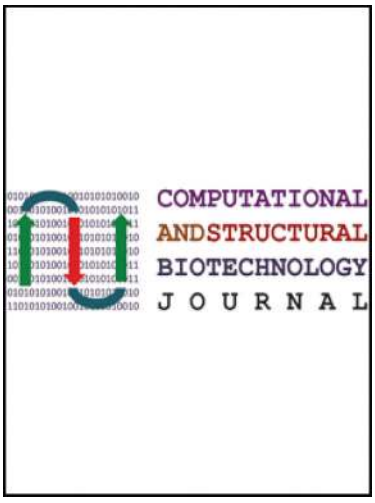
Journal Pre-proofs



Communications

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

PII: S2001-0370(21)00308-1
DOI: https://doi.org/10.1016/j.csbj.2021.07.018
Reference: CSBJ 1124

To appear in: *Computational and Structural Biotechnology Journal*

Received Date: 26 February 2021
Revised Date: 16 July 2021
Accepted Date: 21 July 2021

Please cite this article as: Y. Ajjugal, T. Rathinavelan, 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, *Computational and Structural Biotechnology Journal* (2021), doi: https://doi.org/10.1016/j.csbj.2021.07.018

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 and Thenmalarchelvi Rathinavelan*

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

*For correspondence: tr@bt.iith.ac.in

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(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)$_2$(C<u>A</u>G)(CAG)$_2$.d(CTG)$_2$(C<u>A</u>G).(CTG)$_2$ (1 A…A mismatch) and MutSβ-d(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ (5 mismatches, wherein, A…A occurs periodically) complexes. The interaction of MSH3 residue Tyr$_{245}$ at the minor groove 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(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ distorts the regular B-form geometry. Such an unwinding exposes one of the A…A mismatches (that interacts with Tyr$_{245}$) at the major groove side and also facilitates the on and off hydrogen bonding interaction with Lys$_{535}$ sidechain (MSH2-domain-IV). Nonetheless, kinking of the DNA towards major groove in MutSβ-d(CAG)$_2$(C<u>A</u>G)(CAG)$_2$.d(CTG)$_2$(C<u>A</u>G).(CTG)$_2$ doesn't facilitate such an exposure of the bases towards the major groove. Further, the unwinding of the helix in d(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ enhances the tighter binding between MSH2-domain-I and d(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ at the major groove as well as between MSH3-domain-I and MSH3-domain-IV. Markedly, these interactions are absent in MutSβ-d(CAG)$_2$(C<u>A</u>G)(CAG)$_2$.d(CTG)$_2$(C<u>A</u>G).(CTG)$_2$ that has a single A…A mismatch. Thus, the above-mentioned enhancement in intra- and inter-molecular interactions in d(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ provides the stereochemical rationale for the trapping of MutSβ in CAG repeat expansion disorders.

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, spontaneous deamination *etc.* (Jiricny, 2013, Surtees et al., 2004, Holliday and Grigg, 1993). 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 (Fukui, 2010). MSH2.MSH6 (MutSα) and MSH2.MSH3 (MutSβ) are the two heterodimeric complexes that play the prime role in the eukaryotic mismatch repair process (Kolodner and Marsischky, 1999). While the former recognizes a single base mispair or 1-2 unpaired bases (Warren et al., 2007), the latter recognizes the insertion/deletion 1-15 nucleotides (loops) as well as single base mismatches (Sharma et al., 2014, Harrington and Kolodner, 2007, Owen et al., 2005).

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 (Snell et al., 1993, Cummings and Zoghbi, 2000). The mismatch repair MSH2.MSH3 protein complex is shown to have a major role in the expansion of CAG repeats (Owen et al., 2005). 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 (Moore et al., 1999, Miret et al., 1998). 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 the hairpin and leads to CAG repeat expansion rather than performing the mismatch repair activity (Owen et al., 2005). 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 (Owen et al., 2005). 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 (Lang et al., 2011). 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$(CAG)_2(C\underline{A}G)(CAG)_2$.d$(CTG)_2(C\underline{A}G)$.$(CTG)_2$ (1 mismatch, MutSβ-CAG-1AA) and MSH2.MSH3-d$(C\underline{A}G)_5$.d$(C\underline{A}G)_5$ (5 mismatches, wherein, A…A occurs periodically, MutSβ-CAG-5AA). MD simulations indicate that Tyr_{245} (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(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ leads to an enhancement in the interaction within MutSβ complex as well as with the DNA substrate. These interactions are not found in the case of d(CAG)$_2$(C<u>A</u>G)(CAG)$_2$.d(CTG)$_2$(C<u>A</u>G).(CTG)$_2$ with a single A…A mismatch. Thus, the tighter binding seen in MSH2.MSH3-d(C<u>A</u>G)$_5$.d(C<u>A</u>G)$_5$ complex, perhaps, is the reason behind the trapping of MSH2.MSH3 in the polyQ disorders.

