

Molecular determinants for dsDNA translocation by the transcription-repair coupling and evolvability factor Mfd

Christiane Brugger¹, Cheng Zhang^{2,4}, Margaret M. Suhanovsky^{1,4}, David D. Kim¹, Amy N. Sinclair¹, Dmitry Lyumkis^{2,3} & Alexandra M. Deaconescu¹✉

Mfd couples transcription to nucleotide excision repair, and acts on RNA polymerases when elongation is impeded. Depending on impediment severity, this action results in either transcription termination or elongation rescue, which rely on ATP-dependent Mfd translocation on DNA. Due to its role in antibiotic resistance, Mfd is also emerging as a prime target for developing anti-evolution drugs. Here we report the structure of DNA-bound Mfd, which reveals large DNA-induced structural changes that are linked to the active site via ATPase motif VI. These changes relieve autoinhibitory contacts between the N- and C-termini and unmask UvrA recognition determinants. We also demonstrate that translocation relies on a threonine in motif Ic, widely conserved in translocases, and a family-specific histidine near motif IVa, reminiscent of the “arginine clamp” of RNA helicases. Thus, Mfd employs a mode of DNA recognition that at its core is common to ss/ds translocases that act on DNA or RNA.

¹Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02903, USA. ²Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, CA 92093, USA. ³Department of Computational and Structural Biology, The Scripps Research Institute, La Jolla, CA 92093, USA. ⁴These authors contributed equally: Cheng Zhang, Margaret M. Suhanovsky. ✉email: alexandra_deaconescu@brown.edu

Given that DNA serves as a track for multiple essential molecular machines, including those carrying out replication, transcription, and repair, the movement of nucleic-acid translocases along single-stranded (ss) or double-stranded (ds) DNA is highly regulated spatiotemporally¹. This control has a central role in development, neurodegeneration, aging, and the diseased state², and despite advances, still remains insufficiently understood. Translocation on dsDNA has been challenging to study structurally because many DNA motors have little sequence specificity, are difficult to trap within a crystal, and translocation on dsDNA, as opposed to on ssDNA, leads to no detectable product. These challenges are compounded by the finding that translocases often function while coupled to macromolecular machines. They also display varied processivity and coupling to ATP hydrolysis, and while featuring conserved sequence motifs, have mechanisms of action that are modulated by accessory domains in family-specific ways, evading facile generalizations.

Transcription-repair coupling factors (TRCFs) are large superfamily 2 ATPases that mediate the preferential repair of the transcribed DNA strand (aka transcription-coupled DNA repair, TCR) in organisms ranging from bacteria to humans^{3–5}. These factors provide a complex but poorly understood example of regulated translocation on dsDNA. At its core, TCR appears universally conserved, and relies on (1) the ability of transcription-repair coupling ATPases to release damage-stalled RNA polymerases (RNAPs) off the nucleic-acid template owing to ATP-driven translocation on dsDNA upstream of the transcription bubble, and (2) on their ability to recruit nucleotide excision repair (NER) machinery⁶. In bacteria, a single protein, Mfd (aka TRCF) is necessary and sufficient for the coupling process^{7,8}. Although additional pathways have been implicated in TCR⁹, Mfd remains the only bacterial factor for which both RNAP release and repair enzyme recruitment functions have been demonstrated^{3,8,10–14}. Critically, Mfd associates with RNAP in cells even in the absence of exogenous DNA damage¹⁵, it decreases class II transcriptional pausing¹⁶, promotes strand-specific repair “at a distance” downstream of a transcriptional pause site¹⁷, and dissociates transcription elongation complexes (TECs) stalled not only by DNA damage, but also by protein roadblocks¹⁸, including replication forks colliding head-on with the transcription machinery¹⁹. Thus, given these different contexts in which it acts on TECs (Fig. 1), Mfd is more appropriately viewed as a general transcription factor. Paradoxically, under certain conditions, Mfd acts as an evolvability factor, promoting hypermutation, the accelerated evolution of lagging-strand genes^{20,21} and the rapid development of resistance to multiple, unrelated classes of antibiotics^{22–25}. This makes Mfd an attractive target for the development of a broad-spectrum anti-evolution drug, which could be administered in combination with well-characterized antibiotics to curtail the worldwide crisis of antimicrobial resistance.

Mfd translocation on dsDNA is central to all Mfd functions. This is an ATP-dependent^{13,26–29} and regulatable process¹⁶, which it shares with chromatin remodelers belonging to the same superfamily 2 (SF2) of ATPases. Like chromatin remodelers, Mfd binds and remodels a large macromolecular assembly—the TEC, composed of core RNAP, DNA scaffold, and nascent RNA. TECs activate the Mfd translocase to a level at which processive translocation can readily be detected using classic biochemistry^{26,30}. The structural changes underlying this activation are likely complex and may involve multiple steps²⁷ and interlocked structural elements^{29,31}. Mfd is composed of an ATP-dependent motor core composed of domains D5 and D6 as well as six ancillary domains, including D2 and D7 that pack against each other as an inhibitory “clamp” to restrain the ATPase

(Fig. 2a–c) and mask binding determinants for recruitment of UvrA, a component of early NER^{29,31}. Conformational changes in full-length *Escherichia coli* Mfd occurring during its functional cycle have remained speculative since only a single structure—that of nucleotide-free *E. coli* Mfd³¹—has been reported in peer-reviewed literature. Clamp opening was originally proposed to be prerequisite for translocation and dependent on the interaction with RNAP³⁰ (Fig. 1). Recent and more-sensitive single-molecule studies have demonstrated that in fact Mfd can also translocate on naked DNA, albeit with limited processivity¹⁶, suggesting that Mfd exists in a dynamic conformational equilibrium that can be shifted by TEC binding. However, provided that ATP is supplied, Mfd can make excursions to the translocation-competent form even in the absence of RNAP¹⁶, suggesting that Mfd locates its targets not only by 3D diffusion, but also a more efficient 1-D search along dsDNA, possibly colliding with stalled/paused TECs ahead of it¹⁶. This scenario can lead to two outcomes: RNAP rescue by forward translocation (Fig. 1a) or dissociation (Fig. 1b–d), occurring when RNAP encounters severe hindrances to forward movement¹⁶. It is important to note that Mfd processivity on naked DNA is substantially lower than that of Mfd bound to a RNAP that has been released off the DNA chains, but remains associated with Mfd via interactions with domain D4, and possibly other unknown binding sites¹⁶.

Despite intense research in the last decade, the mechanochemistry of TRCFs remains poorly defined. How do TRCFs recognize DNA and what are the conformational changes associated with translocation on dsDNA? Here, we shed light on these questions by defining the path of DNA across Mfd using electron cryo-microscopy (cryo-EM), structure-guided mutagenesis, and functional assays. We demonstrate that in the absence of RNAP, independent, single substitutions within the two lobes of the Mfd motor core lead to severe DNA-binding defects, and that translocation relies on the existence of a moderate and a strong binding state that correspond to the ATP- and ADP•AlF_x-bound states, respectively. We pinpoint conserved residues within translocase motifs Ic, IV, V, and Vb as contributing to DNA loading, and identify an evolutionary conserved threonine in motif Ic and a family-specific histidine near motif IVa as critical for translocation. We also show that DNA binding to Mfd in the presence of transition state analog ADP•AlF_x leads to large-scale rotation of the UvrB homology module, resulting in disruption of the D2–D7 clamp as well as repositioning of domain D3 of hitherto unknown function. This leads to the breaking of contacts of D3 with D5 and D6 and the establishment of new contacts with D7. Our study provides the first glimpse of a substrate-bound bacterial TRCF, and suggests that these ancient ATPases utilize a mode of DNA recognition that at its core, is common to ss/ds nucleic-acid translocases.

Results

Mfd binds dsDNA tightly with transition state analog ADP•AlF_x. For processive autonomous movement on dsDNA (i.e., not dependent on tethering to RNAP) to occur, Mfd must cycle through several states, and must possess at least two DNA contact points, utilized differentially during the ATPase cycle. Mfd binding to DNA with non-hydrolysable ground ATP mimics, ATPγS and AMPPNP has been well documented^{14,32}. ATP hydrolysis leads to DNA dissociation, e.g., Mfd translocation towards the end of the fragment^{3,26,33}.

Importantly, although ATP binding is thought to reposition D5/D6 and align the ATPase motifs for catalysis³¹, there are no reports of a second state tightly associating with DNA. This would be critical because it would provide the necessary powerstroke for unidirectional movement. We have thus asked

