



Communications

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



Yogeeswar Ajjugal, Thenmalarchelvi Rathinavelan*

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

ARTICLE INFO

Article history:

Received 26 February 2021

Received in revised form 16 July 2021

Accepted 21 July 2021

Available online 24 July 2021

Keywords:

CAG repeat expansion

Polyglutamine diseases

MSH2.MSH3

Molecular dynamics simulations

A...A mismatch

ABSTRACT

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 (Muts β) complex. To understand the mechanism behind the abnormal binding of Muts β with the imperfect hairpin stem structure formed by CAG repeats, molecular dynamics simulations have been carried out for Muts β -d(CAG)₂(CAG)(CAG)₂.d(CTG)₂(CAG)(CTG)₂ (1 A...A mismatch) and Muts β -d(CAG)₅.d(CAG)₅ (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 Muts β , 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)₅.d(CAG)₅ 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 Muts β -d(CAG)₂(CAG)(CAG)₂.d(CTG)₂(CAG)(CTG)₂ doesn't facilitate such an exposure of the bases at the major groove. Further, the unwinding of the helix in d(CAG)₅.d(CAG)₅ enhances the tighter binding between MSH2-domain-I and d(CAG)₅.d(CAG)₅ at the major groove side as well as between MSH3-domain-I and MSH3-domain-IV. Markedly, such enhanced interactions are absent in Muts β -d(CAG)₂(CAG)(CAG)₂.d(CTG)₂(CAG)(CTG)₂ that has a single A...A mismatch. Thus, the above-mentioned enhancement in intra- and inter- molecular interactions in Muts β -d(CAG)₅.d(CAG)₅ provide the stereochemical rationale for the trapping of Muts β 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/>).

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 (Muts α) and MSH2.MSH3 (Muts β) 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 *Huntingtin* (*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 triplet 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

* Corresponding author.

E-mail address: tr@bt.iith.ac.in (T. Rathinavelan).

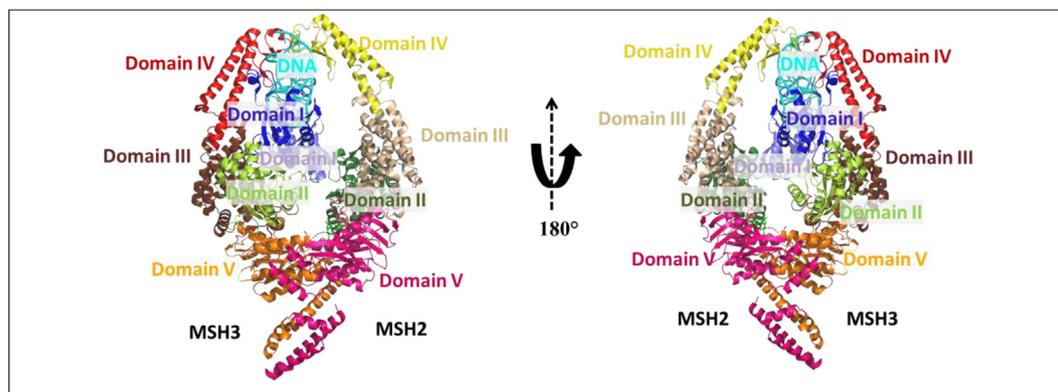


Fig. 1. Cartoon representation of the crystal structure of MSH2.MSH3 and a DNA (colored in cyan) having a bulge (PDB ID: 3THX). Note that the different domains of MSH2 and MSH3 are colored differently.

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)(CAG)₂.d(CTG)₂(CAG)(CTG)₂ (1 mismatch, MutSβ-CAG-1AA) and MSH2.MSH3-d(CAG)₅.d(CAG)₅ (5 mismatches, wherein, A...A occurs periodically, MutSβ-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 MutSβ complex as well as with the DNA substrate. Such enhanced interactions are not found in the case of d(CAG)₂(CAG)(CAG)₂.d(CTG)₂(CAG)(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 (MutSβ) 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 MutSβ-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, 546–547, 646–647, 714–722 and 857–871 residues of MSH2 and, 135–136, 160–168, 262–275, 724–733, 820–836 residues of MSH3. Subsequently, MutSβ-CAG-1AA, MutSβ-CAG-5AA and MutSβ-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

MutSβ-DNA complex models used in the current investigation. Note that the A...A (colored red) mismatch and W&C (colored black) base pairs are represented by "*" and "I" respectively.