Methods

Molecular dynamics simulation protocol

The MSH2.MSH3 (MutSβ) complex in the crystal structure (**PDB ID: 3THX, Figure 1**) was used to dock with 3 different 15-mer DNA substrates (**Schemes (Table 1)**) used in current investigation. Although the 15-mer DNA CAG duplexes, namely, CAG-1AA (has a single A…A), CAG-5AA (has five A…A mismatches) and CAG-WC (has only canonical base pairs) were modeled using molecular modeling tools like 3D-NuS (Patro et al., 2017) and web-3DNA (Zheng et al., 2009), the models obtained from the previous molecular dynamics (MD) simulations (Khan et al., 2015) were used as the starting models. 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 (Fiser and Sali, 2003): 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 and 914-918 residues of MSH3 were built. Subsequently, MutSβ-CAG-1AA and MutSβ-CAG-5AA complexes were generated manually. In all the schemes, adenosine diphosphate (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 (Case et al., 2016). The OL15 and ff14SB force fields were used for the DNA (Zgarbová et al., 2015) and the protein (Maier et al., 2015) 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 10Å cut-off was used for the non-bonded interactions. The long range electrostatic interactions were taken into account by Particle Mesh Ewald method (Essmann et al., 1995) and the SHAKE algorithm was applied to constrain bonds involving hydrogen atoms. A 2fs time step was used during the simulation. All the systems were equilibrated for 50ps (using an NVT ensemble) followed by a 500ns production run with an NPT ensemble, wherein P was kept at 1atm. During the equilibration run, the solute and the solvent were slowly relaxed in several steps as described in earlier studies (Rathinavelan and Yathindra, 2005, Rathinavelan and Yathindra, 2006, Khan et al., 2015, Kolimi et al., 2017, Goldsmith et al., 2016, Ajjugal and Rathinavelan, 2020, Ajjugal et al., 2021b, Thenmalarchelvi and Yathindra, 2005).

Trajectory analysis

The root mean square deviation (RMSD) and protein…DNA interaction analysis of the MD trajectories were calculated using *cpptraj* module (Roe and Cheatham III, 2013) of AMBER suite. GNUPLOT (Williams et al., 2017) software were used for plotting the data. The Pymol

(DeLano, 2002) and VMD (Humphrey et al., 1996) tools were used for the visualization of the trajectories.

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 50ns MD trajectories with a frame size of 50ps. 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 (Case et al., 2016). The end-point binding energy (ΔE_{BE}) between the DNA substrate and MSH2.MSH3 as well as between MSH2 and MSH3 was 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 ($\Delta E_{complex}$), receptor ($\Delta E_{receptor}$) and ligand (ΔE_{ligand}) were estimated using the bond distance, bond angle and dihedral energy terms (ΔE_{int}) as well as van der Waals (ΔE_{vdw}) energy and electrostatic (ΔE_{ele}) energy components using 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.

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 attain a root means square deviation (RMSD) of 4-5Å quite early during the simulation (less than 10ns) (**Supplementary Figure 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 Figure 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.

Tyr_{245} and Lys_{246} 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 A_8 and A_{23} disengage themselves from the hydrogen bonding interaction quite early during the simulation and continues in the same fashion till the end of the simulation (**Figure 2A**). These adenines move out of plane with respect to each other and facilitate the interaction with the MSH3 through the formation of A_{23}(N7)…Tyr_{245}(O) and A_{23}(N6)…Tyr_{245}(O) as well as A_{23}(O4')…Tyr_{245}(N) (**Figure 2B**) hydrogen bonds (**Figure 2C**). The *-syn glycosyl* conformation of A_{23} exposes N6 and N7 to the minor groove side and facilitates this

interaction. A previous mutagenesis study has also shown the importance of Tyr_{245} (equivalent to Tyr_{157}) in MSH2.MSH3 mediated mismatch repair activity in *Saccharomyces cerevisiae* (Dowen et al., 2010).

Although Lys_{246} of MSH3 interacts with the mismatch through the formation of A_{23}(N7)…Lys_{246}(NZ) and A_8(N3)…Lys_{246}(NZ) hydrogen bonds, the interactions are transient in nature (**Figure 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 **(Figure 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, Arg_{313} (MSH3) is involved in hydrogen bonding with the substrate base (**Figure 2E**).

Periodic A…A mismatch in CAG-5AA tightens the interaction between MutSβ and CAG-5AA

In line with the above, Tyr_{245} interacts (which is crucial for the mismatch recognition) with the central A_8…A_{23} mismatch in CAG-5AA albeit the nature of interaction is different from CAG-1AA. In the first place, A_8…A_{23} hydrogen bond is retained majority of the time during the simulation through N6(A_8)…N7(A_{23}) hydrogen bond (**Figure 3A**) unlike in the previous case (**Figure 2A**). Further, N7(A_8) is also engaged in intermittent hydrogen bond formation with Lys_{546}(MSH2) side chain during the simulation (**Figure 3B, C (Right)**). Such interactions are facilitated through the movement of A_{23} (*-syn glycosyl* conformation) towards the major groove. Further, Tyr_{245} (MSH3) is also engaged in N3(A_8)…Tyr_{245}(O) hydrogen bonding interaction (**Figure 3B, C (Left)**). Among the other 2 A…A mismatches (A_5…A_{26} and A_{11}…A_{20}) present in the helix (Note that the remaining two are ignored due to the end fraying effect, Table 1), A_5…A_{26} retains the N6…N7 hydrogen bond (**Figure 3D**). Nonetheless, A_{11}…A_{20} hardly retains the hydrogen bond during the simulation (**Figure 3E**). To our surprise, unwinding of the helix at A_5…A_{26} exposes the N6 atom of A_{26} towards the major groove, facilitating a strong interaction with the MSH2-domain-I mediated by a Na^+ counter ion around 325ns of the simulation (**Figure 4**). Accompanied by the movement of A_{26} towards the major groove, Asp_{41} and Phe_{42} of MSH2-domain-I form a Na^+ coordination network with A_{26} 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 (Lee et al., 2007).

Enhancement in the interaction between domain-I and domain-IV of MSH3 in concomitance with the conformational dynamics of periodic 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 (**Figure 5A**, **Movie S1)** that are far away from each other in the MutSβ-CAG-1AA **(Figure 5B, Movie S2)** as well as in the crystal structure **(Figure 5C)**. These 2 domains interact through hydrophobic interactions. Thus, these bring compactness in the 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 the 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 in the range of 2Å (**Supplementary Figure S3**). In contrast, the RMSD of MutSβ-CAG-5AA and MutSβ-CAG-1AA fall in the range of 4Å (**Supplementary Figure S1**).

