond during the simulation (Fig. 3E). To our surprise, unwinding of the helix at A5…A26 exposes the N6 atom of A26 towards the major groove side, facilitating a strong interaction with the MSH2-domain-I mediated by a Na⁺ counter ion around 325 ns of the simulation (Fig. 4). Accompanied by the movement of A26 towards the major groove, Asp41 and Phe42 of MSH2-domain-I form a Na⁺ coordination network with A26 and, the flanking G…C and C…G base pairs. This eventually, enhances the interaction between the DNA binding domain of MSH2 with the duplex. In line with this, a previous study has pointed out that the deletion of MSH2-domain-I in *Saccharomyces cerevisiae* showed defect in MSH2.MSH3 mediated mismatch repair activity [34].

![Fig. 3. MutSb interaction with the CAG-5AA substrate. A) Time vs hydrogen bond distance plots corresponding to (A) A6(N7)…A6(N6), (B) Ly s546(NZ)…A6(N7) and Tyr245(O)…A6(N3). Note the on and off interaction of Ly s546 and Tyr245 with A6 can occur either simultaneously or individually. C) Snapshots showing the simultaneous Tyr245(O)…AR(N3) (minor groove) and Ly s546(NZ)…A6(N7) (major groove) hydrogen bond formation at 215 ns. (D, E) Hydrogen bond distance plot corresponding to D) A5…A26 and E) A11…A20. Note the short residence time of hydrogen bonds in (E).](image-url)
3.3. Enhancement in the interaction between domain-I and domain-IV of MSH3 in concomitance with the conformational dynamics of periodically occurring A...A mismatch

Strikingly, the periodic occurrence of 5 A...A mismatches in MutS\textsubscript{I}-CAG-5AA influences the interaction among the different domains of MutS\textsubscript{I}. For instance, the MSH3-domain-I (loop region, residue number 298–323) and MSH3-domain-IV (loop region, residue number 730–745) come in close proximity in MutS\textsubscript{I} (Fig. 5A, Movie S1) that are far away from each other in the crystal structure (Fig. 5C). These 2 domains interact through hydrophobic interactions. Thus, these bring compactness in MutS\textsubscript{I}-CAG-5AA complex.

Further, MD simulations carried out by considering d(CAG)\textsubscript{5}.d(CAG)\textsubscript{4} (CTG)\textsubscript{3} duplex (wherein, only canonical base pairs are present) as a substrate for MSH2.MSH3 (Scheme MutS\textsubscript{I}-CAG-WC, Table 1) indicate that duplex doesn’t undergo any structural deformations as seen in the cases of MutS\textsubscript{I}-CAG-5AA and MutS\textsubscript{I}-CAG-1AA. This can be clearly seen in the root mean square deviation (RMSD) of the DNA duplex, which falls around 2 Å (Supplementary Fig. S3). In contrast, the RMSD of MutS\textsubscript{I}-CAG-5AA and MutS\textsubscript{I}-CAG-1AA falls around 4 Å (Supplementary Fig. S1).

3.4. Binding energy estimation

The gas phase binding energy estimated for the MutS\textsubscript{I}-CAG-1AA and MutS\textsubscript{I}-CAG-5AA complexes indicate that the electrostatic energy contribution is favored in the case of the latter compared with the former (Table 2). The electrostatic component of MutS\textsubscript{I}-CAG-5AA complex (−1062.3 k.cal.mol\textsuperscript{−1}) is more favorable compared with MutS\textsubscript{I}-CAG-1AA (−964.7 k.cal.mol\textsuperscript{−1}). In contrast, the van der Waals energy component is more favorable for MutS\textsubscript{I}-CAG-1AA (−154.8 k.cal.mol\textsuperscript{−1}) compared to MutS\textsubscript{I}-CAG-5AA (−122.9 k.cal.mol\textsuperscript{−1}). However, due to a highly favorable electrostatic energy contribution in the case of MutS\textsubscript{I}-CAG-5AA, the gas phase binding energy of MutS\textsubscript{I}-CAG-5AA complex (−1185.3 k.cal.mol\textsuperscript{−1}) is more (about −65 k.cal.mol\textsuperscript{−1}) favorable than MutS\textsubscript{I}-CAG-1AA complex (−1119.5 k.cal.mol\textsuperscript{−1}). Further, the gas phase binding energy (calculated by considering MSH2 as the receptor and MSH3 as the ligand) of MSH2 and MSH3 interaction clearly indicates that CAG-5AA (−1659.7 k.cal.mol\textsuperscript{−1}) enhances the interaction between the two compared to CAG-1AA (−1509.4 k.cal.mol\textsuperscript{−1}). The electrostatic component is the key factor in causing the difference in the gas phase binding energy of MSH2 and MSH3 interaction in the cases of CAG-5AA and CAG-1AA substrates (Table 3). Thus, these results indicate that interaction between MutS\textsubscript{I} and CAG-5AA is more favorable than MutS\textsubscript{I} and CAG-1AA.