S.No	Scheme	Protein	DNA
1	MutSβ-CAG-1AA	MSH2.MSH3	5' C ₁ A ₂ G ₃ C ₄ A ₅ G ₆ C ₇ A ₈ G ₉ C ₁₀ A ₁₁ G ₁₂ C ₁₃ A ₁₄ G ₁₅ 3' 3' G ₃₀ T ₂₉ C ₂₈ G ₂₇ T ₂₆ C ₂₅ G ₂₄ A ₂₃ C ₂₂ G ₂₁ T ₂₀ C ₁₉ G ₁₈ T ₁₇ C ₁₆ 5'
2	MutSβ-CAG-5AA	MSH2.MSH3	5' C ₁ A ₂ G ₃ C ₄ A ₅ G ₆ C ₇ A ₈ G ₉ C ₁₀ A ₁₁ G ₁₂ C ₁₃ A ₁₄ G ₁₅ 3' 3' G ₃₀ A ₂₉ C ₂₈ G ₂₇ A ₂₆ C ₂₅ G ₂₄ A ₂₃ C ₂₂ G ₂₁ A ₂₀ C ₁₉ G ₁₈ A ₁₇ C ₁₆ 5'
3	MutSβ-CAG-WC	MSH2.MSH3	5' C ₁ A ₂ G ₃ C ₄ A ₅ G ₆ C ₇ A ₈ G ₉ C ₁₀ A ₁₁ G ₁₂ C ₁₃ A ₁₄ G ₁₅ 3' 3' G ₃₀ T ₂₉ C ₂₈ G ₂₇ T ₂₆ C ₂₅ G ₂₄ T ₂₃ C ₂₂ G ₂₁ T ₂₀ C ₁₉ G ₁₈ T ₁₇ C ₁₆ 5'

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. GNU PLOT [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:

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

$$\Delta E_{MM} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$$

Note that the energy (ΔE_{MM}) of the complex ($\Delta E_{\text{complex}}$), receptor ($\Delta E_{\text{receptor}}$) 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 means 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

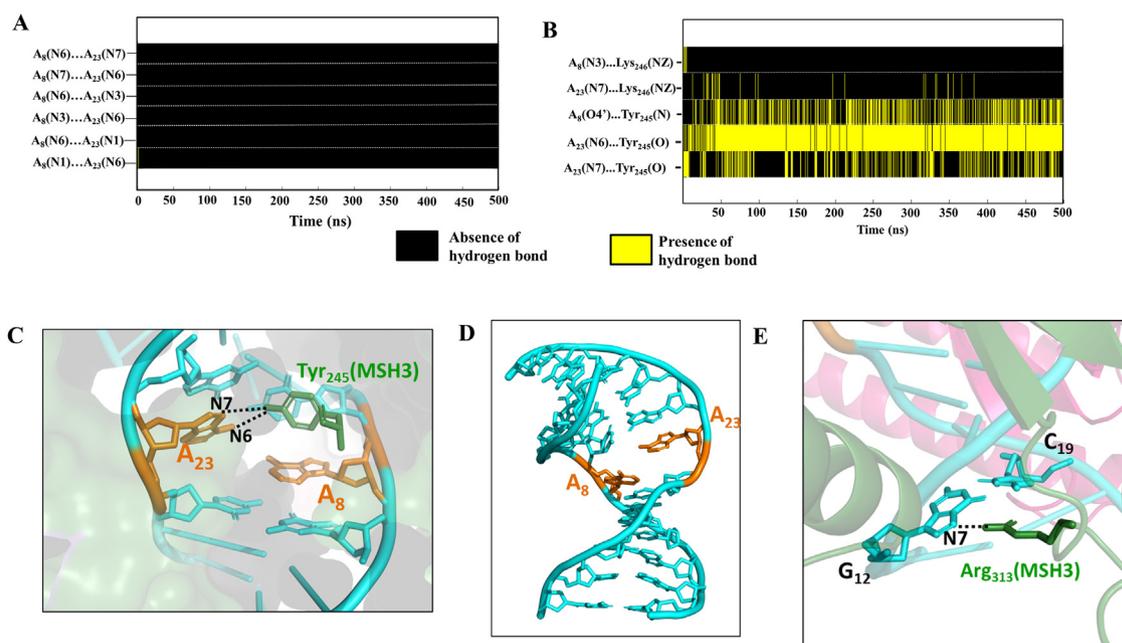


Fig. 2. MutSβ interaction with the CAG-1AA at the mismatch site. (A) Time vs hydrogen bond distance plot showing the complete loss of hydrogen bond between the mismatched A₈ and A₂₃ in the CAG-1AA substrate of MutSβ-CAG-1AA complex. (B) Time vs hydrogen bond distance plot showing the formation of A₂₃(N7)...Tyr₂₄₅(O), A₂₃(N6)...Tyr₂₄₅(O), A₈(O4')...Tyr₂₄₅(N), A₂₃(N7)...Lys₂₄₆(NZ) and A₈(N3)...Lys₂₄₆(NZ) and A₈(N3)...Tyr₂₄₅(N) 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 to a base of the substrate (500 ns).