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 kcal.mol^{-1}) is more favorable compared with MutSβ-CAG-1AA (-964.7 kcal.mol^{-1}). In contrast, the van der Waals energy component is more favorable for MutSβ-CAG-1AA (-154.8 kcal.mol^{-1}) compared to MutSβ-CAG-5AA (-122.9 kcal.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 kcal.mol^{-1}) is more (about -65 kcal.mol^{-1}) favorable than MutSβ-CAG-1AA complex (-1119.5 kcal.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 (-1661.0 kcal.mol^{-1}) enhances the interaction between the two compared to CAG-1AA (-1508.25 kcal.mol^{-1}). The electrostatic component is the key factor in causing the difference in the gas phase binding energy of MSH2 and MSH3 interaction between CAG-5AA and CAG-1AA (Table 3). Thus, these results indicate that the interaction between MutSβ and CAG-5AA is more favorable than MutSβ and CAG-1AA.

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 (Kiliszek et al., 2010, Chan, 2014). Unlike the other 7 non-canonical base pairs (C…C, T…T, G…G, G…T, A…C, T…C and G…A) (Patro et al., 2017), the structural insights about an A…A mismatch in the midst of a 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 (Arnold et al., 1987, Maskos et al., 1993, Gervais et al., 1995) and recent molecular dynamics simulation (Ajjugal et al., 2021b, Ajjugal and Rathinavelan, 2020, Khan et al., 2015, Kolimi et al., 2017) 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 (Ananth et al., 2013). This eventually leads to spontaneous and frequent conformational transitions in the A…A mismatch present in a DNA duplex (Ajjugal et al., 2021b, Ajjugal and Rathinavelan, 2020, Kolimi et al., 2017, Khan et al.,

2015). However, such conformational transitions are absent in the G…G mismatch present in a DNA duplex (Ajjugal et al., 2021a). 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 (Ananth et al., 2013). 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 (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 (**Figures 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 involved in Na^+ mediated coordination with the MSH2-domain-I which is absent in MutSβ-CAG-1AA. The non-isostercity of the A…A mismatch (having a larger diameter compared to the canonical base pairs) (Ananth et al., 2013, Khan et al., 2015) with the flanking canonical base pairs unwinds the helix and pushes one of the adenines towards the major groove, facilitating the abovementioned interaction (**Figure 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 (**Figure 6B**) as also seen in the crystal structure (**Figure 6C**). Intriguingly, the periodic unwinding of the DNA substrate at every A…A mismatch site in MutSβ-CAG-5AA leads to a smooth bending (**Figure 6A**), whereas, a single A…A mismatch in the middle of the DNA substrate in MutSβ-CAG-1AA results in a kink (**Figure 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 MutSβ complex of the schemes MutSβ-CAG-5AA and MutSβ-CAG-1AA. For instance, the conformational changes in MutSβ-CAG-5AA bring compactness between the domains I and IV of MSH3 (**Figure 5**). Although the crystal structure of the A…A mismatch in complex with human MutSβ is not available, the DNA substrate of the *E. coli* MutS has an A…A mismatch (PDB ID:1OH6) and it resembles the kink seen in the MutSβ-CAG-1AA (**Figure 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 and a smooth bending in the DNA duplex having a CAG repeat. 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 MutSβ and, between MutSβ and the DNA substrate. Further, many such tight 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 (Owen et al., 2005).

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 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 MutSβ binds with the expanded CAG repeat (Owen et al., 2005), 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.

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.

Conflict of interests

None

Declaration of interests

None

Supporting information

Attached

Reference

AJJUGAL, Y., KOLIMI, N. & RATHINAVELAN, T. 2021a. Secondary structural choice of DNA and RNA associated with CGG/CCG trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. *Scientific reports,* 11**,** 1-17.

AJJUGAL, Y. & RATHINAVELAN, T. 2020. Sequence dependent influence of an A… A mismatch in a DNA duplex: An insight into the recognition by hZαADAR1 protein. *Journal of Structural Biology***,** 107678.

AJJUGAL, Y., TOMAR, K., RAO, D. K. & RATHINAVELAN, T. 2021b. Spontaneous and frequent conformational dynamics induced by A… A mismatch in d (CAA)· d (TAG) duplex. *Scientific reports,* 11**,** 1-18.

ANANTH, P., GOLDSMITH, G. & YATHINDRA, N. 2013. An innate twist between Crick's wobble and Watson-Crick base pairs. *Rna,* 19**,** 1038-1053.

ARNOLD, F. H., WOLK, S., CRUZ, P. & TINOCO JR, I. 1987. Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes d (CCCAGGG) 2 and d (CCCTGGG) 2. *Biochemistry,* 26**,** 4068-4075.

CASE, D., BETZ, R., CERUTTI, D. S., CHEATHAM, T., DARDEN, T., DUKE, R., GIESE, T. J., GOHLKE, H., GÖTZ, A., HOMEYER, N., IZADI, S., JANOWSKI, P., KAUS, J., KOVALENKO, A., LEE, T.-S., LEGRAND, S., LI, P., LIN, C., LUCHKO, T. & KOLLMAN, P. 2016. *Amber 16, University of California, San Francisco*.

CHAN, H. Y. E. 2014. RNA-mediated pathogenic mechanisms in polyglutamine diseases and amyotrophic lateral sclerosis. *Frontiers in cellular neuroscience,* 8**,** 431.

CUMMINGS, C. & ZOGHBI, H. 2000. Trinucleotide repeats: mechanisms and pathophysiology. *Annual review of genomics and human genetics,* 1**,** 281-328.

DELANO, W. L. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography,* 40**,** 82-92.