4. Discussion

The occurrence of a non-canonical A...A mismatch in the CAG repeat DNA and RNA duplexes plays an important role in the polyglutamine diseases [35,36]. Unlike the other 7 non-canonical base pairs (C...C, T...T, G...G, G...T, A...C, T...C and G...A) [16], the structural insights about an A...A mismatch in the midst of the canonical base pairs in a DNA is not well understood due to its inaccessibility to any experimental technique. Although one can envisage that the occurrence of any non-canonical base pair in the midst of the canonical base pairs may lead to conformational distortions, earlier NMR [37–39] and recent molecular dynamics simulation [15,24,26,27] studies have indicated that the conformational distortions are quite significant in the case of an A...A mismatch. Such a characteristic of an A...A mismatch can readily be attributed to the degree of nonisomorphism which is quite prominent in the case of an A...A mismatch [40]. This eventually leads to spontaneous and frequent conformational transitions when an A...A mismatch is present in a DNA duplex [15,24,26,27]. However, such conformational transitions are absent in the G...G mismatch present in a DNA duplex [41]. Such a differential influence imposed by the A...A and G...G mismatches can readily be attributed to the...
difference in the extent of base pair nonisomorphism between the two [40]. To explore the influence of such A conformational dynamics in trapping the mismatch repair MSH2.MSH3 complex in polyQ diseases, MD simulations of MSH2.MSH3 (MutSβ) in complex with 2 different DNA substrates have been carried out. While one of the substrates has a single A...A mismatch (MutSβ-CAG-1AA), the other has 5 A...A mismatches (MutSβ-CAG-5AA).

While the essential interaction responsible for the recognition and repair of A...A mismatch is retained in both the complexes (Fig. 2B &3C), the nature of interaction is different between the two cases. To our surprise, in the case of MutSβ-CAG-5AA, one of the A...A mismatches is involved in Na+ mediated coordination with the MSH2-domain-I. Such interaction is absent in MutSβ-CAG-1AA. The non-isostercity of the A...A mismatch (having a larger diameter compared to the canonical base pairs) [15,40] with the flanking canonical base pairs unwinds the helix and pushes one of the adenines towards the major groove, facilitating the abovementioned interaction (Fig. 6A). The presence of the canonical base pairs at the equivalent position in MutSβ-CAG-1AA doesn’t expose the base pairs towards the major groove, resulting in the absence of such interaction (Fig. 6B) as also seen in the crystal structure (Fig. 6C). Intriguingly, the periodic unwinding of the DNA substrate at every A...A mismatch site in MutSβ-CAG-5AA leads to a smooth bending (Fig. 6A), whereas, a single A...A mismatch in the middle of the DNA substrate in MutSβ-CAG-1AA results in a kink (Fig. 6B). In fact, the kink in
CAG-1AA towards the major groove prevents the access of the bases to the protein unlike in the case of CAG-5AA. These also lead to significant conformational differences even within the MutSb complex of the schemes MutSb-CAG-5AA and MutSb-CAG-1AA. For instance, the conformational changes in MutSb-CAG-5AA bring compactness between the domains I and IV of MSH3 (Fig. 5) (Movie S1). Although the crystal structure of the A...A mismatch site in complex with human MutSb is not available, the DNA substrate of the E. coli MutS has a A...A mismatch (PDB ID: 1OH6) and it resembles the kink seen in the MutSb-CAG-1AA (Fig. 6C). Further, the conformational distortions seen at the A...A mismatch site of the crystal structure resembles the MD derived structures. Thus, these results clearly pinpoint that the nonisostericity mediated conformational rearrangements in the A...A mismatch leads to an unwinding of the helix at the mismatch site. This further results in a smooth bending in the DNA duplex having a CAG repeat (wherein, A...A occurs periodically). It is noteworthy that the loop region of the hairpin may have some influence on the stem of the hairpin. However, it may not significantly alter the local conformational distortions induced by the A...A mismatch at the MSH2.MSH3 binding site of the DNA duplex. Further, many such tighter binding is expected between MSH2.MSH3 and the DNA substrate in the case of a longer CAG tract, as it has been reported earlier that more than one MSH2.MSH3 binds to the CAG tract [9].
5. Conclusions