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)...Lys₂₄₆(NZ) and A₈(N3)...Lys₂₄₆(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 MutSβ 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 Lys₅₄₆(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].

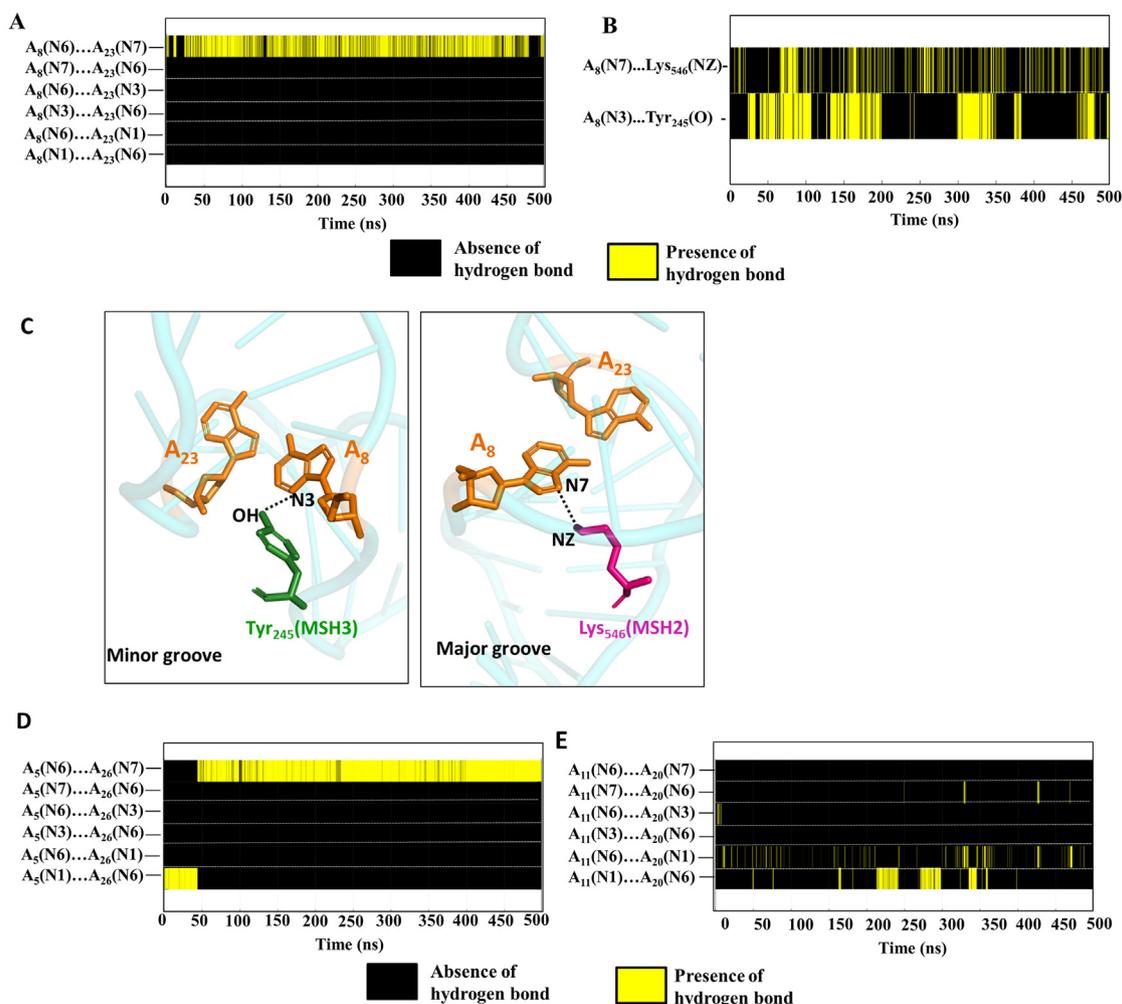


Fig. 3. MutSβ interaction with the CAG-5AA substrate. A) Time vs hydrogen bond distance plots corresponding to (A) A₂₃(N7)...A₈(N6), (B) Lys₅₄₆(NZ)...A₈(N7) and Tyr₂₄₅(O)...A₈(N3). Note the on and off interaction of Lys₅₄₆ and Tyr₂₄₅ with A₈ can occur either simultaneously or individually. C) Snapshots showing the simultaneous Tyr₂₄₅(O)...A₈(N3) (minor groove) and Lys₅₄₆(NZ)...A₈(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).

