Conformational dynamics and DNA recognition by human MutSβ

Jung-Hoon Lee[1]*, Maren Thomsen[1]*, Herwin Daub[1], Stefan C. Steinbacher[1], Agnieszka Sztyler[1], Gabriel Thieulin-Pardo[1], Tobias Neudegger[2], Nikolay V. Plotnikov[3], Ravi R. Iyer[4], Hilary A. Wilkinson[5], Edith Monteagudo[5], Dan P. Felsenfeld[5], Tasir S. Haque[5], Michael Finley[5], Celia Dominguez[5], Thomas F. Vogt[5], and Brinda C. Prasad[5]

[1]Proteros biostructures GmbH, Bunsenstr 7a, 82152 Martinsried, Germany
[2]Current address: Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany
[3]Treeline Biosciences, 11180 Roselle St, San Diego, CA 92121, USA
[4]Script Biosciences, 2929 Arch Street, Philadelphia, PA 19104, USA
[5]CHDI Management/CHDI Foundation, Princeton, NJ 08540, USA

*Contributed equally
Correspondence: Brinda C. Prasad, PhD, brinda.prasad@chdifoundation.org

Abstract

Huntington's disease (HD) is fatal neurodegenerative disorder caused by the expansion of a CAG-repeat tract in the huntingtin (*HTT*) gene. Human and mouse genetics studies have demonstrated a role for DNA mismatch repair (MMR) proteins which control the rate of somatic expansion of the *HTT* CAG repeat and disease onset and progression. MutSβ, a key member of the MMR pathway, is a heterodimeric protein of MSH2 and MSH3 that recognizes and initiates the repair of small insertion or deletion DNA loop outs. Both mouse *Msh3* loss-of-function and reduced-expression alleles of human *MSH3* lead to slower rates of somatic expansion in the *HTT* CAG tract and a delay of disease onset and progression, signifying MSH3 as a promising drug target for HD. Structural biology studies of MutSβ are informative for mechanism, protein structure-function relationships, and guiding small-molecule drug design. Here we report biochemical and cryo-electron microscopy analyses of human MutSβ ensembles, revealing that MutSβ undergoes multiple conformational changes in response to binding and release of nucleotides and DNA. The DNA-free MutSβ-ADP complex adopts an open conformation that is compatible with DNA binding. The conformation of MutSβ in the $(CAG)_2$ DNA-bound open structure most closely resembles the recently identified low-affinity state of MutSα, compared to the

canonical mismatch-bound conformation. The homoduplex-bound and DNA-unbound MutSβ-ATP structures show that MutSβ undergoes an ATP-dependent conformational change towards sliding clamp forms. This study provides a comprehensive understanding of the structural conformational dynamics of MutSβ, insights into the MMR cascade, and a foundation for structure-guided drug discovery.

Introduction

Huntington's disease (HD) is fatal neurodegenerative disorder caused by an expanded CAG repeat in the huntingtin (*HTT*) gene. Individuals who inherit a *HTT* CAG-repeat length greater than 35 are at risk, and the disease is fully penetrant at CAG lengths greater than 39[1]. After the cloning of *HTT* gene[2], differential rates of CAG repeat somatic instability were observed in CNS and peripheral tissues. Recent human genetic studies have shown it is the number of uninterrupted *HTT* CAG triplet repeats, and not simply the number of glutamines, that dictate the timing of the onset of disease[3]. These studies in HD have revealed a number of genetic modifiers of the age onset and disease progression and, remarkably, many of these genes encode proteins in the MMR pathway[4]. These studies intersect with prior studies that showed the role of MMR genes in abrogating *Htt* CAG somatic expansion in mouse models of HD[5]. Both human and mouse genetics have identified *MSH3/Msh3*, encoding a subunit component of the MutSβ heterodimeric complex, as a modifier of *HTT* CAG somatic expansion and disease onset.

The MMR pathway specifically identifies and corrects base-base mismatches and insertion/deletion loops (IDL) that form during DNA replication[6]. Defective MMR activity in humans contributes to cancers such as Lynch syndrome and constitutive MMR deficiency syndrome (CMMR-D)[7]. In contrast, the MMR pathway can be subverted to produce mutations and has been implicated in triplet repeat expansions that underlie several neurodegenerative diseases. Elevated amounts of MSH3 and the ensuring increase in repair activity hastens the onset of HD[4,8]. MutSα and MutSβ, eukaryotic homologues of the prokaryotic MutS protein, recognize mismatches and initiate MMR by recruiting additional members of the pathway to execute the repair[9]. The MutSα heterodimer, composed of MSH2 and MSH6, is similar to the homodimeric bacterial MutS in its ability to detect one or two basepair mismatches. In contrast, MutSβ is capable of recognizing IDL ranging from 1 to 14 nucleotides and DNA with 3' single-stranded extensions[10-12]. Human MutSβ is composed of stably associated MSH2 and MSH3 subunits and directly

interacts with MutLα, an endonuclease that introduces strand breaks, and Proliferating Cell Nuclear Antigen (PCNA), a DNA clamp and strand-directionality factor, to complete MMR[13]. MSH2 and MSH3 are paralogs and share an overall domain organization throughout their lengths. The catalytic ATPase cores of MutSβ contain nucleotide binding pockets that have been identified in members of the ABC (ATP-binding cassette) ATPase superfamily[14]. Previous crystallographic studies showed that MSH2 and MSH3 associate as a pseudo-symmetric heterodimer across the dimer interface in a head-to-tail configuration[15]. The resulting ATPase pocket appears to include nucleotide binding sites from both subunits. Interestingly, a comparable configuration of two ATPase domains that complement each other is seen in the DNA repair protein Rad50, another member of the ABC superfamily[16,17]. This suggests that a composite ATPase active site may be a common characteristic of all members of the ABC superfamily. In addition, mutagenesis and cryo-EM studies of *E. coli* MutS homodimer and human MutSα revealed conformational changes of the two MutS family proteins upon different nucleotide and DNA binding[18-20]. Despite advances in our understanding of bacterial MutS and human MutSα/β structures, precisely defining how the conformational changes of MutSβ are coupled to binding and release of nucleotide and DNA to orchestrate the DNA MMR cascade has been elusive. Here, we present biochemical and structural analysis of MutSβ in response to binding and release of nucleotides and DNA, suggesting a potential mechanism for MMR and its role in the somatic instability of *HTT* CAG repeat.

Results

MutSβ captured by cryo-EM reveals conformational heterogeneity

To determine the structures of human MutSβ, we coexpressed two constructs encoding the full-length *MSH2* and *MSH3* genes in insect cells and copurified them to homogeneity (Extended Data Fig. 1a-c). SEC-MALS confirmed that the reconstituted complex is consistent with a calculated molecular mass of 232.1 kDa, corresponding to the MSH2-MSH3 heterodimeric complex. The catalytic activity of the purified MutSβ was assessed by an *in vitro* ATPase assay (Extended Data Fig. 1d). Surface Plasmon Resonance (SPR) analysis revealed that the full-length wild-type MutSβ efficiently bound DNA (Extended Data Fig. 1e). Initially, MutSβ with and without DNA was applied to a glow-discharged Quantifoil holey carbon grid, and cryo-EM images were recorded on a 200-kV Glacios using a Falcon

4 detector. The vast majority of particles were observed to be denatured. To overcome the instability of the cryo-EM specimen, we added 0.006% Tween-20 to the sample. Cryo-EM analysis of vitrified specimens revealed different conformational states of particles, showing clearly distinguishable two-dimensional (2D) classes (Fig. 1b-g).

The consensus reconstruction of apo MutSβ with 354,534 particles extended to an overall resolution of 3.52 Å, as determined by gold-standard Fourier shell correlation (FSC; see Extended Data Fig. 2). The final map enabled us to unambiguously assign each domain of MSH2 and MSH3 to their corresponding electron densities (Fig. 1b). For model building, crystal structure of the human MutSβ-DNA complex (PDB 3THZ)[15] was rigidly fitted into the cryo-EM map and then manually rebuilt according to the electron densities. The N-terminal domain (NTD) of MSH3 is not defined in the cryo-EM map, presumably because of its conformational flexibility. Intriguingly, MutSβ exhibited two different conformations in the presence of a 61 bp $(CAG)_2$ mismatched DNA. 2D classification and 3D reconstructions of particle images revealed a DNA-bound partially-open form of MutSβ at 3.32 Å that is distinct from a canonical mismatch-bound state at 3.10 Å (Fig. 1c,d). The three additional conformational states of homoduplex DNA-unbound and -bound MutSβ were obtained from the EM specimen in the presence of ATP, showing more compact closed forms compared with those of MutSβ in more relaxed ADP-bound states (Fig. 1e-g, see below).

Apo MutSβ adopts an open, ADP-bound conformation with asymmetric clamps

MutSβ in the absence of DNA adopts a half-opened pretzel shape with the two subunits oppositely coupled at near-central domains and measures ~148 Å in its longest dimension in the high-resolution cryo-EM map (Fig. 2a). The two subunits exhibit pseudo-twofold symmetry with a butterfly-shaped C-terminal ATPase heterodimer linked to lever and clamp domains at the opposite ends, an overall architecture resembling that of bacterial ortholog MutS homodimer (Extended Data Fig. 2h). A reported *E. coli* apo MutS cryo-EM structure (PDB 7OU2)[19] adopts a symmetric open conformation with a common dyad axis for the ATPase homodimer, showing poorly resolved densities for the clamp domains. By contrast, MSH2 in the open state of MutSβ is significantly bent, likely induced by interaction with the mismatch binding domain (MMBD) of MSH3, such that the MSH2 clamp is kinked inward by ~95° toward MSH3. The straight conformation of the MSH3 clamp that does not block the DNA entry

channel of MutSβ may be compatible with DNA binding. The relative orientation of the domains of each subunit observed in the 3.52 Å cryo-EM structure, resembles the previously determined crystal structures of the human MutSβ-DNA complex (PDB 3THZ)[15] with an RMSD of 1.7 Å for 692 Cα atoms in MSH2 and 2.0 Å RMSD for 688 Cα atoms in MSH3. Additionally, we clearly observe the two clamps in this structure which were previously unresolved. The MSH2 and MSH3 subunits in the DNA-free open conformation are tilted across each other by ~42° relative to a canonical mismatch-bound structure, which we hypothesize is coupled to DNA binding (Fig. 2b).

Grid freezing of MutSβ yields the complex with ADP-Mg bound to both subunits. This occurs in the absence of added ADP or ATP during protein purification, suggesting that the ADP originates from ATP hydrolysis during protein expression (Fig. 2c). ADP-Mg is solvent exposed at the α/β ATPase cores of MSH2 and MSH3, causing disordered signature motifs (residues 714-721 and 940-949 of MSH2 and MSH3, respectively) and signature helices located ~11 Å farther away from the bound nucleotides within the active sites. Extending previous studies showing that the ATPase nucleotide binding pocket of MSH3 in four MutSβ-DNA complexes lacks nucleotides[15], we found that the interface between ADP-Mg and MSH3 involves conserved aromatic, hydrophobic, and charged residues and MSH3 Walker motifs[21] that coordinate the ADP phosphate groups and the Mg ion, and it is essentially the same as the MSH2 nucleotide binding pocket (Fig. 2d).

Limited contacts with DNA orient MutSβ in the open state

We determined two distinct conformational states of full-length MutSβ bound to mismatched DNA using cryo-EM. The canonical mismatch-bound state of the 3.10 Å cryo-EM structure resembles published crystal structures of truncated MutSβ-DNA complexes[15] (Fig. 3a,b and Extended Data Fig. 3). The MSH2-MSH3 heterodimer adopts the typical mismatch-bound state, with the ATPase and CTD domains forming an asymmetrical dimeric structure and the clamp and lever domains extending from this structure. The DNA interacts with MMBD of MSH3 and partially with its counterpart in MSH2, similar to previous crystal structures (Fig. 3c). In the cryo-EM structure, the DNA is sharply kinked by ~67° at the insertion-deletion loop (IDL) of $(CAG)_2$ mis-paired nucleotides. The two conserved residues, tyrosine 254 and lysine 255 of MSH3, penetrate the DNA at the IDL causing a localized widening of the DNA strands (Extended Data Fig. 3i). However, a portion of this DNA loop is disordered (Fig. 3a,c).

Consistent with previous crystallography studies[15], ADP was observed to be bound to MSH2 site, even though ADP or ATP was not intentionally included in the sample. However, no density for a nucleotide was detected at the MSH3 nucleotide binding pocket (Fig. 3d).

While examining 2D class averages, we noted that a subset of MutSβ-DNA complex particles appeared to have a distinct characteristic with a clearly distinguishable open conformation of MSH2. We then determined a novel structure of the mismatched DNA-bound MutSβ complex by cryo-EM. We calculated a 3D reconstruction on the basis of an extensive 2D classification of cryo-EM data, interpreted as the MSH2 and MSH3 subunits in complex with a 61 bp $(CAG)_2$ mismatched DNA (Extended Data Fig. 4). The electron density map for MutSβ and the DNA containing the mismatched bases enabled building of a model of the DNA-bound protein complex, guided by the previously determined crystal structures of MutSβ-DNA complexes (PDB 3THZ)[15] and DNA-free open conformation of MutSβ. The MutSβ-DNA complex was initially assembled by rigid-body docking of the DNA-bound and apo structures, whose orientation is uniquely defined by kinked DNA and asymmetric clamps, followed by manual rebuilding and subsequent refinement into the cryo-EM reconstruction at 3.40 Å (Fig. 4a). In the complex, MutSβ embraces the DNA by multiple contacts between DNA and the MMBD and clamp domains of MSH3, exhibiting substantial domain movement of MSH3 clamp and MMBD by ~7 Å and ~5 Å, respectively, upon DNA binding (Fig. 3b,d and Extended Data Fig. 4h). We also observed that, despite not being involved in DNA binding, the MSH2 clamp moves outward by ~27 Å relative to its DNA-free open form when bound to the mismatched DNA. The cryo-EM structure reveals that the DNA undergoes a sharp kink of ~73º at the IDL of $(CAG)_2$ mis-paired nucleotides (Extended Data Fig. 4i). Intriguingly, only the MSH3 subunit of MutSβ heterodimer appears to be involved in DNA binding (Fig. 4c). These DNA contacts with the MMBD of MSH3 are similar to those of the canonical mismatch-bound state of MutSβ, while the interaction between DNA and MSH3 clamp shows a distinct mode, thus representing a partially open conformation of MutSβ assembled onto the mismatched DNA (Fig. 4e). We observed that ADP-Mg occupies both the MSH2 and MSH3 ATPase nucleotide binding pockets (Fig. 4f).

MutSβ undergoes an ATP-dependent conformational change towards a compact structure on DNA

A correlation between the ATP binding of MutS family proteins and their structural changes has been previously suggested. A general model is that upon binding to mismatches, MutS rapidly exchanges ADP for ATP, leading to a conformational change to a sliding clamp state. Recent cryo-EM and biochemical studies of the bacterial MutS have revealed this conformational shift, adopting a compact state in the presence of ATP, AMP-PNP, or ADP vanadate[18,19].