The MD simulations carried out here to explore the influence of the conformational distortions induced by the periodically recurring A...A mismatch in trapping the MutSβ complex in a CAG repeat indicate that the mismatch tightens the interaction not only between the DNA and MutSβ, but also within the domains of MutSβ. The extent of base pair nonisomorphism, which mainly arises from the difference in the diameters of the A...A and conical base pairs, is found to be the origin of such tighter binding as it unwinds the helix and exposes the mismatched adenosines towards either the major or the minor groove. As an earlier experimental investigation has revealed that more than one MutSβ binds with the expanded CAG repeat [9], one can envisage many such tighter binding of MutSβ in different regions of the expanded CAG repeats may influence the trapping of MutSβ as well as the associated recruitment of other proteins involved in the mismatch repair. Thus, this investigation provides the stereochemical rationale for the trapping of MutSβ in polyQ disease. Cryo-electron microscope experiments can further provide a detailed picture about the interaction between longer CAG tracts and multiple MutSβ.

Author contributions

YA carried out the project. YA and TR wrote the manuscript. TR conceptualized and supervised the project.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the National PARAM Supercomputing Facility, Government of India and Indian Institute of Technology Hyderabad for providing the computational facility.

Funding

The work was supported by the Department of Biotechnology, Government of India: IYBA-2012 (D.O.No.BT/06/IYBA/2012) (To TR), Bio-CalRE (SAN.No.102/IFD/SAN/1811/2013-2014) (To TR), R&D (SAN.No.102/IFD/SAN/3426/2013-2014) (To TR), BIRAC-SRISTI (PMU_2017_010) (To TR), BIRAC-SRISTI (PMU_2019_007) (To YA and TR) and Indian Institute of Technology Hyderabad (To TR), The Ministry of Education, Government of India has provided the fellowship to YA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbsij.2021.07.018.