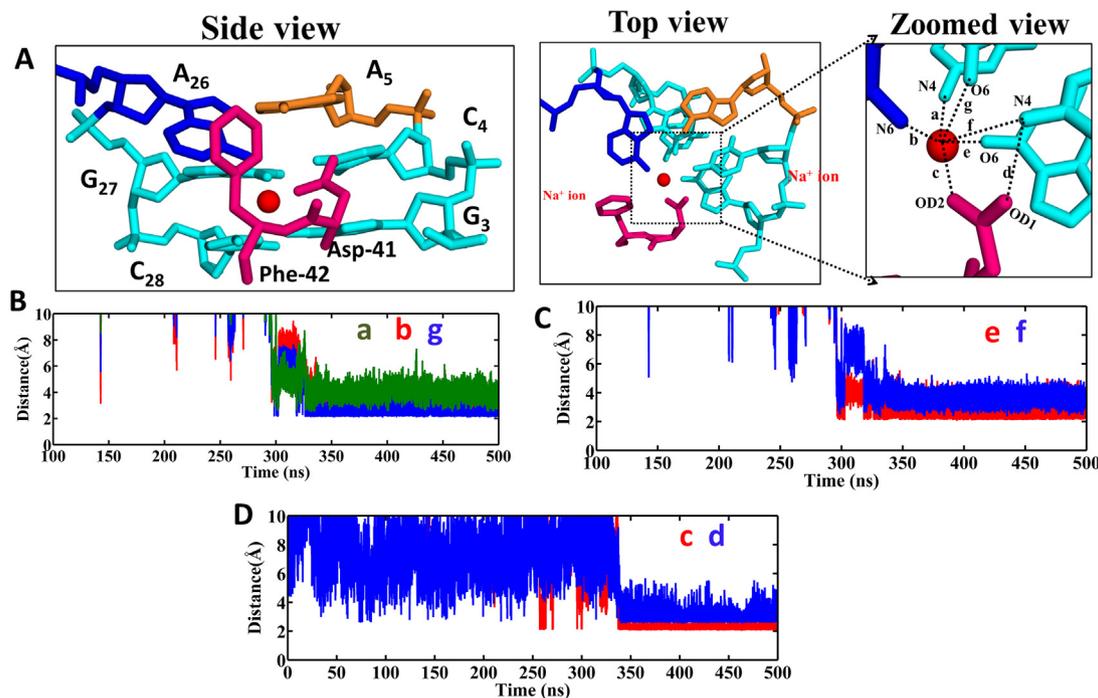


Fig. 4. Na⁺ ion coordination network that tightens the interaction between A₅...A₂₆ and MSH2-domain-I in MutSβ-CAG-5AA. (A) Snapshots showing the Na⁺ mediated network involving A₂₆, G₂₇, C₂₈, G₃, C₄ and Asp-41 residues. (B–D) Distance plots describing the coordination of Na⁺ with DNA/protein residues: Na⁺...C₂₈(N4) (B), Na⁺...A₂₆(N6) (B), Na⁺...G₂₇(O6) (B), Na⁺...G₃(O6) (C), Na⁺...C₄(N4) (C), Na⁺...Asp₄₁(OD2) (D) and Asp₄₁(OD1)...C₄(N4) (D). Note that the coordination distances given in (B–D) represent the interactions (indicated as a, b, c, d, e, f and g) shown in (A).

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 MutSβ-CAG-5AA influences the interaction among the different domains of MutSβ. 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 MutSβ-CAG-5AA (Fig. 5A, Movie S1) that are far away from each other in the MutSβ-CAG-1AA (Fig. 5B, Movie S2) as well as in the crystal structure (Fig. 5C). These 2 domains interact through hydrophobic interactions. Thus, these bring compactness in MutSβ-CAG-5AA complex.

Further, MD simulations carried out by considering d(CAG)₅.d (CTG)₅ duplex (wherein, only canonical base pairs are present) as a substrate for MSH2.MSH3 (Scheme MutSβ-CAG-WC, Table 1) indicate that duplex doesn't undergo any structural deformations as seen in the cases of MutSβ-CAG-5AA and MutSβ-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 MutSβ-CAG-5AA and MutSβ-CAG-1AA falls around 4 Å (Supplementary Fig. S1).