To test whether ATP binding contributes to MutSβ's ability to undergo conformational changes while binding to DNA, we conducted cryo-EM measurements using the full-length MutSβ mixed with a ~1.8 kb linearized plasmid DNA in the presence of ATP (Fig. 5a-c). Due to the presence of long DNA strands stretched across the vitrified holes on the cryo-EM grid, MutSβ displayed a preferential orientation on the DNA. The DNA-bound form, characterized by the markedly recognizable electron density of the DNA, was sufficiently populated to generate well-resolved 2D class averages, whereas the 3D reconstruction appeared to be the 7.08 Å cryo-EM map (Fig. 5c). Despite the low resolution of the map, the homoduplex plasmid DNA-bound MutSβ in the presence of ATP revealed a distinct conformation relative to that of the canonical mismatch-bound structure, which has ADP bound at the MSH2 site. The binding of MutSβ to long homoduplex plasmid DNA in the presence of ATP results in the tilting of its two subunits and bending of the MSH2 clamp, which pushes the DNA downwards in relation to its mismatch-bound conformation and positions it in the central pore of MutSβ (Extended Data Fig. 5). We then determined a 3.29 Å cryo-EM structure of MutSβ bound to a 61 bp end-blocked homoduplex DNA to better define overall architecture and domains of the complex upon ATP binding (Fig. 5d and Extended Data Fig. 6). The consensus reconstruction showed a clear density for MutSβ and its bound ATP, but poorly resolved DNA. In contrast to the heteroduplex DNA bound proteins, the MMBDs of the two subunits and the MSH3 connector are not defined in the cryo-EM map, presumably due to their conformational flexibility. MSH2 had a clear density for its bent clamp towards DNA, positioning the DNA in the central pore of the MutSβ heterodimer. The 3D refinement indicated that the MSH2 connector domain exhibited substantial conformational flexibility, with its relative position rotating downward by ~180°, and the mismatch-binding domain was not observed in the cryo-EM map (Extended Data Fig. 6h). The map displayed a continuous density of DNA extending from the MSH2 connector to its bent clamp and is modeled over a total of 17 bp. Nevertheless, we observed a lower resolution of the DNA compared to the rest of the cryo-EM map, likely due to the inherent flexibility and heterogeneity in the position of the DNA. Despite the limitations of the DNA density, the cryo-EM map

and its structure show that the ATP-bound MutSβ likely binds to DNA primarily through positively charged clusters that are composed of sections of MSH2 connector, lever, and clamp domains (Fig. 5e). Given the two DNA-bound sliding clamp models of MutSβ, we demonstrate that the MSH2 connector in the sliding clamp state can move inward and outward, depending on the position of the DNA (Fig. 5f). When DNA extends beyond the central channel of MutSβ at both termini, the MSH2 connector adopts an outward conformation, resulting in it being positioned ~27 Å away from the DNA. Conversely, when the MSH2 connector is positioned adjacent to one end of the DNA, it partially blocks one end of the DNA-binding channel, resulting in the MSH2 connector adopting an inward conformation. Intriguingly, the ATP-bound MutSβ-DNA complex observed in this study highly resembles the previously determined cryo-EM structure of *E. coli* MutS sliding clamp (PDB 7AIC; see Extended Data Fig. 6i)[18]. The cryo-EM structures of human MutSβ and bacterial MutS sliding clamps display comparable overall shapes and local resemblances essentially throughout their entire length. Notably, a specific subunit's clamp domain (e.g., MSH2 of human MutSβ and one MutS of *E. coli* MutS homodimer) is observed to be sharply kinked towards DNA at the interface between the lever and clamp domain in both cryo-EM structures.

Rearrangement of MutSβ clamp resembles bacterial MutS sliding clamp structures

During the examination of the 2D class images of the MutSβ-61 bp homoduplex DNA-ATP complex in its sliding clamp form, we observed a distinct conformation of one subunit in a subset of particles. The pseudo two-fold symmetry between MSH2 and MSH3, as evident in their domains including the clamps, was analyzed. Two structures of the DNA-unbound MutSβ-ATP complex were obtained, each with a different conformation of the MSH2 clamp (Extended Data Figs. 7 and 8). The first structure is the 3.11 Å cryo-EM structure that shows straight clamps of both subunits bound to ATP-Mg without DNA (Fig. 6a). The second structure of the DNA-unbound form, at 3.43 Å, reveals that the overall architecture of MutSβ, including the bent clamp of MSH2, is nearly identical to that found in the DNA-bound sliding clamp state of the MutSβ-ATP complex (Fig. 6b,c). In the DNA-unbound MutSβ-ATP complex with bent clamp, there is a noticeable structural rearrangement of MutSβ, as the MSH2 connector domain moves outward compared to that in the DNA-bound sliding clamp state (Fig. 6c). An extra density adjacent to the MSH2 connector, corresponding to the N-terminal MMBD, was identified, but due to the limited

quality of the density, the domain structure was not built (Extended Data Fig. 7). The electron densities of the N-terminal, mismatch, and connector domains of MSH3 were not observed, presumably due to their conformational flexibility. In both DNA-unbound forms, the ATPase domains of both subunits are closed by a well-ordered signature loop and helix from the opposite subunit, and ATP-Mg is present in the nucleotide-binding pockets of MSH2 and MSH3 (Fig. 6d).

Intriguingly, a comparison of the two DNA-unbound MutSβ-ATP complex structures with bacterial MutS reveals that the similar conformational change of clamp domain was observed in the DNA-bound sliding clamp structures of *E. coli* MutS in the presence of ATP analog AMP-PNP, suggesting a general mechanism for the conformational change in the sliding clamp state of MutS family proteins (Extended Data Fig. 8h).

Discussion

Here we present six different high-resolution cryo-EM maps of full-length human MutSβ, the most comprehensive assessment reported to date of the conformational heterogeneity of this critical MMR protein. Several structures show significant similarity to structures of the *E. coli* MutS protein. For example, our 3.52 Å apo form of MutSβ demonstrates a similar pseudo-twofold symmetry to a butterfly-shaped C-terminal ATPase heterodimer shared with the reported MutS cryo-EM structure[19]; however, we also resolved the two clamp domains of MutSβ, revealing a clear asymmetry between the kinked MSH2 and extended MSH3 domains (Extended Data Fig. 2). In another example, our 3.11 Å DNA-unbound form of MutSβ exposed to homomeric DNA and ATP shows a striking similarity to mismatched-bound MutS sliding clamp, particularly in the angle adopted between the clamp and lever domains of MSH2 and one of the MutS subunits (Extended Data Fig. 8h). Although our 3.10 Å cryo-EM structure of MutSβ bound to a $(CAG)_2$ IDL resembles the canonical mismatched DNA bound state of the truncated MutSβ-DNA complexes[15] (Extended Data Fig. 3), we also identified a novel 3.32 Å DNA-bound open conformation. Differences between these two conformations highlight the potential for significant domain movements, such as the >10 Å inward movement of the MSH2 and MSH3 clamp domains towards the DNA, required to achieve the DNA-bound canonical form (Fig. 4e).

Despite the improved appreciation of the conformational heterogeneity of the MutSβ complex across both free and DNA-bound forms, we still do not fully understand the relationship between the different

conformations and nucleotide binding or ATP hydrolysis. The previous crystallographic study of MutSβ investigated a truncated human MutSβ bound to IDL DNAs, revealing that MutSβ-IDL complexes exhibit an overall structural similarity to DNA bound human MutSα and bacterial MutS-DNA complexes[15]. These structural analyses highlighted several novel features, including well-defined C-terminal dimerization domains (CTDs) of MSH2 and MSH3, which establish nucleotide-binding asymmetry, as well as characteristic DNA bending that varies depending on the size of mis-paired loop. Notably, these studies demonstrated that all four MutSβ-IDL structures lacked nucleotide in the MSH3 ATPase nucleotide-binding pocket while the MSH2 site was occupied by ADP. Moreover, attempts to soak the crystals with ADP or ATP resulted in reduced X-ray diffraction quality of the protein-DNA complexes. In this study, we observed that both the DNA-free apo MutSβ and its open form bound to $(CAG)_2$ IDL DNA accommodate ADP at the MSH3 ATPase nucleotide-binding pocket. Specifically, the adenosine of ADP is sandwiched between two conserved aromatic residues, Y877 and Y1057, of MSH3. In these two ADP-bound structures of MutSβ, the MSH3 G900 residue in the Walker A motif is repositioned and is opened for phosphate binding. In contrast, the MSH3 Y877 residue occupies the adenine-binding site in the canonical form of MutSβ bound to $(CAG)_2$ IDL DNA, resulting in the absence of nucleotide at the MSH3 site, which is consistent with the previous crystallographic study (Extended Data Fig. 9a). Conversely, the ATP-bound structures exhibit the closure of nucleotide-binding pockets, which is facilitated by the interplay between the negatively charged γ-phosphate of ATP and the positively charged helix dipole, along with the involvement of specific serine residues (S723 in MSH2 and S950 in MSH3) located in the opposite subunit.

In the H-loop of ABC ATPases, a histidine is proposed to control the access of water to the active site and serve as a gatekeeper[14,16]. This regulation can activate a water molecule, facilitating its nucleophilic attack on the γ-phosphate. Intriguingly, a structural comparison of our six different cryo-EM structures of MutSβ reveals that the conformation of the two histidine residues, H783 of MSH2 and H1010 of MSH3, varies among individual chains of the cryo-EM structures depending on the nucleotide status. In the ATP-bound structures, the two histidine residues adopt the “in” conformation required to serve as gatekeepers. In contrast, in the ADP-bound structures, the MSH2 residue H783 adopts the “out” conformation, whereas the MSH3 residue H1010 is shifted inward towards the bound ADP, suggesting that the conformation of the histidine residues in the H-loop in the ADP-bound state is unlikely to be

fixed in only one conformation, but instead can be shifted towards the γ-phosphate when ATP has bound to the active site (Extended Data Fig. 9b).

Recent biochemical and structural studies of human MutSα and its variant, consisting of MSH2 V63E and wild-type MSH6, have proposed that upon initial mismatch binding the MSH2 MMBD moves towards the MSH6 MMBD, forming an interface between the two MMBDs that locks MutSα into a high-affinity state on the mismatched DNA[20]. The MSH2 V63E mutation (which was found in the germline of a patient who developed colorectal cancer at approximately 30 years of age) has been shown to affect DNA binding, likely due to a disrupted interface between the two MMBDs. In the canonical form of the wild-type MutSβ bound to $(CAG)_2$ DNA, MSH2 V63 appears to play a direct role in MSH3 MMBD binding by forming a hydrophobic cluster with key residues in the interface between the two MMBDs (Extended Data Fig. 9c). The MSH2 V63E variant of MutSβ may therefore alter DNA binding more dramatically than the wild-type protein, decreasing the binding affinity of MutSβ to mismatched DNA. Intriguingly, when bound to the $(CAG)_2$ IDL DNA, the wild-type MutSβ was found to adopt a partially open conformation that resembles the low-affinity state of the MutSα MSH2 V63E variant bound to GT-mismatched DNA. Given the observed absence of the MSH2 MMBD in the cryo-EM map of our novel MutSβ conformational state, as well as an open structure of the MSH2 clamp, we suggest that the newly discovered partially-open conformation of DNA-bound MutSβ corresponds to the low-affinity DNA-bound state. Further studies are required to elucidate the precise roles of this partially-open state of MutSβ bound to IDL DNA and their functional implications in DNA MMR.

Our cryo-EM analysis of MutSβ bound to homoduplex DNA revealed two different sliding clamp forms and demonstrated conformational dynamics of the MSH2 connector domain. When bound to long plasmid DNAs that are randomly dispersed in vitreous ice on the cryo-EM grid, the MSH2 connector moves away from the DNA without direct interaction, allowing MutSβ to slide along the DNA until it reaches the open end of the DNA. In the case of the relatively short 61 bp homoduplex DNA binding, the MSH2 connector, which is connected to the compact ATP-bound structure, rotates by almost 180º and partially occludes one end of the central DNA-binding pore, suggesting that the MutSβ sliding clamp is held near the end of the open-ended DNA.

Based on our observations, we propose the following model for MutSβ: in its apo form, MutSβ bound by ADP at both ATPase nucleotide-binding sites adopts a relaxed open conformation suitable for DNA

binding. Upon DNA scanning, MutSβ initially binds mismatched DNA basepairs with low affinity through the mismatch-binding domain of MSH3, causing sharp DNA bending through the insertion of the two conserved residues, tyrosine 254 and lysine 255 of MSH3. During this low-affinity binding state, MutSβ still contains ADP at both nucleotide-binding pockets of the MSH2-MSH3 heterodimer. This is followed by the closing of the mismatch-binding domain of MSH2, which enhances DNA binding and results in the high-affinity binding state of MutSβ to DNA. Rearrangement of local conformation of key residues forming the MSH3 nucleotide-binding site in the high-affinity state induces ADP release in the MSH3 subunit. MutSβ transitions into a sliding clamp conformation by exchanging ADP for ATP at MSH2 and occupying ATP at MSH3, causing DNA to reposition downward towards the central pore of MutSβ and closing the MSH2 and MSH3 ATPase nucleotide-binding pockets. MutSβ binding to DNA is likely coordinated by the DNA clamp loader, Replication Factor C (RFC), and clamp, PCNA, and its repair activity is dependent on recruiting the MutL endonucleases. In the absence of these integral proteins in the MMR repair cycle, we are observing a snapshot of MutSβ recognizing the mismatched DNA and releasing from the mismatch in an ATP-dependent manner.

In summary, our structural and biochemical analyses of MutSβ provide key insights into the architecture of MutSβ and its interaction with DNA, highlighting the conformational dynamics that accompany domain rearrangement upon nucleotide and DNA binding and release. Our study contributes to advancing the understanding of the dynamic conformational changes of MutSβ and its role in the DNA MMR cascade. Improved knowledge of the structure and function of MutSβ will help develop new therapeutic strategies for the treatment for HD and other triplet repeat neurodegenerative disorders.

References

1. Gusella, J. F. & MacDonald, M. E. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* **1**, 109-115 (2000).

2. MacDonald, M. E. et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72,** 971-983 (1993).

3. Wright, G. E. B. et al. Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease. *Am. J. Hum. Genet.* **104**, 1116-1126 (2019).

4. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's Disease onset. *Cell* **178**, 887-900 (2019).

5. Iyer, R. R. & Pluciennik, A. DNA Mismatch Repair and its Role in Huntington's Disease. *J. Huntingtons Dis.* **10**, 75-94 (2021).

6. Martin, S. A., Lord, C. J. & Ashworth, A. Therapeutic targeting of the DNA mismatch repair pathway. *Clin. Cancer Res.* **16**, 5107-5113 (2010).

7. Li, X., Liu, G. & Wu, W. Recent advances in Lynch syndrome. *Exp. Hematol. Oncol.* **10**, 37 (2021).

8. Tomé, S. et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet.* **9**, e1003280 (2013).6. Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **106**, 302-323 (2006).

9. Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **106**, 302-323 (2006).

10. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr. Biol.* **6**, 1185-1187 (1996).

11. Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T. & Jiricny, J. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.* **6**, 1181-1184 (1996).

12. Surtees, J. A. & Alani, E. Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination. *J. Mol. Biol.* **360**, 523-536 (2006).

13. Iyer, R. R., Pluciennik, A., Genschel, J., Tsai, M-S., Beese, L. S. & Modrich, P. MutLalpha and proliferating cell nuclear antigen share binding sites on MutSbeta. *J. Biol. Chem.* **285**, 11730-11739 (2010).

14. Krishnan, A., Burroughs, A. M., Iyer, L. M. & Aravind, L. Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. *Nucleic Acids Res.* **48**, 10045-10075 (2020).

15. Gupta, S., Gellert, M. & Yang, W. Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops. *Nat. Struct. Mol. Biol.* **19**, 72-78 (2011).

16. Hopfner, K. P., Karcher, A., Shin, D., Craig, L., Arthur, L. M., Carney, J. P. & Tainer, J. A. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789-800 (2000).

17. Hopfner, K. P. & Tainer, J. A. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr. Opin. Struct. Biol.* **13**, 249-255 (2003).

18. Fernandez-Leiro, R. et al. The selection process of licensing a DNA mismatch for repair. *Nat. Struct. Mol. Biol.* **28**, 373-381 (2021).