DOWEN, J. M., PUTNAM, C. D. & KOLODNER, R. D. 2010. Functional studies and homology modeling of Msh2-Msh3 predict that mispair recognition involves DNA bending and strand separation. *Molecular and cellular biology,* 30**,** 3321-3328.

ESSMANN, U., PERERA, L., BERKOWITZ, M. L., DARDEN, T., LEE, H. & PEDERSEN, L. G. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics,* 103**,** 8577-8593.

FISER, A. & SALI, A. 2003. ModLoop: automated modeling of loops in protein structures. *Bioinformatics,* 19**,** 2500-2501.

FUKUI, K. 2010. DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids,* 2010.

GERVAIS, V., COGNET, J. A., LE BRET, M., SOWERS, L. C. & FAZAKERLEY, G. V. 1995. Solution Structure of two Mismatches A· A and T· T in the K-ras Gene Context by Nuclear Magnetic Resonance and Molecular Dynamics. *European journal of biochemistry,* 228**,** 279-290.

GOLDSMITH, G., RATHINAVELAN, T. & YATHINDRA, N. 2016. 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,* 11**,** e0152102.

HARRINGTON, J. M. & KOLODNER, R. D. 2007. Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. *Molecular and cellular biology,* 27**,** 6546-6554.

HOLLIDAY, R. & GRIGG, G. 1993. DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 285**,** 61-67.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics,* 14**,** 33-38.

JIRICNY, J. 2013. Postreplicative mismatch repair. *Cold Spring Harbor perspectives in biology,* 5**,** a012633.

KHAN, N., KOLIMI, N. & RATHINAVELAN, T. 2015. Twisting right to left: A… A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. *PLoS Comput Biol,* 11**,** e1004162.

KILISZEK, A., KIERZEK, R., KRZYZOSIAK, W. J. & RYPNIEWSKI, W. 2010. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic acids research,* 38**,** 8370-8376.

KOLIMI, N., AJJUGAL, Y. & RATHINAVELAN, T. 2017. AB–Z junction induced by an A… A mismatch in GAC repeats in the gene for cartilage oligomeric matrix protein promotes binding with the hZαADAR1 protein. *Journal of Biological Chemistry,* 292**,** 18732-18746.

KOLODNER, R. D. & MARSISCHKY, G. T. 1999. Eukaryotic DNA mismatch repair. *Current opinion in genetics & development,* 9**,** 89-96.

LANG, W. H., COATS, J. E., MAJKA, J., HURA, G. L., LIN, Y., RASNIK, I. & MCMURRAY, C. T. 2011. Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. *Proceedings of the National Academy of Sciences,* 108**,** E837-E844.

LEE, S. D., SURTEES, J. A. & ALANI, E. 2007. Saccharomyces cerevisiae MSH2–MSH3 and MSH2–MSH6 complexes display distinct requirements for DNA binding domain I in mismatch recognition. *Journal of molecular biology,* 366**,** 53-66.

MAIER, J. A., MARTINEZ, C., KASAVAJHALA, K., WICKSTROM, L., HAUSER, K. E. & SIMMERLING, C. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation,* 11**,** 3696-3713.

MASKOS, K., GUNN, B. M., LEBLANC, D. A. & MORDEN, K. M. 1993. NMR study of G. cntdot. A and A. cntdot. A pairing in (dGCGAATAAGCG) 2. *Biochemistry,* 32**,** 3583-3595.

MIRET, J. J., PESSOA-BRANDÃO, L. & LAHUE, R. S. 1998. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences,* 95**,** 12438-12443.

MOORE, H., GREENWELL, P. W., LIU, C.-P., ARNHEIM, N. & PETES, T. D. 1999. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proceedings of the National Academy of Sciences,* 96**,** 1504-1509.

OWEN, B. A., YANG, Z., LAI, M., GAJEK, M., BADGER, J. D., HAYES, J. J., EDELMANN, W., KUCHERLAPATI, R., WILSON, T. M. & MCMURRAY, C. T. 2005. (CAG) n-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. *Nature structural & molecular biology,* 12**,** 663-670.

PATRO, L. P. P., KUMAR, A., KOLIMI, N. & RATHINAVELAN, T. 2017. 3D-NuS: a web server for automated modeling and visualization of non-canonical 3-dimensional nucleic acid structures. *Journal of molecular biology,* 429**,** 2438-2448.

RATHINAVELAN, T. & YATHINDRA, N. 2005. 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. *The FEBS journal,* 272**,** 4055-4070.

RATHINAVELAN, T. & YATHINDRA, N. 2006. Base triplet nonisomorphism strongly influences DNA triplex conformation: Effect of nonisomorphic G∗ ⍰ GC and A∗ ⍰ AT triplets and bending of DNA triplexes. *Biopolymers: Original Research on Biomolecules,* 82**,** 443-461.

ROE, D. R. & CHEATHAM III, T. E. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation,* 9**,** 3084-3095.

SHARMA, M., PREDEUS, A. V., KOVACS, N. & FEIG, M. 2014. Differential mismatch recognition specificities of eukaryotic MutS homologs, MutSalpha and MutSbeta. *Biophys J,* 106**,** 2483-92.

SNELL, R. G., MACMILLAN, J. C., CHEADLE, J. P., FENTON, I., LAZAROU, L. P., DAVIES, P., MACDONALD, M. E., GUSELLA, J. F., HARPER, P. S. & SHAW, D. J. 1993. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature genetics,* 4**,** 393-397.

SURTEES, J., ARGUESO, J. & ALANI, E. 2004. Mismatch repair proteins: key regulators of genetic recombination. *Cytogenetic and genome research,* 107**,** 146-159.

THENMALARCHELVI, R. & YATHINDRA, N. 2005. 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 research,* 33**,** 43-55.