3.4. Binding energy estimation

The gas phase binding energy estimated for the MutSβ-CAG-1AA and MutSβ-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 MutSβ-CAG-5AA complex (−1062.3 k.cal.mol^{−1}) is more favorable compared with MutSβ-CAG-1AA (−964.7 k.cal.mol^{−1}). In contrast, the van der Waals energy component is more favorable for MutSβ-CAG-1AA (−154.8 k.cal.mol^{−1}) compared to MutSβ-CAG-5AA (−122.9 k.cal.mol^{−1}). However, due to a highly favorable electrostatic energy contribution in the case of MutSβ-CAG-5AA, the gas

phase binding energy of MutSβ-CAG-5AA complex (−1185.3 k.cal.mol^{−1}) is more (about −65 k.cal.mol^{−1}) favorable than MutSβ-CAG-1AA complex (−1119.5 k.cal.mol^{−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^{−1}) enhances the interaction between the two compared to CAG-1AA (−1509.4 k.cal.mol^{−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 MutSβ and CAG-5AA is more favorable than MutSβ 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

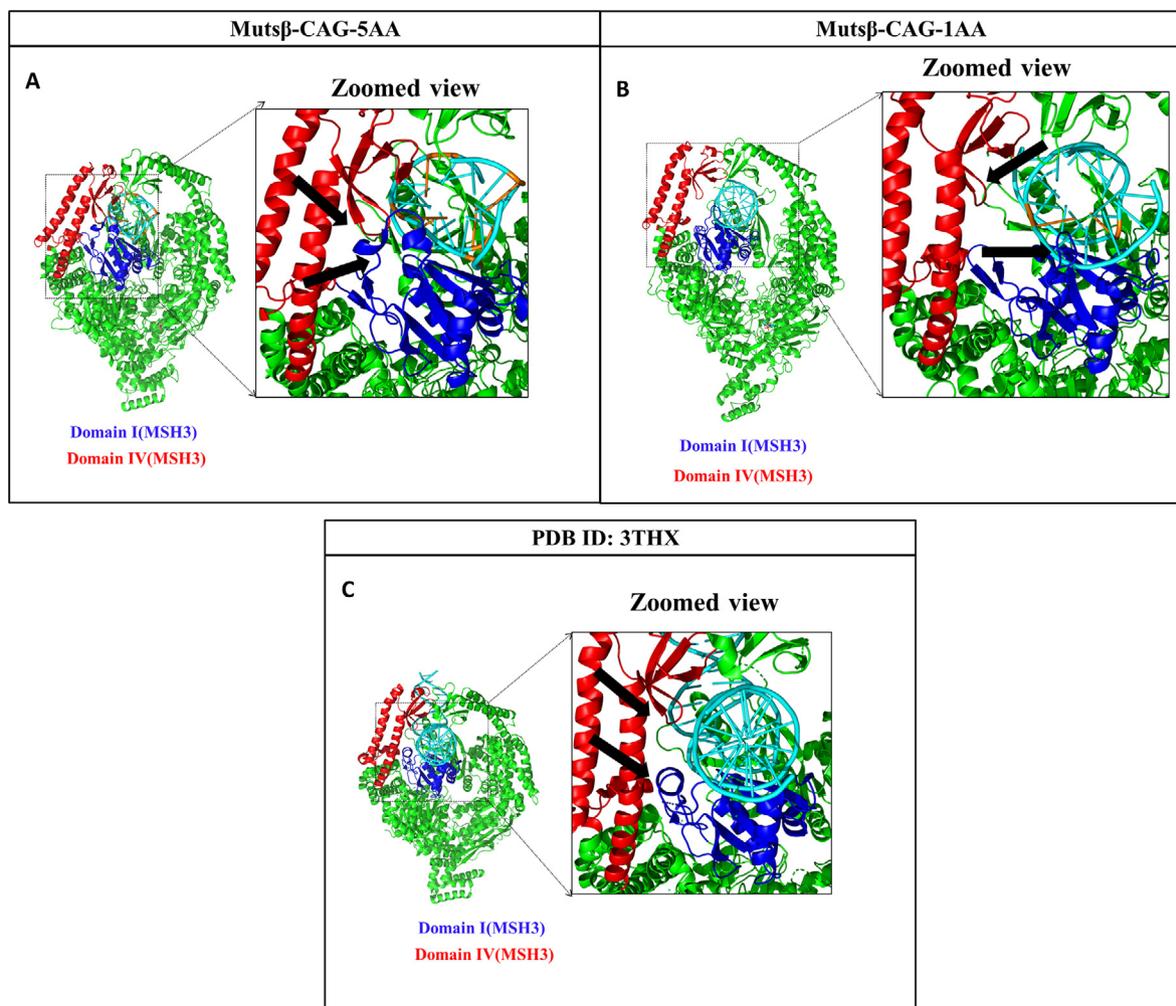


Fig. 5. Cartoon diagram illustrating the nearness (CAG-5AA) or farness (CAG-1AA) of domain-I (colored blue) and domain-IV (colored red) of MSH3. (A-C) MutS β -DNA substrate complex corresponding to (A) MutS β -CAG-5AA (500 ns) and (B) MutS β -CAG-1AA (500 ns) and (C) the crystal structure (PDB ID: 3THX). Note that the arrows indicate (zoomed view) the notable differences seen in the domain movements of the (A-C) three complexes. The proximity of the domain-I and IV can be seen in (A) MutS β -CAG-5AA which is absent in (B) MutS β -CAG-1AA as indicated by the arrows. Note that the DNA substrate is shown in cyan color. See also [Supplementary Movies S1 and S2](#). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Binding energy components of MutS β interaction with CAG-1AA and CAG-5AA calculated from the MD trajectories. Note that MSH2.MSH3 is considered as the receptor and DNA is considered as the ligand.

