Structural basis for transcription complex disruption by the Mfd translocase

Jin Young Kang[1,3,4], Eliza Llewellyn[1,3], James Chen[1], Paul Dominic B. Olinares[2], Joshua Brewer[1], Brian T. Chait[2], Elizabeth A. Campbell[1], Seth A. Darst[1,5,6]

[1]Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY 10065, USA.

[2]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA.

[3]Contributed equally

[4]Present address: Department of Chemistry KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

[5]Lead contact: darst@rockefeller.edu

[6]Correspondence to: darst@rockefeller.edu

Summary

Transcription-coupled repair (TCR) is a sub-pathway of nucleotide excision repair (NER) that preferentially removes lesions from the template-strand (t-strand) that stall RNA polymerase (RNAP) elongation complexes (EC). Mfd mediates TCR in bacteria by removing the stalled RNAP concealing the lesion and recruiting Uvr(A)BC. We used cryo-electron microscopy to visualize Mfd engaging with a stalled EC and attempting to dislodge the RNAP. We visualized seven distinct Mfd-EC complexes in both ATP and ADP-bound states. The structures explain how Mfd is remodeled from its repressed conformation, how the UvrA-interacting surface of Mfd is hidden during most of the remodeling process to prevent premature engagement with the NER pathway, how Mfd alters the RNAP conformation to facilitate disassembly, and how Mfd forms a processive translocation complex after dislodging the RNAP. Our results reveal an elaborate mechanism for how Mfd kinetically discriminates paused from stalled ECs and disassembles stalled ECs to initiate TCR.

INTRODUCTION

DNA template strand (t-strand) lesions that block elongation by RNA polymerase (RNAP), most notably UV-induced cyclobutane dimers (Bohr et al., 1985; Witkin, 1966), are targeted for preferential repair in a process called transcription-coupled repair [TCR; (Hanawalt and Spivak, 2008; Mellon and Hanawalt, 1989; Mellon et al., 1987). In TCR, the stalled elongating RNAP serves as an efficient detector of t-strand lesions that then becomes a privileged entry point into the nucleotide excision repair (NER) pathway through the action of a transcription repair coupling factor (TRCF). While putative TRCFs have been identified in archaea and eukaryotes (Troelstra et al., 1992; Walker et al., 2017; Xu et al., 2017), the bacterial superfamily 2 (SF2) Mfd translocase (Smith et al., 2007) has been shown *in vitro* and *in vivo* to be a bacterial TRCF (Kunala and Brash, 1992; Oller et al., 1992; Selby and Sancar, 1991, 1993; Selby et al., 1991).

Escherichia coli (*Eco*) Mfd (mutation frequency decline - named for its activity to reduce the frequency of suppressor mutations that occur when protein synthesis is inhibited subsequent to mutagenic treatment) was discovered through genetic analysis (Bockrath et al., 1987; Witkin, 1966), then identified as a TRCF, purified, and biochemically characterized (Selby and Sancar, 1993, 1994, 1995a, 1995b). These experiments showed that Mfd was necessary and sufficient for TCR *in vivo* and *in vitro* and that it expressed two major activities:

1) Relief of inhibition of NER by recognition and ATP-dependent removal of the stalled RNAP elongation complex (EC) otherwise concealing the damaged DNA (Brueckner et al., 2007; Selby and Sancar, 1990).

2) After RNAP displacement, stimulation of the rate of DNA repair by direct recruitment of the Uvr(A)BC endonuclease to the lesion (Deaconescu et al., 2012; Selby and Sancar, 1990, 1991, 1993).

Subsequent biochemical, biophysical, and structural analyses have painted a more detailed outline of the Mfd functional cycle as a TRCF:

1) Mfd engages with stalled or paused ECs (the specific means of EC stalling isn't important) through two interactions: i) Mfd binds to the RNAP β subunit via its RNAP-Interacting Domain [RID; (Deaconescu et al., 2006; Park et al., 2002; Selby and Sancar, 1995b; Smith and Savery, 2005; Westblade et al., 2010), and ii) Mfd binds to duplex DNA upstream of the EC in an ATP-dependent manner via its RecG-like SF2 translocase domains (Chambers, 2003; Deaconescu et al., 2006; Gorbalenya et al., 1989; Mahdi et al., 2003; Selby and Sancar, 1995b).

2) The Mfd SF2 ATPase translocates on the upstream duplex DNA, pushing the RNAP in the downstream direction (same direction as transcription) via multiple cycles of ATP hydrolysis (Howan et al., 2012; Park et al., 2002). Backtracked and arrested ECs (Komissarova and Kashlev, 1997a, 1997b) can thus be rescued into productive elongation (Park et al., 2002). However, if a t-strand lesion or other type of roadblock prevents RNAP forward translocation, the continued translocase activity of Mfd overwinds the upstream region of the transcription bubble, facilitating displacement of the RNA transcript, transcription bubble reannealing, and removal of the RNAP from the DNA (Chambers, 2003; Deaconescu et al., 2012; Howan et al., 2012; Manelyte et al., 2010; Murphy et al., 2009; Park and Roberts, 2006; Park et al., 2002; Proshkin and Mironov, 2016; Selby and Sancar, 1995a; Smith and Savery, 2005; Smith et al., 2007).

3) After disruption of the EC, Mfd remains on the DNA and continues to slowly translocate in the downstream direction in a highly processive manner over thousands of base pairs, carrying the RNAP along for the ride (Graves et al., 2015; Haines et al., 2014; Howan et al., 2012).

4) The processively translocating Mfd-RNAP complex disassembles when it interacts with $UvrA_2B$, leaving behind the $UvrA_2B$ complex to recruit UvrC and complete the NER pathway (Fan et al., 2016; Selby, 2017).

Most of the biochemical activities of Mfd, such as ATPase activity, DNA binding and translocation, and interaction with UvrA, are strongly repressed in free Mfd (apo-Mfd) (Manelyte et al., 2010; Murphy et al., 2009; Smith et al., 2007). The apo-Mfd X-ray crystal structure revealed a protein with six structural modules connected by flexible linkers but held in a compact, inactive conformation (Deaconescu et al., 2006). Mfd activities are 'unleashed' by engagement with the EC, which is expected to be accompanied by profound conformational changes in Mfd (Srivastava and Darst, 2011).

Here, we used single particle cryo-electron microscopy (cryo-EM) to visualize how Mfd engages and displaces a stalled EC. We observe Mfd undergoing its ATP hydrolysis cycle attempting to release the RNA transcript and dislodge the RNAP from the DNA template. The ECs were stalled by nucleotide deprivation on a DNA scaffold containing a non-complementary transcription bubble that cannot rewind. Thus, despite engaging in cycles of ATP hydrolysis, Mfd was unable to efficiently dislodge the RNAP from the nucleic acids, facilitating the visualization of intermediates.

Using image classification approaches, we visualized seven distinct Mfd-EC complexes, some with ATP and others with ADP. Features of the structures allow their

placement in a pathway that provides a structural basis for understanding the extensive remodeling of Mfd upon its engagement with the EC and displacement of the RNAP. The structures explain how Mfd is remodeled from the repressed conformation (Deaconescu et al., 2006), how the UvrA-interacting surface of Mfd is hidden during most of the remodeling process to prevent premature engagement with the NER pathway, and how Mfd ultimately forms the processive translocation complex after dislodging the RNAP (Graves et al., 2015). Our results reveal the elaborate mechanism for how the Mfd translocase disassembles an otherwise stable EC and provide insight into the molecular motions that initiate TCR.

RESULTS

Mfd with ATP forms a stable complex with an RNAP EC assembled on a non-complementary transcription bubble

We assembled *Eco* RNAP ECs on a nucleic acid scaffold (Komissarova et al., 2003) with the following features: i) a 20-mer RNA transcript designed to generate a post-translocated 9 base pair RNA/DNA hybrid with the DNA t-strand, ii) a mostly non-complementary transcription bubble to inhibit Mfd-mediated displacement of the RNAP (Park and Roberts, 2006), and iii) 40 base pairs of duplex DNA upstream of the transcription bubble to allow Mfd function (Park et al., 2002) (Figure 1A). Native electrophoretic mobility shift analysis (EMSA) showed a unique band containing protein and nucleic acid that appeared only in the presence of the pre-formed EC, Mfd, and

ATP (labeled '*' in Figure 1B). Subsequent analysis of the contents of the *-band by SDS-polyacrylamide gel electrophoresis showed that it contained both RNAP and Mfd (Figure 1C, lane 5). Stable ECs and Mfd-ECs with 1:1 stoichiometry were also detected by native mass spectrometry (nMS; Figure 1D).

Seven structures in the Mfd activity cycle

In the crystal structure of the 130 kDa *Eco* apo-Mfd (Deaconescu et al., 2006), domains D1a-D2-D1b form a structural unit with similarity to the namesake elements of UvrB that interact with UvrA (the 'UvrB homology module'; Figure 2A). Indeed, mutagenesis and structural studies establish that UvrB and Mfd share a common mode of UvrA recognition (Deaconescu et al., 2006, 2012; Manelyte et al., 2010; Pakotiprapha et al., 2009, 2012; Selby and Sancar, 1995b). D3 is a non-conserved, lineage-specific domain with unknown function. D4 (RNAP interacting domain, or RID) is necessary and sufficient for Mfd interaction with the RNAP βprotrusion (Deaconescu et al., 2006; Park et al., 2002; Selby and Sancar, 1995b; Smith and Savery, 2005; Westblade et al., 2010). D5 and D6 comprise the RecG-like SF2 translocase domains (TD1/TD2). The interaction between the C-terminal D7 and the UvrA-interacting surface of D2 (Figure 2A) maintains apo-Mfd in its repressed state (Deaconescu et al., 2006; Manelyte et al., 2010; Murphy et al., 2009; Smith et al., 2007).

To visualize the expected conformational changes in Mfd upon EC engagement and de-repression, we analysed the Mfd-EC complexes (Figure 1) by single particle cryo-EM. Steps of maximum-likelihood classification (Scheres, 2012) revealed seven Mfd-EC structures (L1, L2, C1-C5; Figure 2B-H) ranging from 3.4 to 4.1 Å nominal

153 resolution (Figures S1, S2, Table S1). With the exception of L1, the cryo-EM maps were
154 of sufficient quality to directly observe the nucleotide status of each structure. C1, C2,
155 and C5 contained either ATP or ADP•P (where the hydrolysed γ-phosphate has not
156 been released); L2, C3 and C4 contained ADP (Figure 2C-H). Thus, Mfd was trapped
157 progressing through its nucleotide hydrolysis cycle.

158 In all seven structures, the ECs have similar RNAP and nucleic acid
159 conformations (root-mean-square-deviation for superimposed RNAP α-carbons;
160 0.59 Å < rmsd < 3.64 Å; Table S2A. See Methods for a description of the
161 superimposition procedure. See Figure S3A-G for examples of cryo-EM density) and
162 the Mfd-D4(RID) maintains its interactions with the RNAP βprotrusion (Figure 2B-H)
163 (Westblade et al., 2010). By contrast, the conformation of Mfd and the disposition of the
164 upstream duplex DNA varies dramatically (Figure 2, Table S2B). As expected, all of the
165 Mfd/DNA interactions occur through the DNA phosphate backbone.

166 In most of the structures, the cryo-EM density in the region around the upstream
167 edge of the transcription bubble is poor and difficult to interpret, likely due to
168 heterogeneity in that region. As a consequence, the precise register of the upstream
169 duplex DNA in the models is tentative. However, gross features of the upstream DNA,
170 such as the paths of the DNA backbone and of the overall helical axes, are clear
171 (Figure S3A-G). In this regard, the engagement of Mfd significantly distorts the
172 upstream duplex DNA. Mfd induces a localized kink in the DNA, roughly centered in the
173 footprint of the Mfd translocation module [D5(TD1)/D6(TD2)] on the DNA. The Mfd
174 translocation module engages with the duplex DNA from the minor groove; the kink,
175 which ranges from 7°-15° away from Mfd, is accompanied by significant widening of the

DNA minor groove (Figure S3H). The entire length of the upstream duplex DNAs bends away from the Mfd translocation module by 16° to 45° (Figure S3A-G). Based on single molecule observations, the Mfd interaction with DNA was proposed to induce bending or wrapping of the DNA (Howan et al., 2012), consistent with these structural observations.

One ATP hydrolysis cycle corresponds to translocation by one base pair

Sequence analysis identifies Mfd as an SF2 ATPase (Gorbalenya and Koonin, 1993) with a RecA-type catalytic core that is most closely related to RecG (Chambers, 2003; Mahdi et al., 2003). Mfd and RecG are unique among SF2 ATPases in harboring a conserved 'TRG' (translocation in RecG) motif following the seven SF2 ATPase signature motifs. Mutations in conserved residues of the TRG motif uncouple ATP hydrolysis from duplex DNA translocation (Chambers, 2003; Mahdi et al., 2003). It has not been possible to understand the relationship between duplex DNA binding, the nucleotide status of the RecA catalytic core, and duplex DNA translocation mediated by the TRG motif due to the lack of structures of an SF2 translocase bound to duplex DNA in different nucleotide states. The series of Mfd-EC structures determined here help in this understanding.

The seven SF2 ATPase signature motifs cluster together at the interface between TD1 and TD2 where the nucleotide binds (Figure 3A). Since some of the structural states contain ATP (C1, C2, C5) while others contain ADP (L2, C3, C4; Figure 2B-H), we can compare the disposition of the translocation domains with respect to each other, and correlate with the nucleotide status.

We superimposed α-carbons of residues 580-780 of TD1 (excluding the relay helix, which undergoes very large structural changes), yielding rmsds ranging between 0.282 - 0.894 Å (Figure 3B). The TD1 superimposition yielded two discrete positions of TD2, one of which corresponded with the ATP-bound structures (C1, C2, C5; green in Figure 3B), and the other with the ADP-bound structures (L2, C3, C4; blue in Figure 3B). L1, the complex in which the resolution of the cryo-EM map in the region of the Mfd nucleotide binding site (Figure S2C) was insufficient to directly assign the nucleotide status, clearly groups with the ATP-bound structures (red in Figure 3B). Therefore, we infer that L1 contains Mfd(ATP).

In the ATP-containing structures, the carbonyl oxygen of G874 (motif V), and the side chains of R902 (motif VI) and R905 (just beyond motif VI), all in TD2, form polar interactions with the ATP □-phosphate (Figure 3A, left). These three residues are absolutely conserved in an alignment of 65 Mfd sequences (Deaconescu et al., 2006). With the release of the γ-phosphate in the ADP structures, these interactions are lost and TD2 rotates away from TD1 [Figures 3A(right), 3B]. The movement of TD2 with respect to TD1 on transitioning from the ATP- to the ADP-bound state corresponds to an ~16° rotation about an axis roughly perpendicular to the helical axis of the upstream duplex DNA (Figure 3B). As a result, the center of mass of TD2 shifts parallel to the DNA helical axis about 3.5 Å in the downstream direction (Figure 3B), corresponding to one base pair rise of B-form DNA. These observations are suggestive of an 'inchworm' model for translocation, as proposed for related SF1 helicase translocation on single-stranded nucleic acids (Lohman et al., 2008; Singleton et al., 2007; Velankar et al., 1999; Yarranton and Gefter, 1979) and for Mfd based on single-molecule observations

(Le et al., 2018). In this model (schematized in Figure 3C), upon hydrolysis and release of the γ-phosphate, TD2 moves forward on the duplex DNA and forms tight interactions one base pair downstream. Next, the TD1/DNA interactions loosen and ADP must exchange for ATP in the nucleotide-binding site. This allows TD1 to close towards TD2, reestablishing interactions with the ATP γ-phosphate, now with both TD1 and TD2 translocated one base pair downstream on the duplex DNA (Figure 3C; Video S1). The model predicts a translocation stepsize of one nucleotide per ATP hydrolysis cycle, consistent with measurements for SF1 helicases (Dillingham et al., 2000; Tomko et al., 2007).