19. Borsellini, A. Kunetsky, V. Fredhoff, R. & Lamers, M. H. Cryogenic electron microscopy structures reveal how ATP and DNA binding in MutS coordinates sequential steps of DNA mismatch repair. *Nat. Struct. Mol. Biol.* **29**, 59-66 (2022).

20. Bruekner, S. R. et al. Unexpected moves: a conformational change in MutSα enables high-affinity DNA mismatch binding. *Nucleic Acids Res.* **51**, 1173-1188 (2023).

21. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951 (1982).

22. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).

23. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

24. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

25. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

26. Wood, C., Burnley, T., Patwardhan, A., Scheres, S., Topf, M., Roseman, A. & Winn, M. D. Collaborative Computational Project for Electron cryo-Microscopy, *Acta Cryst.* **D71**, 123-126 (2015).

27. Chen, V. B., Arendall III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

28. DeLano, W.L. (2002). The PyMOL User's Manual (San Carlos, CA: DeLano Scientific).

29. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

30. Tan, Y. Z., et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods.* **14**, 793-796 (2017).

Figures and figure legends

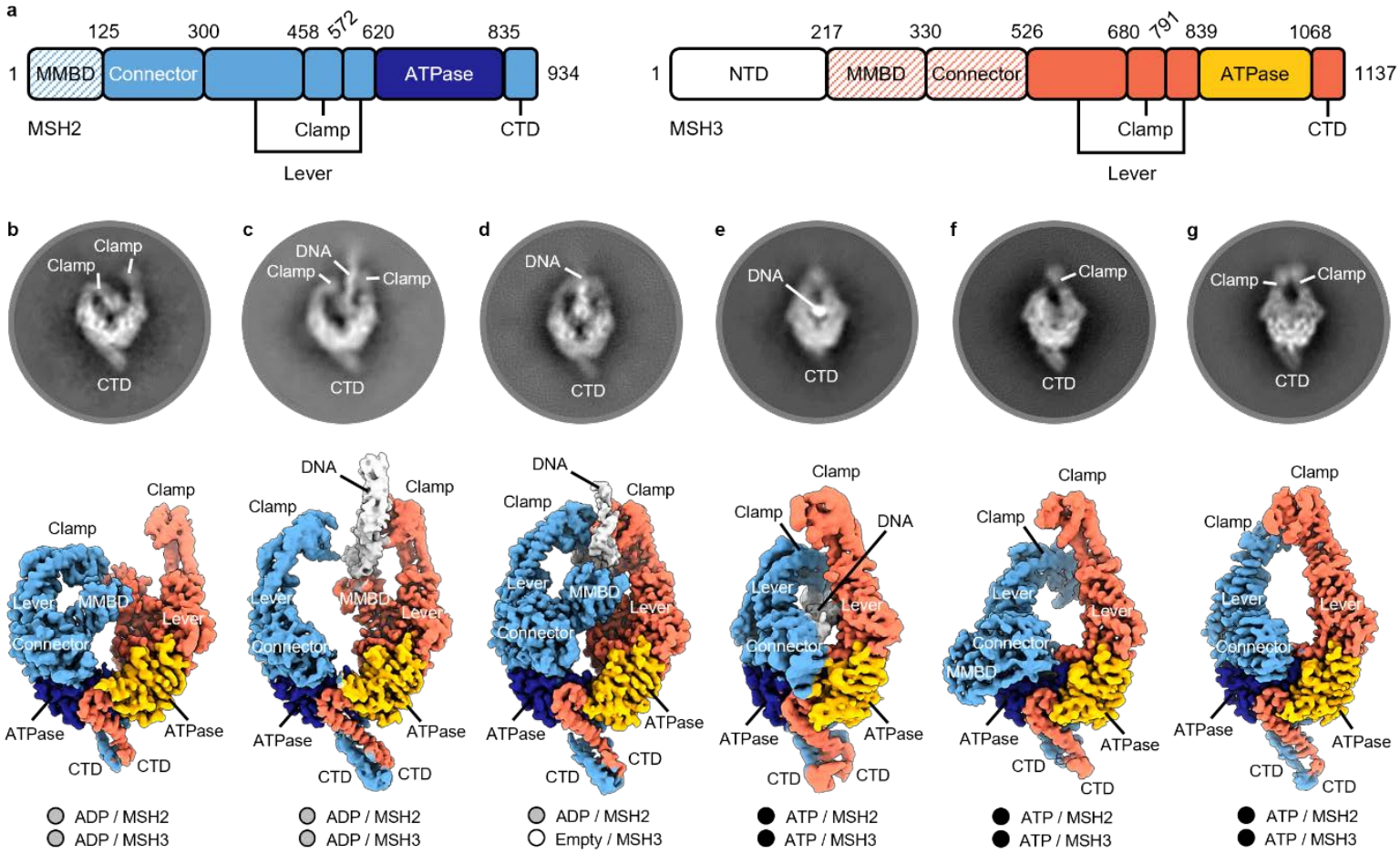


Fig. 1 l Human MutSβ heterodimer and its conformational heterogeneity captured by cryo-EM. a, Domain organization of human MutSβ subunits. In all six cryo-EM maps, the MSH3 NTD is not visible. The hatching pattern indicates the absence of density corresponding to the domains in several cryo-EM maps, presumably because of their flexibility (see results). **b-g,** Representative 2D class averages (top) and cryo-EM maps (bottom) of six different conformers of apo and DNA-bound MutSβ. The same subunit color code is used throughout.

Table 1. Cryo-EM data collection, refinement, and validation statistics

	Open form	Mismatch-bound open form	Canonical mismatch-bound form	Sliding clamp form, homoduplex-bound	Kinked MSH2 clamp, DNA-unbound	Straight MSH2 clamp, DNA-unbound	Sliding clamp form, plasmid-bound
	EMDB-16969	EMDB-16971	EMDB-16964	EMDB-16972	EMDB-16973	EMDB-16974	EMDB-16975
	PDB 8OM5	PDB 8OM9	PDB 8OLX	PDB 8OMA	PDB 8OMO	PDB 8OMQ	
Data collection and processing							
Microscope	Glacios	Glacios	Glacios	Glacios	Glacios	Glacios	Glacios
Detector	Falcon 4	Falcon 4	Falcon 4	Falcon 4	Falcon 4	Falcon 4	Falcon 4
Magnification	15,000	15,000	15,000	15,000	15,000	15,000	15,000
Voltage (kV)	200	200	200	200	200	200	200
Electron exposure ($e^-/Å^2$)	57.29	55.95	55.95	56.80	57.38	56.80	59.93
Defocus range (μm)	0.7 – 2.5	0.7 – 2.5	0.7 – 2.5	0.7 – 2.5	0.7 – 2.5	0.7 – 2.5	0.7 – 2.5
Pixel size (Å)	0.9142	0.9142	0.9142	0.9142	0.9142	0.9142	0.9142
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1
Initial particle images (no.)							
Final particle images (no.)	354,534	240,958	379,788	226,539	263,343	260,322	11,791
Map resolution (Å)	3.52	3.32	3.10	3.29	3.43	3.11	7.08
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.40 – 10.70	3.27 – 13.21	3.04 – 6.36	3.17 – 7.82	3.23 – 7.27	2.97 – 8.29	-
Map sharpening B factor (Å)	-118.45	-57.57	-160.3	-164.2	-177.4	-88.40	-402.4
Refinement							
Model resolution (Å)	3.52	3.32	3.10	3.29	3.43	3.11	
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	
Initial model used (PDB code)	3THZ	3THZ, open form	3THZ	Kinked MSH2 clamp, DNA-unbound	Straight MSH2 clamp, DNA-unbound	3THZ	
Model composition							
Non-hydrogen atoms	13,130	13,314	14,490	11,379	10,290	10,584	
Protein residues	1,639	1,548	1,699	1,337	1,288	1,322	
Nucleotide residues	-	44 (61bp $(CAG)_2$ DNA)	45 (61bp $(CAG)_2$ DNA)	34 (61bp homoduplex DNA)	- (61bp $(CAG)_2$ DNA)	- (61bp homoduplex DNA)	
Ligands	4 (2 Mg, 2 ADP)	4 (2 Mg, 2 ADP)	2 (1 Mg, 1 ADP)	3 (1 Mg, 2 ATP)	4 (2 Mg, 2 ATP)	4 (2 Mg, 2 ATP)	
R.m.s. deviations							
Bond lengths (Å)	0.004	0.005	0.004	0.005	0.005	0.004	
Bond angles (°)	1.039	1.166	0.994	1.169	1.155	0.988	
Validation							
MolProbity score	2.49	2.65	2.31	2.75	2.81	2.55	
Clashscore	19.84	25.33	15.93	26.33	21.28	16.42	
Poor rotamers (%)	1.59	2.83	2.00	2.97	4.59	3.07	
Ramachandran plot							
Favored (%)	90.07	93.74	94.56	92.13	91.97	92.71	
Allowed (%)	9.81	6.26	5.44	7.87	8.03	7.29	
Disallowed (%)	0.12	0.00	0.00	0.00	0.00	0.00	

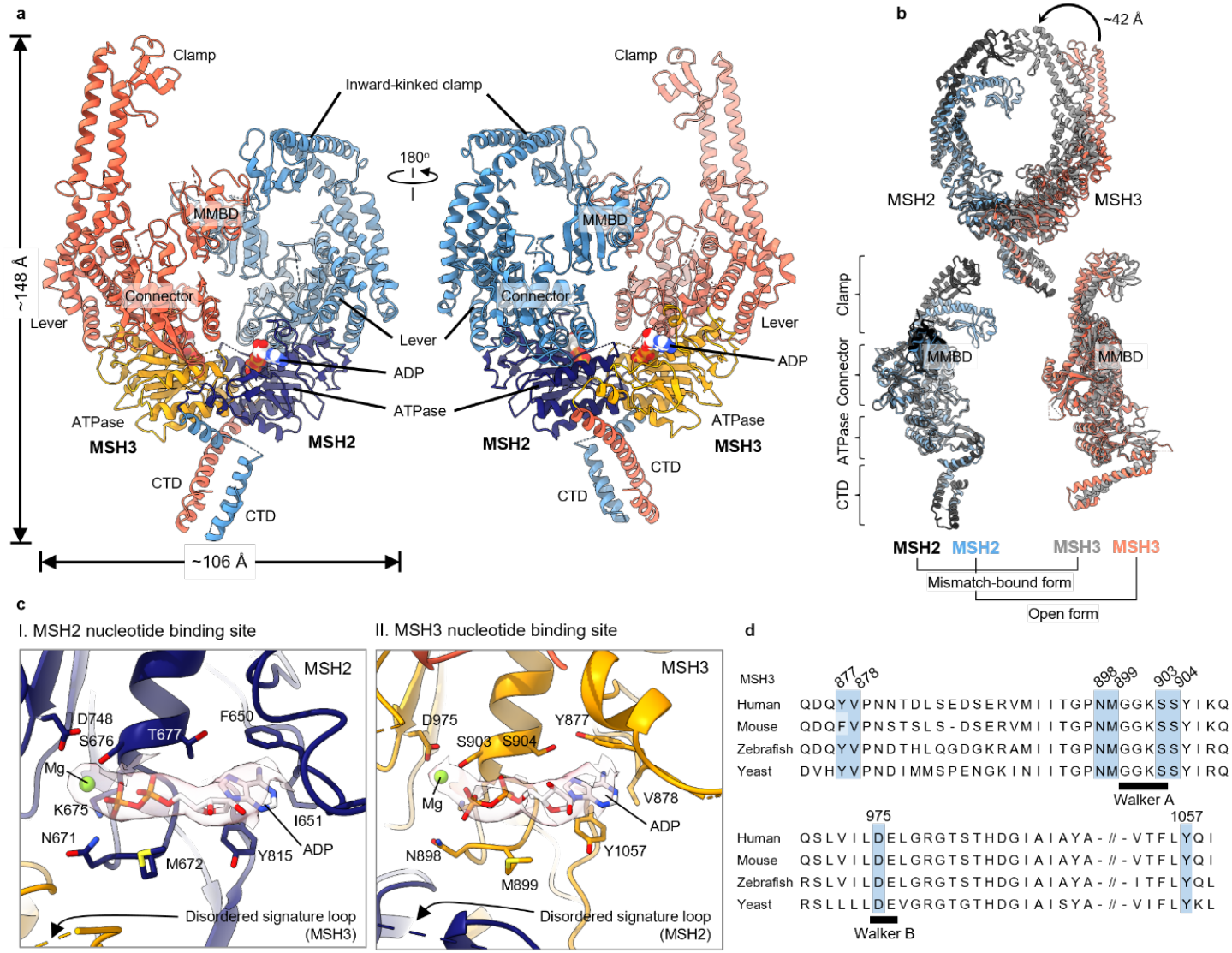


Fig. 2 l Apo MutSβ adopts an open, ADP-bound conformation with asymmetric clamps. a, Two views of a cartoon representation of MutSβ-ADP complex in the open conformation at 3.52 Å. **b,** Relative domain movements of the open conformation of MutSβ compared to the canonical mismatch-bound form (PDB

3THZ)[15] (DNA, MMBD and connector are omitted for clarity). **c,** Magnified views showing nucleotide binding sites of MutSβ in the open form. Electron densities for ADP-Mg are shown in pink (transparent). **d,** Multiple sequence alignment of MSH3 from yeast and human is shown. Key residues coordinating ADP and Mg are highlighted in blue.

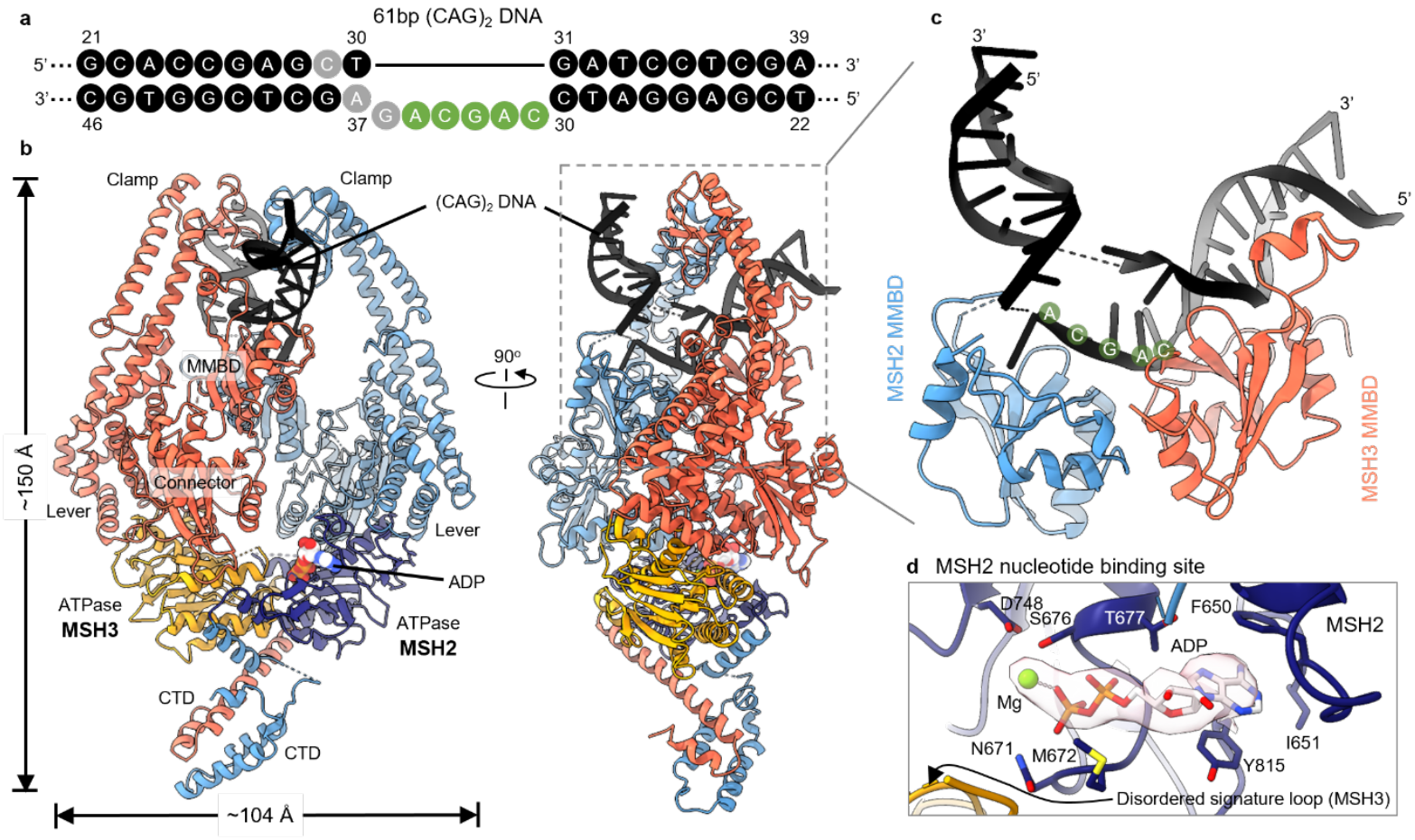


Fig. 3 l MutSβ bound to (CAG)$_2$ DNA in a canonical mismatch-bound conformation. a, Nucleotides are represented as solid spheres in black (modelled) or grey (disordered). Green spheres indicate the modelled mis-paired bases of two CAG repeats. **b,** Two orthogonal views of a cartoon representation of the MutSβ-ADP complex bound to (CAG)$_2$ mismatched DNA. **c,** A magnified side view of the DNA and two MMBDs. **d,** Closed-up view of the MSH3 nucleotide binding pocket of MutSβ in the mismatch-bound state. Electron density for ADP-Mg is shown in pink (transparent).