Energy terms	MutS β -CAG-1AA (kcal.mol ⁻¹)	MutS β -CAG-5AA (kcal.mol ⁻¹)
ΔE_{ele}	-964.7 (182)	-1062.3 (144)
ΔE_{vdw}	-154.8 (98.1)	-122.9 (8.4)
ΔE_{BE}	-1119.5 (182)	-1185.3 (146)

Table 3

Binding energy components of MSH2 interaction with MSH3 calculated from the MD trajectories of CAG-1AA and CAG-5AA. Note that MSH2 is considered as the receptor and MSH3 is considered as the ligand.

Energy terms	MutS β -CAG-1AA (kcal.mol ⁻¹)	MutS β -CAG-5AA (kcal.mol ⁻¹)
ΔE_{ele}	-942.4 (95)	-1083.1 (111)
ΔE_{vdw}	-567 (14)	-576.5 (16)
ΔE_{BE}	-1509.4 (95)	-1659.7 (114)

difference in the extent of base pair nonisomorphism between the two [40]. To explore the influence of such A...A conformational dynamics in trapping the mismatch repair MSH2.MSH3 complex

in polyQ diseases, MD simulations of MSH2.MSH3 (MutS β) in complex with 2 different DNA substrates have been carried out. While one of the substrates has a single A...A mismatch (MutS β -CAG-1AA), the other has 5 A...A mismatches (MutS β -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 MutS β -CAG-5AA, one of the A...A mismatches is involved in Na⁺ mediated coordination with the MSH2-domain-I. Such interaction is absent in MutS β -CAG-1AA. The non-isostericity 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 MutS β -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 MutS β -CAG-5AA leads to a smooth bending (Fig. 6A), whereas, a single A...A mismatch in the middle of the DNA substrate in MutS β -CAG-1AA results in a kink (Fig. 6B). In fact, the kink in