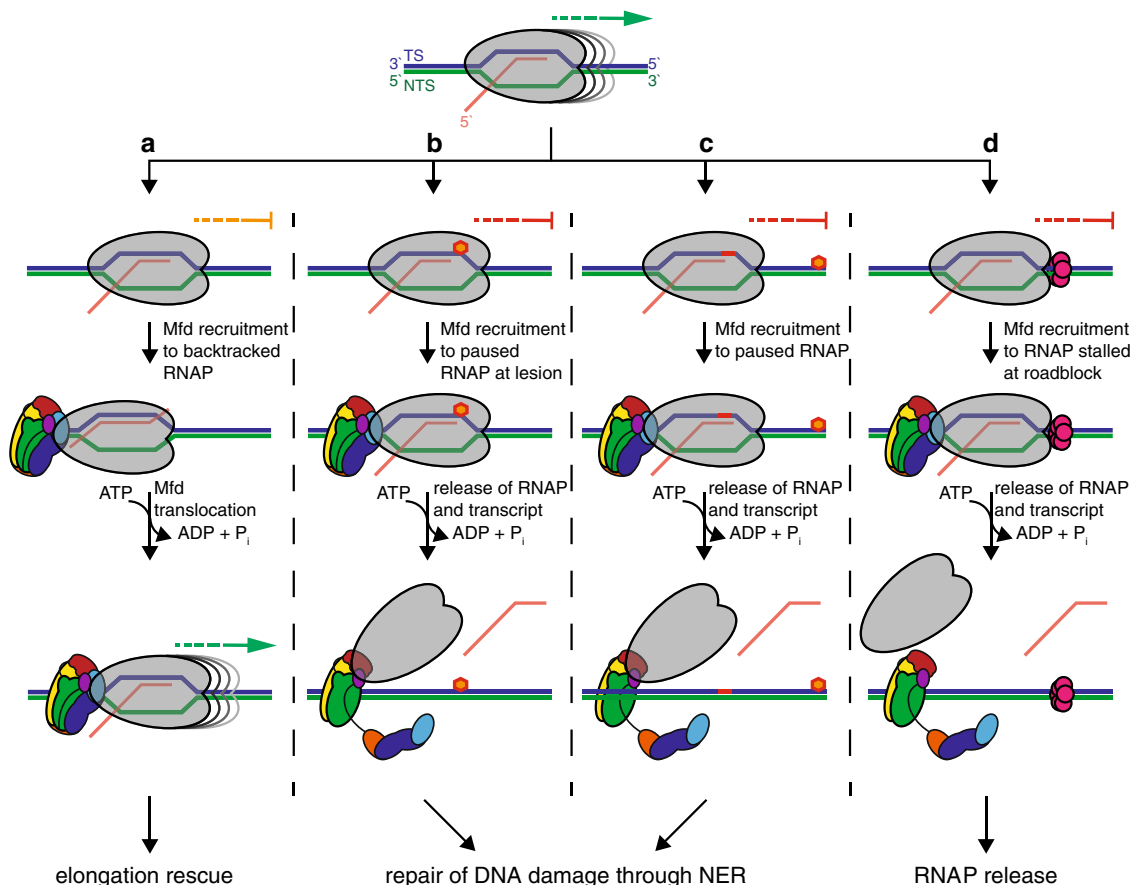


Fig. 1 All Mfd functions are dependent on translocation on DNA or action upstream of transcription elongation complexes. Shown at the top is a schematic of an elongating RNAP (gray) with labeled nucleic-acid moieties (RNA, red; NTS, non-template strand, green; TS, template strand, blue) that can either become stalled or temporarily paused as indicated in **a–d**. **a** In Mfd-dependent rescue of class II transcriptional pausing, Mfd binds to backtracked RNAPs and promotes their forward translocation thereby rescuing transcript elongation^{16,26}. **b** In canonical TCR, RNAP becomes stalled at a lesion (orange hexagon) on the TS and recruits Mfd (colored by domain, with D1a and D1b in blue, D2 in cyan, D3 in orange, D4 (RID) in magenta, D5 in yellow, D6 in green and D7 in red), either through 3D-diffusion or 1D-diffusion, enabled by a catch-up-then release mechanism dependent on ATP hydrolysis¹⁶, disruption of the D2–D7 clamp, RNA release, and UvrAB recruitment^{29,31}. **c** In repair “at a distance”, Mfd is recruited to RNAP paused at class II pause signals (red line), which is first released off the nucleic-acid chains and acts as a processivity factor for Mfd to translocate toward downstream lesions in the TS, and initiate strand-specific repair¹⁷. **d** When RNAPs collide with protein roadblocks head-on, including replisomes, Mfd will release RNAP off the nucleic-acid chains, thereby freeing the DNA for other DNA-based processes^{18,19}.

whether trapping such an alternate state might be possible using ADP and ADP•AlF_x. During ATP hydrolysis, the γ -phosphate of ATP passes through a trigonal planar configuration with three equatorial oxygen atoms and two axial ligands, an oxygen from the β -phosphate and the attacking water molecule. To mimic this state, we used Mg²⁺•ADP•AlF_x, the most extensively employed analog of the ATP hydrolysis transition state. This binds as an AlF₄[−] species to most ATPases and adopts an octahedral geometry when complexed to the β -phosphate and the nucleophilic water moiety³⁴. To quantify DNA binding under equilibrium conditions, we used a fluorescence anisotropy assay with a hexachlorofluorescein (HEX)-labeled 40mer dsDNA fragment. In the presence of ATP γ S, the equilibrium dissociation constant, K_d , was 151 ± 3 nM (Fig. 2d, Supplementary Table 1), in agreement with previous determinations²⁹. We detected much reduced binding with ADP and enhanced binding with ADP•AlF_x ($K_d^{\text{ADP}\cdot\text{AlF}_x} = 24 \pm 2$ nM). In agreement with several reports in *E. coli*²⁹, but inconsistent with reports of oligomeric *Mycobacterium tuberculosis* Mfd³⁵, we observed no oligomerization by gel filtration (Supplementary Fig. 1a), suggesting that the increase in affinity in the presence of the transition state is owing to intramolecular

rearrangements rather than oligomerization, as in other SF2 nucleic-acid translocases³⁶.

The primary DNA-binding site in Mfd is its motor core. Both our study and previous work have not detected stable binding of Mfd to dsDNA in the absence of nucleotide¹⁴. Modeling of a duplex DNA–Mfd complex based on similarity with other SF2 proteins such as *Sulfolobus solfataricus* Snf2 is consistent with this finding. In the nucleotide-free form, the motor core of Mfd is more open. dsDNA modeled onto nucleotide-free Mfd contacts a highly conserved and positively charged D5 surface patch centered around invariant K712, but is nevertheless positioned too far away from D6 for direct contacts with this domain (Supplementary Fig. 1b–d), supporting that affinity is dictated by nucleotide status and suggesting that D6 swings down to clamp the DNA (Supplementary Fig. 1b–d). However, the conformation of the ATPase motor module in many translocases/helicases, and by extension, in Mfd, is modulated not only by nucleotide status but also interactions with family-specific accessory domains³⁷, suggesting that changes that are more global rather than restricted to the motor module.

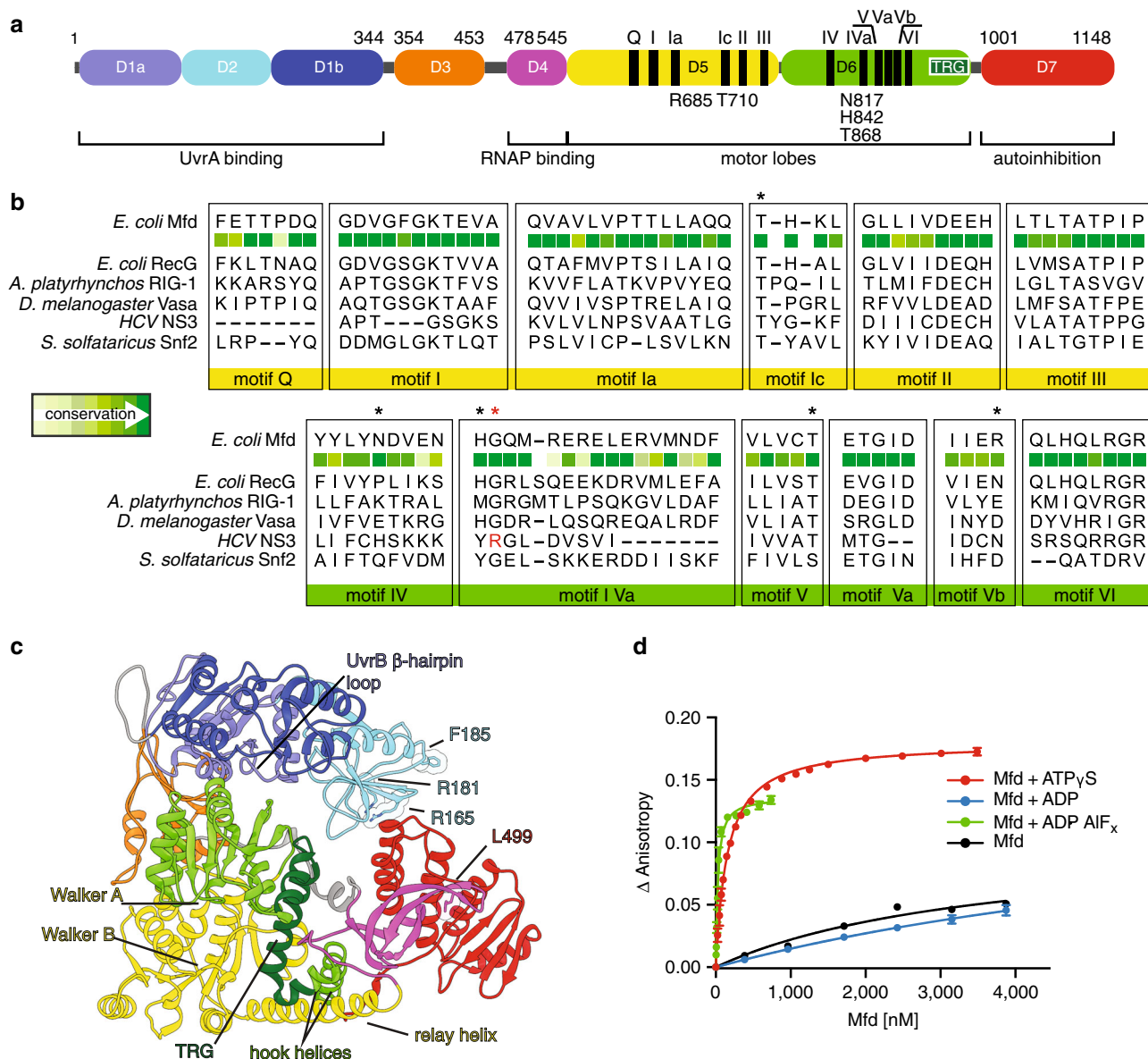


Fig. 2 DNA binding by Mfd becomes tightest in the presence of transition state analog ADP•AlF_x. **a** Domain organization of *E. coli* Mfd with annotated conserved sequence motifs and key residues mutated in this study. **b** Structure-based sequence alignment of the ATPase motifs of Mfd and other dsDNA and ssDNA translocases used for structural modeling purposes. Mutated residues are indicated by asterisks and sequence conservation, based on an alignment of Mfd proteins is color-coded from dark green (strongly conserved) to light green (variable). ATPase domain motifs are shown underneath for *Escherichia coli* RecG, *Sulfolobus solfataricus* Swi2/Snf2, *Drosophila melanogaster* Vasa, *HCV* NS3, and *Anas platyrhynchos* RIG-1. The arginine clamp R393 of *HCV* NS3 is shown in red. **c** Structural overview of nucleotide-free *E. coli* Mfd (PDB ID 2EYQ) colored by domain and with annotated ATPase motifs and key structural elements. Shown as CPK models are residues implicated in UvrA and RNAP binding^{29,31}. **d** Fluorescence anisotropy DNA-binding curves for Mfd in distinct nucleotide states: nucleotide-free (black), with 2 mM ATP γ S (red), ADP (blue), or ADP•AlF_x (green). Error bars represent S.D.M. ($n = 3$) and are often smaller than symbols. Curves obtained in the absence/presence of ATP γ S and ADP are replotted for convenience from the study by Le et al.¹⁶ Source data are provided as a Data Source file.