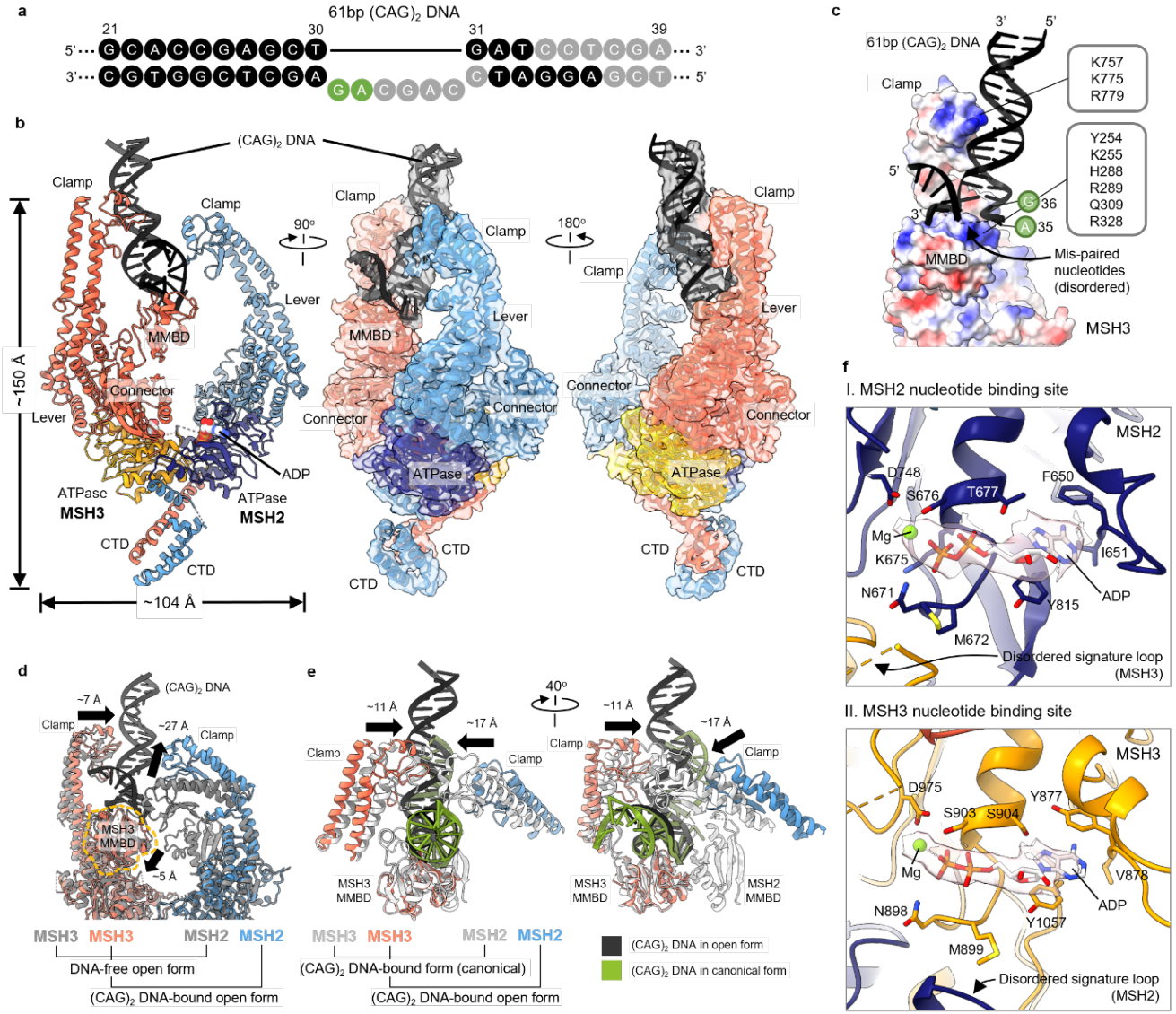


Fig. 4 | Limited contacts with the mismatched DNA orient MutSβ in the open state. a, Nucleotides are represented as solid spheres in black (modelled) or grey (disordered). Green spheres indicate the modelled mis-paired bases of two CAG repeats. **b,** A ribbon diagram of the $(CAG)_2$ DNA-bound open conformation

of the MutSβ-ADP complex (left) and two views of cryo-EM reconstruction at 3.32 Å overlaid with the model (right). **c,** Closed-up views of the interfaces between MutSβ and DNA. Key residues that are involved in DNA binding are highlighted in boxes. **d,** Superposition of the DNA-free open conformation of MutSβ with its $(CAG)_2$ DNA-bound open state. Substantial domain movements are indicated by arrows. **e,** Two views of superposition of the two $(CAG)_2$ DNA-bound MutSβ cryo-EM structures. The arrows indicate the inward movement of the two clamps in the transition to the canonical high-affinity state. **f,** Magnified views showing nucleotide binding sites of MutSβ in the DNA-bound open form. Electron densities for ADP-Mg are shown in pink (transparent).

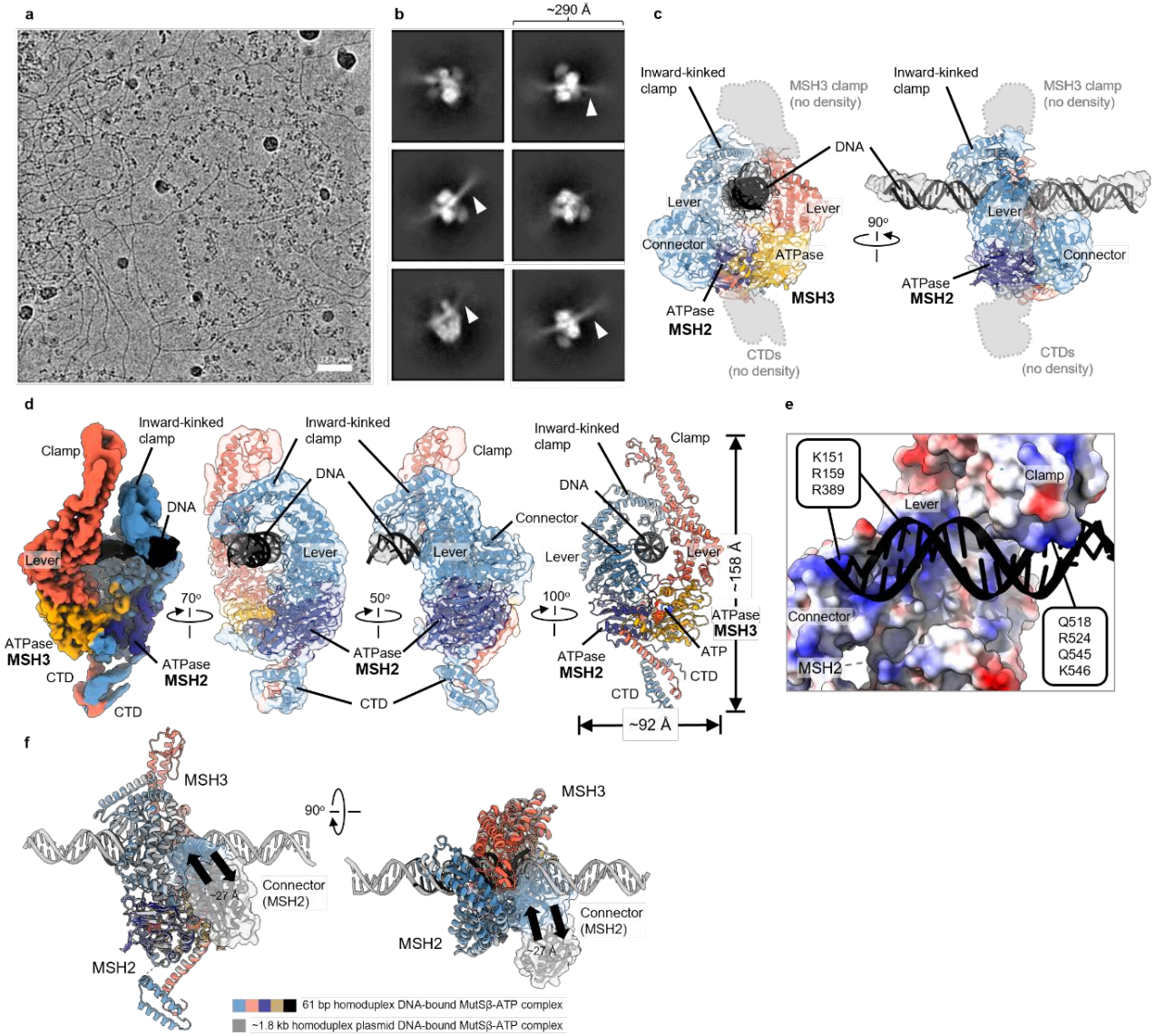


Fig. 5 l MutSβ undergoes ATP-dependent conformational change towards compact structure on DNA. a, Representative raw particles from an original micrograph. **b,** Representative reference-free 2D class averages of MutSβ-ATP complex bound to a 1.8 kb homoduplex plasmid. DNA is highlighted in arrow.

c, Model of MutSβ-ATP-DNA complex fitted into a 7.08 Å cryo-EM map. **d,** Four views of cryo-EM reconstruction of MutSβ bound to a 61 bp homoduplex DNA at a resolution of 3.29 Å (far left) and ribbon diagram of the structure (far right) with cryo-EM map (middle). **e**, A close-up view of the interface between MSH2 and bound DNA. Key residues of MSH2 that are involved in DNA binding are highlighted in boxes. **f,** Two orthogonal images showing the overlay of two structures of MutSβ-DNA-ATP complexes. The arrows indicate substantial domain movement of MSH2 connector.

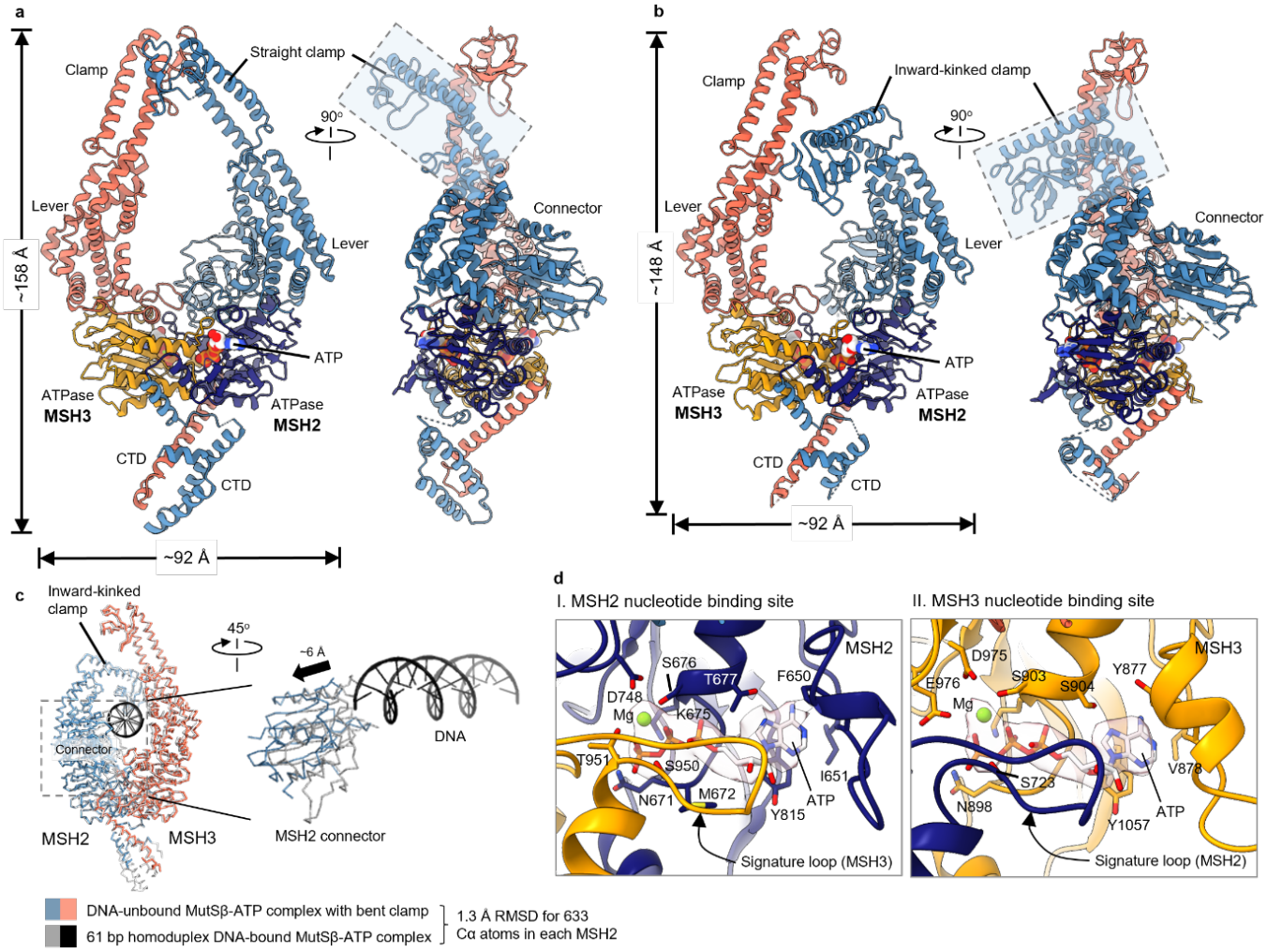


Fig. 6 I Cryo-EM structures of DNA-unbound MutSβ-ATP complexes. a-b, Cartoon representation of the MutSβ-ATP complex showing orthogonal views with a bent (**a**) or straight (**b**) MSH2 clamp. **c,** Superposition of the two MutSβ-ATP complexes in the absence and presence of 61 bp homoduplex DNA. The arrow indicates the domain movement of MSH2 connector. **d,** Close-up views showing nucleotide binding pockets of MutSβ in the DNA-unbound MutSβ-ATP complex with a straight MSH2 clamp.