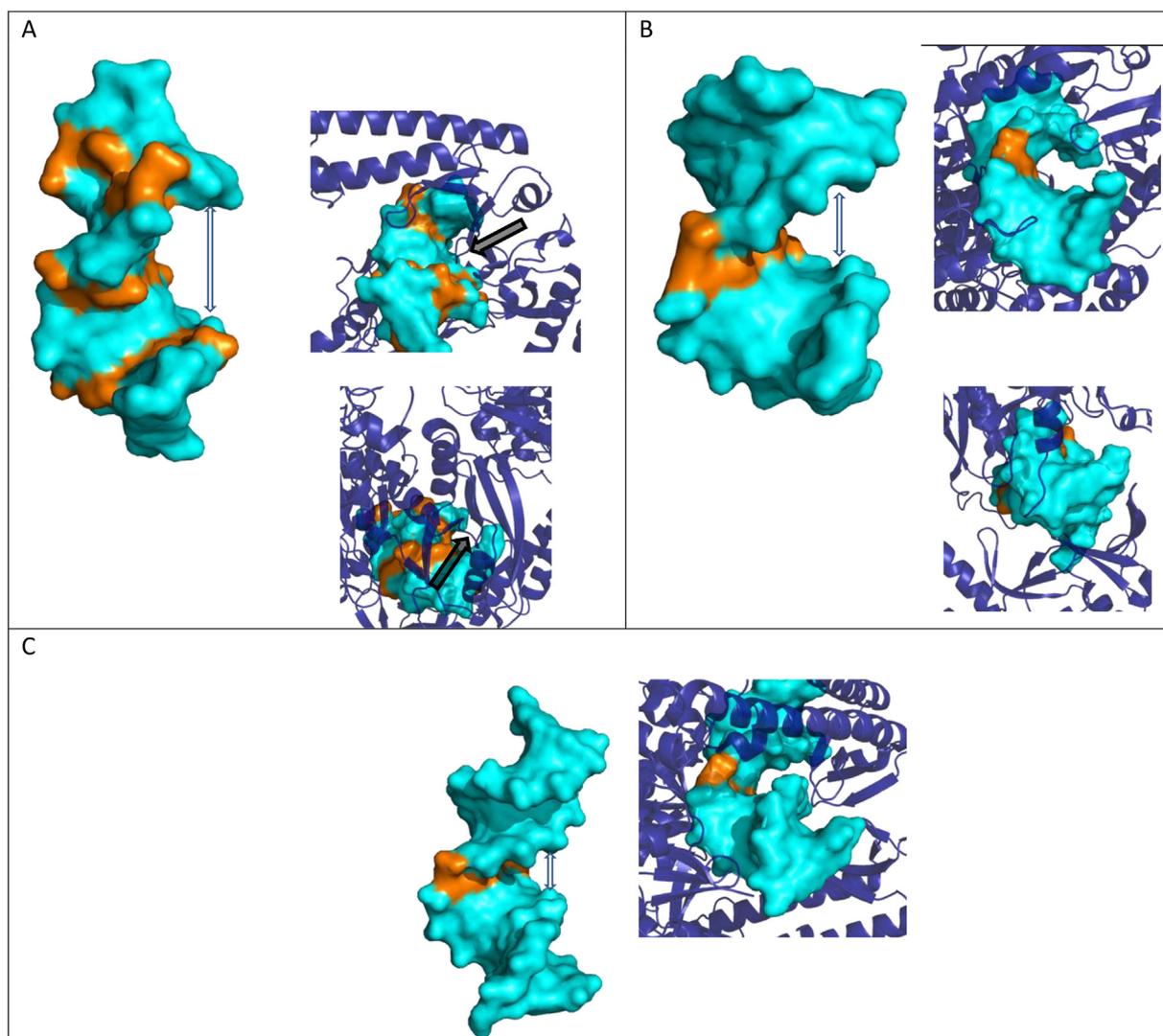


Fig. 6. Exposure of the bases towards the major groove in MutS β -CAG-5AA and its absence in MutS β -CAG-1AA illustrated by considering 500 ns structure as the representative structure. (A) Extension and (B) compression of the DNA substrate (Left, cyan surface) and the consequent exposure of the bases towards the major groove in (A) MutS β -CAG-5AA and its absence in (B) MutS β -CAG-1AA can be seen at the mismatch site. The double headed arrows indicate the extension and compression of the substrates. Note the kink in the DNA towards the major groove in MutS β -CAG-1AA doesn't expose the bases to MutS β (B, Top-Right, Bottom-Right), whereas the exposure of the bases toward the major groove in MutS β -CAG-5AA facilitates its interaction with MutS β (A, Top-Left, Bottom-Left, indicated by single headed arrows). Note that the terminal 2 base pairs on both the sides of the DNA substrates are not shown due to the end fraying effect. (C) The crystal structure of 16-mer DNA substrate (Left, cyan surface) (with an A...A mismatch) and its complex with *E. coli* (PDB ID: 1OH6) homologue of human MutS β mismatch repair complex is shown for comparison. A compression at the A...A mismatch site as seen in (B) and the consequent inaccessibility of the bases to the protein can readily be seen. Note that the human MutS β -DNA complex (PDB ID:3THX) is available only with a DNA loop region (viz., not with an A...A or any other mismatches) and thus, is not shown here. The A...A mismatch is indicated in the golden color and the protein is shown in the blue color cartoon. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 MutS β complex of the schemes MutS β -CAG-5AA and MutS β -CAG-1AA. For instance, the conformational changes in MutS β -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 in complex with human MutS β 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 MutS β -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. In any case, the conformational rearrangements induced by the periodic A...A mismatch facilitates the tighter binding within different domains of MutS β and, between MutS β and the DNA substrate. 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 MutS β complex in a CAG repeat indicate that the mismatch tightens the interaction not only between the DNA and MutS β , but also within the domains of MutS β . The extent of base pair nonisomorphism, which mainly arises from the difference in the diameters of the A...A and canonical base pairs, is found to be the origin of such tighter binding as it unwinds the helix and exposes the mismatched adenines towards either the major or the minor groove. As an earlier experimental investigation has revealed that more than one MutS β binds with the expanded CAG repeat [9], one can envisage many such tighter binding of MutS β in different regions of the expanded CAG repeats may influence the trapping of MutS β 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 MutS β in polyQ disease. Cryo-electron microscope experiments can further provide a detailed picture about the interaction between longer CAG tracts and multiple MutS β .