Our observation that ADP•AlF_x enhances the affinity for DNA allowed us to reconstitute stable Mfd-DNA complexes, and prepare specimens suitable for single-particle cryo-EM for structure determination. For our DNA substrate, we chose a 21mer blunt dsDNA fragment, consistent with previously determined Mfd footprints of ~26bp^{16,26,29}. This allowed us to restrict binding to one Mfd molecule per DNA fragment. Initial reconstructions obtained from untilted specimens were severely affected by the preferred orientation of the particles at the air-water interface, limiting resolution. We thus acquired images at 30° tilts, which allowed us to obtain a more isotropic map

resolved to 5.5 Å (Methods, Supplementary Fig. 2, and Supplementary Table 2). This enabled us to readily dock all the individual domains of Mfd³¹ into the map and perform real-space refinement to obtain the final model. At this resolution, we do not observe density that can be assigned to the transition state analog moiety. The model (Fig. 3) affords several critical observations.

First, our reconstruction reveals an unexpected large-scale swiveling motion of the UvrB homology module (D1a-D2-D1b), which results in a ~49 Å displacement and an 82° rotation of D2. Importantly, D1a-D2-D1b appears to move in concert, as a rigid body, which is consistent with previous structures of D1aD2D1b

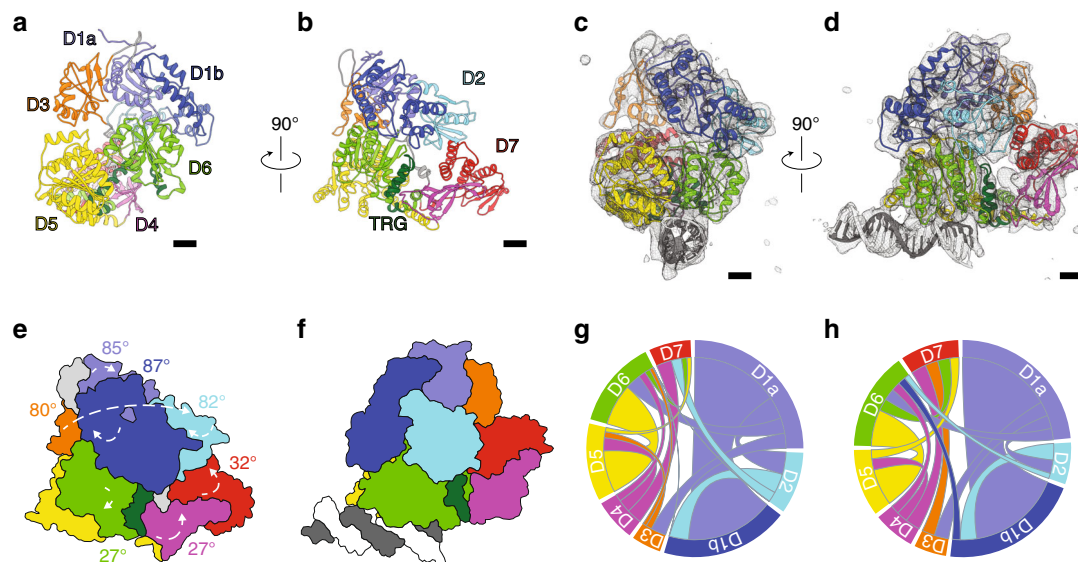


Fig. 3 Cryo-EM reconstruction of dsDNA-bound Mfd reveals a mobile D3 module and the interaction of dsDNA with the motor lobes. **a, b** Orthogonal views of nucleotide-free Mfd (PDB ID 2EYQ). Scale bar is equivalent to 10 Å. **c, d** Orthogonal views of the cryo-EM reconstruction (gray surface) with the fitted Mfd model shown as a cartoon and colored as in **a**. DNA is shown in gray. **e, f** Cartoons of Mfd structures in the nucleotide free **e** and DNA-bound form **f** after superposition on the C α trace of D5 and highlighting the relative rotation of Mfd domains in the free and tight DNA-bound state. **g, h** Chord plots highlighting intramolecular rearrangements in Mfd before **g** and after **h** DNA binding. In all panels, Mfd is colored by domain as in Fig. 2a.

and D1aD2D1bD3 truncations, which assumed the same structure in isolation as in the context of the full-length, nucleotide-free protein (r.m.s.d. of 0.8 Å between PDB ID 2EYQ and PDB ID 2B2N and 0.6 Å between PDB ID 2EYQ and PDB ID 3HJH)^{32,38}. Limited proteolysis of Mfd in the absence/presence of nucleotides and Mg•ADP•AlF_x did not reveal differences in protease susceptibility (Supplementary Fig. 3), suggesting that the conformational changes we observe are induced by DNA binding and not by the progression from the ATP ground to the transition state. Recruitment of UvrA might thus preferentially occur in the ADP-Pi state and might promote product release and dissociation of Mfd off the DNA, leaving behind an exonuclease loaded in a strand-specific manner.

Second, particularly striking is the repositioning of domain of unknown function D3, which swings from the ventral side laterally toward D7 upon DNA binding (Fig. 3 and Supplementary Movie 1), establishing new bridging interactions with the UvrB homology module and D7. This represents a large scale ~47 Å displacement of D3 enabled by its flanking flexible linkers connecting D3 to D1b and D4 (Figs. 2a–b, 3). The distance between the termini of the fitted D3 and D4 (36 Å) is fully compatible with the length of this partially disordered D3–D4 linker. Consistent with our data, the sequence corresponding to D3 in the structure of D1aD2D1bD3 was completely disordered³², pointing to it forming a dynamic and separate structural unit. Recent structures of mycobacterial Mfd in the presence of ADP and ADP•Pi also revealed movement of D3³⁹, but this is distinct from what we observe. In mycobacterial Mfd, sharing ~33% identity with *E. coli* Mfd, D3 is repositioned when ADP and Pi are bound such that it contacts D1a and D5 (Fig. 4a–c). These contacts are presumably important for nucleotide binding, as they are accompanied by a flip-flop action of the D3–D4 linker (Fig. 4 and Supplemental Fig. 5)³⁹. Movement of this linker relieves steric occlusion of the conserved phenylalanine of motif Q (F641 mycobacterial Mfd, F599 in *Eco* Mfd), which becomes free to stack against the adenine base of ADP. In contrast, when Mfd is bound to DNA, D3 makes contacts with D1a and D7 (Fig. 3), swapping position with D2. The UvrA-binding surface on D2, including residues critical for binding such as R165 and R181²⁹

(Fig. 2d–f) becomes fully exposed. Contacts of D2 with conserved residues in D7, D0148, and F1050 are lost.

Last, but not least, our reconstruction confirms the long-hypothesized role of the TRG motif in DNA translocation. The TRG (translocation in RecG) motif is common to both the RecG and Mfd subfamilies, and consists of a helical hairpin engaging a helical “hook” C-terminal to it (Fig. 2c). This hook was partially disordered in substrate-bound RecG, but helical in Mfd and wrapped around yet another helix—the relay helix, connecting D4 to the motor core (Fig. 2c). In our reconstruction, the two helices of the TRG snap together and are likely stabilized by ADP•AlF_x binding. In contrast, in nucleotide-free Mfd, the TRG helices are splayed open by motif VI helix ATP sensor³¹. The closed TRG motif engages part of the hook structure but releases motif VI, which undergoes a 46 Å displacement upon rotation of D6, and indirectly modulates the conformation of the relay helix, which becomes slightly bent and rotates about a hinge located in the loop connecting it to D5 (Fig. 5 and Supplementary Movie 1). Thus, we conclude that domain D3 of hitherto unknown function plays a critical role in unmasking of the UvrA binding site and that the interlocking of motif VI, TRG, hook and relay helices enable the complex and long-range conformational changes associated with DNA binding.

Linchpin residues for DNA loading span both lobes of the motor core. For translocation to occur, Mfd must first load the DNA. Our reconstruction suggests that this involves D5 and D6 exclusively, and no other contacts (Fig. 3a–f), although secondary binding determinants might be unmasked under very specific conditions.