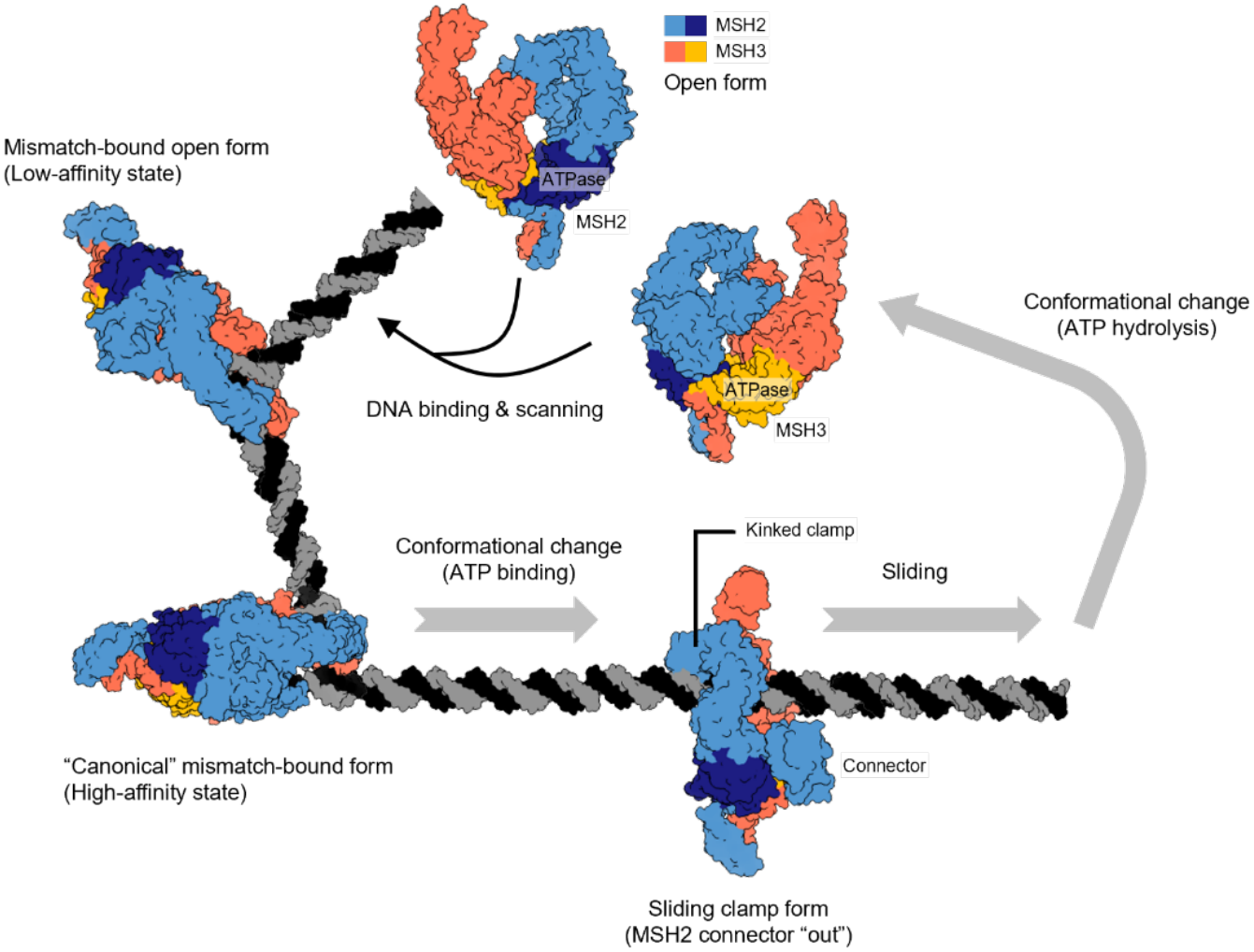
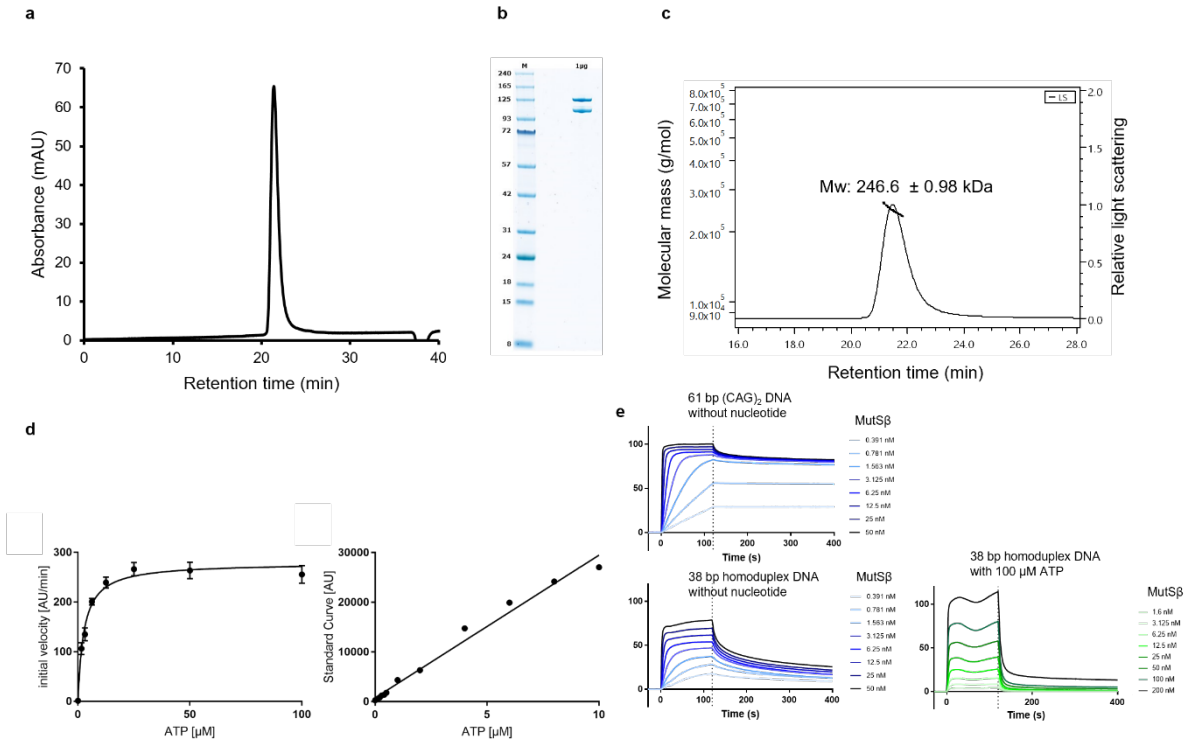
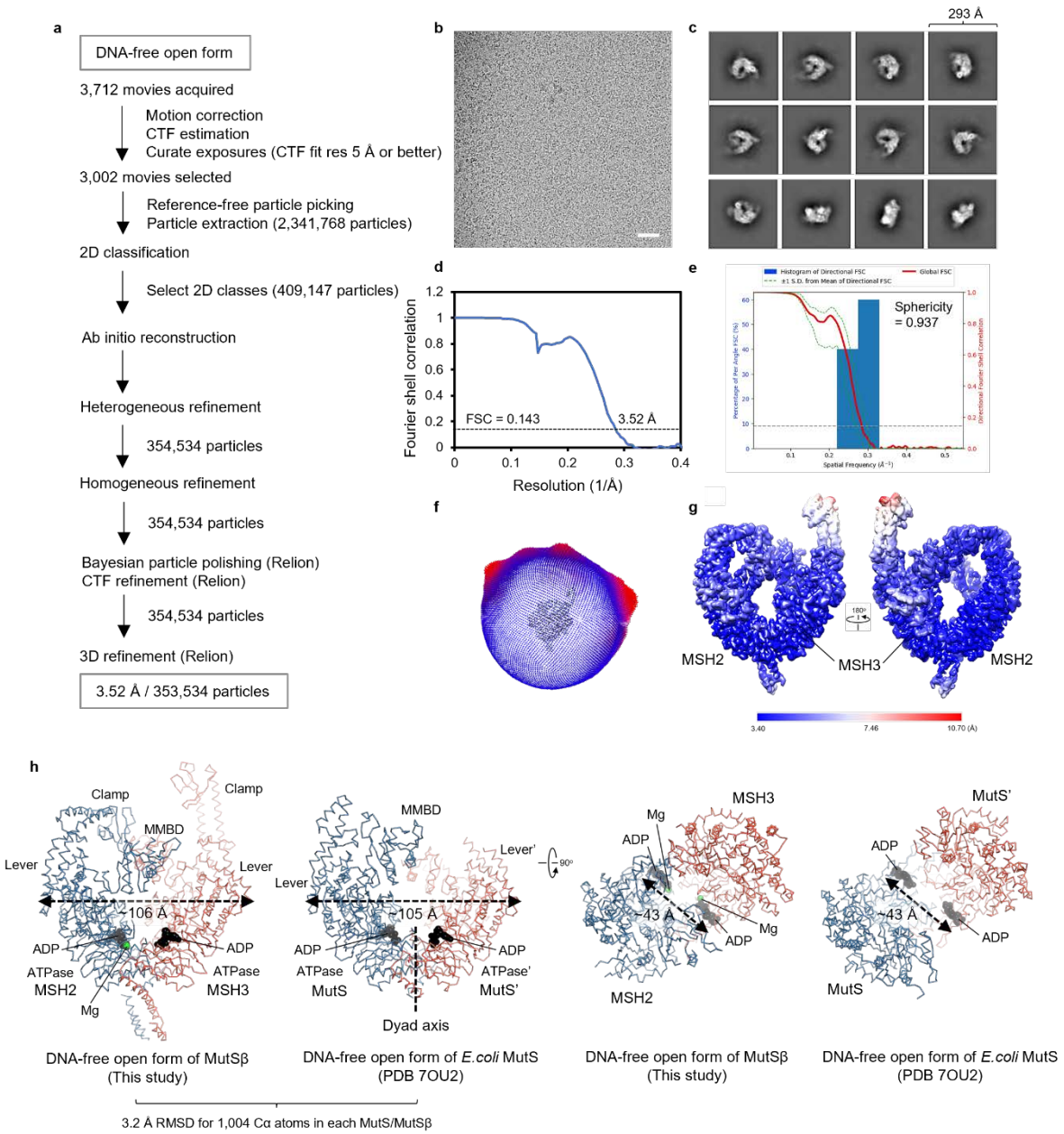


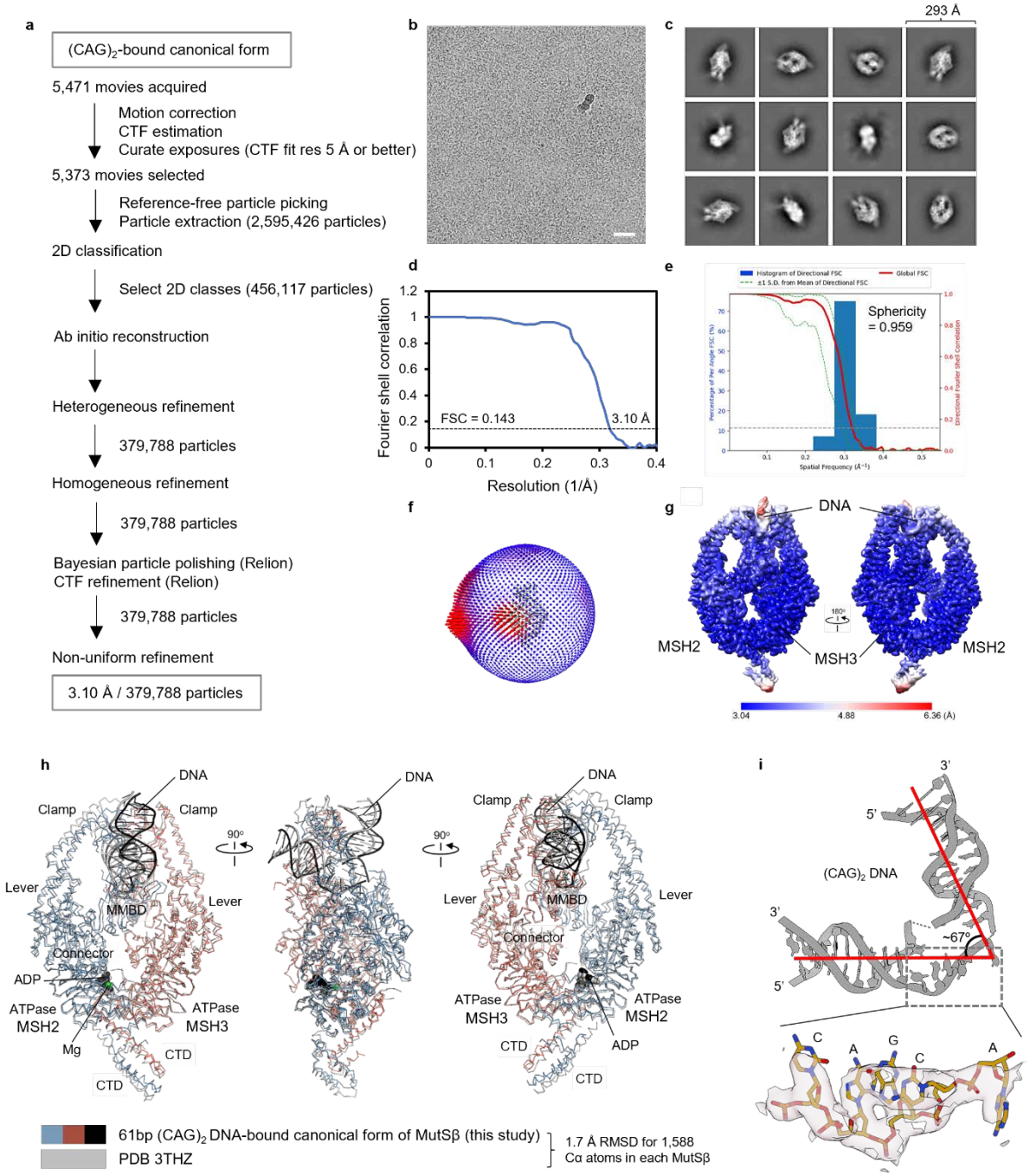
Fig. 7 | Proposed working model of DNA MMR by MutSβ. Our cryo-EM models depict individual steps of MMR by MutSβ, based on biochemical and structural analyses. The proposed working model builds upon our findings and integrates them with current understanding of MMR.



Extended Data Fig. 1 | Biochemical analysis of full-length MutSβ. a, SEC analysis of purified full-length MutSβ. **b,** SDS-PAGE of purified MutSβ for cryo-EM grid preparation. **c,** SEC-MALS analysis of purified full-length MutSβ. **d,** ATPase activity determination of MutSβ by ADPglo in absence of DNA. Background subtracted initial velocity of ATP dilution series is plotted over ATP concentration (left). Non-linear fit of reaction time yields a Vmax of 279 AU/min or 0.1 µM/min, converted using a slope of 2,872 AU/[µM] from linear regression (right; V_{max} = 0.12 µM/min; K_M = 6.9µM; kcat = 1.22 min^{-1}). **e,** DNA binding of MutSβ to immobilized DNA as measured by SPR.