The TRG motif, which couples ATP hydrolysis to translocation, contains a helical hairpin motif followed by a meandering loop structure (Figure 3D). Each of these structural elements harbors one of the three conserved TRG residues that are critical for translocation: R929 (1st helix), R953 (2nd helix), and Q963 (loop) (Chambers, 2003; Mahdi et al., 2003). All three of these residues interact with the DNA phosphate backbone. The rotation axis of the TD2(ATP) -> TD2(ADP) transition passes directly through the linker connecting the helical hairpin (Figure 3D). Examining the structures more closely, the structural elements of TD2 C-terminal to the linker, which includes the second TRG hairpin-helix, the TRG loop, and the hook-helix, make extensive interactions with TD1 and move with TD1 as a rigid body, while the N-terminal portion of TD2 (residues 781-939) opens and closes depending on the nucleotide status. Thus, the linker connecting the TRG helical hairpin acts as the hinge (centered near absolutely conserved G942) for the TD1/TD2 conformational change in response to

nucleotide status. Opening of TD2 in the ADP state results in closing of the TRG helical-hairpin (Figure 3D).

Pathway for Mfd function

For the purpose of analyzing and discussing the structures, it is useful to place them in an ordered pathway. To begin the pathway, apo-Mfd from solution (Figure 2A) interacts with an EC. Comparing the Mfd component of each complex with the apo-Mfd structure (Deaconescu et al., 2006)], only L1(ATP) has an rmsd < 10 Å (all the others are >> 30 Å; Table S2B). Therefore we place L1(ATP) as the first structure in the pathway (Figure S4A).

To order the rest of the structures, we superimposed α-carbons of each complete structure (Mfd and RNAP α-carbons) with α-carbons of every other structure and calculated the rmsd of α-carbon positions, generating a table of 21 pair-wise rmsd values (Table S3A). L1(ATP) and L2(ADP) clearly stand apart from the other structures; L1 and L2 compared with every other structure exhibit rmsds between 19 Å to 47 Å, while C1-C5 compared with each other exhibit rmsds between 3.7 Å and 11.4 Å. We therefore propose that L1(ATP) and L2(ADP) are Mfd 'Loading' complexes, while C1(ATP), C2(ATP), C3(ADP), C4(ADP), and C5(ATP) are related structures looping in the Mfd nucleotide hydrolysis Cycle. Starting with L1(ATP), the path of transitions from one structure to the next that gives the smallest cumulative rmsd (Table S3A) is shown in Figure S4A.

Mfd loading requires multiple rounds of ATP hydrolysis and is accompanied by profound conformational changes

The Mfd component of L1(ATP) [Mfd(ATP)$_{L1}$], the first experimental structure in the pathway (Figure S4A), is most similar to the structure of apo-Mfd [PDB 2EYQ; (Deaconescu et al., 2006)]. We modeled a presumed precursor to L1, [L0] (square brackets denote a structural model), by superimposing the apo-Mfd-D4(RID) structure onto the Mfd(ATP)$_{L1}$-D4(RID) (Figure 4A). This reveals that the [L0] -> L1 transition involves large translations and rotations of TD1 (11 Å translation, 43° rotation) and TD2 (16 Å translation, 37° rotation; Table S3B), bringing the two ATPase domains into alignment to bind ATP and engage with the DNA (Figures 4B, C).

Mfd requires ~30 base pairs of duplex DNA upstream of the EC transcription bubble [to about -40 (Park et al., 2002)]. Mfd/DNA interactions occur between about -34 to -14 with one exception, Mfd(ATP)$_{L1}$ interacts with the DNA further upstream, from about -38 to -27 (Figures 4C, S4B), explaining the result of (Park et al., 2002) and also confirming that L1(ATP) is an obligate intermediate in the Mfd loading pathway.

In apo-Mfd, The D4(RID) is connected to the first RecA ATPase domain (TD1) by a 30-residue α-helix, the Relay Helix (RH, residues 548-577; Figure 4D). The RH at the N-terminus of TD1 interacts with the hook helices at the very C-terminus of TD2. In the [L0] -> L1(ATP) transition, the translations and rotations of TD1 and TD2 result in unfolding of seven residues in the middle of the RH (561-567), and kinking of the RH ~112° around the second hook helix. The first hook helix also completely unfolds (Figure 4E).

The transition from L1(ATP) -> L2(ADP) involves remarkable rearrangements of the Mfd structural modules. Other than the D4(RID), which stays anchored to the RNAP βprotrusion, the minimum center-of-gravity translation of an Mfd structural module [D1-D3, D5(TD1), D6(TD2), D7] is 60 Å, while the minimum rotation is 148° (Figures 4F, 5; Table S3B). As a result of the large conformational rearrangement of Mfd, $Mfd(ADP)_{L2}$ is topologically 'wrapped' around the DNA (Figure 5), likely explaining how Mfd (with RNAP in tow) translocates processively over many kilobases of DNA (Fan et al., 2016; Graves et al., 2015).

Although it is difficult to imagine the choreography of the Mfd structural modules in the L1(ATP) -> L2(ADP) transition without parts of Mfd passing through itself or through the DNA, a pathway exists. First, D1-D3 must dissociate from its position in L1(ATP), generating a hypothetical intermediate [L1.5a] (Figure 5). The movement of D1-D3 could be triggered by initial rounds of ATP hydrolysis/translocation by the translocation module and is facilitated by the 25 amino acid linker connecting D3 with D4(RID) (Figures 2A, 5). The release of D1-D3 now opens a path for the Mfd translocation module to 'walk' along the DNA, corkscrewing in the downstream direction (clockwise in the view of Figure 5) until it bumps into the RNAP at its position in L2 (Figures 4F, 5). As the Mfd translocation module corkscrews along the DNA, the N-terminal part of the RH (residues 548-560) completely unfolds and is dragged around the DNA, forming part of the topological link of Mfd on the DNA (Figure 4F).

In L1(ATP), the UvrA-interacting surface of D2 is occluded through its interaction with D7, as it is in apo-Mfd [Figures 4D, E; (Deaconescu et al., 2006, 2012)]. During the large rearrangements in the L1(ATP) -> L2(ADP) transition, D7 travels with the

translocation module, separating it from D2 (Figure 4F). However, in L2(ADP), the UvrA-interacting surface of D2 is still occluded by the DNA (Figure 4F).

The L2(ADP) -> C1(ATP) transition involves another large rearrangement of D1-D3 (45 Å translation, 102° rotation; Figure 5; Table S3B), which accommodates into its position seen in C1(ATP) -> C5(ATP). In this configuration, the UvrA-interacting surface of D2 is finally exposed to solution (Figure 5). The other Mfd structural modules make relatively small motions and Mfd remains topologically wrapped around the DNA (Figure 5).

The alternating nucleotide states on the transition from L1(ATP) -> L2(ADP) -> C1(ATP) indicate that the complete loading of Mfd involves rounds of ATP hydrolysis. Furthermore, since the translocation module traverses nine base pairs on the DNA in the L1(ATP) -> L2(ADP) transition, at least nine molecules of ATP must be hydrolysed (Figure S4A). Note this is before Mfd has entered the nucleotide hydrolysis cycle that serves to displace the RNAP (Figure S4A).

To test the structure-based hypothesis that ATP hydrolysis is required for Mfd loading to achieve a stable Mfd-EC complex, we incubated ECs and Mfd with vanadate (VO_4^{3-}) and either ADP or ATP. With ADP in the nucleotide-binding site, vanadate can bind in the position normally occupied by the γ-phosphate; the ADP-VO_4 complex is thought to mimic the ATP hydrolysis transition state and is an effective inhibitor of ATP binding and hydrolysis (Davies and Hol, 2004). ADP and vanadate from solution bind directly in the nucleotide-binding site without any rounds of ATP hydrolysis since no ATP is present. With ATP and vanadate, on the other hand, at least one round of ATP hydrolysis can occur. Following ATP hydrolysis, vanadate substitutes for the leaving

334 inorganic phosphate before ADP can be released, inhibiting further ATP hydrolysis
335 (Oldham and Chen, 2011; Shimizu and Johnson, 1983).

336 Recall that incubating Mfd and ATP with ECs formed on a nucleic acid scaffold
337 containing a partially non-complementary transcription bubble gave rise to a unique
338 complex observed by EMSA (Figures 1, 5B, band labeled '*' in lane 1). ADP + vanadate
339 does not support complex formation, while ATP + vanadate does (Figure 5B,
340 lanes 2, 3). The same concentrations of ATP + vanadate completely inhibited Mfd
341 function in an EC displacement assay (Chambers, 2003)(Figure S4C). These
342 experiments establish that at least one round of ATP hydrolysis is required for Mfd to
343 form a stable complex with an EC.

344

345 **Mfd manipulates the RNAP conformation during its nucleotide hydrolysis cycle**

346 After the Mfd loading pathway results in the formation of C1(ATP), we propose that Mfd
347 then loops through an ATP hydrolysis cycle, captured in the five distinct structural states
348 [C1(ATP) <-> C2(ATP) <-> C3(ADP) <-> C4(ADP) <-> C5(ATP) <-> C1(ATP) ...;
349 Figure 6]. Compared to the profound conformational changes that occur through the
350 loading pathway (Figures 4, 5), the conformations of Mfd in C1 - C5 are similar to each
351 other (Table S2B). Comparing the Mfd loading conformations of L1(ATP), L2(ADP), and
352 C1(ATP) gives rmsd values ranging from 26.0 Å to 38.4 Å, while comparing C1(ATP)
353 through C5(ATP) gives rmsd values ranging from 3.95 Å to 7.68 Å (Table S2B).
354 Although the Mfd conformations in C1(ATP) through C5(ATP) are similar, Mfd as a
355 whole, along with the upstream duplex DNA, wobbles back and forth with respect to the
356 EC by more than 30° (Figure 6, Table S3B).

357 During its nucleotide hydrolysis cycle (C1-C2-C3-C4-C5-C1...), Mfd translocates
358 on the upstream duplex DNA in the downstream direction, tracking in the minor groove
359 (Figure 3, Video S1). If the RNAP is unable to translocate forward, this results in
360 overwinding of the upstream DNA, facilitating collapse of the transcription bubble and
361 release of the RNA transcript (Park and Roberts, 2006). However, the motions of Mfd
362 during the nucleotide hydrolysis cycle also induce significant conformational changes to
363 the RNAP itself: clamp opening and βlobe-Si1 rotation (Figure 6).

364 The conformation of RNAP in C1(ATP) is very similar to the conformation of
365 RNAP in an EC (Kang et al., 2017). Using C1(ATP) as a reference conformation for
366 RNAP (clamp opening = βlobe-Si1 rotation = 0°), the conformational changes of Mfd
367 during its nucleotide hydrolysis cycle cause opening and closing of the clamp, with the
368 peak of clamp opening, 17.3°, at C3(ADP) (Figure 6). Similar magnitudes of clamp
369 opening have been observed in many structures, without (Zhang et al., 1999) and with
370 nucleic acids (Tagami et al., 2010; Weixlbaumer et al., 2013).

371 The Mfd conformational changes through the nucleotide hydrolysis cycle also
372 cause a sideways rotation of the βlobe-Si1 module, with the peak of βlobe-Si1 rotation,
373 14.3°, at C5(ATP) (Figure 6). The direction and magnitude of the βlobe-Si1 rotation are
374 very similar to an RNAP conformational change induced by TraR binding (Chen et al.,
375 2019a). In the case of TraR, this conformational change occurs in the context of an
376 initiation complex with promoter DNA and has multiple effects depending on the
377 promoter context (Chen et al., 2019a, 2020). In the context of an EC, we propose that
378 the βlobe-Si1 rotation alters the RNAP-downstream DNA duplex contacts in the RNAP

cleft, destabilizing these interactions and thereby destabilizing the EC (Nudler et al., 1996). The conformational transitions during the Mfd-NHC are visualized in Video S2.

DISCUSSION

The crystal structure of apo-Mfd revealed a compact but complex arrangement of structural modules connected by long, flexible linkers [Figure 2A; (Deaconescu et al., 2006)]. Interdomain interactions maintain Mfd in a repressed state, where its activities of ATP hydrolysis, DNA translocation, and UvrA binding are suppressed (Manelyte et al., 2010; Murphy et al., 2009; Smith et al., 2007). Amino acid substitutions that disrupt key interdomain interactions cause profound conformational changes in Mfd and unleash its ATPase and DNA translocase activities (Manelyte et al., 2010; Murphy et al., 2009; Smith et al., 2007). In this work, we observed seven distinct structures (Figure 2B-H) that delineate how the initial interaction of Mfd with an EC triggers a stepwise series of dynamic conformational changes, culminating in the stable engagement of Mfd with the EC and then ATP-hydrolysis powered disruption of the EC. Key structural features and conformational changes are highlighted in the context of the transition path from one structure to the next (Figure S4A) in the summary Figure 7.

Initial binding of the Mfd-RID to the RNAP βprotrusion to generate the hypothetical state [L0] (Figures 4A, 4B, 7) does not appear to require conformational changes as presumed, but rather tethers the Mfd translocation module in close proximity to the upstream duplex DNA of the EC (Figure 4B). The disposition of TD1 and TD2 in [L0] is not conducive to DNA or ATP binding (Deaconescu et al., 2006), but

thermal breathing of the Mfd domains could transiently align TD1 and TD2 to allow ATP binding and stable engagement of the upstream duplex DNA in L1(ATP) (Figures 4C, 7). In L1(ATP), Mfd engages with the upstream duplex DNA to -38, explaining why Mfd requires upstream duplex DNA to about -40 to displace the EC (Park et al., 2002).

The alignment of TD1 and TD2 and ATP binding allows ATP hydrolysis in L1(ATP), which initiates inchworming of the Mfd translocation module in the downstream direction (Figure 3). We hypothesize that this induces the displacement of D1-D3 (which is still tethered), clearing a path for continued translocation of TD1/TD2 (Figure 5A). TD1/TD2 walk along the duplex DNA (Figure 3), corkscrewing around the DNA for nine base pairs and in the process leave the unfolded relay-helix polypeptide wrapped around the DNA in L2(ADP) (Figures 4F, 5A). This ATP-hydrolysis-driven choreography results in Mfd completely encircling the upstream duplex DNA, explaining the remarkable processivity of the translocating Mfd-RNAP complex subsequent to EC disruption (Fan et al., 2016; Graves et al., 2015).

After walking freely in the downstream direction for nine base pairs, the Mfd translocation module butts up against the RNAP (Figure 4F). The D1-D3 module accommodates itself to the new molecular environment it finds itself in, finally exposing the previously occluded UvrA-binding determinant (Figure 5A, 7).

At this point, if the RNAP is backtracked (Komissarova and Kashlev, 1997a, 1997b; Nudler et al., 1997), continued Mfd translocation in turn forward translocates the RNAP until the active configuration of the EC, with the RNA transcript 3'-end in the RNAP active site, is reached (Park et al., 2002). In the presence of NTP substrates,

425 RNA chain elongation by the RNAP can resume and the RNAP, which translocates at
426 ~15-20 nucleotides/s, runs away from Mfd, which translocates at only ~4 nucleotides/s
427 (Howan et al., 2012; Le et al., 2018).