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-CaRE (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.csbj.2021.07.018>.

References

- [1] Jiricny J. Postreplicative mismatch repair. *Cold Spring Harbor Perspect Biol* 2013;5(4):a012633.
- [2] Surtees J, Argueso J, Alani E. Mismatch repair proteins: key regulators of genetic recombination. *Cytogenetic Genome Res* 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, Marsischky GT. Eukaryotic DNA mismatch repair. *Curr Opin Genet Dev* 1999;9(1):89–96.

- [6] Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese L. Structure of the human MutS α DNA lesion recognition complex. *Mol Cell* 2007;26(4):579–92.
- [7] Sharma M et al. Differential mismatch recognition specificities of eukaryotic MutS homologs, MutS α and MutS β . *Biophys J* 2014;106(11):2483–92.
- [8] Harrington JM, Kolodner RD. Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mismatches. *Mol Cell Biol* 2007;27(18):6546–54.
- [9] Owen BAL, Yang Z, Lai M, Gajek M, Badger JD, Hayes JJ, et al. (CAG) n-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* 2005;12(8):663–70.
- [10] Snell RG, MacMillan JC, Cheadle JP, Fenton 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 pathophysiology. *Annu Rev Genomics Hum Genet* 2000;1(1):281–328.
- [12] Moore H, Greenwell PW, Liu C-P, Arnheim N, Petes TD. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc Natl Acad Sci* 1999;96(4):1504–9.
- [13] Miret JJ, Pessoa-Brandao L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proc Natl Acad Sci* 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* 2011;108(42):E837–44.
- [15] Khan N, Kolimi N, Rathinavelan T, MacKerell A. Twisting right to left: A...A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. *PLoS Comput Biol* 2015;11(4):e1004162.
- [16] Patro LPP, 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] Zgarbová M, Šponer J, Otyepka M, Cheatham TE, Galindo-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, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;11(8):3696–713.
- [21] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103(19):8577–93.
- [22] Rathinavelan T, Yathindra N. Molecular dynamics structures of peptide nucleic acid–DNA hybrid in the wild-type and mutated alleles of Ki-ras proto-oncogene: Stereochemical 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 nonisomorphism G*GC and A*AT triplets and bending of DNA triplexes. *Biopolymers: Original Res Biomol* 2006;82(5):443–61.
- [24] Kolimi N, Ajjugal 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 hZ α ADAR1 protein. *J Biol Chem* 2017;292(46):18732–46.
- [25] Goldsmith G, Rathinavelan T, Yathindra N, Salsbury 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):e0152102.
- [26] Ajjugal Y, Rathinavelan T. Sequence dependent influence of an A...A mismatch in a DNA duplex: an insight into the recognition by hZ α ADAR1 protein. *J Struct Biol* 2021;213(1):107678.
- [27] Ajjugal 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] Thenmalarchelvi R, Yathindra N. New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acids Res* 2005;33(1):43–55.
- [29] Roe DR, Cheatham 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. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter Protein Crystallography* 2002;40(1):82–92.
- [32] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14(1):33–8.
- [33] Downen JM, Putnam CD, Kolodner RD. Functional studies and homology modeling of Msh2-Msh3 predict that mismatch recognition involves DNA bending and strand separation. *Mol Cell Biol* 2010;30(13):3321–8.
- [34] Lee SD, Surtees JA, Alani E. Saccharomyces cerevisiae MSH2–MSH3 and MSH2–MSH6 complexes display distinct requirements for DNA binding domain I in mismatch recognition. *J Mol Biol* 2007;366(1):53–66.
- [35] Kilişzek A, Kierzek R, Krzyżosiak WJ, Rypniewski W. Atomic resolution structure of CAG RNA repeats: 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 (CCCAGGG) 2 and d (CCCTGGG) 2. *Biochemistry* 1987;26(13):4068–75.
- [38] Maskos K, Gunn BM, LeBlanc DA, Morden KM. NMR study of G. cntdot. A and A. cntdot. 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- A and T- 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 trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. *Sci Rep* 2021;11(1):1–17.