We have thus asked if isolated Mfd domains might contribute to ss or dsDNA binding. We have purified Mfd functional modules as defined by the crystal structure of *E. coli* Mfd (UvrB homology module Mfd^{D1aD2D1b}, domain of unknown function Mfd^{D3}, RNAP interaction domain Mfd^{D4}, and clamp domain Mfd^{D7} (Fig. 2a and Supplementary Fig. 4) with the exception of Mfd^{D5D6}. We then assessed DNA binding with either a 40mer HEX-labeled dsDNA or a shorter Cy3-labeled double- and single-stranded 27mer. As expected, we observed no stable binding to

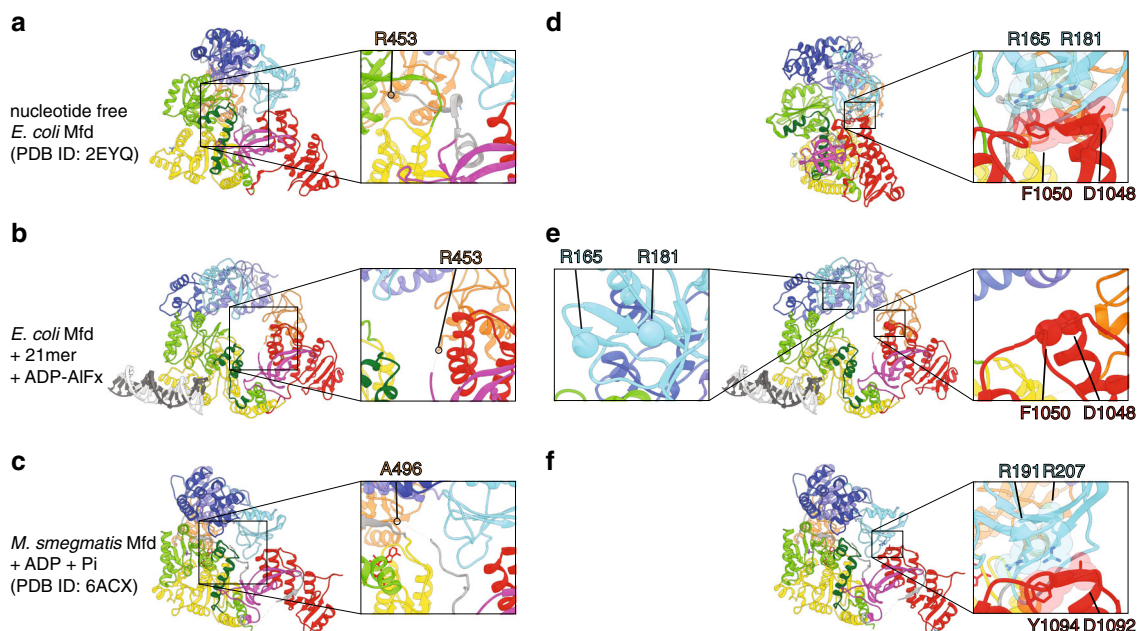


Fig. 4 DNA binding unmasks the UvrA recruitment surface within domain D2 of Mfd. **a–c** Top view of nucleotide-free *E. coli* Mfd **a**, *E. coli* Mfd complexed to DNA and ADP-AIF_x **b** and *M. smegmatis* Mfd bound to ADP and Pi **c** (PDB ID 6ACX). Insets show the D3–D4 linker, and C-terminal residues in D3 as positional markers. The D3–D4 linker is not visible in our map of DNA-bound Mfd **b** or in *M. smegmatis* Mfd **c**. **d–f** Overall views of Mfd with insets highlighting the disruption of the D2–D7 interaction upon DNA binding. Amino acids involved in domain–domain interactions (R165, R181, D1048, F1050) are shown as CPK with stick models. Middle inset highlights the unmasking of R165 and R181 (spheres), important for UvrA recruitment. In all panels, Mfd is colored by domain as in Fig. 2a.

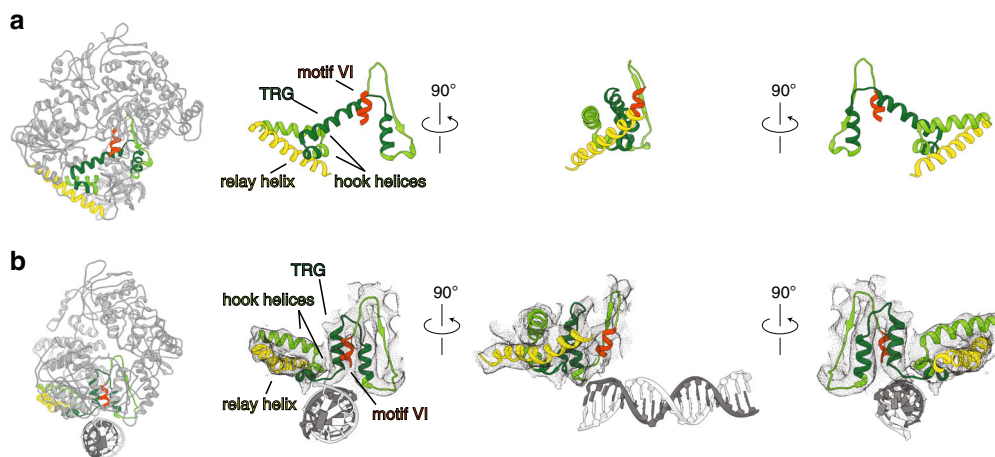


Fig. 5 Structural interlocking of the TRG motif, motif VI, and relay helix. **a** View of motif VI (red), TRG motif (dark green), hook helices (green), and relay helix (yellow) in the context of full-length nucleotide-free *E. coli* Mfd (left) and, for emphasis, as close-ups. **b** Similar views of DNA-bound *E. coli* Mfd highlighting conformational changes in this region. The TRG helical hairpin closes and a bend is introduced into the relay helix.

any of the purified modules (Supplementary Fig. 4). However, we cannot rule out lower affinity, secondary binding sites that might require initial contacts with Mfd^{D5D6}, a DNA fragment longer than what was used for our EM reconstruction, or a particular DNA structure, akin to an open or partially open bubble. On longer cellular DNA substrates, multiple domains may be required for DNA binding and contacts with individual modules, as defined here, may be too weak to be detected.

Although the EM reconstruction allowed us to identify residues proximal to DNA, specific contacts with the nucleic acid are not resolved with confidence at this resolution. To identify residues critical for DNA loading in both the ATP- and ADP•AIF_x states, we used our EM reconstruction but also took advantage of evolutionary relationships and structural superpositions with

well-characterized nucleic acid bound ATPases. Structural alignments with RecG, the helicase most closely related to Mfd, proved unfruitful, as crystallized RecG only contains a short fork DNA that does not reach across the helicase domains⁴⁰. Instead, we extended our analyses to other ss/ds nucleic-acid translocases of known structure. We focused on the similarity with Snf2, an SF2 chromatin remodeler for which several structural models in different functional states are available. These structures include models of Snf2 bound to dsDNA⁴¹, akin to the DNA moiety found in our specimen, but also curved DNA, as found in the structure of Snf2 bound to nucleosomes^{42,43}. We combined the information derived from our EM reconstruction and these superpositions (Supplementary Figure 6), and selected a set of well-conserved residues, highlighted in Fig. 6a and the