428 If RNAP forward translocation is blocked, either by the absence of NTP
429 substrates, a bulky lesion in the t-strand DNA (such as a CPD dimer), or a roadblock
430 such as a DNA binding protein, continued translocation by the Mfd translocation module
431 will essentially pull and torque the DNA out the back of the RNAP. The torquing action
432 of Mfd positively supercoils the DNA between the Mfd translocation module and the tight
433 grip of the RNAP on the downstream duplex DNA, causing the nt-strand DNA of the
434 transcription bubble to reanneal with the t-strand DNA, displacing the RNA transcript
435 (Park and Roberts, 2006). The stability of the EC is dependent on the RNA/DNA hybrid
436 (Kireeva et al., 2000), so release of the RNA transcript results in release of the RNAP
437 from the DNA template.

438 The particles that gave rise to all seven of the Mfd-EC structures came from two
439 separate samples that were prepared in the same way (Figure S1), so the relative
440 numbers of particles that belong to each structural class are related to the relative
441 stability of that class - the more particles in a structural class, the more stable that
442 structure is. The first structure in our pathway [L1(ATP)] is also the least stable
443 (Figure S4A), indicating that this first step of engagement of Mfd with the DNA is
444 reversible and that L1(ATP) is likely in a dynamic equilibrium with [L0] and/or Mfd in
445 solution. The single molecule analysis of (Howan et al., 2012) found that engagement of
446 Mfd with a stalled EC was characterized by weak initial binding of Mfd (Mfd molecules

attempt binding to the stalled EC many times before engaging productively), consistent with our findings.

The initial weak binding of Mfd to the stalled EC was followed by a very slow catalytic step (k = 0.059 s^{-1}) that required multiple rounds of ATP hydrolysis (Howan et al., 2012). The L1(ATP) -> L2(ADP) transition requires 9 rounds of ATP hydrolysis. The reversible equilibrium at the first step, and the presumably slow, multi-step transition required to ultimately reach the NHC, are consistent with proposals that Mfd kinetically discriminates a stalled EC (which Mfd is charged with displacing) and an EC that is simply paused (Kang et al., 2019; Landick, 2006) and should not be displaced. The slow, reversible loading pathway ensures that only the very long-lived stalled EC becomes a target for Mfd displacement and recruitment of UvrA. The kinetic discrimination of stalled vs. paused (but still productive) ECs explains why overexpression of Mfd is not toxic to cells (Deaconescu et al., 2006; Roberts and Park, 2004; Selby and Sancar, 1993; Smith et al., 2012).

Based on the populations of the states (Figure S4A), the NHC states (C1-C5) are considerably more stable than the loading states (L1-L2), indicating that once the complex passes through the loading pathway and enters the NHC, it is essentially irreversibly trapped, where it attempts to translocate on the DNA against the immovable EC, imparting positive torque on the DNA and ultimately collapsing the transcription bubble, releasing the RNA transcript, and displacing the RNAP from the DNA. During this process, Mfd manhandles the RNAP, prying open the RNAP clamp and βlobe-Si1 to cause RNAP to lose its grip on the nucleic acids (Figure 6).

Thus, entry into the NHC is essentially like a molecular switch. In the reversible, slow loading pathway, Mfd probes the lifetime of the stalled/paused EC. During these stages, the Mfd UvrA-binding determinant is occluded; premature recruitment of UvrA would disrupt the Mfd-EC complex, short circuiting the TCR pathway. It could also counterproductively recruit NER components to sites on the genome that are not damaged. The elaborate conformational choreography of Mfd loading and EC displacement is thus evolutionarily tuned to prevent disruption of normal regulation of transcription elongation but to engage the NER pathway when a lesion is likely encountered.

Acknowledgments We thank M. Oldham and A.J. Smith for assistance and advice with experimental procedures, M. Ebrahim and J. Sotiris at The Rockefeller University Evelyn Gruss Lipper Cryo-electron Microscopy Resource Center for help with cryo-EM data collection, and N. Savery, T. Strick, and members of the Darst/Campbell laboratory for helpful discussion on the manuscript. This work was supported by NIH grants P41 GM109824 and P41 GM103314 to B.T.C., R01 GM114450 to E.A.C., and R35 GM118130 to S.A.D.

Author contributions J.Y.K., E.L. and J.B. expressed and purified proteins and performed biochemical assays. J.Y.K., E.L., P.D.B.O., and B.T.C. performed native mass spectrometry analyses. J.Y.K. and E.L. prepared cryo-EM grids and collected cryo-EM data. J.Y.K., E.L., and J.C. processed cryo-EM data. S.A.D. built, refined, and

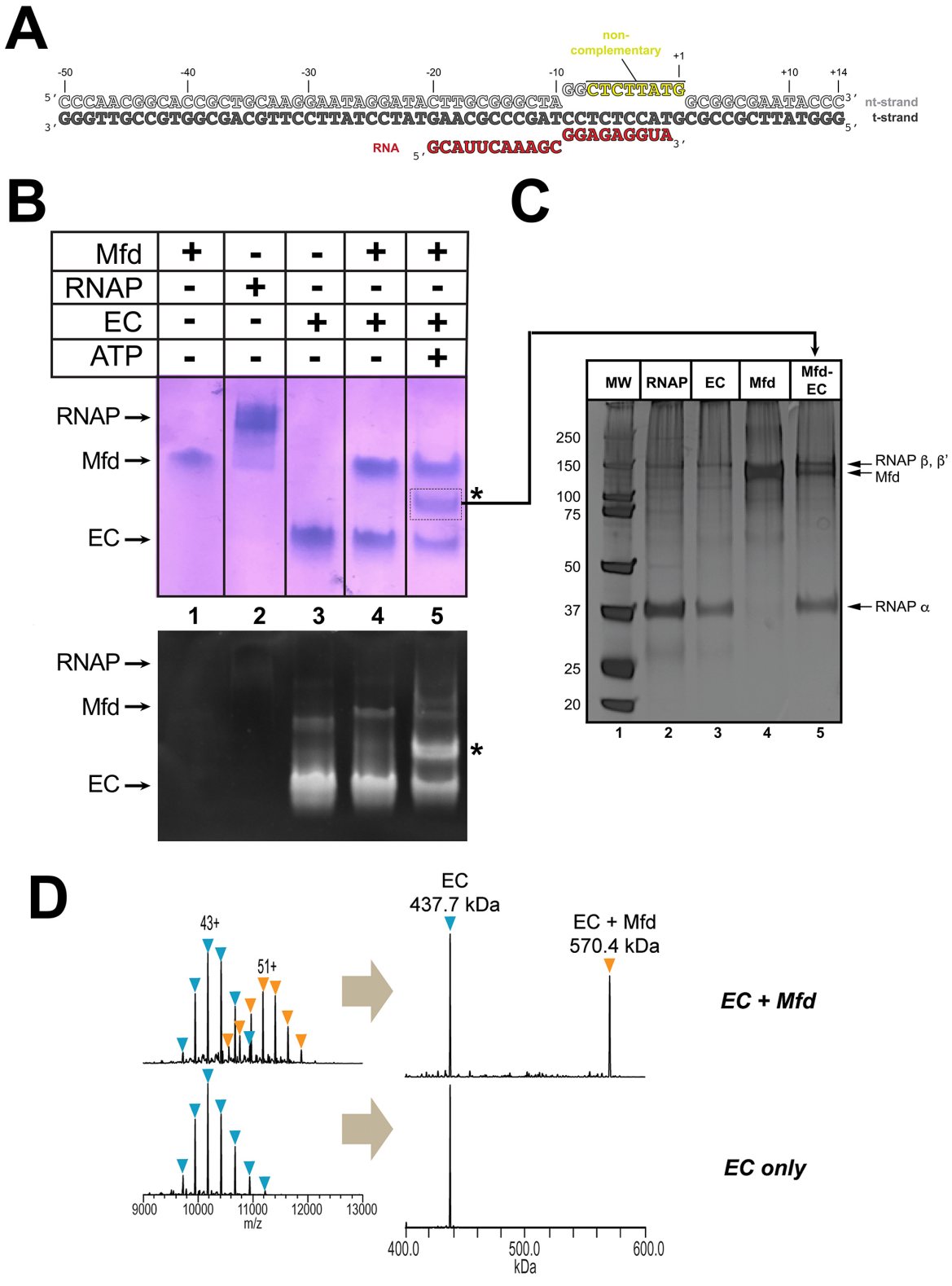
491 validated models. B.T.C., E.A.C., and S.A.D. supervised the project. All authors

492 prepared and revised the manuscript.

493

494 **Competing interests** The authors declare there are no competing interests.

495



Kang et al, Figure 1

496

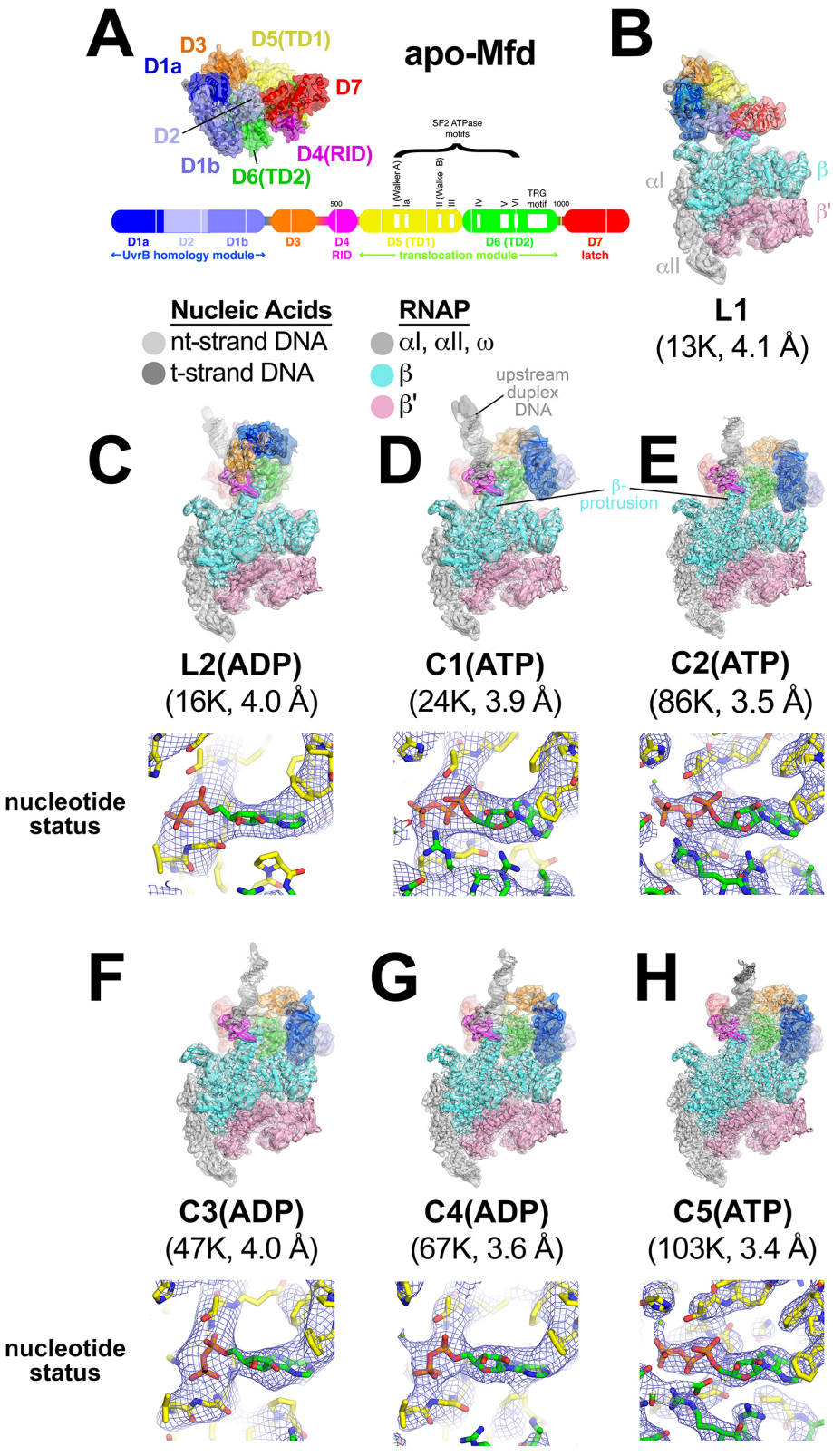
Figure 1. Formation of stable complexes between Mfd and an RNAP elongation complex.

A. The nucleic acid scaffold used for native mass spectrometry (nMS) and cryo-EM. The region of the nt-strand DNA shown in yellow is non-complementary with the t-strand.

B. Native gel electrophoretic mobility shift assay reveals the formation of a unique complex in the presence of an *Eco* RNAP EC (formed on the nucleic acid scaffold shown in A), Mfd, and 2 mM ATP (band labeled '*', lane 5). The same gel was stained with Coomassie blue to reveal protein (top panel) or Gel Red to reveal nucleic acids (bottom panel).

C. The band labeled '*' (panel B, lane 5) was excised from the gel and analyzed by SDS-polyacrylamide gel electrophoresis with silver staining, revealing the presence of RNAP and Mfd (lane 5).

D. nMS spectra and the corresponding deconvolved spectra for EC and Mfd-EC samples with the nucleic acid scaffold shown in (A). RNAP forms a stable EC with 1:1 stoichiometry (437.7 kDa complex). Upon incubation of this complex with sub-stoichiometric Mfd (Mfd:EC ratio of 1:2) and 2 mM ATP, about 40% of the EC was converted to the Mfd-EC complex (570.4 kDa) with 1:1 stoichiometry. Excess EC was mixed with Mfd because unbound Mfd yielded extremely high relative peak signals that dominated the spectrum, making interpretation and quantitation difficult.



Kang et al., Figure 2

518

519 **Figure 2. Mfd-EC structures.**

520 A. The structure of apo-Mfd [PDB 2EYQ; (Deaconescu et al., 2006)] is shown as a
521 transparent molecular surface surrounding a backbone ribbon. The domain
522 nomenclature and color-coding are schematically represented by the horizontal bar
523 below, which represents the 1148-residue *Eco* Mfd primary sequence (every
524 100 residues are marked by a vertical white line). Structural domains are shown as thick
525 bars, thin bars represent connecting linkers. The UvrB homology module (D1a-D2-D1b)
526 is structurally homologous to the namesake elements of UvrB (Deaconescu et al., 2006;
527 Selby and Sancar, 1993). D4 is the RNAP Interacting Domain (RID). D5
528 (Translocation Domain 1, or TD1) and D6 (TD2) contain the seven SF2 ATPase motifs
529 [denoted by white boxes and labeled (Gorbalenya and Koonin, 1993)] as well as the
530 TRG motif (Chambers, 2003; Mahdi et al., 2003), and together comprise the translocation
531 module.

532 B. - H. (*top*) Overall structures of Mfd-EC complexes obtained by cryo-EM. The cryo-EM
533 density maps [low-pass filtered to the local resolution (Cardone et al., 2013) are shown
534 as transparent surfaces with final models superimposed. Mfd is colored as shown in (A);
535 the RNAP and nucleic acids are colored according to the key.

536 (*bottom*) Cryo-EM density (blue mesh) and superimposed models in the region around
537 the Mfd nucleotide-binding site. Except for L1 (B), the maps were of sufficient quality to
538 identify the nucleotide status, either ADP or ATP (or ATP•P).