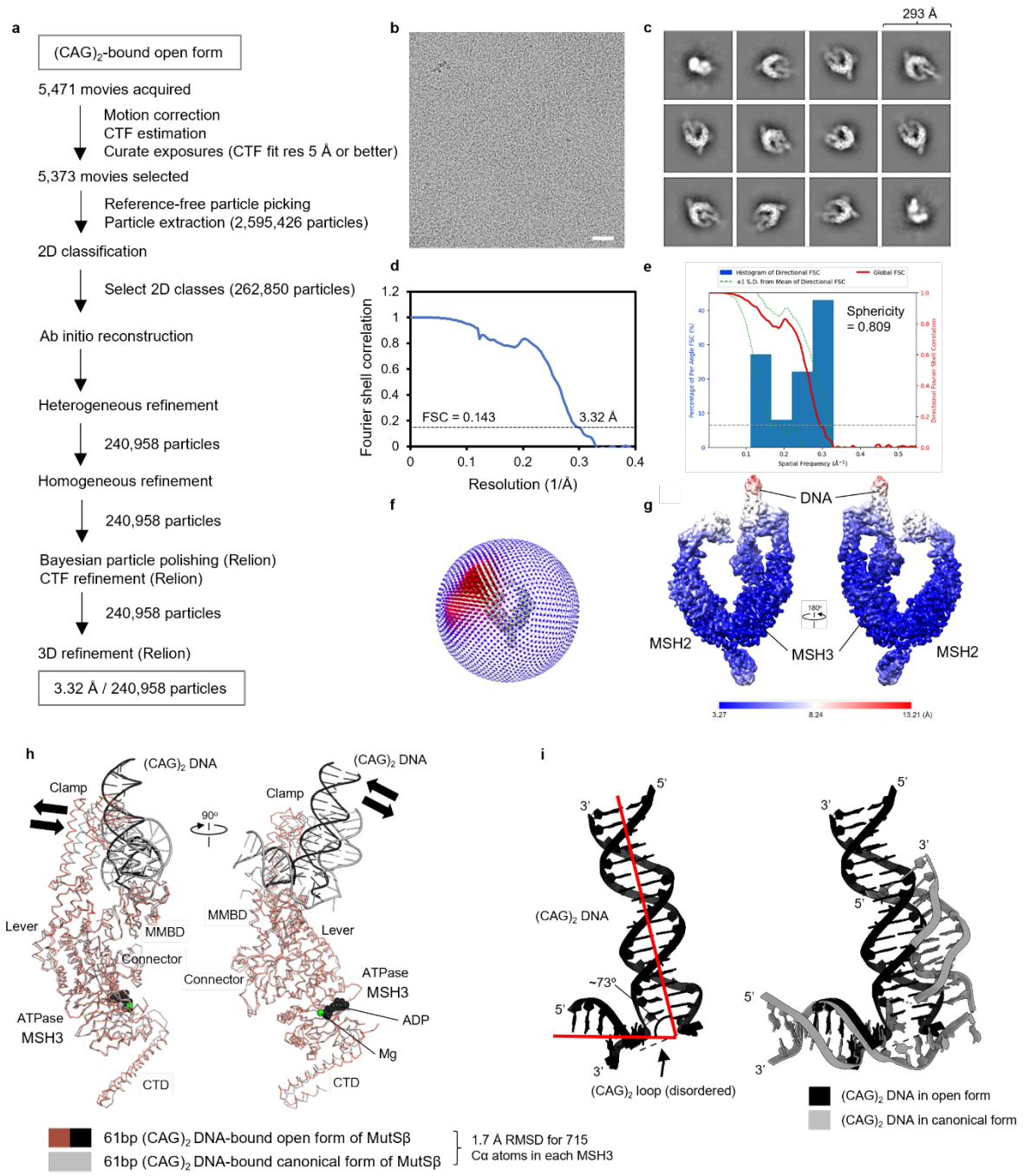


Extended Data Fig. 2 | Cryo-EM analysis of DNA-free open form of MutSβ-ADP complex. a, Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of DNA-free open form of MutSβ. **d,** Gold-standard FSC curve for the density map of DNA-free open form of MutSβ. **e,** 3D FSC[30] plot for the density map of DNA-free open form of MutSβ. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Ribbon diagrams of human MutSβ and *E. coli* MutS (PDB 7OU2)[19] are shown in their DNA-free open form, based on superposition.



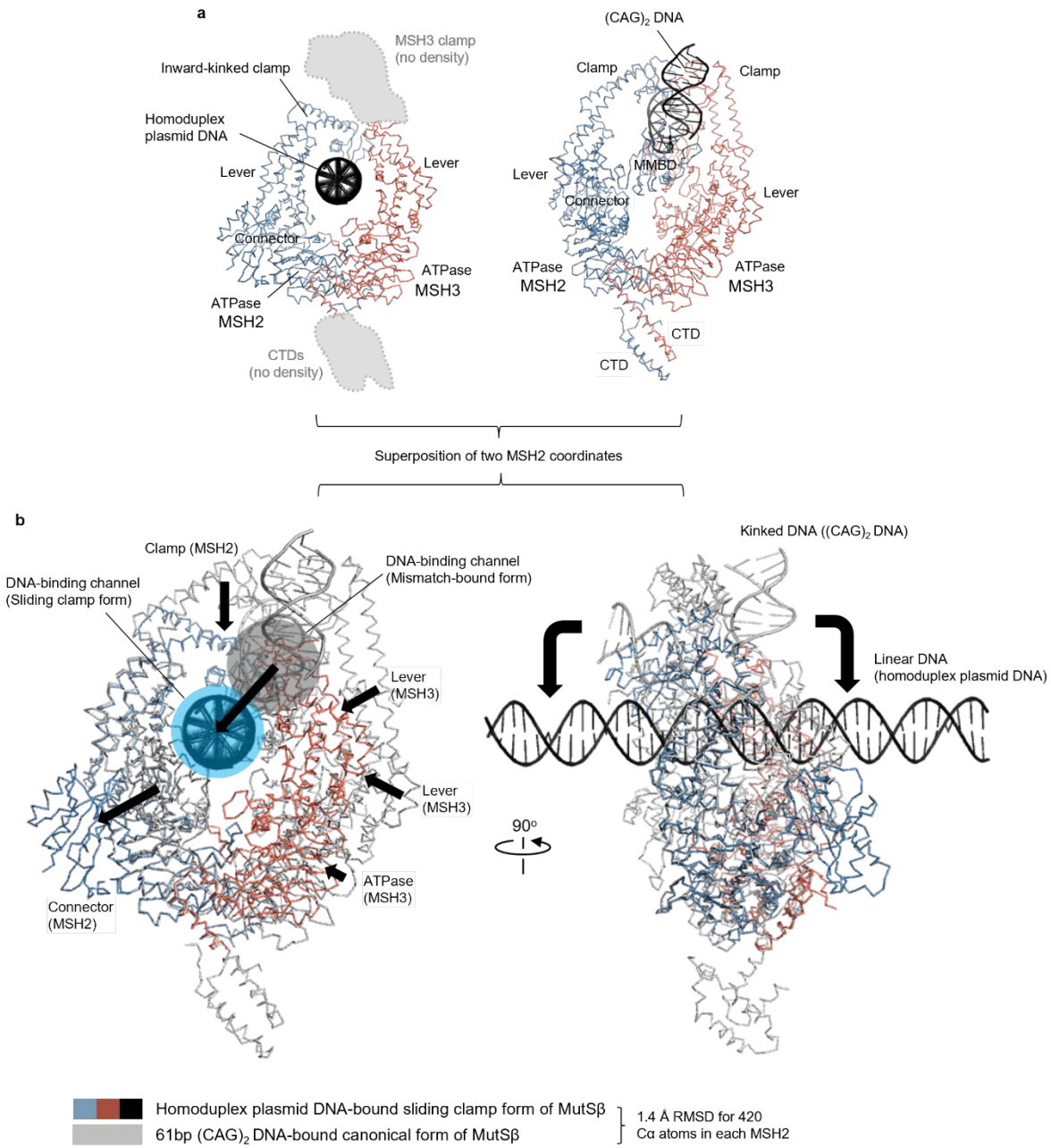
Extended Data Fig. 3 | Cryo-EM analysis of $(CAG)_2$ DNA-bound canonical form of MutSβ-ADP complex. a, Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of $(CAG)_2$ DNA-bound canonical form of MutSβ. **d,** Gold-standard FSC curve for the density map of $(CAG)_2$ DNA-bound canonical form of MutSβ. **e,** 3D FSC[30] plot for the density map of $(CAG)_2$ DNA-bound canonical form of MutSβ. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Ribbon diagrams of the mismatch-bound canonical structure of MutSβ, determined by cryo-

EM and crystallography (PDB 3THZ)[15], are shown based on superposition. **i,** The $(CAG)_2$ DNA structure in the canonical mismatch-bound state of MutSβ, including a close-up view of mis-paired bases overlaid with electron density (transparent pink) is shown.

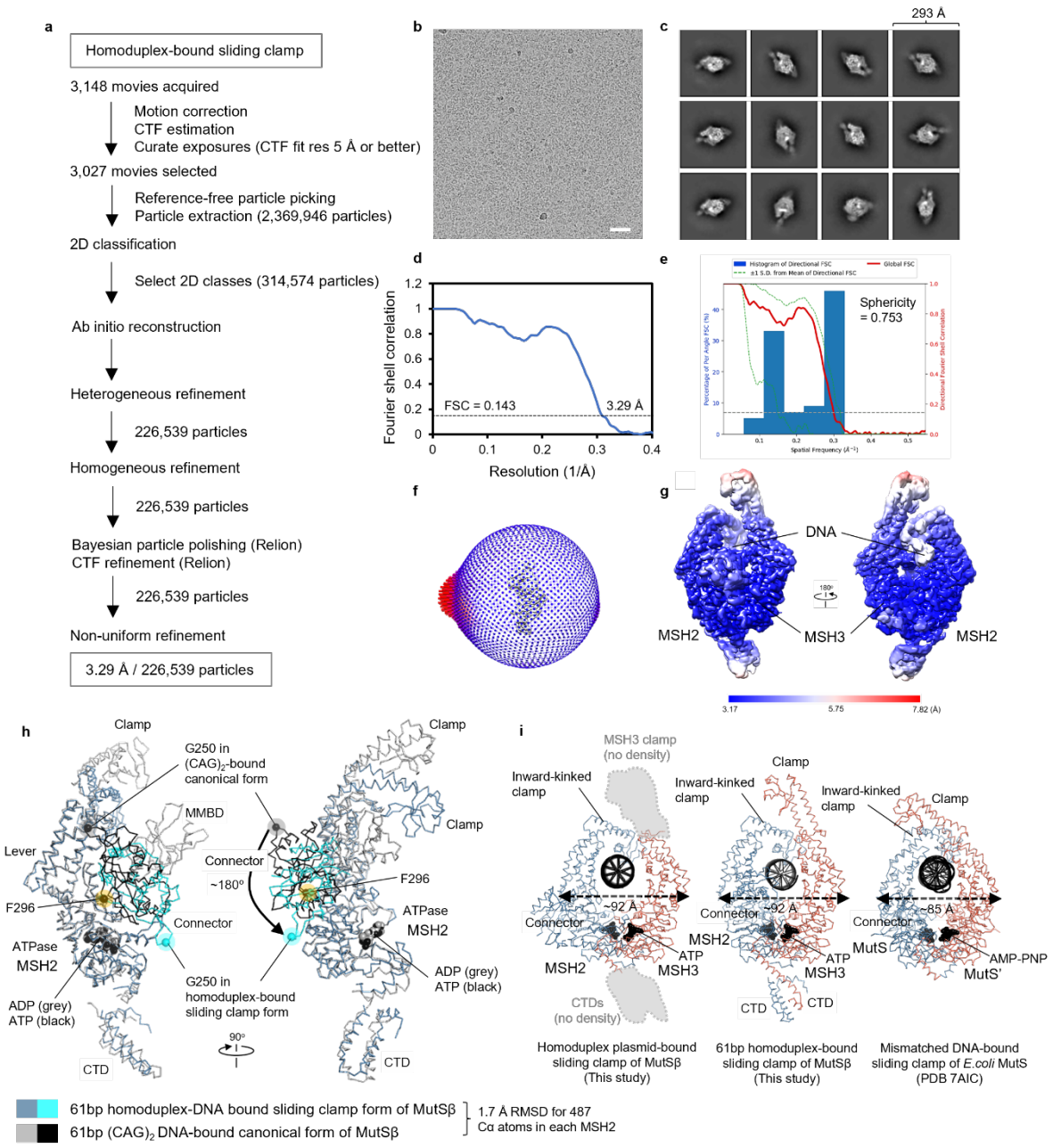


Extended Data Fig. 4 | Cryo-EM analysis of $(CAG)_2$ DNA-bound open form of MutSβ-ADP complex. **a,** Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of $(CAG)_2$ DNA-bound open form of MutSβ. **d,** Gold-standard FSC curve for the density map of $(CAG)_2$ DNA-bound open form of MutSβ. **e,** 3D FSC[30] plot for the density map of $(CAG)_2$ DNA-bound open form of MutSβ. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Ribbon diagrams of the canonical and open forms of MutSβ bound to $(CAG)_2$ DNA are shown based on superposition.

The black arrows indicate substantial movement of MSH3 clamp and DNA in the two different conformations of MutSβ-$(CAG)_2$ DNA complex. **i,** The $(CAG)_2$ DNA structure in the mismatch-bound open state of MutSβ (left) and its overlay with the same DNA in the canonical mismatch-bound state of MutSβ (right) are shown.

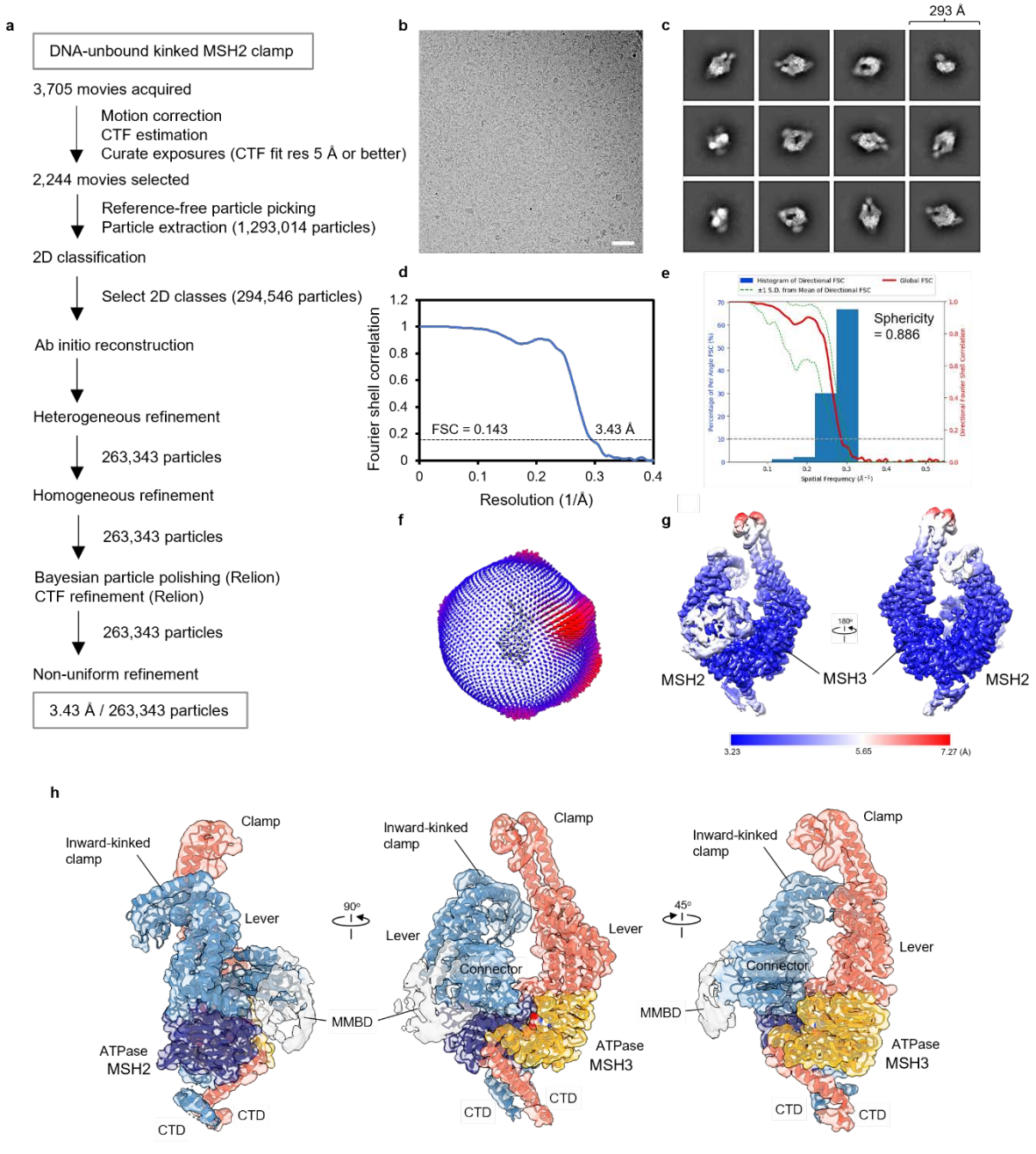


Extended Data Fig. 5 | Structural comparison of homoduplex plasmid DNA-bound sliding clamp form of MutSβ with its canonical form bound to $(CAG)_2$ DNA. a, Ribbon diagrams of human MutSβ in its DNA-bound form are shown for two different conditions: (1) bound to 1.8 kb homoduplex plasmid DNA in the presence of ATP (left) and (2) bound to 61 bp $(CAG)_2$ DNA in the absence of added nucleotide (right), based on superposition. **b,** Two orthogonal views of superimposed cryo-EM structures show conformational differences between two structures. The black arrows indicate domain movements of MSH2, MSH3, and DNA towards the homoduplex plasmid DNA-bound sliding clamp state of Mutβ.