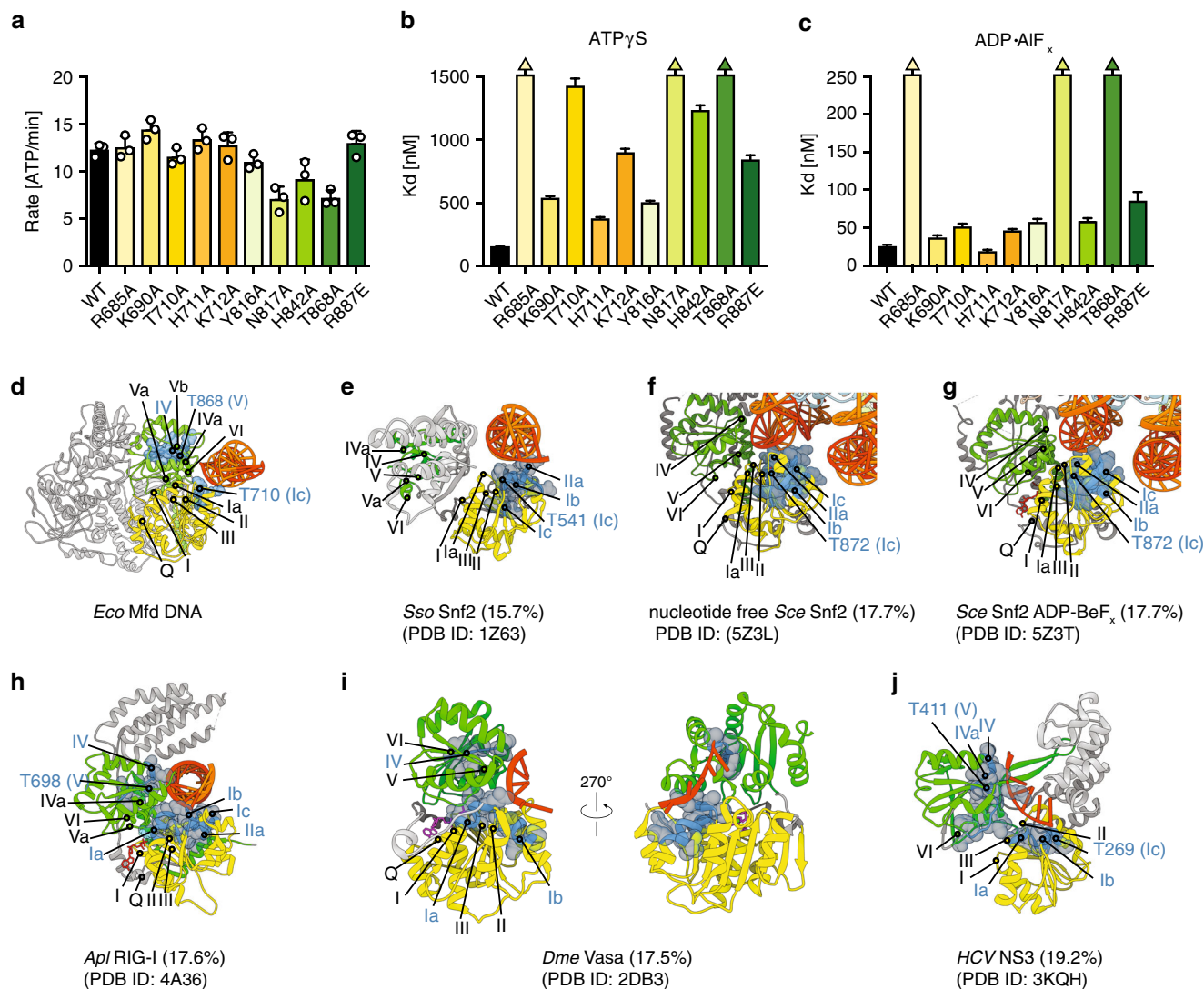


Fig. 6 Key roles for R685, N817, and R868 in DNA loading. **a** Steady-state ATPase activities of single Mfd variants. ATP hydrolysis rates were measured using an ATP/NADH-coupled ATPase assay. Data shown are the means \pm SD ($n = 3$). Variants are color-coded by domain as in Fig. 2a (D5 variants, yellow hues; D6 variants, green hues). Data for Mfd^{R685A} and Mfd^{N817A} are replotted for convenience from Le et al.¹⁶ **b–c** Equilibrium dissociation constants for the Mfd-dsDNA interaction derived from fluorescence anisotropy binding curves. D5 and D6 variants are color-coded by domain as in Fig. 2a. Data represent mean values \pm SD ($n = 3$). Error bars are often smaller than symbols. Arrows indicate data sets that could not be reliably fit to a binding model owing to compromised affinity for DNA, and likely represent an overestimation of the true affinity. Binding curves are shown in Supplementary Fig. 7. **d–j** Structural comparison between double-stranded nucleic-acid translocases and *Escherichia coli* (*Eco*) Mfd (**d**) using superpositions restricted to the α trace of D5. Shown are *Sulfolobus solfataricus* (*Sso*) Snf2 (r.m.s.d. of 1.1 Å, **e**), *Saccharomyces cerevisiae* (*Sce*) Snf2 in its nucleotide-free (r.m.s.d. of 1.4 Å, **f**) and ADP•BeF $_x$ -bound state (r.m.s.d. of 1.4, **g**), RIG-I (r.m.s.d. of 1.0 Å, **h**) and ssRNA translocases, *Drosophila melanogaster* (*Dme*) Vasa (r.m.s.d. of 0.9 Å, **i**) and *HCV* NS3 (r.m.s.d. of 1.1 Å, **j**). Bound Mg $^{2+}$ is shown as a dark green sphere, analog ADP•AlF $_3^-$ as an aquamarine CPK model, bound nucleotide as magenta sticks, ADP as a red CPK model, BeF $_x$ as a dark red CPK model and nucleic-acid strands are colored red and orange. The two lobes of the ATPase core are in yellow (N-terminal lobe) and green (C-terminal lobe), except for DNA-binding residues, which are highlighted as slate CPK models. Other domains are shown as a gray ribbon with annotated ATPase motifs. Those motifs contacting DNA are color-coded in slate font and some are not visible in these views. Indicated in parenthesis is the degree of conservation between the various motor cores and D5 and D6 of *E. coli* Mfd. Source data are provided as a Source Data file.

Supplementary Movie 1, as likely key for DNA recognition. We constructed Mfd variants carrying substitutions in either D5 (Mfd^{R685A}, Mfd^{K690A}, Mfd^{T710A}, Mfd^{H711}, Mfd^{K712A}) or D6 (Mfd^{Y816A}, Mfd^{N817A}, Mfd^{H842A}, Mfd^{T868A}, Mfd^{R887E}) (Fig. 2a). All were active in ATPase assays at levels comparable to wild type (Fig. 6a), indicating they are competent for nucleotide binding and hydrolysis, and consistent with proper folding, which was assessed using circular dichroism (Supplementary Fig. 7b). We then assessed binding to the same 40mer HEX-labeled dsDNA used above. The variants displayed various degrees of DNA-

binding defects in the presence of both ATP γ S (Fig. 6b, Supplementary Fig. 7c and Supplementary Table 1) and ADP•AlF $_x$ (Fig. 6c, Supplementary Fig. 7d and Supplementary Table 1). Interactions with DNA were more sensitive to salt in the ATP γ S than the ADP•AlF $_x$ state (Supplementary Fig. 7e, f), pointing to the formation of non-electrostatic contacts as partially accounting for the gain in affinity in the transition state.

Mfd^{R685A}, Mfd^{N817A}, and Mfd^{T868A} were severely impaired with both ATP γ S and ADP•AlF $_x$, indicating a critical role in loading DNA. T868 is particularly intriguing, as it is located in

conserved helicase motif V (Fig. 2a–b), which in several cases has been shown to contribute to either ATP or oligonucleotide binding⁴⁴. In the case of Mfd, substitution of T868 affects DNA binding severely, and less so ATP binding or hydrolysis (Fig. 6a–c). This is not without precedent. The structure of duck RIG-I RNA helicase bound to 19mer dsRNA and ADP•AlF_x, shows this highly conserved threonine (T698 in duck RIG-I) in direct contact with a phosphate of the RNA backbone (Fig. 6h), making a hydrogen bond with the phosphoryl oxygen atom⁴⁵. However, mutation of this conserved residue to alanine in human RIG-I eliminates dsRNA binding while retaining ATP hydrolysis activity⁴⁶, suggesting that this residue confers affinity that is uncoupled from the splitting of the phosphoanhydride bond. In *Thermotoga maritima* RecG, the mutation of T478 reduces fork reversal (i.e., translocation) without affecting ATP hydrolysis⁴⁷. A structurally equivalent threonine, T546, also contacts the RNA phosphate backbone in the structure of Vasa RNA helicase bound to ssRNA and AMPPNP (PDB ID 2DB3, Fig. 6i)⁴⁸. At last, the SF2 helicase domain of virally encoded DExH helicase NS3 also has a similar conserved threonine, and this was seen to contact oligonucleotide in multiple structural studies^{49,50}. As with Mfd, mutation of this threonine (T411 in NS3h, Fig. 6j) does not affect its basal ATPase activity, but abolishes oligonucleotide binding and duplex unwinding activity⁵¹.

Mfd mechanism of translocation is reminiscent of that used by RNA helicases. In the presence of ADP•AlF_x, some of our variants (Mfd^{K690A}, Mfd^{T710A}, Mfd^{K712A}, Mfd^{Y816A}, Mfd^{H842A}) showed only a slight impairment in function, or no DNA-binding defect at all, as was the case for Mfd^{H711A} (Fig. 6 and Supplementary Table 1). Most striking are the substitutions of T710A and H842A (Fig. 6a), which lead to severe loss of affinity in the ATPγS state, but had a comparatively minor effect on DNA affinity in the presence of ADP•AlF_x. Using optical trapping, we have previously shown that in the context of a nucleotide-starved TEC, Mfd binds to the TEC such that D5 is located in front (adjacent to the TEC), followed by D6, and that upon ATPγS binding D6 moves 12 bp toward a stationary D5, approaching the RNAP footprint¹⁶. As ATP hydrolysis is required for processive translocation, this suggested to us that upon hydrolysis and/or product release, D5 may invade the RNAP footprint. We thus reasoned that the residues with nucleotide-dependent roles might be critical for cycling through all the states required for translocation. To probe translocation, we developed a fluorescent triplex displacement assay, based on the ability of dsDNA translocases to displace a fluorescently labeled triplex forming oligonucleotide (TFO) off a DNA duplex. Triplex displacement assays have been well established for constitutively hyperactive Mfd variants, such as Mfd^{D7-AAA}, carrying the E1045A, D1048A, and R1049A substitutions in D7 to disrupt the inhibitory clamp interaction (Fig. 7a)⁵². Given the recently determined short processivity of Mfd of ~200 bp¹⁶, we employed a 70 bp substrate, shorter than the Mfd processivity length, in conjunction with the hyperactive Mfd^{D7-AAA} variant. We included in our reaction non-labeled ssDNA, for which Mfd has a weaker affinity¹⁴, but acts as an efficient trap to inhibit a contaminating endonuclease activity that initially severely reduced the signal corresponding to full-length TFO (not shown). In the presence of Mfd^{D7-AAA}, free TFO accumulated over time, and this activity was inhibited by non-hydrolysable analog, AMPPNP and the absence of nucleotide (Fig. 7b–c and Supplementary Fig. 8). Both mutants, Mfd^{D7-AAA}; T710A and particularly Mfd^{D7-AAA}; H842A displayed defects in translocation and loading, but not ATP hydrolysis (Fig. 7f), and featured circular dichroism spectra similar to Mfd^{D7-AAA} (Supplementary Fig. 8b). The loading defects of these complex

variants are somewhat different than the defects on the single T710A and H842A variants in the wild-type background (Figs. 6a–c, 7), and suggest that artificial disruption of the clamp interaction partially uncouples DNA binding from nucleotide status, consistent with a now derepressed enzyme³⁰. As such, these variants display severe translocation defects, with Mfd^{H842A} and Mfd^{T710} displaying no detectable translocation activity (Fig. 7b, c). At higher protein concentrations used in an attempt to overcome the loss in affinity of the D7-AAA mutants, both in the presence of ATPγS and ADP•AlF_x, the apparent translocation activity of both Mfd^{D7-AAA} and its two variants decreases (Supplementary Fig. 8), likely owing to multiple loading molecules interfering with translocation, which complicates the analysis. Interestingly, although wild-type Mfd is not stimulated by the addition of herring sperm DNA, Mfd^{D7-AAA} and to a lesser extent Mfd^{D7-AAA}; T710A, and Mfd^{D7-AAA}; H842 show an increase in ATPase activity (Fig. 7f and Supplementary Fig. 8f). Thus, whereas deregulation via clamp disruption hyperactivates Mfd, mutations within motifs Ic and IVa may repress Mfd activity.