WARREN, J. J., POHLHAUS, T. J., CHANGELA, A., IYER, R. R., MODRICH, P. L. & BEESE, L. S. 2007. Structure of the human MutSα DNA lesion recognition complex. *Molecular cell,* 26**,** 579-592.

WILLIAMS, T., KELLEY, C., BERSCH, C., BRÖKER, H.-B., CAMPBELL, J., CUNNINGHAM, R., DENHOLM, D., ELBER, G., FEARICK, R. & GRAMMES, C. 2017. gnuplot 5.2.

ZGARBOVÁ, M., SPONER, J., OTYEPKA, M., CHEATHAM III, T. E., GALINDO-MURILLO, R. & JURECKA, P. 2015. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. *Journal of chemical theory and computation,* 11**,** 5723-5736.

ZHENG, G., LU, X.-J. & OLSON, W. K. 2009. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic acids research,* 37**,** W240-W246.

Tables

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 "|" 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′

Table 2. Binding energy components of Mutsβ 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	MutSβ-CAG-1AA (kcal.mol^{-1})	MutSβ-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	MutSβ-CAG-1AA (kcal.mol^{-1})	MutSβ-CAG-5AA (kcal.mol^{-1})
ΔE_{ele}	-941.3 (96)	-1084.2 (112)
ΔE_{vdw}	-566.9 (14)	-576.8 (16)
ΔE_{BE}	-1508.2 (96)	-1661 (115)

Figures

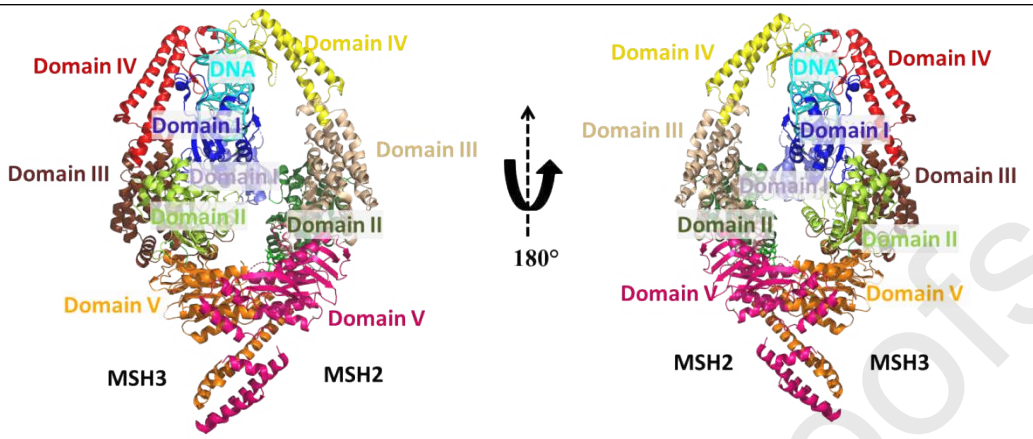


Figure 1. Cartoon representation of the crystal structure of MSH2.MSH3 and a DNA substrate having a bulge (PDB ID: 3THX). Note that the different domains of MSH2 and MSH3 are colored differently.

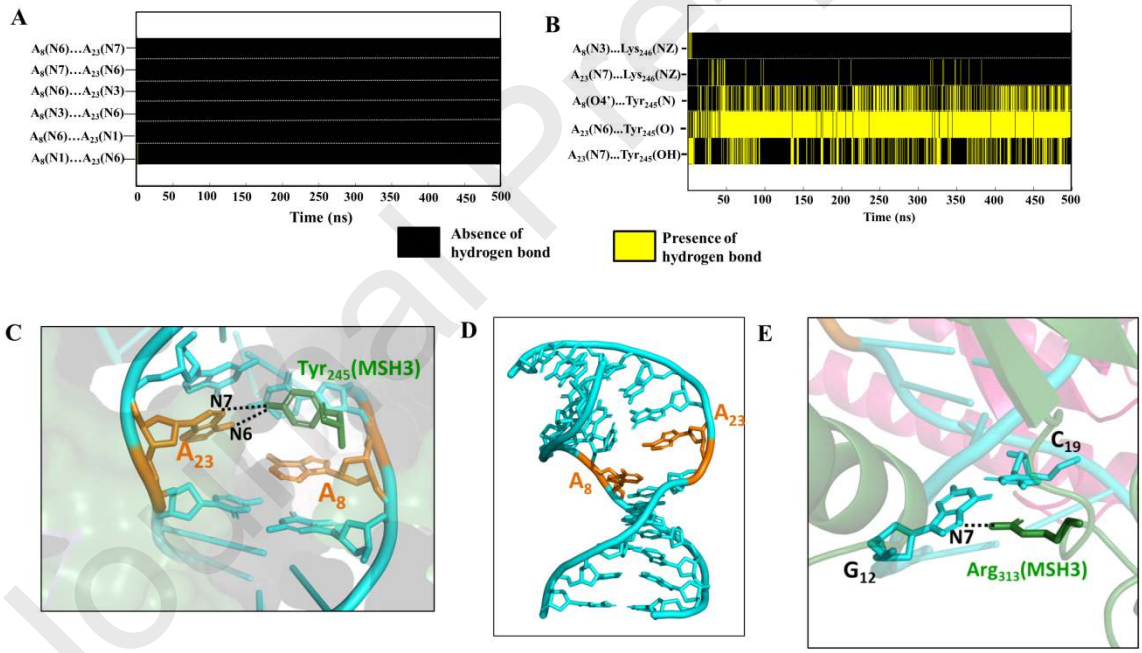


Figure 2. MutSβ interaction with the CAG-1AA at the mismatch site. (A) Time vs hydrogen bond distance plot showing the complete loss of hydrogen bond between the mismatched A_8 and A_{23} in the CAG-1AA substrate of MutSβ-CAG-1AA complex. (B) Time vs hydrogen bond distance plot showing the formation of A_{23}(N7)…Tyr_{245}(OH), A_{23}(N6)…Tyr_{245}(O), A_8(O4')…Tyr_{245}(N), A_{23}(N7)…Lys_{246}(NZ) and A_8(N3)…Lys_{246}(NZ) hydrogen bonds. (C, D) A snapshot showing (C) the interaction of Tyr_{245} with A_8 and A_{23} and, (D) the kink at the mismatch site of the DNA substrate at 500ns. (E) A snapshot illustrating the interaction of Arg_{313}(MSH3) to a base of the substrate (500ns).