539 B. L1.

540 C. L2(ADP).

541 D. C1(ATP).

542 E. C2(ATP).

543 F. C3(ADP).

544 G. C4(ADP).

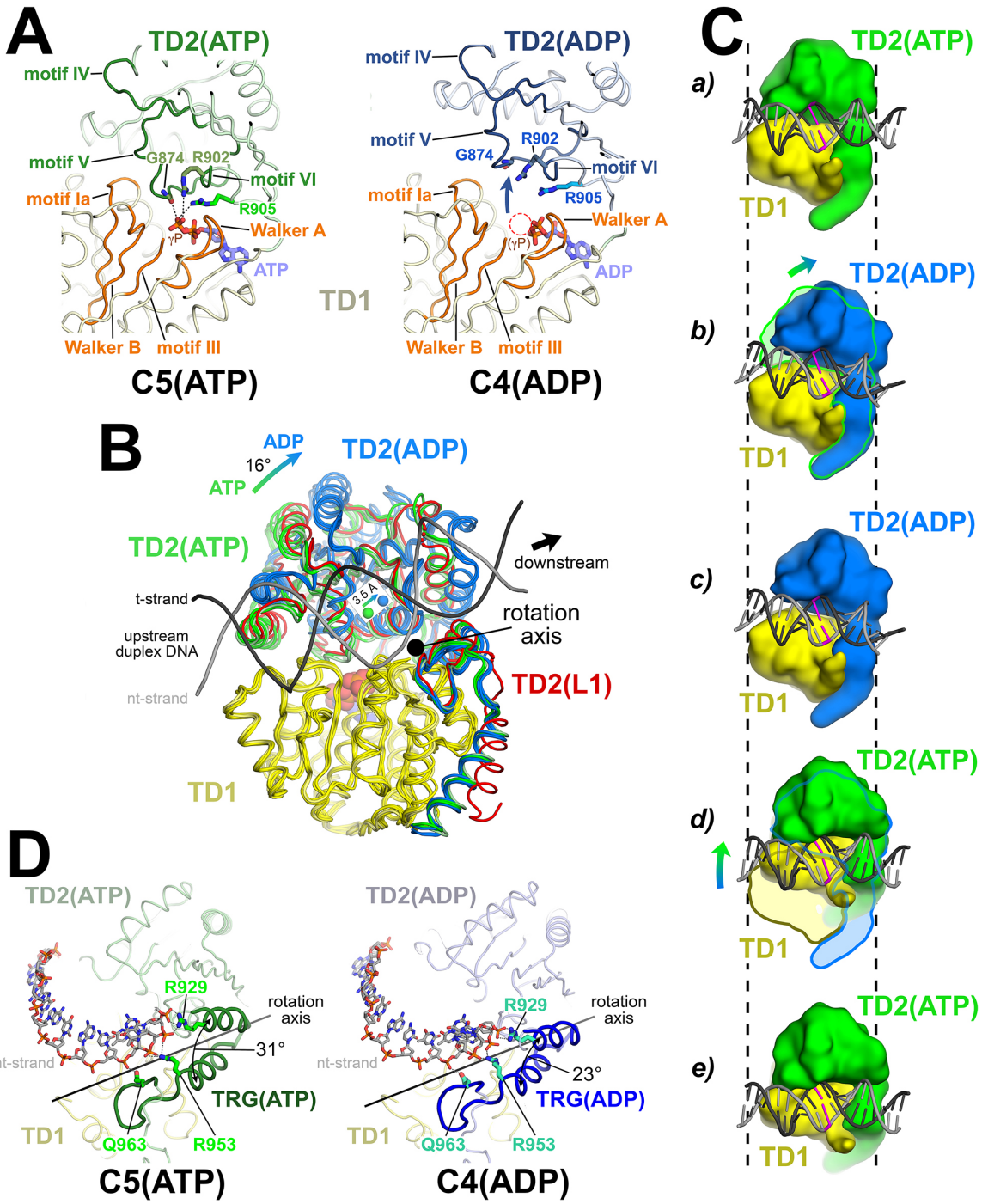
545 H. C5(ATP).

546 See also Figures S1 - S3 and Table S1.

547

548

549



Kang et al., Figure 3

552 **Figure 3. The Mfd translocation module and the DNA translocation mechanism.**

553 A. Conformational changes of the Mfd translocation module induced by ATP hydrolysis
554 and Pi release. The structural environments of the ATP [left; C5(ATP)] or ADP [right;
555 C4(ADP)] binding sites are shown. The protein is shown as a backbone worm; TD1 is
556 colored pale yellow but the SF2 ATPase motifs of TD1 (Walker A, motif Ia, Walker B,
557 motif III) are colored orange; TD2(ATP) or (ADP) are colored pale green or light blue,
558 respectively, but the SF2 ATPase motifs of TD2 (motifs IV, V, and VI) are colored dark
559 green or dark blue. The nucleotide is shown in stick format with blue carbon atoms. The
560 side chain or backbone atoms of three key residues, G874 (motif V), R902 (motif VI),
561 and R905 (just beyond motif vi) are also shown. The backbone carbonyl of G874 and
562 the side chains of R902 and R905 form polar interactions with the ATP γ-phosphate
563 (denoted as gray dashed lines). In the ADP structure, these interactions are lost due to
564 the missing γ-phosphate (denoted by the dashed red circle), causing TD2 to swing away
565 from TD1 (denoted by the thick arrow).

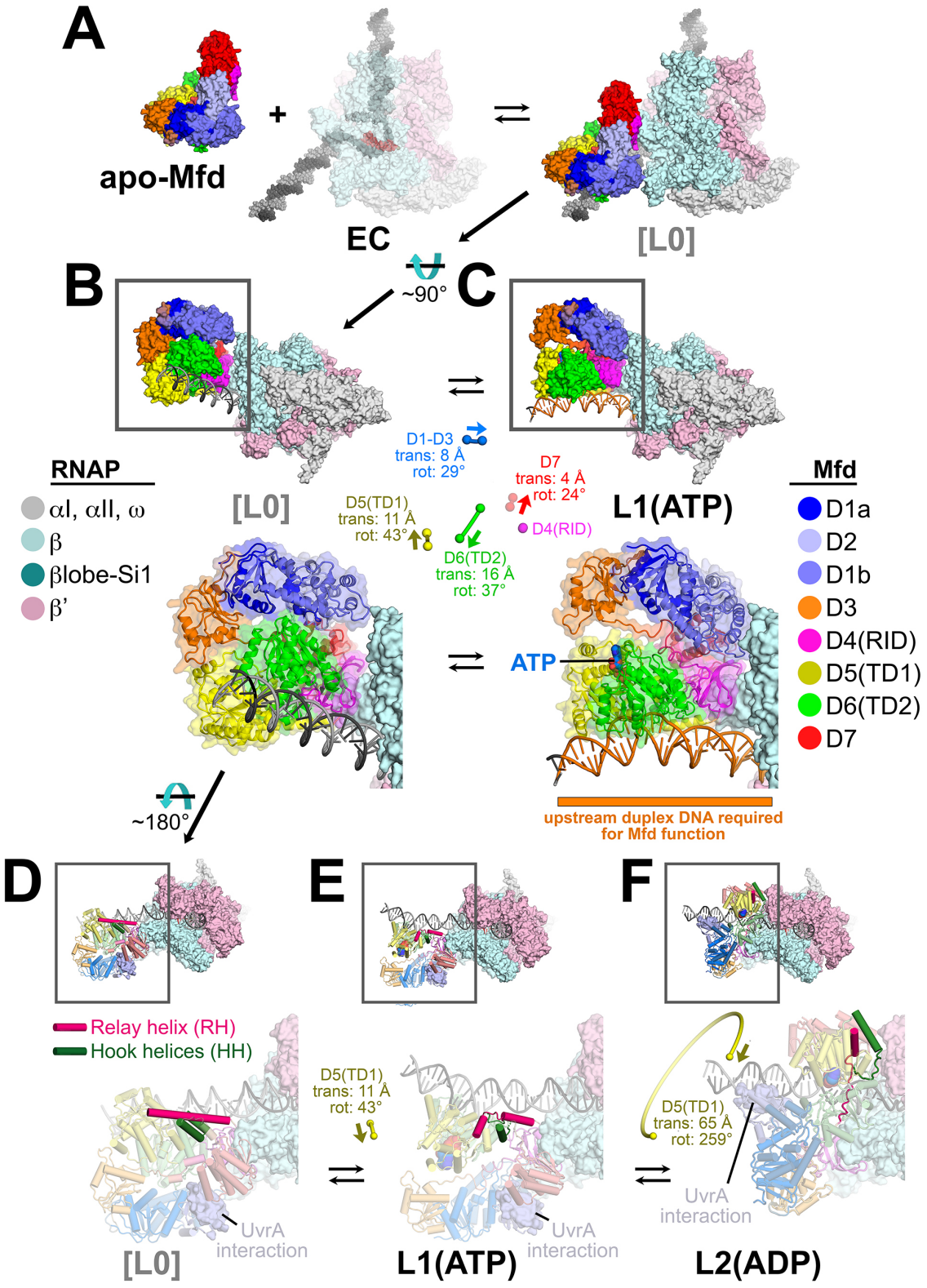
566 B. The translocation modules of all seven Mfd-EC structures were superimposed by
567 alignment of TD1 (colored yellow) α-carbons. The resulting positions of TD2 clustered
568 into two groups, those with ATP (TD2 colored green) or ADP (TD2 colored blue). TD2 of
569 L1 is shown in red and clusters with the ATP-bound structures. The relative disposition
570 of the upstream duplex DNA is also shown (gray phosphate backbone worms).
571 TD2(ATP) and TD2(ADP) are related by an ~16° rotation (denoted by the thick arrow,
572 upper left) about an axis roughly perpendicular to the DNA helical axis (denoted by the
573 black dot), resulting in a 3.5 Å shift of the TD2 center of mass roughly parallel to the
574 DNA helical axis [center of mass positions for TD2(ATP) and TD2(ADP) denoted by the

575 green and blue spheres, respectively], corresponding to one base pair rise of B-form
576 DNA.

577 C. Inchworm model for duplex DNA translocation. Duplex DNA is shown as a cartoon
578 (for reference, a central base pair is colored magenta). TD1 is colored yellow, while TD2
579 is colored green (ATP) or blue (ADP). In *a*, both TD1 and TD2(ATP) interact with the
580 duplex DNA (the initial positions of TD1 and TD2 on the DNA are denoted by the
581 vertical dashed reference lines). Upon ATP hydrolysis and Pi release, TD2(ADP) rotates
582 away from TD1 (*b*) and interacts with the DNA one base pair downstream (to the right,
583 *c*). With the release of ADP, ATP binding induces TD1 to rotate towards TD2 (*d*). In *e*,
584 TD1 and TD2(ATP) both interact with the duplex DNA but one base pair to the right.

585 D. Conformational changes of the TRG motif. Protein is shown as a backbone worm;
586 TD1 is colored pale yellow; TD2(ATP) or (ADP) are colored pale green or light blue,
587 respectively, but the TRG motifs are colored dark green or dark blue. The nt-strand of
588 the upstream duplex DNA is shown in stick format (the t-strand of the DNA is not shown
589 for clarity). Three key TRG motif residues interact with the nt-strand DNA backbone,
590 R929, R953, and Q963 (side chains shown, polar interactions with the DNA denoted by
591 the gray dashed lines). The rotation axis of the TD2(ATP) -> TD2(ADP) conformational
592 change passes directly through the TRG motif helical hairpin linker, which serves as the
593 hinge. Opening of TD2(ADP) causes the TRG helical hairpin to pinch closed nearly
594 10 Å.

595



Kang et al., Figure 4

596

Figure 4. Initial stages of the Mfd loading pathway.

Color-coding of RNAP subunits and Mfd domains are shown in the keys on the left and right, respectively.

A. Apo-Mfd [PDB 2EYQ (Deaconescu et al., 2006) combines with an EC [PDB 6ALF (Kang et al., 2017) with upstream and downstream duplex DNA extended] to form a putative initial encounter complex [L0], which was modeled by superimposing apo-Mfd D4(RID) onto the Mfd_{L1}-D4(RID) and adjusting the trajectory of the upstream duplex DNA.

B. - C. The [L0] -> L1(ATP) transition is shown. In this view, the downstream duplex DNA (and the direction of transcription) points away from the viewer.

(*top*) The Mfd-EC structures are shown as molecular surfaces with DNA shown in cartoon format. The boxed regions are magnified below.

(*bottom)* Mfd is shown as a transparent molecular surface surrounding a backbone ribbon. In the middle, the colored spheres denote the relative positions of the Mfd domain center-of-masses (com), with connecting lines denoting the motions from the L[0] -> L1(ATP) transition (the translations of the com's, as well as the relative rotation of the domains, are listed. The D4(RID) motion is negligible; also see Table S3B).

B. [L0].

C. L1(ATP): The region of the upstream duplex DNA colored orange and denoted by the orange stripe was found to be required for Mfd function on an EC (Park et al., 2002).

D. - F. View of the [L0] -> L1(ATP) -> L2(ADP) transition, highlighting the structural changes in the Mfd relay helix (RH) and hook helices (HH). In this view, the Mfd-EC

619 complex is rotated ~180° about a horizontal axis, so the downstream duplex DNA (and
620 direction of transcription) is towards the viewer.

621 (*top*) The RNAP is shown as a molecular surface, with nucleic acids shown in cartoon
622 format. Mfd is shown with cylindrical helices. Color-coding is as above but the RH is
623 colored hot pink, the HHs are colored dark green. The boxed region is magnified below.

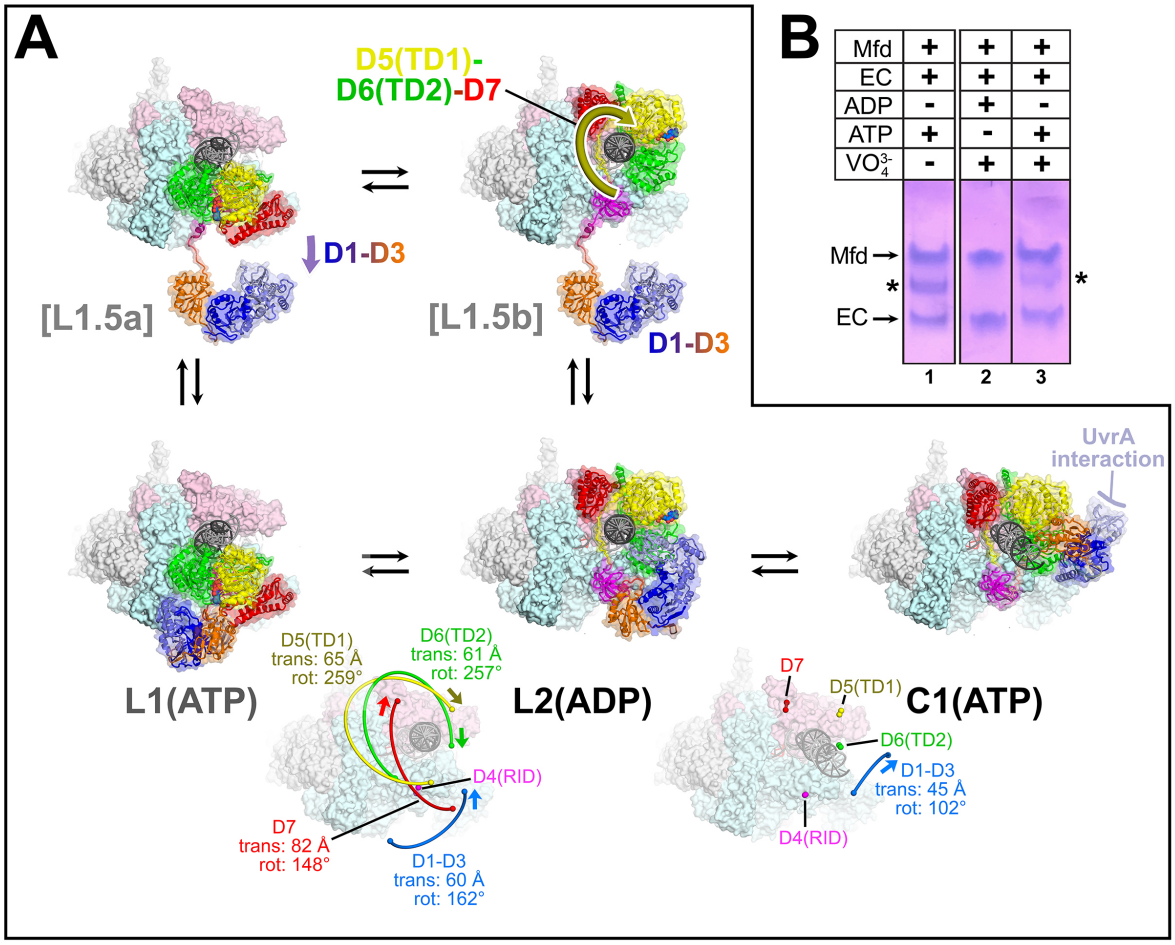
624 (*bottom*) The complexes are shown in faded colors except for the RH and HHs. Also
625 shown as a molecular surface are the residues of Mfd-D2 that interact with UvrA
626 [determined from PDB 4DFC (Deaconescu et al., 2012)].