References

[1] Jiricny J. Postreplicative mismatch repair. Cold Spring Harbor Perspect Biol 2013;5(4):a012633.
[2] Suresse J, Argueso J, Alani E. Mismatch repair proteins: key regulators of genetic recombination. Cytogenetic and Genome Research 2004;107(3–4):146–59.
[3] Holliday R, Grigg GW. DNA methylation and mutation. Mutat Res/Fundam Mol Mech Mutagenesis 1993;285(1):61–7.
[4] Fukui K. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids 2010;2010:1–16.
[5] Kolodner RD, Marisichky GT. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev 1999;9(1):89–96.
[6] Warren JJ, Poliharu TJ, Changela A, Iyer RR, Modrich PL, Beese L. Structure of MutSα homologs, MutSα and MutSβ. Biophy J 2014;106(11):2483–92.
[7] Sharma M et al. Differential mismatch recognition specificities of eukaryotic MutS homologs, MutSα and MutSβ. Biophy J 2014;106(11):2483–92.
[8] Harrington JM, Kolodner RD. Saccharomyces cerevisiae Msh2-Msh3 acts in mendelian trinucleotide repeat base-base mispair resolution. Mol Cell Biol 2002;22(8):2654–6.
[9] Owen BAL, Yang Z, Liu C-P, Arnheim N, Petes TD. Duplex repeat forms secondary structures that escape DNA repair in yeast. Proc Natl Acad Sci USA 1999;96(4):1504–9.
[10] Snaid R, MacMillan JC, Chand J, Pienton I, Lazarou LP, Davies P, et al. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington’s disease. Nat Genet 1993;4(4):393–7.
[11] Cummings C, Zoghbi H. Trinucleotide repeats: mechanisms and pathologies. Annu Rev Physiol 2000;62(1):281–328.
[12] Moore H, Greenwell PW, Liu C-P, Arnheim N, Petes TD. Duplex repeat forms secondary structures that escape DNA repair in yeast. Proc Natl Acad Sci USA 1999;96(4):1504–9.
[13] Miret JJ, Pessoa-Brandao L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTCG/CAG trinucleotide repeats in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 1998;95(21):12438–43.
[14] Lang WH, Coats JE, Majka J, Hura GL, Lin Y, Rasnik I, et al. Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. Proc Natl Acad Sci USA 2011;108(42):E837–44.
[15] Khan N, Kolimi N, Rathinavelan T, Mackerrall A. Twisting right to left: A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. PLoS ONE 2015;10(4):e0100416.
[16] Patro LLP, Kumar A, Kolimi N, Rathinavelan T. 3D-NuS: a web server for automated modeling and visualization of non-canonical 3-dimensional nucleic acid structures. J Mol Biol 2017;429(16):2438–48.
[17] Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. Bioinformatics 2003;19(18):2500–1.
[18] Case D et al. Amber 16. San Francisco: University of California; 2016.
[19] Zgarbóvá M, Sponer J, Otyepka M, Chetham TE, Salindo-Murillo R, Jurečka P. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. J Chem Theory Comput 2015;11(12):5723–36.
[20] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Smilinger CM. F14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;11(8):3696–713.
[21] Szostak JW, Caulfield MP. A specific particle mesh Ewald method. J Chem Phys 1995;103(19):8577–93.
[22] Rathinavelan T, Yathindra N. Molecular dynamics structures of peptide nucleic acid hybrid in the wild-type and mutated alleles of Ki-ras prototype oncogene: Stereoechemical rationale for the low affinity of PNA in the presence of an A...C mismatch. FEBS J 2005;272(16):4055–70.
[23] Rathinavelan T, Yathindra N. Base triplet nonisomorphism strongly influences DNA triplex conformation: effect of nonisomorphous G-GC and A-AT triplets and bending of DNA triplexes. Biopolymers: Original Res Biomol 2006;82(5):443–61.
[24] Kolimi N, Ajugal Y, Rathinavelan T. AB-Z junction induced by an A...A mismatch in GAC repeats in the gene for cartilage oligomeric matrix protein promotes binding with the hZdADAR1 protein. J Biol Chem 2017;292(46):18732–46.
[25] Goldsmith G, Rathinavelan T, Yathindra N, Salisbury F. Selective preference of parallel DNA triplexes is due to the disruption of Hoogsteen hydrogen bonds caused by the severe nonisostericity between the G* GC and T* AT triplets. PLoS ONE 2016;11(3):0152102.
[26] Ajugal Y, Rathinavelan T. Sequence dependent influence of an A...A mismatch in a DNA duplex: an insight into the recognition by hZdADAR1 protein. J Struct Biol 2021;213(1):107678.
[27] Ajugal Y, Tomar K, Rao DK, Rathinavelan T. Spontaneous and frequent conformational dynamics induced by A...A mismatch in d(CAA). d(TAG) duplex. Sci Rep 2021;11(1):1–18.
[28] Themnalarveli R, Yathindra N. New insights into DNA triplexes: residual twist and radial differences as measures of base pair triple non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. Nucleic Acids Res 2005;33(1):43–53.
[29] Roe DR, Chetham TE, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 2013;9(7):3084–95.
[30] Williams T et al., gnuplot 5.2; 2017.
[31] DeLano WL. The PyMOL Molecular Graphics System. CCP4 Newsletter Protein Crystallography 2002;40:82–92.
[32] Humphrey W, Dalke A, Schulten K. Visual molecular dynamics. J Mol Graph 1996;14(1):33–8.
[33] Nunes JM, Putnam CD, Kolodner RD. Functional studies and homology modeling of Msh2-Msh3 predict that mispair recognition involves DNA binding and strand separation. Mol Cell Biol 2010;30(13):3321–8.
[34] Lee SD, Suresse J, Alani E. Saccharomyces cerevisiae MSH2–MSH3 and MSH2– MSH6 complexes display distinct requirements for DNA binding domain 1 in mismatch recognition. J Mol Biol 2007;366(1):53–66.
[35] Kilszik A, Kierzek R, Krzyzowski WJ, Rynpiwski W. Atomic resolution structure of CAG RNA repeat: structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Res 2010;38(22):8370–6.
[36] Chan HYE. RNA-mediated pathogenic mechanisms in polyglutamine diseases and amyotrophic lateral sclerosis. Front Cell Neurosci 2014;8:431.

[37] Arnold FH, Wolk S, Cruz P, Tinoco I. Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes d (CCCCAGGG) 2 and d (CCCTGGG) 2. Biochemistry 1987;26(13):4068–75.

[38] Maskos K, Gunn BM, LeBlanc DA, Morden KM. NMR study of G, cndot, A and A, cndot. A pairing in (dGCGAATAAGCG). Biochemistry 1993;32 (14):3583–95.

[39] Gervais V, Cognet JAH, Bret M, Sowers LC, Fazakerley GV. Solution structure of two mismatches A, cndot, A and T, cndot, T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. Eur J Biochem 1995;228(2):279–90.

[40] Ananth P, Goldsmith G, Yathindra N. An innate twist between Crick’s wobble and Watson-Crick base pairs. RNA 2013;19(8):1038–53.

[41] Ajjugal Y, Kolimi N, Rathinavelan T. Secondary structural choice of DNA and RNA associated with CGG/CCG trimucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. Sci Rep 2021;11(1):1–17.