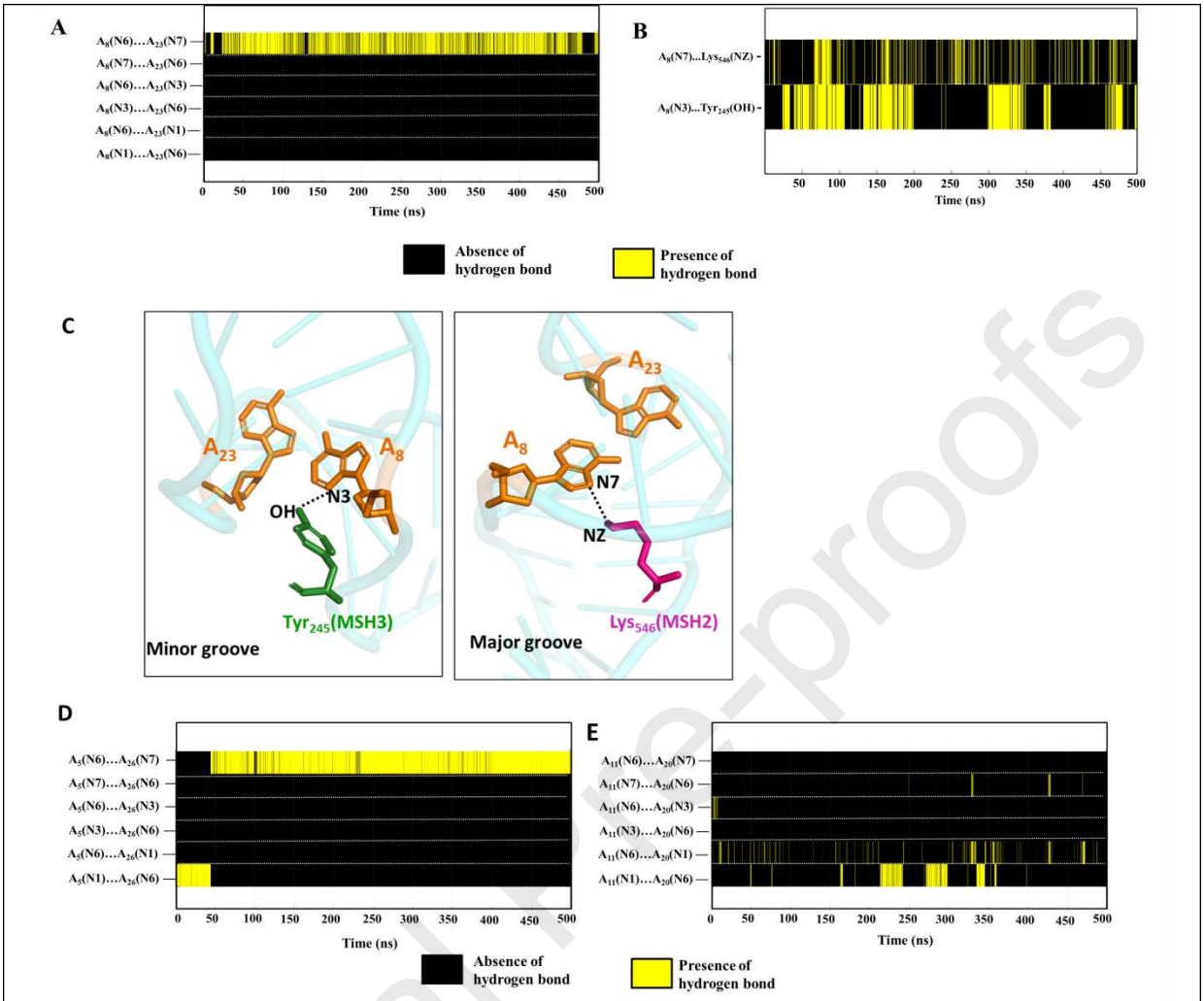


Figure 3. MutSβ interaction with the CAG-5AA substrate. A) Time vs hydrogen bond distance plots corresponding to (A) A_{23}(N7)…A_8(N6), (B) Lys_{546}(NZ)… A_8(N7) and Tyr_{245}(OH)…A_8(N3). Note the on and off interaction of Lys_{546} and Tyr_{245} with A_8 can occur either simultaneously or individually. C) Snapshots showing the simultaneous Tyr_{245}(O)…A8(N3) (minor groove) and Lys_{546}(NZ)…A_8(N7) (major groove) hydrogen bond formation at 215ns. (D, E) Hydrogen bond distance plot corresponding to D) A_5…A_{26} and E) A_{11}…A_{20}. Note the short residence time of hydrogen bonds in (E).