627 D. L[0]: The RH at the very N-terminus of TD1 extends for 30 residues and is
628 surrounded by the HH's at the very C-terminus of TD2. The UvrA-interacting surface of
629 Mfd-D2 is occluded by D7 (Deaconescu et al., 2006).

630 E. L1(ATP): The middle portion of the RH helix unfolds and the RH kinks about 112°
631 around the second HH due to the translation/rotation of TD1 (denoted) and also TD2.
632 The UvrA-interacting surface of Mfd-D2 is still occluded by D7.

633 F. L2(ADP): The transition from L1(ATP) -> L2(ADP) involves a 259° rotation of TD1
634 around the backside of the DNA, as well as a 65 Å translation towards the RNAP
635 (denoted). This is likely accomplished by ATP-hydrolysis-dependent walking of the Mfd
636 translocation module and D7 along the DNA until it bumps into the RNAP. The
637 corkscrewing translocation module unfolds the N-terminal half of the RH, wrapping it
638 around the DNA as it goes. In this process D2 is separated from D7 but the UvrA-
639 interacting surface of D2 is now occluded by the DNA.

640 See also Figure S4 and Tables S2 and S3.



Kang et al., Figure 5

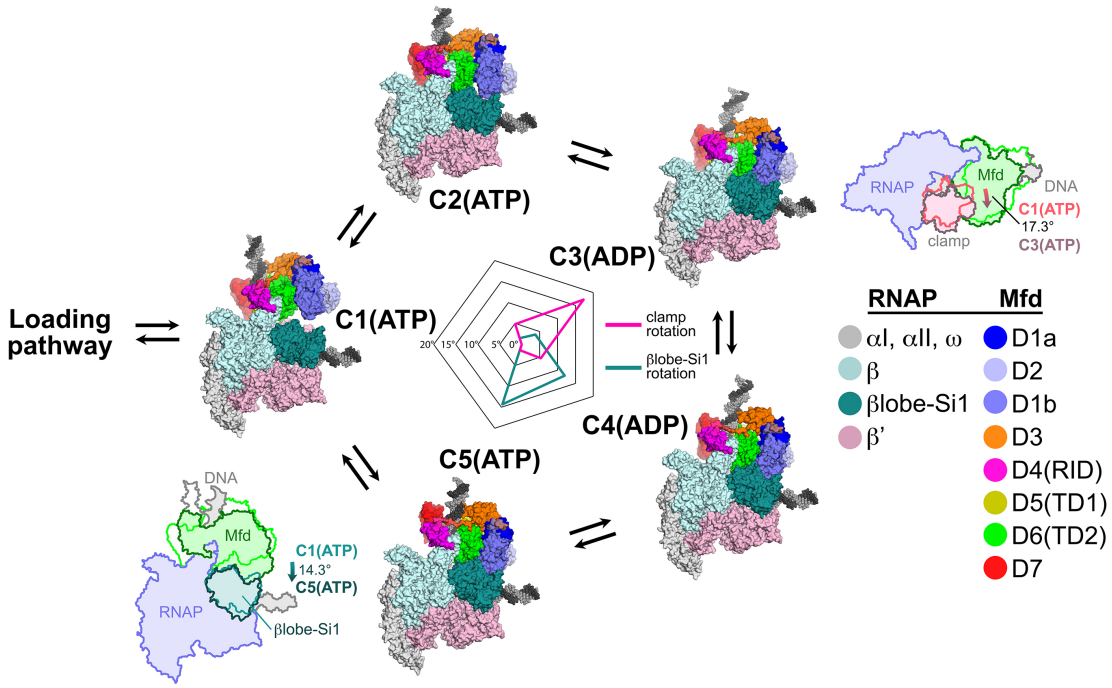
641
642

643 **Figure 5. The L1(ATP) -> L2(ADP) -> I(ATP) transition.**

644 A. A back view of the Mfd-EC structures is shown, viewing down the axis of the
645 upstream duplex DNA (the direction of transcription would be to the right). The RNAP is
646 shown as a molecular surface with nucleic acids shown in cartoon format. Mfd is shown
647 as transparent molecular surfaces surrounding the backbone ribbon. The bottom row
648 shows the transition through the experimental structures [L1(ATP) -> L2(ADP) ->
649 C1(ATP)]. Below that is denoted the translations and rotations of the Mfd domains
650 associated with each transition (superimposed on the EC structure). The large
651 translation/rotation of Mfd D5(TD1)-D6(TD2)-D7 in the L1(ATP) -> L2(ADP) transition
652 must occur by clockwise corkscrewing around the DNA minor groove in order to leave
653 behind the unfolded portion of the RH, which wraps around the DNA. The configuration
654 of Mfd D1-D3 in L1(ATP) would block this transition and also result in the entanglement
655 of linkers; thus, we propose that this transition is facilitated by transient intermediates
656 [L1.5a] and [L1.5b], which have been modeled with displaced Mfd D1-D3 tethered by
657 the long linker connecting D3 with the D4(RID). This allows the unencumbered
658 transition of Mfd D5(TD1)-D6(TD2)-D7 from [L1.5a] to [L1.5b] (illustrated by the thick
659 yellow arrow). Mfd D1-D3 then accomodates in a new configuration in L2(ADP). The
660 L2(ADP) -> C1(ATP) transition involves another large translation and rotation of D1-D3
661 (denoted), which finally exposes the Mfd-D2 UvrA-interacting surface.

662 B. Native gel electrophoretic mobility shift assay shows that a stable Mfd-EC complex
663 (denoted by '*') requires a minimum of one round of ATP hydrolysis (lane 3).

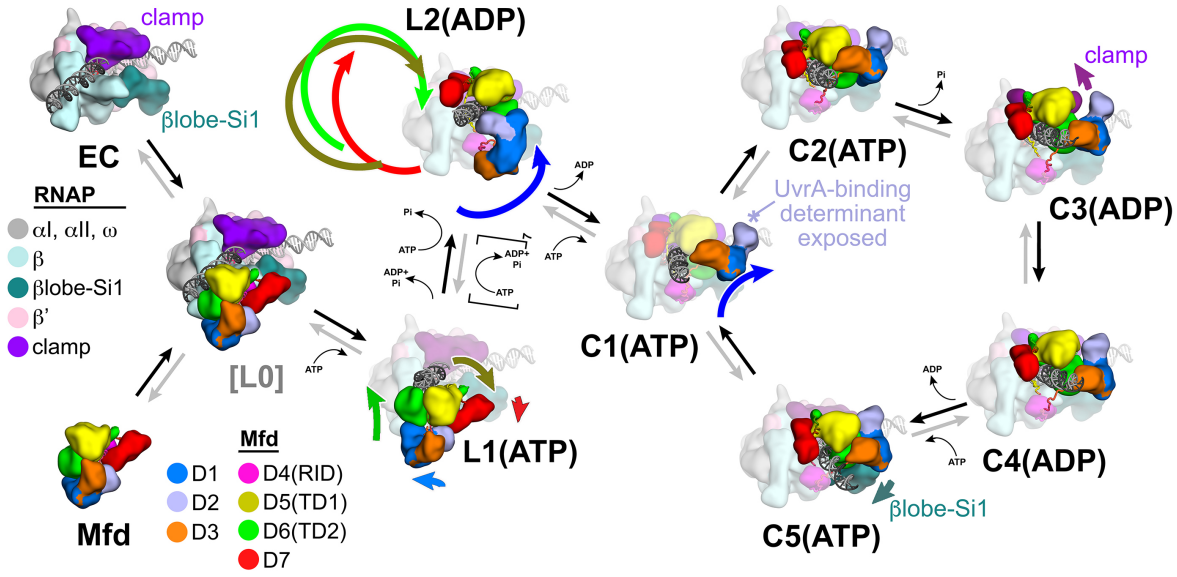
664 See also Figure S4 and Tables S2 and S3.



Kang et al., Figure 6

Figure 6. Mfd motions during its nucleotide hydrolysis cycle cause significant RNAP conformational changes.

The completion of the Mfd loading pathway culminates in the formation of C1(ATP) (Table S3A). Mfd then cycles through five distinct states in the order proposed here (also see Table S3A and Figure S4A). In looping through this cycle, internal conformational changes of Mfd are relatively small (involving primarily the nucleotide-dependent shifts of D5(TD1) and D6(TD2) with respect to each other; see Figure 3), but Mfd and the upstream duplex DNA as a whole wobble back and forth by about 30° on the upstream face of the RNAP. These motions cause significant RNAP conformational changes quantified in the radar plot in the middle. Using C1(ATP) as a reference structure, the RNAP clamp of C3(ADP) is opened 17.3° (schematically illustrated in the cartoon inset). The βlobe-Si1 domain of C5(ATP) is rotated 14.3° as illustrated.



Kang et al., Figure 7

678
679

Figure 7. Structural pathway for RNAP EC displacement by Mfd.

Putative ordered pathway for Mfd-mediated displacement of the RNAP EC (see Figure S4A). Structures are shown as cartoons, with RNAP and Mfd-domain color-coding shown in the keys on the left. For most of the structures, domains that undergo the most significant conformational changes at each step are highlighted, and the relative direction and magnitude of the conformational changes are indicated by the thick, colored arrows. Starting at the left, the RNAP EC (top left) and apo-Mfd (bottom left) combine to form the putative encounter complex [L0] (see Figures 4A, B). In the [L0] -> L1(ATP) transition (Figure 4B), Mfd domains D5(TD1) and D6(TD2) rearrange to bind a molecule of ATP and engage fully with the upstream duplex DNA. D1-D3 and D7 also make small rearrangements. In the L1(ATP) -> L2(ADP) transition, the Mfd translocation module [D5(TD1)/D6(TD2)] walks nine base-pairs in the downstream direction (into the page) by clockwise corkscrewing around the duplex DNA approximately 260°, hydrolyzing nine molecules of ATP in the process. The transition to C1(ATP) involves another accommodation of D1-D3 which finally exposes the UvrA-binding determinant of D2. We propose that the Mfd-EC complex then loops through an ATP hydrolysis cycle (C1 -> C2 -> C3 -> C4 -> C5 -> C1 -> ...) that works to overwind the upstream duplex DNA, facilitating RNA transcript release and transcription bubble collapse to displace the RNAP from the nucleic acids. During this cycle, movements of Mfd on the upstream duplex DNA also force conformational changes in the RNAP; clamp opening in C3(ADP), and βlobe-Si1 opening in C5(ATP). These RNAP conformational changes also serve to weaken the RNAP-nucleic acid contacts, facilitating RNAP release.

STAR METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

All unique/stable reagents generated in this study are available without restriction from the Lead Contact, Seth A. Darst (darst@rockefeller.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

RNAP core ($\alpha_2\beta\beta'\omega$) and Mfd are proteins found in *Eco*. For protein expression, *Eco* BL21(DE3) [*Eco* str. B F^- *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S)] was used.

METHOD DETAILS

Structural biology software was accessed through the SBGrid consortium (Morin et al., 2013).

Protein Expression and Purification

Eco RNAP (harboring full-length α-subunits) was purified as described previously (Chen et al., 2019a). A pET-based plasmid overexpressing each subunit of *Eco* RNAP (full-length α, β, ω) as well as β'-PPX-His10 (PPX; PreScission protease site, LEVLFQGP, GE Healthcare Life Sciences) was co-transformed with a pACYCDuet-1 plasmid containing *Eco* rpoZ (encoding ω) into *Eco* BL21(DE3) (Novagen). Protein expression

was induced with 1 mM isopropyl ß-D-thiogalactopyranoside (IPTG) for 4 hr at 30°C. Cells were harvested and lysed with a French Press (Avestin) at 4ºC. Lysate was precipitated using polyethyleneimine [PEI, 10% (w/v), pH 8.0, Acros Organics]. Pellets were washed and RNAP was eluted. The PEI elutions were precipitated with ammonium sulfate. Pellets were harvested, resuspended and loaded on to HiTrap IMAC HP columns (GE Healthcare Life Sciences) for purification by Ni^{2+}-affinity chromatography. Bound RNAP was washed on column, eluted and dialyzed. Dialyzed RNAP was loaded onto a Biorex-70 column (Bio-Rad) for purification by ion exchange chromatography. Eluted RNAP was concentrated by centrifugal filtration, then loaded onto a HiLoad 26/600 Superdex 200 column (GE Healthcare Life Sciences) for purification by size exclusion chromatography. Purified RNAP was supplemented with glycerol to 20% (v/v), flash frozen in liquid N_2, and stored at −80°C.

Eco Mfd was purified as described previously (Deaconescu and Darst, 2005). A pET-based plasmid overexpressing N-terminal His_6-tagged *Eco* Mfd was transformed into Rosetta(DE3)pLysS cells (Novagen). Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 hour at 30 ºC, and the cells were harvested and lysed in a buffer containing 50 mM Tris, pH 8.0, 500 mM NaCl, 15 mM imidazole, 10% (v/v) glycerol, 2 mM β-mercaptoethanol (β-ME), 1 mM PMSF, and protease inhibitor cocktail (cOmplete, EDTA-free protease inhibitor cocktail, Roche). For purification, the clarified lysate was loaded onto a Ni^{2+}-charged Hitrap IMAC (GE Healthcare Life Sciences) column and the protein was eluted using 0 – 200 mM imidazole gradient. The elutions were dialyzed in a buffer containing 20 mM Tris, pH 8.0, 100 mM NaCl, 10% (v/v) glycerol, 5 mM EDTA, and 10 mM β-ME, and loaded

747 onto a Hitrap Heparin (GE Healthcare Life Sciences) column. The protein was eluted
748 with a 100 mM – 2 M NaCl gradient, and further purified by size exclusion
749 chromatography on a HiLoad 16/600 Superdex200 (GE Healthcare Life Sciences)
750 column in a buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, and 10 mM DTT.
751 Purified Mfd was supplemented with glycerol to 20% (v/v), flash frozen in liquid N_2, and
752 stored at −80°C.