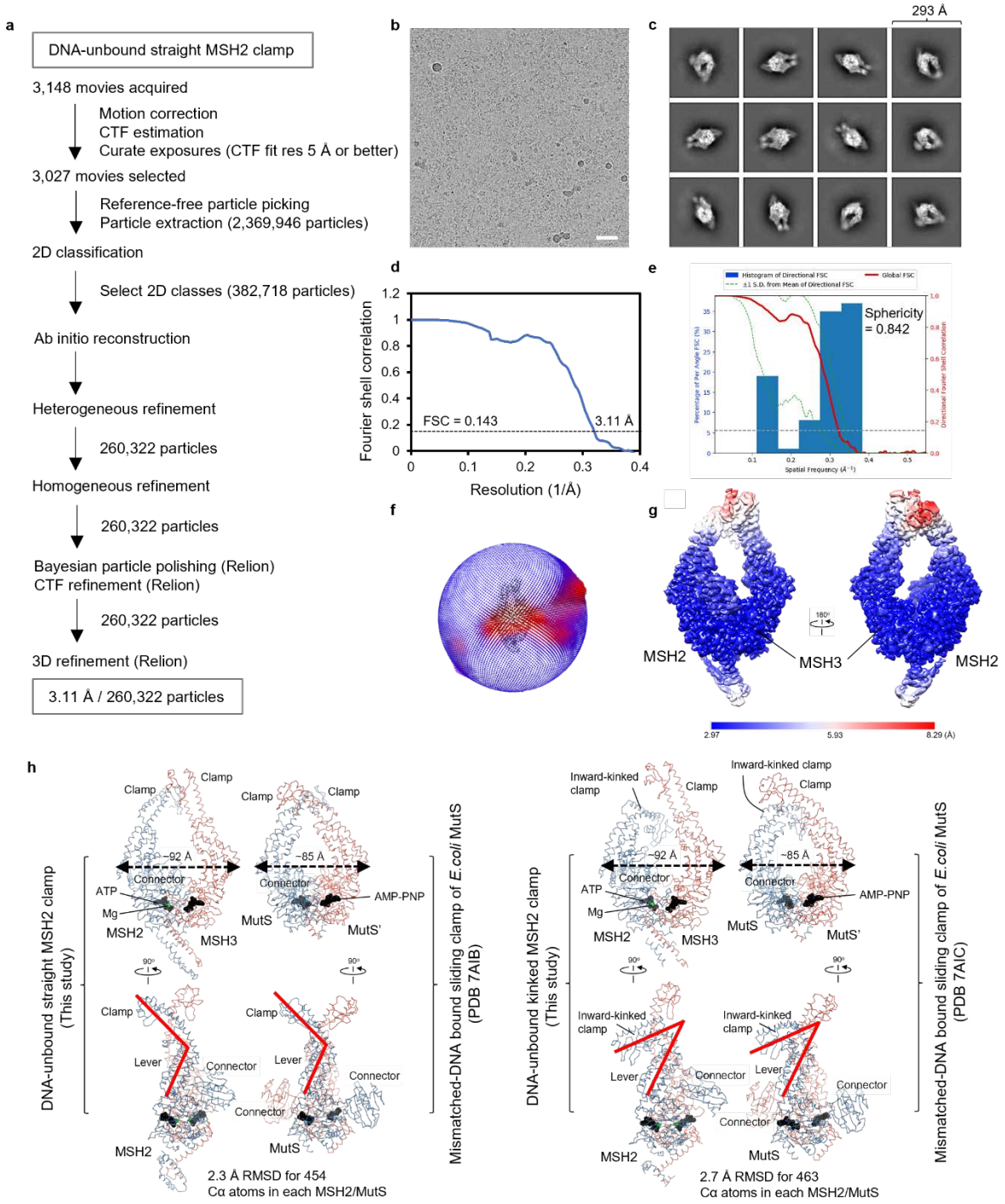


Extended Data Fig. 6 | Cryo-EM analysis of 61 bp homoduplex DNA-bound sliding clamp form of MutSβ-ATP complex. a, Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of 61 bp homoduplex-bound sliding clamp form of MutSβ. **d,** Gold-standard FSC curve for the density map of 61 bp homoduplex-bound sliding clamp form of MutSβ. **e,** 3D FSC[30] plot for the density map of 61 bp homoduplex-bound sliding clamp form of MutSβ. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Ribbon diagrams of the canonical and sliding clamp forms of MutSβ bound to DNA are shown based on superposition. The MSH2 connector rotates

around residue F296 compared to the $(CAG)_2$-bound canonical form relative to other domains. The black arrows indicate substantial domain movement of MSH2 connector. **i,** Ribbon diagrams of human MutSβ and *E. coli* MutS (PDB 7AIC)[18] are shown in their DNA-bound sliding clamp form, based on superposition.

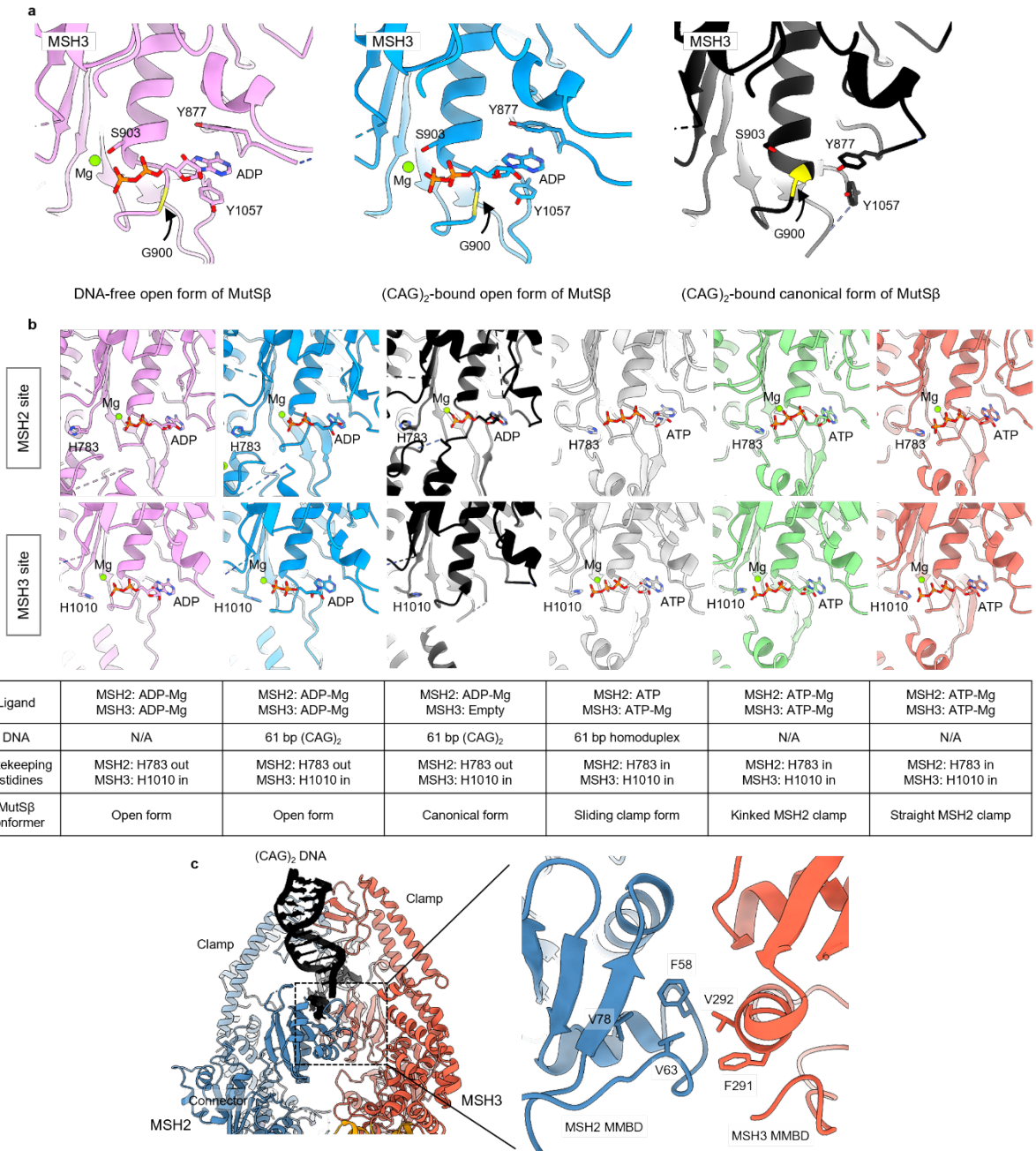


Extended Data Fig. 7 | Cryo-EM analysis of DNA-unbound MutSβ-ATP complex with bent clamp of MSH2. a, Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of DNA-unbound MutSβ-ATP complex. **d,** Gold-standard FSC curve for the density map of DNA-unbound MutSβ-ATP complex. **e,** 3D FSC[30] plot for the density map of DNA-unbound MutSβ-ATP complex. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Three views of the DNA-unbound MutSβ-ATP complex overlaid with the 3.43 Å cryo-EM map are shown. The MSH2 MMBD is not defined, as the limited quality of the electron density is represented by transparent white.



Extended Data Fig. 8 | Cryo-EM analysis of DNA-unbound MutSβ-ATP complex with straight clamps. a, Summary of the image processing workflow. **b,** Representative micrograph. The scale bar in white indicates 37.3 nm. **c,** Representative 2D classes of DNA-unbound MutSβ-ATP complex. **d,** Gold-standard FSC curve for the density map of DNA-unbound MutSβ-ATP complex. **e,** 3D FSC[30] plot for the density map of DNA-unbound MutSβ-ATP complex. **f,** Heat map showing particle orientation distribution. **g,** Local resolution represented by a heat map on the density contour. **h,** Structural

comparison of human MutSβ and *E. coli* MutS homodimer (PDB 7AIB and 7AIC; DNA and MutL are omitted for clarity)[18] are shown in their ATP- or AMP-PNP-bound forms, based on superposition. A sharp kinking of the clamp domain of MSH2 and one subunit of MutS (indicated in red on the right) is shown.



Ligand	MSH2: ADP-Mg MSH3: ADP-Mg	MSH2: ADP-Mg MSH3: ADP-Mg	MSH2: ADP-Mg MSH3: Empty	MSH2: ATP MSH3: ATP-Mg	MSH2: ATP-Mg MSH3: ATP-Mg	MSH2: ATP-Mg MSH3: ATP-Mg
DNA	N/A	61 bp $(CAG)_2$	61 bp $(CAG)_2$	61 bp homoduplex	N/A	N/A
Gatekeeping histidines	MSH2: H783 out MSH3: H1010 in	MSH2: H783 out MSH3: H1010 in	MSH2: H783 out MSH3: H1010 in	MSH2: H783 in MSH3: H1010 in	MSH2: H783 in MSH3: H1010 in	MSH2: H783 in MSH3: H1010 in
MutSβ conformer	Open form	Open form	Canonical form	Sliding clamp form	Kinked MSH2 clamp	Straight MSH2 clamp

Extended Data Fig. 9 | Structural analysis of different conformational states of human MutSβ. a, Ribbon diagrams of the MSH3 site of the DNA-free open form of MutSβ (left) and its $(CAG)_2$ DNA-bound form (middle and right) are shown based on superposition. The MSH3 G900 residue in the Walker A motif is highlighted in yellow. In the canonical form of MutSβ complexed with $(CAG)_2$ DNA, the MSH3 Y877 residue occupies the adenine-binding site, leading to a lack of nucleotide at the MSH3 nucleotide-binding pocket. **b,** Structural comparison of the conformation of gatekeeping histidine residues at the nucleotide-binding pockets of the MSH2 and MSH3 ATPase cores in the six different conformational

states of MutSβ. **c,** Hydrophobic interface between MSH2 and MSH3 MMBDs in the $(CAG)_2$ DNA-bound canonical form of MutSβ.

Methods

Protein expression and purification

Pellet from baculovirus infected Sf9 cells expressing the human MutSβ complex were resuspended in lysis buffer (25 mM HEPES pH 8, 1 M NaCl, 10 % glycerol, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease Inhibitors on ice. Cell lysis was achieved by Turrax homogenization on ice. The lysate was centrifuged at 30,000 g for 10 minutes at 10 °C. The supernatant (clarified lysate) was collected. The clarified lysate was loaded onto a 20 mL DEAE column pre-equilibrated in buffer A (25 mM HEPES pH 8, 1 M NaCl, 10 % glycerol, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The bound proteins were eluted from the column by a step of 100 % buffer B (25 mM HEPES pH 7.6, 2 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The fractions were analyzed by SDS-PAGE and the fractions containing the MSH2-MSH3 complex were pooled. The pooled fractions from the previous chromatography step were diluted to a final NaCl concentration of 250 mM. The resulting sample was loaded onto a 5 mL Heparin column pre-equilibrated in buffer C (25 mM HEPES pH 7.6, 250 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The bound proteins were eluted from the column by a gradient (0-100 % over 5 column volumes) of buffer D (25 mM HEPES pH 7.6,1 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The fractions were analyzed by SDS-PAGE and the fractions containing the MSH2-MSH3 complex were pooled. The pooled fractions from the previous chromatography step were diluted to a final NaCl concentration of 100 mM. The resulting sample was loaded onto a 6 mL Resource Q column pre-equilibrated in buffer E (25 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The bound proteins were eluted from the column by a gradient (0-100 % over 5 column volumes) of buffer D (25 mM HEPES pH 7.6, 1 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The fractions were analyzed by SDS-PAGE and the fractions containing the MSH2-MSH3 complex were pooled. The pooled fractions from the previous chromatography step were diluted to a final NaCl concentration of 100 mM. The resulting sample was loaded onto a 1 mL Mono Q column pre-equilibrated in buffer E (25 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The bound proteins were eluted from the column by a gradient (0-100% over 5 column volumes) of buffer D (25 mM HEPES pH 7.6, 1 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol) supplemented with protease inhibitors. The fractions were

analyzed by SDS-PAGE and the fractions containing the MSH2-MSH3 complex were pooled. The pool from the previous purification step was concentrated to a volume of 10 mL and loaded onto a S200 26/60 column pre-equilibrated in buffer G (25 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM TCEP, 5% glycerol). The fractions were analyzed by SDS-PAGE and the fractions containing the MSH2-MSH3 complex were pooled. The fractions pooled from the size-exclusion chromatography step were concentrated using an Amicon concentrator (MWCO 50 kDa) to a final concentration of 13 mg/mL.