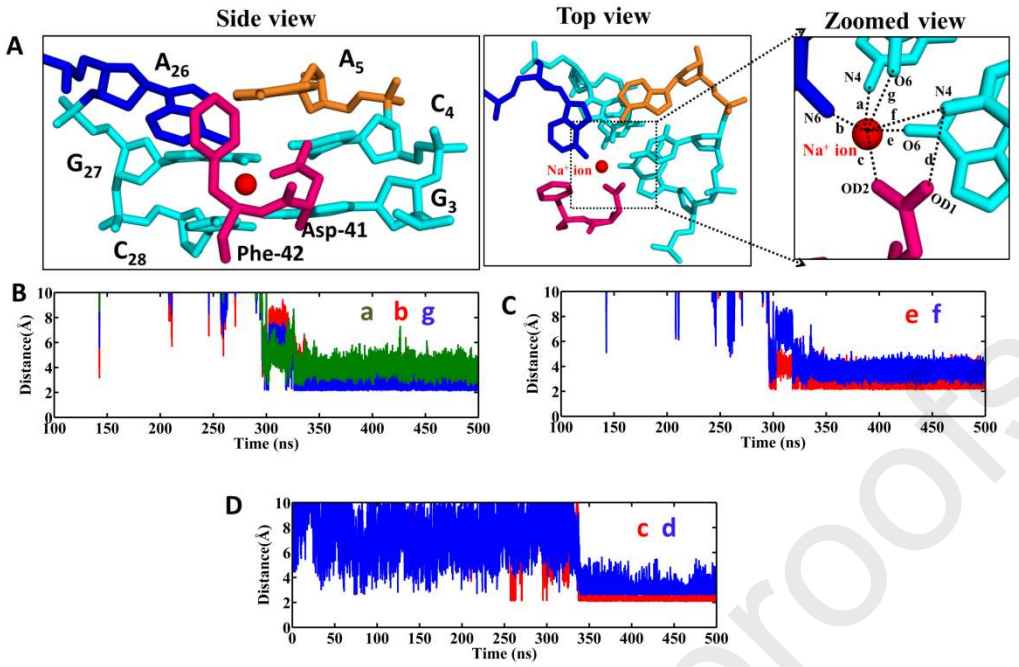


Figure 4. Na^+ ion coordination network that tightens the interaction between A_5…A_{26} and MSH2-domain-I in MutSβ-CAG-5AA. (A) Snapshots showing the Na^+ mediated network involving A_{26}, G_{27}, C_{28}, G_3, C_4 and Asp-41 residues. (B-D) Distance plots describing the coordination of Na^+ with DNA/protein residues: Na^+…C_{28} (N4) (B), Na^+…A_{26}(N6) (B), Na^+…G_{27}(O6) (B), Na^+…G_3(O6) (C), Na^+…C_4(N4) (C), Na+…Asp_{41}(OD2) (D) and Asp_{41}(OD1)…C_4(N4) (D).

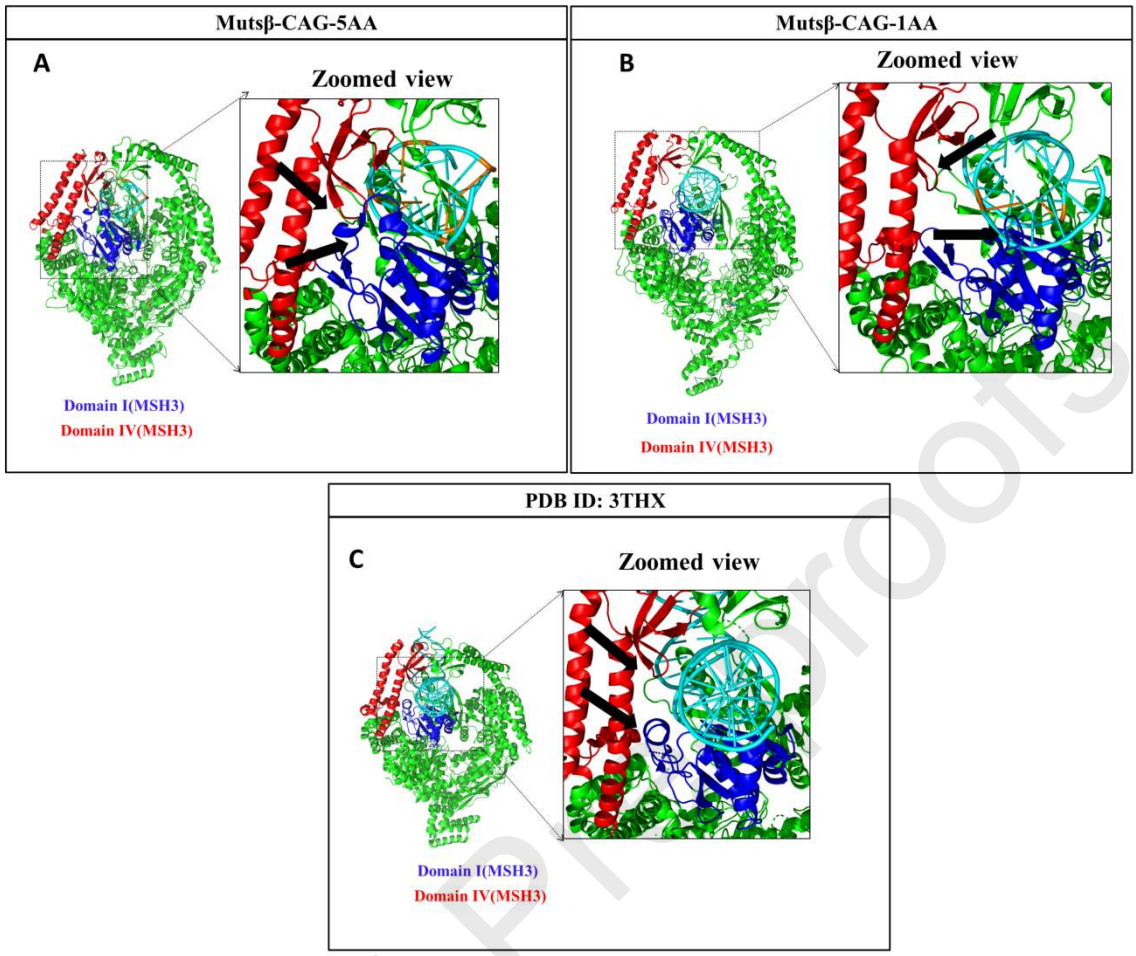


Figure 5: Cartoon diagram illustrating the nearness (CAG-5AA) or farness (CAG-1AA) of domain-I (colored blue) and domain-IV (colored red) of MSH3. (A-C) MutSβ-DNA substrate complex corresponding to (A) MutSβ-CAG-5AA (500ns) and (B) MutSβ-CAG-1AA (500ns) and (C) the crystal structure (PDB ID: 3THX). Note 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) MutSβ-CAG-5AA which is absent in (B) MutSβ-CAG-1AA as indicated by the arrows. Note that the DNA substrate is shown in cyan color. See also Supplementary Movies S1 and S2.