753

754 **Assembly of ECs and Mfd-EC complexes**

755 To assemble ECs, the RNA and the t-strand DNA (Figure 1A) were annealed at a 1:1
756 molar ratio in a thermocycler (95 ºC for 2 min, 75 ºC for 2 min, 45 ºC for 5 min, followed
757 by cooling to 25 ºC at 1 ºC/min). Purified *Eco* RNAP was buffer exchanged into
758 transcription buffer (20 mM Tris, pH 8.0, 150 mM KCl, 10 mM $MgCl_2$, 5 mM DTT). The
759 annealed DNA-RNA hybrid was added to the buffer-exchanged *Eco* RNAP at a 1.2:1
760 molar ratio, and incubated for 15 minutes at room temperature. Then the nt-strand DNA
761 was added and incubated for 10 minutes. To assemble Mfd-EC complexes, 2 mM ATP
762 was added to the EC, and purified Mfd was added at a 1:2 ratio.

763

764 **Biochemical analysis of Mfd/EC complexes**

765 EC-Mfd complexes were visualized by EMSA. ECs were incubated with 2-fold molar
766 excess of Mfd and 2 mM ATP at room temperature for five minutes in transcription
767 buffer. Samples were diluted with 6X loading dye containing 20 mM Tris, pH 8.0, xylene
768 cyanol, bromophenol blue, and 30% (v/v) glycerol and loaded onto 4.5% (29:1

acrylamide:bis-acrylamide) polyacrylamide native gels at 4°C in 1X TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA).

To prepare VO_4^{3-}, sodium orthovanadate (Sigma-Aldrich) was dissolved in water, and the pH was adjusted to 10 using HCl, resulting in a dark orange color. The solution was boiled for 2 min until it turned colorless, then cooled to room temperature. The process of adjusting the pH and boiling was repeated two more times. The VO_4^{3-} concentration was determined spectroscopically using its molar extinction coefficient of 2,925 $M^{-1}cm^{-1}$. The solution was stored frozen at -80°C until further use. Reactions with ATP-VO_4^{3-} contained 8 mM sodium orthovanadate (pH 10). Alternatively, a 10 mM mixture of ADP-VO_4^{3-} was added to the ECs and Mfd for a final concentration of 2 mM.

For *in vitro* EC displacement assays (Figure S4C), we followed the procedures of (Chambers, 2003) with some modifications. Stalled ECs were formed by nucleotide starvation on a 529 base-pair RasI-SmaI fragment of plasmid pAR1707 (Levin et al., 1987) end-labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP. ECs stalled at +20 were formed by initiating with the dinucleotide ApU and ATP, CTP, and GTP (excluding UTP), and with the addition of the chain-terminator 3'-deoxy-UTP. Complexes were analysed by EMSAs using 4.5% polyacrylamide (37.5:1 acrylamide/bisacrylamide) in 1X TBE at 4°C. Radiolabelled bands were detected using a phosphor screen and quantified using Imagequant software (Molecular Dynamics). *Eco* RNAP holoenzyme was assembled by incubating *Eco* RNAP (250 nM final) with σ^{70} (1.25 μM final) at 37°C for 20 min. The ^{32}P-labeled dsDNA linear scaffold (0.4 nM final) was combined with holoenzyme (10 nM final) at 37°C for 10 minutes before ApU dinucleotide (200 μM final), ATP (2 mM final), CTP (50 μM final), GTP (50 μM final), and 3'-deoxy-

UTP (100 μM final). Heparin (Sigma-Aldrich) was also added (10 μg/mL final) to ensure single-round transcription. After 10 minutes of incubation at 37°C, Mfd (250 nM final) was added to activate stalled RNAP displacement and translocation. Samples were collected at different timepoints, combined with excess EDTA to stop ATPase activity, and placed on ice before gel loading. The assays were conducted without and with the presence of VO_4^{3-} (20 mM final).

Native mass spectrometry (nMS) analysis

For the EC + Mfd sample, 7 μM EC was incubated with 3.5 μM Mfd (2:1 molar ratio) in transcription buffer with 2 mM ATP for 1 min at RT. The nonvolatile buffer components, including ATP, needed to be removed prior to nMS analysis because these components can form nonspecific adductions on protein complexes and degrade spectral quality. The EC and EC + Mfd samples were then buffer-exchanged into nMS solution (500 mM ammonium acetate, pH 7.5, 0.01% Tween-20) using Zeba microspin desalting columns (Thermo Fisher Scientific) with a 40-kDa MWCO (Olinares et al., 2016).

For nMS analysis, 2–3 μL of the buffer-exchanged sample was loaded into a gold-coated quartz emitter that was prepared in-house and then electrosprayed into an Exactive Plus EMR instrument (Thermo Fisher Scientific) with a static nanospray source (Olinares and Chait, 2019). The MS parameters used include: spray voltage, 1.2–1.3 kV; capillary temperature, 125 °C; in-source dissociation, 10 V; S-lens RF level, 200; resolving power, 8,750 at *m/z* of 200; AGC target, 1 x 10^6; maximum injection time, 200 ms; number of microscans, 5; injection flatapole, 8 V; interflatapole, 4 V; bent flatapole, 4 V; high energy collision dissociation (HCD), 180 – 200 V; ultrahigh vacuum

pressure, 6 – 7 × 10^{-10} mbar; total number of scans, at least 100. Mass calibration in positive EMR mode was performed using cesium iodide. For data processing, the acquired MS spectra were visualized using Thermo Xcalibur Qual Browser (versions 3.0.63 and 4.2.47). MS spectra deconvolution was performed either manually or using the software UniDec versions 3.2 and 4.1 (Marty et al., 2015; Reid et al., 2019). The deconvolved spectra obtained from UniDec were plotted using the m/z software (Proteometrics LLC). Experimental masses were reported as the average mass ± standard deviation (S.D.) across all the calculated mass values within the observed charge state series. Mass accuracies were calculated as the % difference between the measured and expected masses relative to the expected mass. The measured masses for the corresponding complexes (with mass accuracies reported in parentheses) include EC: 437,680 ± 20 Da (0.016%) for the EC only sample and EC + 1 Mfd: 570,360 ± 20 Da (0.03%), EC: 437,700 ± 20 Da (0.02%), $\alpha_2\beta$ subcomplex: 223,700 ± 3 Da (0.02%), and Mfd: 132,582 ± 8 Da (0.003%) for the EC sample incubated with Mg-ATP and Mfd.

Preparation of Mfd-EC Cryo-EM grids

For cryo-EM grid preparation, ECs prepared as above were purified over a Superose 6 INCREASE gel filtration column (GE Healthcare Life Sciences) equilibrated with transcription buffer. ATP (2 mM) and 2-fold excess Mfd were added and incubated for 1 min before adding {3-([3-cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate} (CHAPSO; Anatrace) to a final concentration of 8 mM (Chen et al., 2019b). The final buffer condition for all the cryo-EM samples was the same as

838 transcription buffer. C-flat holey carbon grids (CF-1.2/1.3-4Au, Protochips) were glow-
839 discharged for 20 sec prior to the application of 3.5 µL of the samples. Using a
840 Vitrobot Mark IV (Thermo Fisher Scientific Electron Microscopy), grids were blotted and
841 plunge-froze into liquid ethane with 100% chamber humidity at 22ºC.

842

843 **Cryo-EM data acquisition and processing**

844 For the cryo-EM structure determination of the Mfd-EC, two datasets were collected and
845 combined (Figure S1). In each data collection, grids were imaged using a 300 keV Titan
846 Krios (Thermo Fisher Scientific Electron Microscopy) equipped with a K2 Summit direct
847 electron detector (Gatan, Pleasanton, CA). Images were recorded with Serial EM
848 (Mastronarde, 2005) with a pixel size of 1.3 Å over a defocus range of -0.8 to -2.4 µm.
849 Movies were recorded in super-resolution mode at 8 electrons/physical pixel/s in dose-
850 fractionation mode with subframes of 0.2 s over a 10 s exposure (50 frames) to give a
851 total dose of 80 electrons/physical pixel or 47.3 electrons/Å^2. Dose-fractionated movies
852 were gain-normalized, drift-corrected, binned, summed, and dose-weighted using
853 MotionCor2 (Zheng et al., 2017). The contrast transfer function was estimated for each
854 summed image using Gctf (Zhang, 2016). Gautomatch (developed by K. Zhang, MRC
855 Laboratory of Molecular Biology, Cambridge, UK, http://www.mrc-
856 lmb.cam.ac.uk/kzhang/Gautomatch) was used to pick particles without templates.
857 Picked particles from each dataset were extracted from the dose-weighted images in
858 RELION (Scheres, 2012) using a box size of 300 pixels and subjected to
859 2D classification in RELION to exclude false particles. The selected particles from 2D
860 classification were transferred to CryoSPARC (Punjani et al., 2017) to perform

heterogeneous refinement with multiple 3D references to separate Mfd-EC complexes from isolated EC and Mfd particles. For the heterogeneous refinement, cryo-EM maps of *Eco* EC [EMD-8585; (Kang et al., 2017), Mfd-EC, and isolated Mfd (the latter two generated from a subset of the data by *ab initio* recontruction) were used as templates (each low-pass filtered to 30 Å resolution). The class representing Mfd-EC was further refined in CryoSPARC homogeneous refinement, yielding 3.9 Å and 3.6 Å nominal resolution maps from the first and second datasets, respectively. The refined particles from the first dataset were transferred to RELION for 3D auto-refinement, CTF refinement, Bayesian polishing, and 3D autorefinement before being combined with the particles from the second dataset. The resulting 594,435 particles were 3D autorefined in RELION for a consensus refinement, yielding a 3.86 Å nominal resolution map. A second round of CTF refinement, Bayesian polishing, and 3D autorefinement improved the resolution to 3.64 Å. By using focused classification around the Mfd region, eight classes were generated with distinct Mfd conformations. Among the eight classes, two classes were discarded because they could not be refined further. Four of the classes were further refined using RELION 3D autorefinement to yield C1, C3, C4, and C5 (Figure S1). The remaining two classes were further sorted using partial signal subtraction (Bai et al., 2015) of the region outside of Mfd and the RNAP □-protrusion, then classifying these subtracted particles using a mask encompassing Mfd and the RNAP β-protrusion density, resulting in L1, L2, and C2 (Figure S1). In total, the seven resulting maps showed well-defined EC density but variable quality maps for the Mfd component, indicating that the alignments were dominated by the EC portion of the complex. Therefore, to better resolve the density around Mfd, particles from each class

were imported into cryoSPARC and refined using cryoSPARC Non-uniform Refinement (Punjani et al., 2019). Using the resulting maps, masks around Mfd were generated for cryoSPARC Local Refinement. The fulcrum points (alignment centers) for each of these masks were defined using 'Volume Tracer' in UCSF Chimera (Pettersen et al., 2004). Each class from the Non-uniform Refinement was subjected to Local Refinement using their respective Mfd mask and fulcrum point. The local refinements resulted in better resolved Mfd density for each class with the following nominal resolutions in this region: L1 (ATP, 6.6 Å), L2 (ADP, 6.7 Å), C1 (ATP, 5.2 Å), C2 (ADP, 3.5 Å), C3 (ADP, 4.4 Å), C4 (ADP, 3.8 Å) and C5 (ADP, 3.3 Å).

The EC-centered cryo-EM maps from RELION and the Mfd-centered maps from cryoSPARC were combined using the PHENIX combine_focused_maps command (Adams et al., 2010). The procedure that gave the highest quality maps was as follows: 1) Coordinates were rigid body refined into each map. For the combination step: 1) For the EC-centered coordinates, the occupancies of the EC-RNAP, the EC nucleic acids, and Mfd-D4(RID) were set to 1.0, while the occupancies for the rest of Mfd and the upstream duplex DNA were set to 0; 2) For the Mfd-centered coordinates, the occupancies of the entire Mfd, the RNAP β-protrusion, and the upstream duplex DNA were set to 1.0, while the rest of the EC was set to occupancy=0. Thus, in the combined maps, the EC density and EC nucleic acids came from the EC-centered cryo-EM maps, while Mfd and the upstream duplex DNA density came from the Mfd-centered maps, and the density for the Mfd-D4(RID) and the RNAP β-protrusion were weighted according to the combine_focused_maps algorithm. These combined maps were the most interpretable over the entirety of each complex and were therefore used for

907 building, refinement, statistics reporting (Table S1), and deposition in the Electron
908 Microscopy Data Bank (EMDB). RELION 3D auto-refinement and post-processing of
909 the polished particles resulted in structures with the following nominal resolutions:
910 L1 (ATP, 4.1 Å), L2 (ADP, 4.0 Å), C1 (ATP, 3.9 Å), C2 (ADP, 3.9 Å), C3 (ADP, 3.2 Å),
911 C4 (ADP, 3.6 Å) and C5 (ADP, 3.3 Å). Local resolution calculations were done using
912 blocres and blocfilt from the Bsoft package (Cardone et al., 2013).

913

914 **Model building and refinement**

915 For initial models of the complexes, the *Eco* EC structure [PDB ID 6ALF (Kang et al.,
916 2017)] was manually fit into the combined cryo-EM density maps using Chimera
917 (Pettersen et al., 2004) and real-space refined using Phenix (Adams et al., 2010). The
918 DNAs and Mfd domains [taken from 2EYQ (Deaconescu et al., 2006)were mostly built
919 *de novo* based on the density maps. For real-space refinement, rigid body refinement
920 was followed by all-atom and B-factor refinement with Ramachandran and secondary
921 structure restraints. Refined models were inspected and modified in Coot (Emsley and
922 Cowtan, 2004).

923

924 **Superimposition of structures and calculation of rmsds**

925 For the statistics presented in Tables S2A, S2B, and S3A, α-carbons of the specified
926 structural components were superimposed using the PyMOL align command (the
927 resulting rmsd is listed in the 'align' column, with the number of α-carbon atoms used for
928 the rmsd calculation listed underneath in parentheses). The rmsd for all of the α-carbon

929 atoms was then determined using the PyMOL 'rms_cur' command (listed under the
930 'rms_cur' column).

931

932 **Quantification and statistical analysis**

933 The nMS spectra were visualized using Thermo Xcalibur Qual Browser (versions 3.0.63
934 and 4.2.27), deconvolved using UniDec versions 3.2 and 4.1 (Marty et al., 2015; Reid et
935 al., 2019) and plotted using the m/z software (Proteometrics LLC, New York, NY).
936 Experimental masses (Figure 1D) were reported as the average mass ± standard
937 deviation across all the calculated mass values obtained within the observed charge
938 state distribution.

939 ImageQuant 5.2 (GE Healthcare, Pittsburgh PA) was used to visualize and
940 quantify gels. To quantify the EC displacement assays (Figure S4C), mean values and
941 the standard error of the mean from at three independent measurements were
942 calculated.

943 Structural biology software was accessed through the SBGrid consortium (Morin
944 et al., 2013). The local resolution of the cryo-EM maps (Figure S2) was estimated using
945 blocres (Cardone et al., 2013) with the following parameters: box size 15, verbose 7,
946 sampling 1.3, and cutoff 0.5. The quantification and statistical analyses for model
947 refinement and validation were generated using MolProbity (Chen et al., 2010) and
948 PHENIX (Adams et al., 2010).