H842 maps to a loop N-terminal to motif IVa and is conserved within the Mfd family, but not across translocases (Fig. 2a). However, based on a structural alignment, it is located close to where R393 in HCV NS3 maps (Supplementary Fig. 9). This has been coined the arginine clamp and is key for nucleic-acid binding and translocation⁵³. In agreement with our findings, a set of structures of HCV translocase NS3h bound to DNA alone and in complex with ground-state ATP and transition state mimics revealed contacts of nucleic acid with R393 as well as a series of nucleotide-dependent structural transitions involving the equivalent of T710 (T269 in NS3h, Fig. 6j) as well as NS3h T411, the equivalent of T868⁵⁰. T269 is in helicase motif Ic (aka TxGx motif), which while not originally identified as a conserved sequence, has been shown to be involved in nucleotide-dependent oligonucleotide binding in a variety of helicases, including NS3^{44,54}. The NS3 structure obtained in the absence of ATP shows these threonine residues three nucleotides apart, however they are only two nucleotides apart in the structure obtained with the ATP transition state analog (Supplementary Fig. 9b–d)^{49,50}. T411 maintains contact with the same nucleotide throughout the ATP hydrolysis cycle, whereas T269 releases the DNA in the ATP-bound state and makes a new contact with the DNA one nucleotide away when NS3h is in complex with the ATP transition-state analog. Conservation of these two threonine residues in Mfd, the nucleotide-dependent role of T710 and the confirmation that T868 is critical for DNA binding by Mfd regardless of the status of the phosphoanhydride bond point to common features of translocation in NS3h and Mfd, despite Mfd lacking helicase activity and binding preferentially to DNA rather than RNA.

Discussion

Mfd-dependent processes rely heavily on its translocation on dsDNA. This is important not only for early events required to locate RNAP targets on the chromosome, but also for destabilizing TECs halted by DNA damage, pause signals, or various roadblocks⁶. Mfd action on transcription complexes effectively builds up torque to reanneal the upstream edge of the transcription bubble, shorten the bubble, and release the RNA transcript^{26,55}. As a first step in shedding light on these events, here we describe the first structure of DNA-bound *E. coli* Mfd as well as complementary functional analyses. Our data allow for the first model for how nucleotide status and DNA binding control the breaking and formation of interdomain contacts and also reveal the molecular basis for DNA recognition by Mfd proteins.

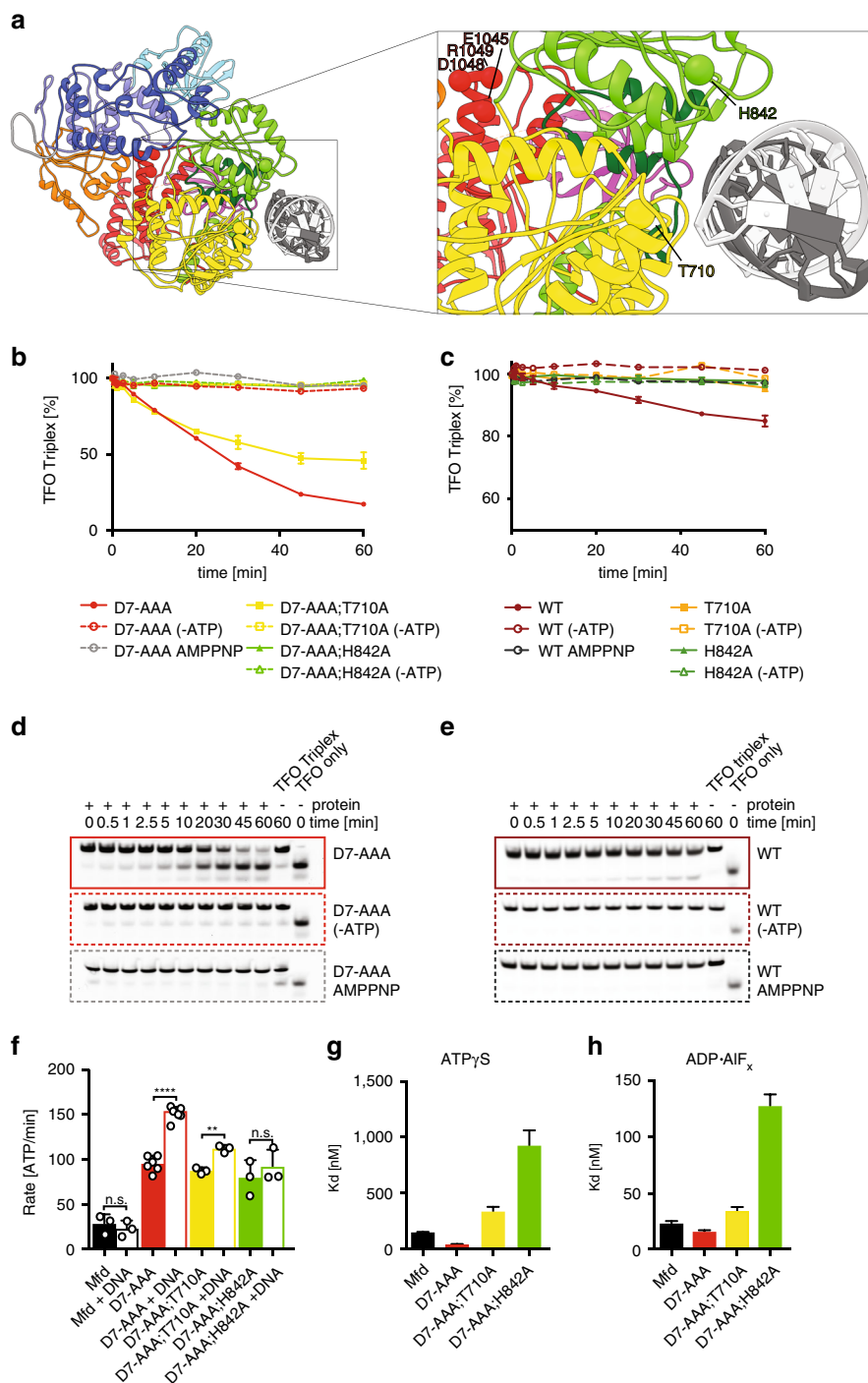


Fig. 7 T710 and H842 are critical regulators of translocation by Mfd. **a** EM-based model of DNA-bound Mfd with boxed region highlighting the structural environment of T710 and H842 residues. The resolution of our reconstruction precludes precise determination of the conformation of motif IVa loop and of amino-acid rotameric forms. Crucial residues are indicated as α spheres, H842 in green, T710A in yellow and E1045, D1048, and R1049, mutated in Mfd D7^{AAA} in red. **b-c** TFO displacement assays carried out with Mfd^{D7-AAA} variants. Percentage of remaining TFO-Triplex is plotted vs time to show translocation of Mfd variants. Plot shows means \pm SD ($n = 3$), with error bars often smaller than symbols. **d-e** Electrophoretic separation of TFO displacement assays carried out with Mfd^{D7-AAA} variants. Representative gels for Mfd^{D7-AAA} **d** and Mfd^{D7-AAA} **e** +/-ATP and +AMPPNP are shown. Example gels for Mfd, Mfd^{D7-AAA;T710A} and Mfd^{D7AAA;H842A} variants are shown in Supplementary Fig. 8 ($n = 3$). **f** Steady-state ATPase activities of Mfd variants normalized to wild-type Mfd activity. Data represent means \pm SD ($n \geq 3$). **denotes $p < 0.01$, **** $p < 0.0001$ and ns, $p > 0.05$ (unpaired two-tailed t test). The p value for Mfd^{D7AAA;T710} with/without DNA was 0.0012. **g, h** Equilibrium dissociation constants for the Mfd-dsDNA interaction derived from fluorescence anisotropy binding curves obtained by titrating increasing amounts of protein into a mixture of 10 nM HEX-labeled 40mer DNA and 2.0 mM ATP γ S **e** and ADP \bullet AlF $_x$. Data represent mean values \pm SD ($n = 3$). K_D values are tabulated in Supplementary Table 1. Source data are provided as a Source Data file.

We observe that DNA is bound by the two lobes of the motor core, consistent with a previous truncation analysis, which determined that in the presence of ATP γ S, the DNA interaction determinants are restricted to D5 and most of D6¹⁴. In this study, domain boundary definitions were not based on structural information, and notably truncated ATPase motif VI and the TRG motif (Fig. 2a–c), with crucial roles in coupling nucleotide status to translocation³³. However, even with our structure-based truncations, we detect no binding of ds/ssDNA to domains flanking the motor core. Previous studies noted a striking similarity between D1aD2D1b and a region in UvrB that contacts DNA during NER, and also locally melts it for damage recognition using a highly conserved beta hairpin at the interface between D1a and D1b³⁸. This beta hairpin extension is absent in Mfd, is replaced by a short loop, which lacks the aromatics involved in nucleotide flipping in UvrB. This loop is located ~50 Å away from the D5/D6 binding site (Fig. 2c), further supporting that the similarity might be strictly architectural. Our reconstruction suggests that on DNA substrates longer than used here, the DNA fragment would likely approach the C-terminal region of the relay helix, where previous studies have identified one substitution (Mfd^{W50A}) that enhances DNA binding and motor activity⁵⁶. This could involve bending of the DNA across the protein surface, and/or modest rearrangements of the hook and TRG motifs, located in close proximity to the DNA.