Generation of dsDNA substrates

The 61 bp single-stranded DNAs were synthesized by Metabion (Metabion International AG, Planegg, Germany), the sequences can be found in table 2. Equimolar amounts of each ssDNA molecule were mixed to a final concentration of 500 µM each in TE Buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). Annealing was performed by heating the DNA sample to 95 °C for 2 minutes and subsequently cooling from 95 °C to 20 °C with a 1 °C/min ramp (Labcycler, SensoQuest GmbH). In the case of the biotinylated dsDNA substrates, pre-saturation of Streptavidin with Biotin was achieved by mixing Streptavidin (AnaSpec Inc.) to Biotin (Merck Millipore) in a molar ratio of 1:3.5 at 500 µM Streptavidin and 1,750 µM Biotin in 100 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, incubating for 3 h at ambient temperature. Subsequently, biotinylated dsDNA samples were end-blocked by addition of 1 eq. of 61 bp 5'-biotinylated homoduplex dsDNA to 16 eq. of pre-saturated Streptavidin at a total volume of 36 µL, incubated over night at 4 °C. This corresponds to a final DNA concentration of 27.78 µM with a final Streptavidin concentration of 444.44 µM. Purification of the Strep-capped DNA was achieved by loading the sample mixture on a Pierce™ Strong Anion Exchange Spin Column and washing with 400 µL of 50 mM Tris pH 8.0, 5 mM $MgCl_2$ with increasing NaCl concentration, starting from 200 mM NaCl. Spin columns were equilibrated and handled according to manufacturer's instructions. Columns were conditioned twice using 400 µL of 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM $MgCl_2$. Samples were pre-washed by addition of 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$ with increased NaCl concentrations from 200 to 600 mM in 100 mM increments. From 600 mM to 800 mM, the sodium chloride concentration was increased in 25 mM increments to elute the Streptavidin-end-blocked DNA from the column.

Absorbance spectra of elution fractions were measured and pooled according to an absorbance ratio of A260/A280 of 1.00 to 1.65.

Table 2. Sequences of the 61 bp dsDNA substrates used in this study

dsDNA substrate	Strand	Sequence	Modification
61 bp homoduplex, end-capped	Top strand	5'-CTGAAGCTTAGCTTAGGATCATCGAGGATCAGCTCGGTGCAATTCAGCGGTACCCAATTC-3'	5' Biotin
	Bottom strand	5'-GAATTGGGTACCGCTGAATTGCACCGAGCTGATCCTCGATGATCCTAAGCTAAGCTTCAG-3'	5' Biotin
61 bp $(CAG)_2$, end-capped	Top strand	5'-CTGAAGCTTAGCTTAGGATCATCGAGGATCCAGCAGAGCTCGGTGCAATTCAGCGGTACCCAATTC-3'	5' Biotin
	Bottom strand	5'-GAATTGGGTACCGCTGAATTGCACCGAGCTGATCCTCGATGATCCTAAGCTAAGCTTCAG-3'	5' Biotin
61 bp $(CAG)_2$	Top strand	5'-CTGAAGCTTAGCTTAGGATCATCGAGGATCCAGCAGAGCTCGGTGCAATTCAGCGGTACCCAATTC-3'	-
	Bottom strand	5'-GAATTGGGTACCGCTGAATTGCACCGAGCTGATCCTCGATGATCCTAAGCTAAGCTTCAG-3'	-

The linear plasmid substrate was generated by PCR from a pET16b vector to generate a 1,783 bp fragment comprising the ORI and AmpR gene and promoter of the pET16b vector.

SEC-MALS analysis

For the determination of the molecular mass of MutSβ, SEC-MALS was applied. Therefore, an Ultimate 3000 HPLC system (Thermo Fisher Scientific) equipped with a degasser, a quaternary pump, an autosampler (kept at 10 °C) and a VWD detector was coupled to a miniDawn TREOS II 3-angle light-scattering detector and an Optilab RI detector (Wyatt Technology). The HPLC system was controlled by Chromeleon 7.3.1 (Thermo Fisher Scientific) and the MALS as well as the RI detector with Astra 7.3.2.21 (Wyatt Technology). For the analysis the protein solution (100 µL, 1 mg/mL) was injected on a Superdex 200 Increase 10/300 GL (30 cm × 10 mm, 8.6 µm particle size) SEC column. As a mobile phase the following buffer was used: 25 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 0.5 mM EDTA, pH7.5. All measurements were conducted at room temperature. Data analysis was performed using Chromeleon 7.3.1 (HPLC-UV data) and Astra 7.3.2.21 (MALS data).

ATPase activity assay

ATPase activity was determined by ADP-Glo Kinase Assay Kit (Promega, V9102), following the manufacturer's protocol. Initially, a 384-well plate (Corning® 384-well Low Volume White, product number 4513) was prepared for reaction stop. All wells were prefilled with 4 µl ADP-GloTM reagent. Storage at room temperature. Then, a master mix plate was prepared for the ATPase reaction. A serial two-fold dilution of ATP was prepared in Eppendorf tubes (maximum ATP conc. = 200 µM, 2-fold dilution series, eight titration points). 30 µl of ATP dilutions were filled in the respective columns of Row A and Row B of a second 384-well plate (Greiner product number 781270) with columns 22-24 = maximum ATP concentration, columns 1-3 = 0 mM ATP. MutSβ samples were diluted in reaction buffer to a concentration of 200 nM in an Eppendorf tube. The reaction was started by transferring 30 µl of MutSβ samples to the preplated ATP dilution series in Row A using a Multipette (Eppendorf), hereafter referred to as "reaction mix". For the background control section, 30 µl buffer were added to row B, hereafter name "background control mix". Prepared plates were thoroughly mixed using MixMate (Eppendorf), followed by incubation at 37 °C. Final enzyme concentration was 100 nM, with a maximum ATP concentration of 100 µM. For the 0 min reaction time point, 4 µl ATP from the dilution series prepared for the master mix plate (Eppendorf tubes with maximum ATP conc. = 200 µM) were added to row A and row I on reaction stop plate. Transfer of 4 µl of prediluted 200 nM MutSβ protein solution to row A or of buffer to row I, followed by thorough mixing using MixMate (Eppendorf). For all further reaction time points, 8 µl of ATP-ATPase reaction mix were transferred from row A on master mix plate to respective row (B to H) on reaction stop plate using a Matrix™ Multichannel Pipette (Thermo Fisher). Additionally, transfer of 8 µl background control mix from row B of master mix plate to respective row (J to P) on reaction stop plate, followed by thorough mixing using MixMate (Eppendorf). Samples were taken at 15, 30, 45, and 60 min timepoints. Subsequently, the detection reaction was performed by 40 min incubation at room temperature with ADP-Glo reagent, followed by addition of 6 µl kinase detection reagent. Plates were thoroughly mixed using MixMate (Eppendorf). After another incubation for 40 min at room temperature, the readout of luminescence signal was recorded using a LUMIstar Omega (BMG Labtech). For each run, a standard curve was created that represents the luminescence signal corresponding to the ATP converted to ADP. Therefore, ATP and ADP stock solutions that was provided with the Assay kit were mixed in different ratios to represent 100 % conversion at 10 µM ADP/0 µM ATP

and 0 % conversion at 0 µM ATP/10 µM ADP, as well as linear combinations of both ADP and ATP in between. Collected data was first corrected for background control signal by subtraction from the reaction mix signal at the respective time point. The background subtracted ADP-Glo Signal was plotted versus time for each concentration of ATP and fitted to a linear regression model. Linearity was granted for R-squared ≥ 0.95. At low ATP concentrations, this criterion was not met for long reaction time points as all ATP was consumed such that a plateau was reached. These data points were excluded from data evaluation. The slope represents the reaction velocity in luminescence counts/min for each ATP concentration. The slope was plotted versus ATP concentration and fitted to the Michaelis Menten Equation (Equation 1).

$$Y = \frac{(V_{max} \times X)}{(K_m + X)}$$

Equation 1: K_m represents the ATP concentration at half maximal reaction velocity. V_{max} represents the maximum reaction velocity in luminescence counts/min. The standard curve was fitted to a linear regression and the slope in luminescence counts/µM was used to convert the V_{max} value to µM/min. The turnover number k_{cat} in min^{-1} was determined by dividing the V_{max} value by the enzyme concentration.

Surface plasmon resonance

SPR binding experiments were performed on a Biacore 8K instrument (Cytiva) at T = 25°C. Biotinylated DNA oligos were annealed to dsDNA in solution and subsequently immobilized on a streptavidin-coated SA chip Series S (Cytiva). Immobilization was performed using a DNA concentration of 1 nM. Immobilization levels were kept between 10 and 30 RUs, depending on DNA oligo length. Running buffer was 25 mM HEPES, pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.05% Tween and 0.05% BSA. Samples were kept at 20°C until injection. Binding studies with MutSβ are performed using 2-fold dilution series. Experimental parameters for kinetics experiments were chosen as follows: MutSβ in presence of ATP [1.56 – 200 nM] in absence of ATP [0.39 – 50 nM] was injected for 120 s and dissociation was monitored for 300 s at a flow rate of 30 µl/min. After each MutSβ injection, the chip surface was regenerated with a solution of 0.5% SDS. For assessment of ATP effects, running buffer was supplemented with 0.1 mM ATP. Kinetics data from SPR measurements were analyzed using

Biacore Insight Evaluation software v4.0.8.20368. Dissociation constants (KD_{eq}) were obtained from steady state affinity global fitting. Information on kinetic rate constants and residence times were extracted using a 1:1 binding model. During the experiments, we checked that calculated maximum binding (theoretical R_{max}) was in line with immobilized oligonucleotide. Graphs were created using GraphPad Prism v7.05.

Cryo-EM sample preparation

Four microliters of MutSβ at a concentration of 3 µM were mixed with DNA (3.5 µM of a 61 bp $(CAG)_2$ DNA, 50 nM of linearized plasmid DNA, or 3.5 µM of a 61 bp homoduplex DNA). No DNA or nucleotide were added to the protein for the DNA-free apo form of MutSβ. For the DNA-unbound kinked and straight MSH2 clamp forms and DNA-bound sliding clamp forms, 1 mM ATP was added to the protein-DNA mixtures. All cryo-EM samples mixed with 0.006 % Tween-20 were applied onto a glow-discharged R1.2/1.3 cooper or gold 300-mesh holey carbon grid (Quantifoil) and were immediately frozen in a liquid ethane-propane mixture using a Vitrobot Mark IV (Thermo Fisher Scientific) with the settings at 4 ºC, 95 % humidity, and 3-6 seconds of blot time.

Cryo-EM data processing and image processing

All cryo-EM data were acquired at the cryo-EM facility at Proteros Biostructures GmbH (Table 1). The data were acquired on a Glacios transmission electron microscope (Thermo Fisher Scientific) operated at 200 keV and equipped with a Falcon IV electron detector. The data were collected at a nominal dose of 55.95-59.93 $e^-/Å^2$ with 40 frames per movie and a pixel size of 0.9142 Å. The target defocus range was between 0.7 and 2.5 µm. A total of 3,712, 5,471, 5,471, 571, 3,148, 3,705, and 3,148 movies were collected for the DNA-free apo form, 61 bp $(CAG)_2$-bound open form, 61 bp $(CAG)_2$-bound canonical form, ~1.8 kb linearized homoduplex plasmid-bound sliding clamp form, 61 bp homoduplex DNA-bound sliding clamp form, DNA-unbound kinked MSH2 clamp, and DNA-unbound straight MSH2 clamp, respectively. All cryo-EM data were recorded using EPU (Thermo Fisher Scientific). The individual dataset was imported into Relion[22], and the movie frames were aligned using Relion's MOTIONCOR implementation. The motion-corrected micrographs were then fed into cryoSPARC[23], and the CTF

(Contrast Transfer Function) was estimated using CTFFIND. Micrographs that were unsuitable for image analysis (e.g. due to significant drift or heavy contamination with crystalline ice) were removed by manual inspection. Particles were picked with cryoSPARC's reference-free blob picker from the selected micrographs. An initial set of particles was extracted, and two rounds of 2D classification were performed to clean up the individual dataset. Ab initio model generation was carried out using cryoSPARC, followed by hetero refinement, resulting in one highly populated class and several "junk" classes. A subset of particles attributed to the best class was subjected to homo refinement using cryoSPARC, followed by Bayesian polishing and CTF refinement using Relion. The final set of particles was refined using either cryoSPARC's non-uniform refinement or Relion's 3D refinement.

Model building and refinement

To build atomic structures of six different conformations of MutSβ, the crystal structure of the human MutSβ-DNA complex (PDB 3THZ)[15] was fitted into the refined 3D reconstruction maps using UCSF (University of California at San Francisco) Chimera[24]. The structures were then manually rebuilt in COOT[25] to fit the densities, guided mainly by bulky side-chain residues for sequence assignment. The final atomic structures were refined using CCPEM Refmac[26] and validated using MolProbity[27]. Structure analysis was performed in COOT[25], and figures were prepared using PyMOL[28] and ChimeraX[29].

To generate a low-resolution model of the homoduplex plasmid DNA-bound sliding clamp form of MutSβ, the 3.43 Å cryo-EM structure of the DNA-unbound kinked MSH2 clamp structure was rigid-body fitted into the map. The MSH2 connector was individually fitted into the density. The MSH3 clamp and two CTD domains, which are not defined in the low-resolution map, were removed. Next, the ideal B-form dsDNA was generated using COOT[25] and fitted into the density of the central channel of MutSβ.

Data availability

Cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-16969 (open form), EMD-16971 (mismatch-bound open form), EMD-16964 (canonical mismatch-bound form), EMD-16972 (homoduplex-bound sliding clamp form), EMD-16975 (plasmid-bound sliding clamp form), EMD-16973 (DNA-unbound kinked MSH2 clamp), and EMD-16974 (DNA-

unbound straight MSH2 clamp). Model coordinates have been deposited in the Protein Data Bank (PDB) under accession codes 8OM5 (open form), 8OM9 (mismatch-bound open form), 8OLX (canonical mismatch-bound form), 8OMA (homoduplex-bound sliding clamp form), 8OMO (DNA-unbound kinked MSH2 clamp), and 8OMQ (DNA-unbound straight MSH2 clamp).

Acknowledgements

We thank Simon Noble for comments.

Author contributions

M.T., J.-H.L., H.W., R.R.I., N.P., D.P.F, T.S., M.F. and B.P. conceived the study. G.T. cloned and purified all described proteins and prepared linearized plasmid DNA. H.D. prepared end-blocked dsDNA samples and performed all biophysical experiments in this study. A.S. performed ADPglo experiments. J.-H.L. carried out grid freezing, data collection and data processing of all described structures in this study. All authors were involved in analyzing data and manuscript writing.

Competing interests

This work was supported by the nonprofit CHDI Foundation Inc. CHDI Foundation is a nonprofit biomedical research organization exclusively dedicated to collaboratively developing therapeutics that substantially improve the lives of those affected by Huntington's disease. CHDI Foundation conducts research in a number of different ways; for the purposes of this manuscript, all research was conceptualized, planned, and directed by all authors and conducted at Proteros.

Correspondence and requests for materials should be addressed to Brinda Prasad (brinda.prasad@chdifoundation.org).