949

950

Data and code availability

The cryo-EM density maps have been deposited in the EMDataBank under accession codes EMD-21996 [L1(ATP)], EMD-22006 [L2(ADP)], EMD-22012 [I(ATP)], EMD-22039 [II(ATP)], EMD-22043 [III(ADP)], EMD-22044 [IV(ADP)], and EMD-22045 [V(ATP)]. The atomic coordinates have been deposited in the Protein Data Bank under accession codes 6X26 [L1(ATP)], 6X2F [L2(ADP)], 6X2N [I(ATP)], 6X43 [II(ATP)], 6X4W [III(ADP)], 6XYY [IV(ADP)], and 6X5Q [V(ATP)].

SUPPLEMENTAL INFORMATION

Structural basis for transcription complex disruption by the Mfd translocase

Jin Young Kang[1,3,4], Eliza Llewellyn[1,3], James Chen[1], Paul Dominic B. Olinares[2], Joshua Brewer[1], Brian T. Chait[2], Elizabeth A. Campbell[1], Seth A. Darst[1,5,6]

[1]Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY 10065, USA.

[2]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA.

[3]Contributed equally

[4]Present address: Department of Chemistry KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

[5]Lead contact: darst@rockefeller.edu

[6]Correspondence to: darst@rockefeller.edu

Supplemental Information includes 4 figures, 3 tables, and 2 movies.

977 **Table S1 | Cryo-EM data collection, refinement and validation statistics. Related to Figure 2.**
978

Data collection and processing							
Voltage (kV)				300			
Electron exposure (e–/Å^2)				60			
Defocus range (μm)				-0.8 to -2.4			
Pixel size (Å)				1.3			
Symmetry imposed				C1			
Initial particle images (no.)							
Structure	L1	L2(ADP)	I(ATP)	II(ATP)	III(ADP)	IV(ADP)	V(ATP)
Final particle images (no.)	13,013	16,067	23,856	85,840	47,312	66,617	102,741
Map resolution (Å)	4.1	4.0	3.9	3.6	3.8	3.6	3.3
FSC threshold 0.143							
Map resolution range (Å)	3.4 - 10	3.5 - 10	3.3 - 10	3.0 - 9	3.3 - 10	3.1 - 9.5	2.8 - 8
EMDB accession code	EMD-21996	EMD-22006	EMD-22012	EMD-22039	EMD-22043	EMD-22044	EMD-22045
Refinement							
Model composition							
Non-hydrogen atoms	35,968	36,058	36,078	36,055	36,019	36,086	36,075
Protein residues	4,314	4,316	4,316	4,316	4,316	4,316	4,315
Nucleic acid residues	113	113	113	111	112	113	113
Ligands	4 (1 Mg^{2+}, 2 Zn^{2+}, 1 ATP)	4 (1 Mg^{2+}, 2 Zn^{2+}, 1 ADP)	4 (1 Mg^{2+}, 2 Zn^{2+}, 1 ATP)	5 (2 Mg^{2+}, 2 Zn^{2+}, 1 ATP)	4 (1 Mg^{2+}, 2 Zn^{2+}, 1 ADP)	5 (2 Mg^{2+}, 2 Zn^{2+}, 1 ADP)	5 (2 Mg^{2+}, 2 Zn^{2+}, 1 ATP)
B factors (Å^2)							
Protein	142.0	97.54	89.23	52.86	127.5	82.27	66.68
Nucleic acid	202.1	182.9	145.0	110.4	202.9	148.5	113.0
Ligands	258.7	89.94	87.36	36.59	121.0	79.73	47.05
R.m.s. deviations							
Bond lengths (Å)	0.006	0.007	0.010	0.007	0.010	0.009	0.004
Bond angles (°)	0.839	0.873	1.003	0.783	0.975	0.825	0.652
Validation							
MolProbity score	2.53	2.54	2.61	2.22	2.59	2.32	2.72
Clashscore	27.07	27.91	29.14	14.4	25.79	18.05	11.93
Poor rotamers (%)	0.70	0.47	0.55	0.47	0.75	0.88	7.74
Ramachandran plot[a]							
Favored (%)	88.96	88.54	86.21	89.66	87.64	91.1	92.2
Allowed (%)	10.71	11.06	13.70	10.15	12.08	8.86	7.57
Outliers (%)	0.33	0.40	0.09	0.19	0.28	0.40	0.23
PDB accession code	6X26	6X2F	6X2N	6X43	6X4W	6XYY	6X5Q

979 [a] Refinement/validation parameters as calculated by PHENIX real_space_refine (Adams et al., 2010) and MOLPROBITY (Chen et al., 2010)
980

Table S2. Conformational changes for the RNAP (A) and Mfd (B) components of the Mfd-EC structures. Related to Figure 2.

Table S2A. Conformational changes for the RNAP component of the Mfd-EC structures.

	L2(ADP)		C1(ATP)		C2(ATP)		C3(ADP)		C4(ADP)		C5(ATP)	
	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur
L1(ATP)	0.556 (2798)	0.82 (3171)	0.596 (2832)	0.877 (3172)	0.661 (2536)	1.46 (3169)	0.707 (2362)	3.638 (3164)	0.642 (2318)	2.203 (3171)	0.553 (2583)	2.188 (3172)
L2(ADP)			0.456 (2860)	0.585 (3171)	0.586 (2409)	1.318 (3168)	0.711 (2322)	3.61 (3163)	0.603 (2332)	2.244 (3170)	0.565 (2625)	2.313 (3171)
C1(ATP)					0.544 (2393)	1.24 (3169)	0.693 (2279)	3.514 (3164)	0.548 (2328)	2.177 (3171)	0.522 (2592)	2.3 (3172)
C2ATP)							0.546 (2343)	2.69 (3167)	0.303 (2480)	1.818 (3168)	0.346 (2319)	2.244 (3169)
C3(ADP)									0.57 (2326)	2.669 (3163)	0.585 (2351)	3.518 (3164)
C4(ADP)											0.393 (2492)	1.044 (3171)

Table S2B. Conformational changes for the Mfd component of the Mfd-EC structures.

	L1(ATP)		L2(ADP)		C1(ATP)		C2(ATP)		C3(ADP)		C4(ADP)		C5(ATP)	
	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur
2EYQ	4.998 (916)	9.179 (1113)	19.606 (889)	34 (1106)	12.002 (897)	36.909 (1107)	13.519 (907)	37.193 (1115)	18.508 (940)	36.746 (1109)	16.586 (928)	35.991 (1109)	10.253 (891)	36.375 (1109)
L1(ATP)			20.303 (910)	32.199 (1115)	13.454 (901)	37.08 (1110)	16.895 (930)	36.536 (1114)	14.825 (905)	38.634 (1114)	13.717 (900)	37.494 (1112)	12.331 (898)	36.58 (1112)
L2(ADP)					20.605 (1032)	26.011 (1135)	22.779 (1075)	25.75 (1131)	20.421 (1076)	23.088 (1139)	21.076 (1140)	21.205 (1139)	24.73 (1140)	24.818 (1135)
C1(ATP)							0.827 (930)	3.954 (1134)	3.686 (1020)	5.674 (1138)	4.714 (976)	7.676 (1136)	0.747 (925)	7.109 (1136)
C2(ATP)									3.888 (1031)	5.171 (1136)	4.449 (1003)	7.294 (1134)	0.667 (915)	5.769 (1132)
C3(ADP)											1.453 (953)	5.61 (1140)	3.637 (938)	7.392 (1138)
C4(ADP)													4.946 (1122)	5.098 (1136)

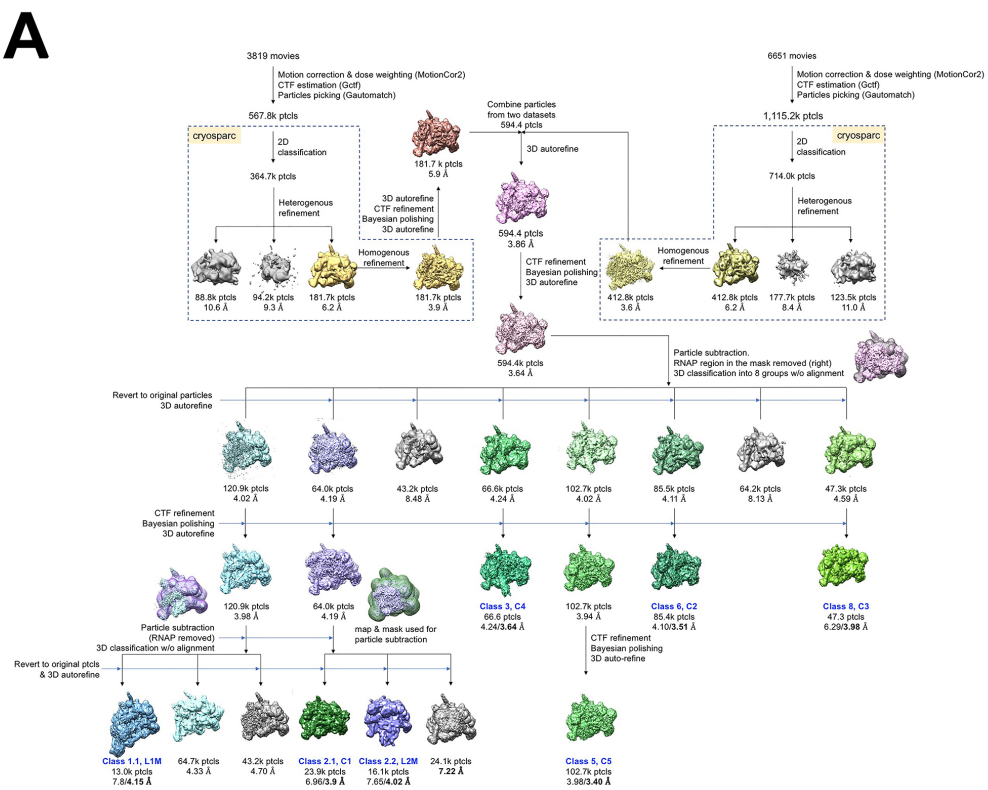
Table S3. Conformational changes for the entire complex (S3A), and conformational transitions (translation and rotation) of Mfd domains from one state to the next. Related to Figures 2-7.

Table S3A. Conformational changes for the entire Mfd-EC complexes.

	L2(ADP)		I(ATP)		II(ATP)		III_V(ADP)		IV(ADP)		V_III(ATP)	
	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur	align	rms_cur
L1(ATP)	0.747 (3143)	35.736 (4286)	0.851 (3177)	42.301 (4282)	1.119 (3071)	44.219 (4283)	1.934 (3025)	46.889 (4278)	1.631 (3034)	44.535 (4283)	1.185 (3061)	44.162 (4284)
L2(ADP)			0.534 (3135)	19.00 (4306)	1.037 (3033)	22.844 (4299)	1.521 (2977)	24.626 (4302)	1.269 (3009)	23.003 (4309)	0.879 (3066)	24.152 (4306)
I(ATP)					0.967 (3034)	6.839 (4303)	1.450 (2957)	10.137 (4302)	1.067 (2937)	9.960 (4307)	0.788 (3033)	11.428 (4308)
II(ATP)							2.127 (3247)	5.428 (4303)	0.458 (2928)	4.699 (4302)	0.793 (2926)	7.738 (4301)
III_V(ADP)									2.765 (4186)	3.736 (4303)	2.905 (3540)	6.433 (4302)
IV(ADP)											0.633 (3071)	4.491 (4307)

Table S3B. Conformational transitions (translation of center-of-gravity and rotation) for Mfd domains.

	Structural transition					
	L(0) -> L1(ATP)		L1(ATP) -> L2(ADP)		L2(ADP) -> I(ATP)	
Mfd structural domains	cog translation (Å)	rotation (°)	cog translation (Å)	rotation (°)	cog translation (Å)	rotation (°)
D1-D3	7.7	28.7	59.6	162.0	44.5	102.0
D4(RID)	0.37	2.8	3.9	11.0	0.7	1.6
D5(TD1)	10.9	42.5	64.5	259.1	2.8	16.6
D6(TD2)	15.9	36.6	61.0	256.7	2.7	10.0
D7	4.3	24.4	81.8	147.7	6.8	14.0
Mfd overall	cog translation (Å)	rotation (°)				
I(ATP) -> II(ATP)	9.8	15.5				
II(ATP) -> III(ADP)	7.9	14.8				
III(ADP) -> IV(ADP)	8.3	1.9				
IV(ADP) -> V(ATP)	5.1	10.4				
V(ATP) -> I(ATP)	12.6	30.9				

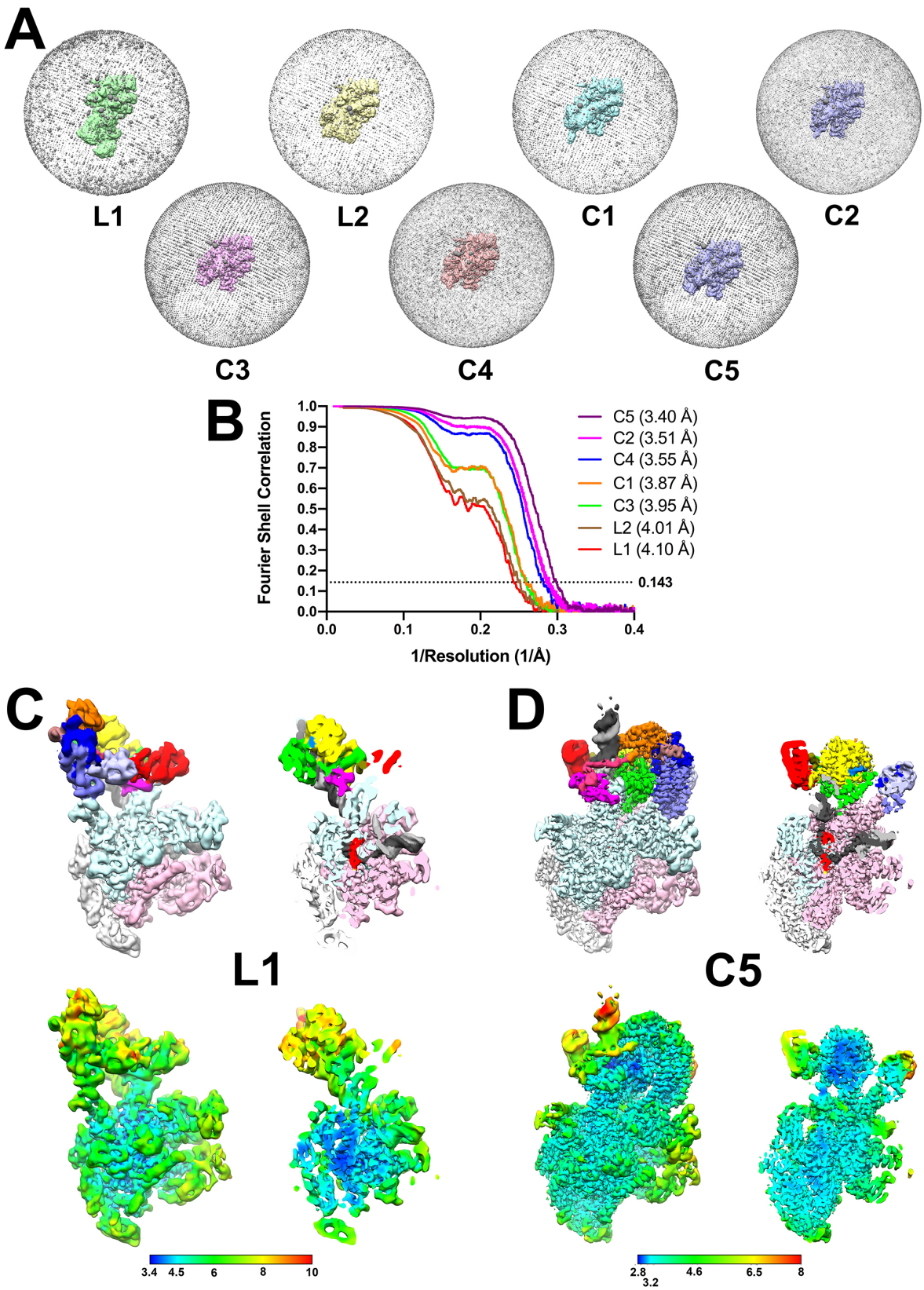


Kang et al., Figure S1

Figure S1. Cryo-EM processing pipeline for Mfd-EC complexes. Related to Figure 2.