We thus conceptualize DNA binding by Mfd in terms of two clusters of residues located at the tip of the D5 and D6 pincers that establish contacts with the substrate, primarily the tracking strand, and cycle between a high-affinity state (nucleotide-bound) and a low affinity (nucleotide-free) state. This shares features with findings from the HCV NS3 system, but bears a distinction in that NS3 cycles between a low affinity state when nucleotide-bound to a high-affinity state when nucleotide-free. Cluster I is comprised of R685 and T710, whereas cluster II is comprised of R865, N817, T868, and H842. Within these clusters, we distinguish between residues key for loading (R685, T868) and for translocation (T710 and H842), the latter having markedly nucleotide-dependent roles as also observed in the HCV NS3 system. It will be interesting to determine if Mfd operates via a Brownian ratchet that involves translocation in 1 bp steps (like in NS3), and whether and how the binding of RNAP might rectify this ratchet to achieve high processivity on the order of hundreds of base pairs, detected using single-molecule techniques^{16,28}. We demonstrate that although there is no biochemical evidence for the TRG helices being directly involved in DNA binding, this structural element in direct communication with ATPase motif VI and the hook is mobile and likely used as a pawl during translocation. Here, we present the first structural evidence for the hypothesized interlocking of nucleotide-responsive elements (the TRG, the hook and relay helices, seen in Fig. 2c, 5) that is driven by movement of ATPase motif VI.

Translocases that move along dsDNA often make more extensive interactions with one DNA strand, termed the tracking strand. In the structure of *S. solfataricus* Snf2, more extensive interactions exist between the tracking strand and motifs I, Ia, Ib, Ic, and IIa⁴¹, located in the N-terminal lobe (Fig. 2a, c), whereas recent studies of Snf2 bound to a nucleosome suggest tighter interactions with the motifs in the C-terminal lobe^{42,43}. It is impossible for us to determine at this stage which of the strand represents the “tracking” strand. This remains an important outstanding question since it underlies strand-specific probing of damage, for whose recognition UvrA, UvrB and Mfd are required⁵².

Our identification of an Mfd family-specific residue important for translocation (H842, located in motif IVa) provides an attractive avenue to pursue in the development of an Mfd

inhibitor of high selectivity. Mfd deletion limits the acquisition of antimicrobial resistance in multiple bacteria, including *E. coli*²⁵, *Salmonella typhimurium*²⁵, *Bacillus subtilis*²⁵, *Helicobacter pylori*²³, *Campylobacter jejuni*²², and difficult to treat pathogens such as *M. tuberculosis*²⁵. Furthermore, in *Bacillus cereus* and *Shigella flexneri*, recent experiments have implicated Mfd in resistance to the nitric oxide response generated by the immune response during infection²⁴. Thus, an Mfd inhibitor would provide a novel strategy for antibiotic development. To block function while achieving selectivity, this could target highly variable motif IVa, rather than one of the other, more conserved ATPase motifs. In the NS3 system, the arginine clamp in motif IVa has been shown to be druggable⁵⁷. We note that at the current resolution of our EM reconstruction, establishing whether H842 directly contacts DNA or may exert its effect indirectly by modulating the conformation of the motif IVa loop remains challenging.

Finally, our work provides the first evidence that domain D3 of unknown function might be key for TCR regulation in *E. coli* and related species. While D3 is sandwiched between D1a and D5/D6 in nucleotide-free Mfd, it is repositioned in the presence of DNA and transition state analog to make bridging interactions with D7 (Supplementary Movie 1). Motion of D3 is coupled to a peppermill-like motion of the UvrB homology module relative to the C-terminal region of Mfd (D4D5D6D7), which in turn undergoes local rearrangements to close the D5–D7 interdomain cleft, whereas D4–D7 interactions remain largely unperturbed and the UvrA recruitment surface is unmasked. Our work establishes that these conformational changes are coupled to the active site, primarily through motif VI, but also movements of the TRG, hook and relay helices, which remain unresolved at the residue level and await higher resolution structural and mechanistic analyses.

Methods

Cloning, protein expression, and purification. Constructs pMS1–pMS8 were obtained via site-directed mutagenesis of pAD6 (encoding N-terminally His-tagged *E. coli* Mfd⁵⁸) using the method of Edelheit et al.⁵⁹. All constructs were verified by DNA sequencing, and in select cases, purified protein was verified using mass spectrometry. Overexpression and purification was performed according to published protocols by a succession of Ni²⁺-affinity chromatography, His-tag cleavage with Precision protease, heparin chromatography and gel filtration⁵⁹. For constructs pMS2 and pMS5, HiTrap Q-sepharose column (GE Healthcare) was used instead of heparin affinity as these variants failed to bind to heparin affinity matrix. Size-exclusion chromatography was carried out on a Superdex 200 10/30 Increase column (GE Healthcare) using a buffer consisting of 20 mM Tris pH 8.0, 0.10 M NaCl, 5 mM DTT. Mfd truncations were expressed as N-terminal fusions to polyhistidine tags (Supplementary Table 3), which were cleaved using Precision protease after the initial immobilized metal affinity chromatography step.

Triplex displacement assays. To generate the TFO-Triplex, linear ds 70mer DNA (50 nm) was mixed with 5' Alexa647-labeled 21mer TFO oligo (25 nm) in an opaque tube in triplex reconstitution buffer (10 mM MES pH 5.5, 12.5 mM MgCl₂) and heated to 57 °C for 15 min. The reaction was then placed at 4 °C and allowed to thermally equilibrate overnight. Protein samples were dialyzed against 10 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM MgCl₂, 2 mM β -mercaptoethanol. Reactions were assembled by mixing: 10 mM TFO-triplex, 50 nm 44 bp ssDNA trap in reaction buffer (50 mM Tris-HCl pH 8, 10 mM MgCl₂, 1 mM DTT) and protein (450 nM for constructs used in Fig. 7b–c; 4.5 μ M for the experiments of Supplementary Fig. 8d). Triplex displacement was initiated by addition of 2 mM ATP. Samples were withdrawn at indicated timepoints, stopped by addition of GSM buffer (15% (w/v) glucose, 3% (w/v) SDS, 250 mM MOPS pH 5.5), and loaded on 3–8% Tris-Acetate gels (Invitrogen, Carlsbad, CA), which were pre-run for 60 min at 75 V using Tris-Acetate Running buffer (40 mM Tris-Acetate pH 5.5, 5 mM sodium acetate, 1 mM MgCl₂) at 4 °C. Samples were electrophoresed for 180 min at 75 V and imaged using a ChemiDoc system (Biorad, Hercules, CA). The disappearance of TFO-Triplex over time was quantified using BioRad ImageLab and plotted using the Graphpad Prism software package.

Specimen preparation for cryo-EM. Size-exclusion chromatography (Superdex 200 10/30, GE Healthcare) was performed prior to complex reconstitution in a buffer containing 20 mM Tris-HCl pH 8, 100 mM NaCl, 20 mM MgCl₂, 2 mM TCEP.

For complex formation, Mfd was incubated with 21mer dsDNA in a 1:1.5 molar excess in the presence of 2 mM ADP, 10 mM NaF, and 1 mM AlCl₃. 2.5 µl Mfd-DNA complex at 1.0 mg/mL was applied to UltrAuFoil R1.2/1.3 300 mesh grids (Quantifoil) that were previously plasma-cleaned using a Gatan Solarus (75% argon/25% oxygen atmosphere, 15 W for 7 s), then manually blotted with a Whatman No. 1 filter paper in a cold room with >80% humidity, and plunged into liquid ethane using a manual plunger. The dsDNA used for reconstitution had the following sequence: 5'-ATAGGATACCTTACAGCCATCG-3'.

Cryo-EM data collection. Automated EM image acquisition was performed with Legion⁶⁰. Data collection was carried out on a Talos Arctic microscope (FEI) operating at 200 kV and equipped with a K2 Summit direct electron detector (Gatan) at a nominal magnification of ×36,000 and a defocus range from 2.0 µm to 4.0 µm giving a pixel size of 1.15 Å at the specimen level. To ameliorate preferential specimen orientation, images were collected using 30° tilts, as previously described⁶¹. Other than setting the nominal tilt angle during data acquisition, standard procedures were employed. Details are in Supplementary Table 2.

Cryo-EM image processing. Movies were aligned and dose-weighted using MotionCor2⁶². CTF estimation and particle picking was performed in Warp⁶³, which resulted in a stack of 69,721 particles that were imported into Relion (version 3.06)⁶⁴ for 2D classification. One round of 2D classification was performed, and bad particles were excluded. Subsequently, a round of 3D classification, using five classes was run in Relion (Supplementary Fig. 2). The initial model used for 3D classification was generated from a small subset of the data, during data collection using cisTEM⁶⁵. The best class was selected from 3D classification for moving forward. The selected particle stack from Relion contained 19,083 particles and was then input into cryoSPARC for iterative 2D and 3D classification. This cleaned stack containing 9822 particles was then imported into cisTEM and refined using the Auto-Refine procedure, whereby only frequencies up to 80% of Nyquist were included. This resulted in a final resolution of 5.47 Å, assessed by the conventional Fourier Shell Correlation (FSC) criteria. The refined map was sharpened in cisTEM by flattening the amplitude spectrum between 8 Å and the nominal resolution of 5.5 Å. Directional resolution volumes were generated using the 3D FSC tool⁶¹, whereas the local resolution was calculated using sxlocres.py, which is implemented within the Sparx processing package⁶⁶. Details on model building and refinement are in Supplementary Methods.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The reconstruction of DNA-bound Mfd was deposited in the EMDB under accession code EMD-22146 and model coordinates into the PDB under PDB ID 6XEO. Other data reported here can be obtained upon request from the authors. Source data are provided with this paper.

Code availability

Computer code used for calculating sampling distributions and the sampling compensation factor value is available upon request from D.L.