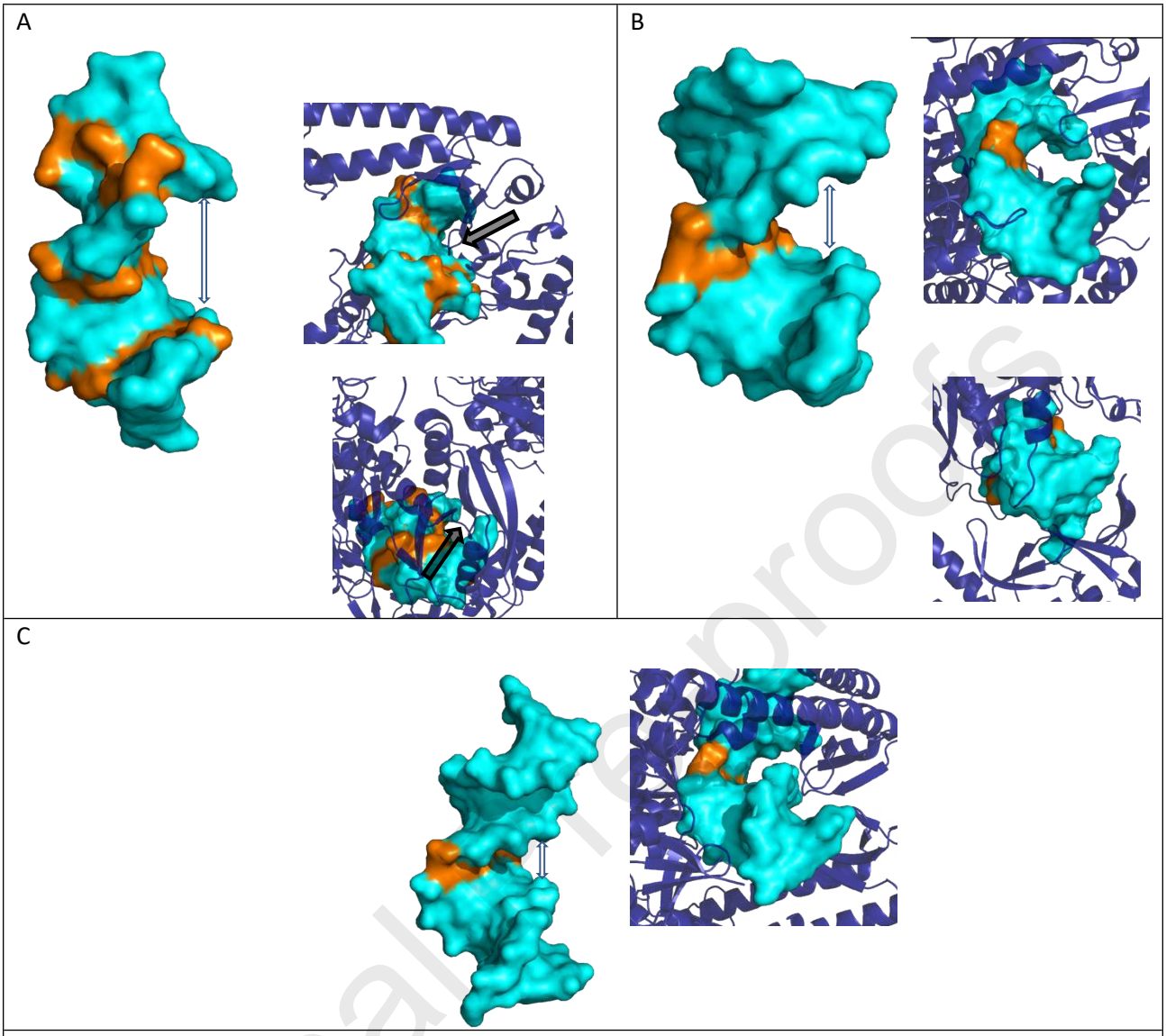


Figure 6. Exposure of the bases towards the major groove in MutSβ-CAG-5AA and its absence in MutSβ-CAG-1AA illustrated by considering 500ns structure as the representative structure. (A) Extension and (B) compression of the DNA substrate (Left, cyan surface) and the consequent exposure of the bases towards the major groove in (A) MutSβ-CAG-5AA and its absence in (B) MutSβ-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 MutSβ-CAG-1AA doesn't expose the bases to MutSβ (B, Top-Right, Bottom-Right), whereas the exposure of the bases towards the major groove in MutSβ-CAG-5AA facilitates the interaction with MutSβ (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 present in the *E. coli* (PDB ID: 1OH6) homologue of human MutSβ 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 MutSβ-DNA complex (PDB ID:3THX) is available only with a loop region (*viz.*, not with an A…A or any other mismatches) and thus, not shown here. The A…A mismatch is indicated in the golden color and the protein in shown in the blue color cartoon.

Author's contribution:

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

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

Affiliation:

All the authors are affiliated with Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana State-502285, India.