Received: 14 January 2020; Accepted: 1 July 2020;

Published online: 27 July 2020

References

- Lohman, T. M., Tomko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.* **9**, 391–401 (2008).
- Suhasini, A. N. & Brosh, R. M. Jr. DNA helicases associated with genetic instability, cancer, and aging. *Adv. Exp. Med. Biol.* **767**, 123–144 (2013).
- Selby, C. P. & Sancar, A. Molecular mechanism of transcription-repair coupling. *Science* **260**, 53–58 (1993).
- Bohr, V. A., Smith, C. A., Okumoto, D. S. & Hanawalt, P. C. DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* **40**, 359–369 (1985).
- Stantial, N., Dumpe, J., Pietrosimone, K., Baltazar, F. & Crowley, D. J. Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst.)* **41**, 63–68 (2016).
- Deaconescu, A. M. & Suhanovsky, M. M. From Mfd to TRCF and back again: A perspective on bacterial transcription-coupled nucleotide excision repair. *Photochem. Photobiol.* **93**, 268–279 (2017).
- Adebali, O., Sancar, A. & Selby, C. P. Mfd translocase is necessary and sufficient for transcription-coupled repair in *Escherichia coli*. *J. Biol. Chem.* **292**, 18386–18391 (2017).
- Selby, C. P., Witkin, E. M. & Sancar, A. *Escherichia coli* mfd mutant deficient in “mutation frequency decline” lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proc. Natl. Acad. Sci. USA* **88**, 11574–11578 (1991).
- Epshtein, V. et al. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372–377 (2014).
- Selby, C. P. & Sancar, A. Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *J. Biol. Chem.* **265**, 21330–21336 (1990).
- Selby, C. P. & Sancar, A. Gene- and strand-specific repair in vitro: partial purification of a transcription-repair coupling factor. *Proc. Natl. Acad. Sci. USA* **88**, 8232–8236 (1991).
- Selby, C. P. & Sancar, A. Mechanisms of transcription-repair coupling and mutation frequency decline. *Microbiol. Rev.* **58**, 317–329 (1994).
- Selby, C. P. & Sancar, A. Structure and function of transcription-repair coupling factor. II. Catalytic properties. *J. Biol. Chem.* **270**, 4890–4895 (1995).
- Selby, C. P. & Sancar, A. Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J. Biol. Chem.* **270**, 4882–4889 (1995).
- Ho, H. N., van Oijen, A. M. & Ghodke, H. The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage. *Nat. Commun.* **9**, 1570 (2018).
- Le, T. T. et al. Mfd dynamically regulates transcription via a release and catch-up mechanism. *Cell* **172**, 344–357 e15 (2018).
- Haines, N. M., Kim, Y. I., Smith, A. J. & Savery, N. J. Stalled transcription complexes promote DNA repair at a distance. *Proc. Natl. Acad. Sci. USA* **111**, 4037–4042 (2014).
- Washburn, R. S., Wang, Y. & Gottesman, M. E. Role of *E. coli* transcription-repair coupling factor Mfd in Nun-mediated transcription termination. *J. Mol. Biol.* **329**, 655–662 (2003).
- Pomerantz, R. T. & O'Donnell, M. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* **327**, 590–592 (2010).
- Million-Weaver, S. et al. An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **112**, E1096–E1105 (2015).
- Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikkh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**, 512–515 (2013).
- Han, J., Sahin, O., Barton, Y. W. & Zhang, Q. Key role of Mfd in the development of fluoroquinolone resistance in *Campylobacter jejuni*. *PLoS Pathog.* **4**, e1000083 (2008).
- Lee, G. H. et al. The *Helicobacter pylori* Mfd protein is important for antibiotic resistance and DNA repair. *Diagn. Microbiol. Infect. Dis.* **65**, 454–456 (2009).
- Guillemet, E. et al. The bacterial DNA repair protein Mfd confers resistance to the host nitrogen immune response. *Sci. Rep.* **6**, 29349 (2016).
- Ragheb, M. N. et al. Inhibiting the evolution of antibiotic resistance. *Mol. Cell* **73**, 157–+ (2019).
- Park, J.-S., Marr, M. T. & Roberts, J. W. *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**, 757–767 (2002).
- Fan, J., Leroux-Coyau, M., Savery, N. J. & Strick, T. R. Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. *Nature* **536**, 234–237 (2016).
- Graves, E. T. et al. A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence. *Nat. Struct. Mol. Biol.* **22**, 452–457 (2015).
- Deaconescu, A. M., Sevostyanova, A., Artsimovitch, I. & Grigorieff, N. Nucleotide excision repair (NER) machinery recruitment by the transcription-repair coupling factor involves unmasking of a conserved intramolecular interface. *Proc. Natl. Acad. Sci. USA* **109**, 3353–3358 (2012).
- Smith, A. J., Szczelkun, M. D. & Savery, N. J. Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. *Nucleic Acids Res.* **35**, 1802–1811 (2007).
- Deaconescu, A. M. et al. Structural basis for bacterial transcription-coupled DNA repair. *Cell* **124**, 507–520 (2006).
- Murphy, M. N. et al. An N-terminal clamp restrains the motor domains of the bacterial transcription-repair coupling factor Mfd. *Nucleic Acids Res.* **37**, 6042–6053 (2009).
- Chambers, A. L., Smith, A. J. & Savery, N. J. A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd. *Nucleic Acids Res.* **31**, 6409–6418 (2003).
- Wittinghofer, A. Signaling mechanistics: aluminum fluoride for molecule of the year. *Curr. Biol.* **7**, R682–R685 (1997).

35. Prabha, S., Rao, D. N. & Nagaraja, V. Distinct properties of hexameric but functionally conserved Mycobacterium tuberculosis transcription-repair coupling factor. *PLoS ONE* **6**, e19131 (2011).
36. Racki, L. R. et al. The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature* **462**, 1016–1021 (2009).
37. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev. Biochem* **76**, 23–50 (2007).
38. Assenmacher, N., Wenig, K., Lammens, A. & Hopfner, K. P. Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. *J. Mol. Biol.* **355**, 675–683 (2006).
39. Sivasankar Putta, S. P., et al. Structural Insights into the molecular mechanisms of the Mycobacterium evolvability factor Mfd. Preprint at <https://www.biorxiv.org/content/10.1101/728246v1> (2019).
40. Singleton, M. R., Scaife, S. & Wigley, D. B. Structural analysis of DNA replication fork reversal by RecG. *Cell* **107**, 79–89 (2001).
41. Durr, H., Korner, C., Muller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
42. Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445 (2017).
43. Li, M. et al. Mechanism of DNA translocation underlying chromatin remodelling by Snf2. *Nature* **567**, 409–413 (2019).
44. Caruthers, J. M. & McKay, D. B. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123–133 (2002).
45. Kowalinski, E. et al. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423–435 (2011).
46. Bamming, D. & Horvath, C. M. Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J. Biol. Chem.* **284**, 9700–9712 (2009).
47. Warren, G. M., Stein, R. A., McHaourab, H. S. & Eichman, B. F. Movement of the RecG motor domain upon DNA binding is required for efficient fork reversal. *Int. J. Mol. Sci.* **19**, 3049 (2018).
48. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* **125**, 287–300 (2006).
49. Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. USA* **107**, 521–528 (2010).
50. Kim, J. L. et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89–100 (1998).
51. Lin, C. & Kim, J. L. Structure-based mutagenesis study of hepatitis C virus NS3 helicase. *J. Virol.* **73**, 8798–8807 (1999).
52. Manelyte, L., Kim, Y. I., Smith, A. J., Smith, R. M. & Savery, N. J. Regulation and rate enhancement during transcription-coupled DNA repair. *Mol. Cell* **40**, 714–724 (2010).
53. Lam, A. M., Keeney, D. & Frick, D. N. Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action. *J. Biol. Chem.* **278**, 44514–44524 (2003).
54. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
55. Park, J. S. & Roberts, J. W. Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci. USA* **103**, 4870–4875 (2006).
56. Smith, A. J., Pernstich, C. & Savery, N. J. Multipartite control of the DNA translocase, Mfd. *Nucleic Acids Res.* **40**, 10408–10416 (2012).
57. Maga, G. et al. Specific targeting of hepatitis C virus NS3 RNA helicase. Discovery of the potent and selective competitive nucleotide-mimicking inhibitor QU663. *Biochemistry* **44**, 9637–9644 (2005).
58. Deaconescu, A. M. & Darst, S. A. Crystallization and preliminary structure determination of Escherichia coli Mfd, the transcription-repair coupling factor. *Acta Crystallogr Sect. F. Struct. Biol. Cryst. Commun.* **61**, 1062–1064 (2005).
59. Edelheit, O., Hanukoglu, A. & Hanukoglu, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* **9**, 61 (2009).
60. Suloway, C. et al. Automated molecular microscopy: the new Legimon system. *J. Struct. Biol.* **151**, 41–60 (2005).
61. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
62. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
63. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146–1152(2019).
64. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
65. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife* **7**, e35383 (2018).
66. Hohn, M. et al. SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).

Acknowledgements

A.M.D. is a recipient of a Medical Research Grant from The Rhode Island Foundation, of RI-INBRE Pilot Project and Career Development Awards under grant P20GM103430 and of R01 grant GM121975 from the National Institute of General Medical Sciences of the National Institutes of Health. Research in the laboratory of D.L. was supported by the Office of the Director grant from the NIH, DP5 OD021396.

Author contributions

C.B. and M.M.S. performed cloning, protein purification, functional assays, and data analysis. D.D.K. and A.N.S. performed cloning and protein expression and purification. C.Z. carried out cryo-EM work, image analysis, rigid-body docking, and initial refinement. C.B. performed model fitting and real-space refinement. D.L. supervised EM work, and A.M.D. conceptualized and designed the overall study, designed protocols, supervised research, analyzed data, and wrote the manuscript with contributions from M.M.S. and figure contributions from C.B. and C.Z. The manuscript was reviewed by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-17457-1>.

Correspondence and requests for materials should be addressed to A.M.D.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020