A. Cryo-EM processing pipeline for Mfd-EC complexes.

B. EC-centered and Mfd-centered maps were combined using the PHENIX combine_focused maps command (Adams et al., 2010).



Kang et al, Figure S2

1001

Figure S2. Cryo-EM of Mfd-EC complexes. Related to Figure 2.

A. Angular distribution of particle projections for each structural class.

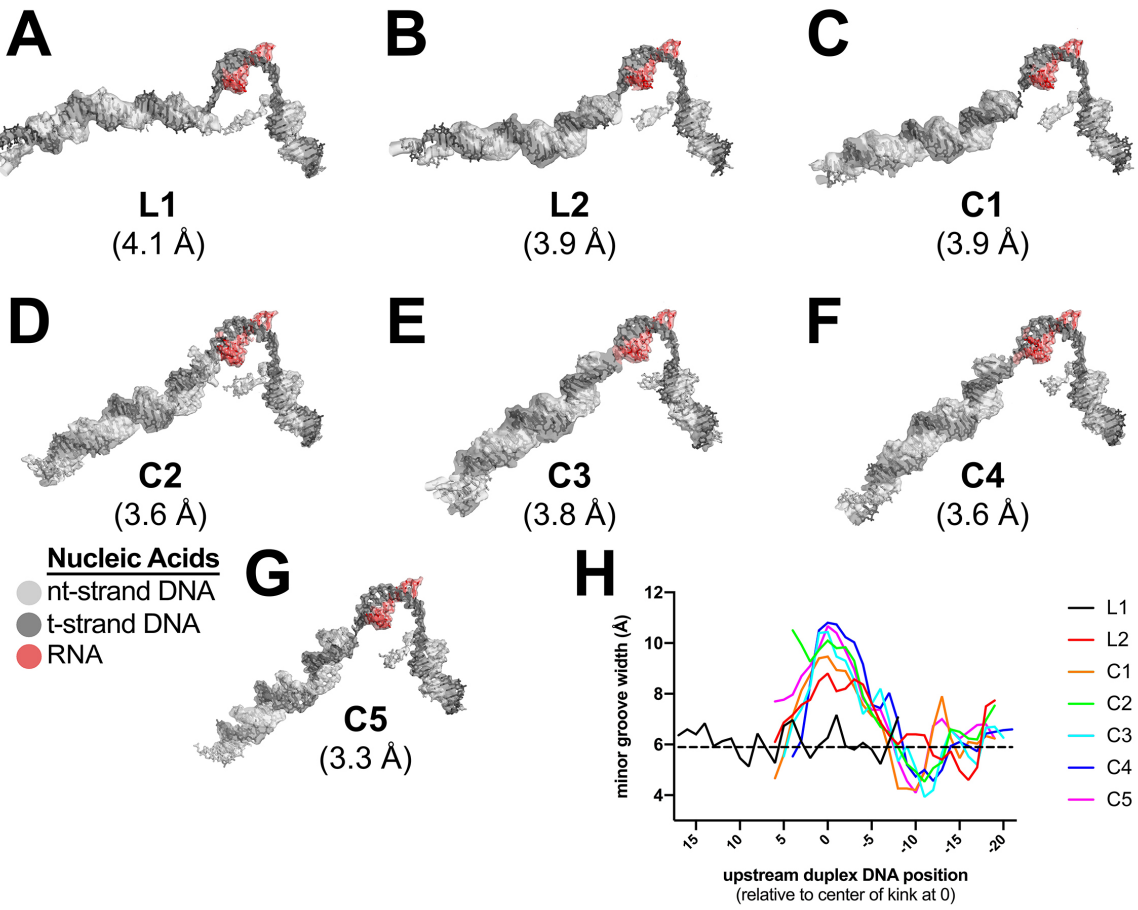
B. Gold-stndard FSC for the EC-centered/Mfd-centered composite maps, calculated by comparing the two independently determined half-maps using the MTRIAGE module (Afonine et al., 2018) of PHENIX (Adams et al., 2010). The dotted line represents the 0.143 FSC cutoff.

C-D. (*top*) Views of the cryo-EM density map, colored according to the key of Figure 2. The right view is a cross-section through the center of the left view.

(*bottom*) Same views as on top but colored by local resolution (Cardone et al., 2013).

C. L1, the lowest resolution structural class.

D. C5, the highest resolution structural class.



Kang et al., Figure S3

Figure S3. Examples of cryo-EM density and Mfd-induced DNA kink.

A.-G. Cryo-EM density maps, filtered according to the local resolution (Cardone et al., 2013), corresponding to the nucleic acids, are shown as transparent surfaces with the final refined models superimposed. The color-coding is shown in the key.

A. L1.

B. L2.

C. C1.

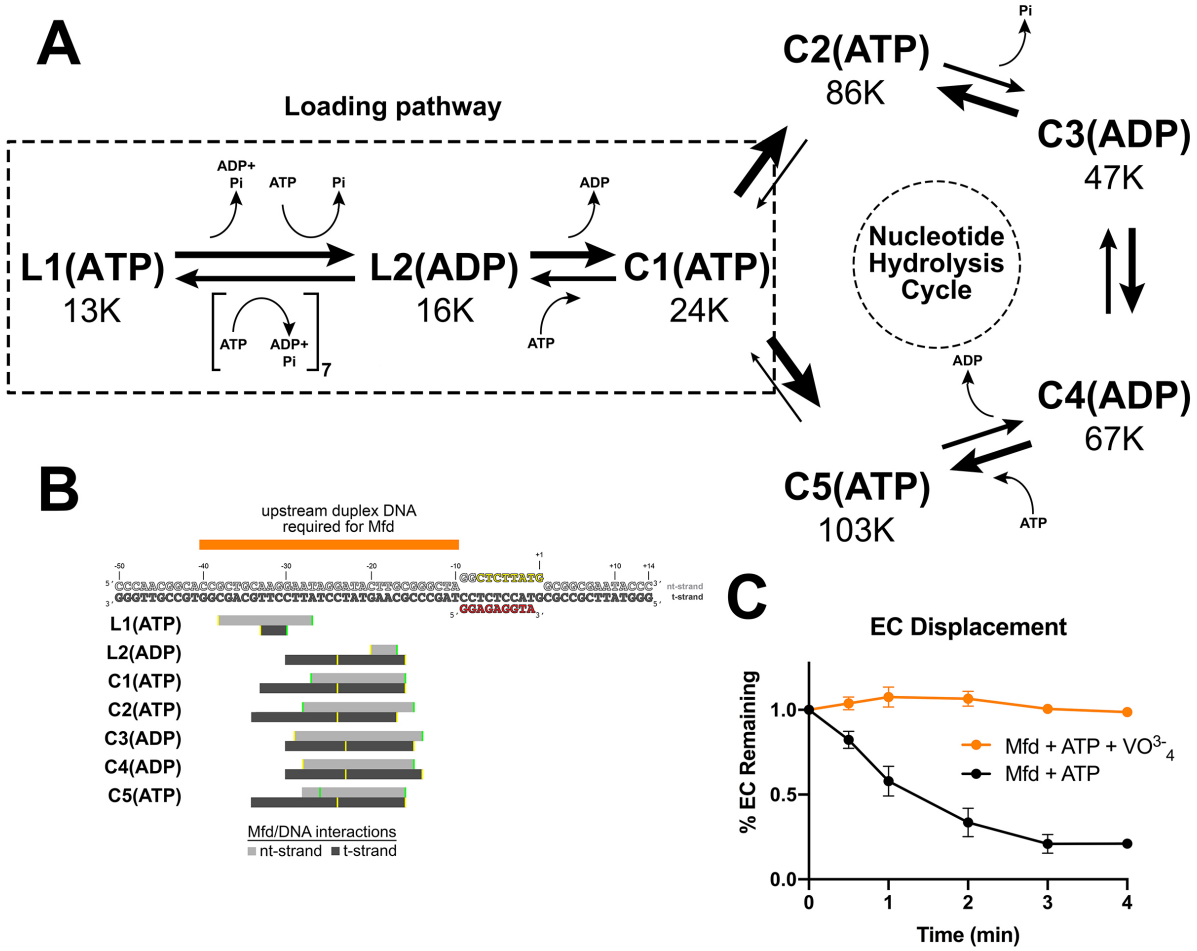
D. C2.

E. C3.

F. C4.

G. C5.

H. Plot showing the minor groove width [calculated using Curves+ (Lavery et al., 2009)] of the upstream duplex DNA, aligned by the center of the kink (defined as position '0').



Kang et al., Figure S4

Figure S4. Putative structural pathway for Mfd activity, Mfd/DNA interactions, and inhibition of EC displacement by VO^{3-}_4.

A. Putative ordered pathway of seven Mfd-EC structures (Figures 2B-H). Each structure is shown in it's position along with the number of particles that gave rise to that structural class (Figure S1). All the structures came from identically prepared samples so the particle numbers likely represent the stability of each complex. L1(ATP) was placed first in the pathway because the Mfd component [$Mfd(ATP)_{L1}$ most closely superimposes with apo-Mfd (Table S2B). The structures can be grouped into two main groups, L1(ATP) and L2(ADP), which don't superimpose well with any of the other structures, and C1(ATP), C2(ATP), C3(ADP), C4(ADP), and C5(ATP), which are all relativey similar to each other (Tables S2B, S3A). We thus place L1(ATP) and L2(ADP) in a loading pathway, while C1-C5 represent the fully EC-engaged nucleotide hydrolysis cycle (NHC) for Mfd. As described in the text, the complete loading pathway requires a minimum of ten ATP hydrolysis events, while each cycle of the NHC requires one ATP hydrolysis.

B. (*top*) The nucleic acid scaffold is shown (same as Figure 1A except the upstream single-stranded RNA is not shown for clarity). Above the sequences, the orange bar denotes the extent of upstream duplex DNA required for Mfd function (Park et al., 2002). The gray bars below denote the extent of Mfd/DNA interactions in the seven Mfd-EC structures (light gray bar, nt-strand interacts; dark gray bar, t-strand). The interactions of $Mfd(ATP)_{L1}$ with the DNA explain the requirement for ~40 base pairs of upstream duplex DNA (Park and Roberts, 2006) and indicate that L1(ATP) is an obligate intermediate in the pathway.

C. Displacement of stalled *Eco* ECs from end-labeled DNA fragments by Mfd was monitored by EMSA and quantified using a phosphorimager and Imagequant software (Chambers, 2003). Data shown are the average of three independent experiments and are expressed as a percentage of the amount of EC present prior to the addition of Mfd. Error bars indicate standard deviation.

SUPPLEMENTAL VIDEO TITLES AND LEGENDS

Video S1. Mfd translocation. Related to Figure 3.

The video illustrates the Mfd translocation module inchworm translocation model. The Mfd translocation module [D5(TD1), yellow; D6(TD2), green] transitions between the closed ATP-bound state and the open ADP-bound state as it translocates on duplex DNA. In the first section of the video, the Mfd translocation module inchworms on stationary DNA. The second section of the video simulates the effect of blocking the forward (left-to-right) motion of the protein but with continued ATP hydrolysis - the protein continues to translocate with respect to the DNA, but since the protein can't move, the DNA moves (from right-to-left) instead. A reference base pair in the DNA is colored magenta.

Video S2. The Mfd nucleotide hydrolysis cycle. Related to Figure 6.

The video starts with two orthogonal views of C1 (color-coded as in Figure 6, with the RNAP βlobe-Si1 colored in teal, and the RNAP clamp colored in purple, and each domain highlighted with a transparent molecular surface). The video cycles through the five structures of the Mfd NHC (C1 -> C2 -> C3 -> C4 -> C5). The current structure is highlighted in red (lower left). At C2 -> C3 the motions of Mfd open the RNAP clamp, and at C4 -> C5 the motions of Mfd twist the βlobe-Si1. After two cycles, two more cycles are shown but Mfd and the nucleic acids are removed to further highlight the motions of RNAP.

1083 **REFERENCES**

1084

1085 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd,
1086 J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a
1087 comprehensive Python-based system for macromolecular structure solution. Acta
1088 Crystallographica Section D Biological Crystallography *66*, 213–221.

1089 Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C.,
1090 Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of
1091 cryo-EM maps and atomic models. Acta Crystallographica. Section D, Structural Biology
1092 *74*, 814–840.

1093 Bai, X., Rajendra, E., Yang, G., Shi, Y., and Scheres, S.H.W. (2015). Sampling the
1094 conformational space of the catalytic subunit of human γ-secretase. ELife *4*, e11182.

1095 Bockrath, R., Barlow, A., and Engstrom, J. (1987). Mutation frequency decline in
1096 Escherichia coli B/r after mutagenesis with ethyl methanesulfonate. Mutation Research
1097 *183*, 241–247.

1098 Bohr, V.A., Smith, C.A., Okumoto, D.S., and Hanawalt, P.C. (1985). DNA repair in an
1099 active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much
1100 more efficient than in the genome overall. Cell *40*, 359–369.

1101 Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD Damage
1102 Recognition by Transcribing RNA Polymerase II. Science *315*, 859–862.

1103 Cardone, G., Heymann, J.B., and Steven, A.C. (2013). One number does not fit all:
1104 mapping local variations in resolution in cryo-EM reconstructions. Journal of Structural
1105 Biology *184*, 226–236.

1106 Chambers, A.L. (2003). A DNA translocation motif in the bacterial transcription-repair
1107 coupling factor, Mfd. Nucleic Acids Research *31*, 6409–6418.

1108 Chen, J., Gopalkrishnan, S., Chiu, C., Chen, A.Y., Campbell, E.A., Gourse, R.L., Ross,
1109 W., and Darst, S.A. (2019a). E. coli TraR allosterically regulates transcription initiation
1110 by altering RNA polymerase conformation. ELife *8*.

1111 Chen, J., Noble, A.J., Kang, J.Y., and Darst, S.A. (2019b). Eliminating effects of particle
1112 adsorption to the air/water interface in single-particle cryo-electron microscopy_
1113 Bacterial RNA polymerase and CHAPSO. Journal of Structural Biology: X *1*, 100005.

1114 Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A.Y., Olinares, P.D.B., Saecker, R.M.,
1115 Winkelman, J.T., Maloney, M.F., Chait, B.T., Ross, W., et al. (2020). Stepwise Promoter
1116 Melting by Bacterial RNA Polymerase. Mol Cell.

1117 Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J.,
1118 Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom
1119 structure validation for macromolecular crystallography. Acta Crystallographica Section
1120 D Biological Crystallography *66*, 12–21.

1121 Davies, D.R., and Hol, W.G.J. (2004). The power of vanadate in crystallographic
1122 investigations of phosphoryl transfer enzymes. FEBS Letters *577*, 315–321.

1123 Deaconescu, A.M., and Darst, S.A. (2005). Crystallization and preliminary structure
1124 determination of Escherichia coliMfd, the transcription-repair coupling factor. Acta
1125 Crystallographica Section F Structural Biology and Crystallization Communications *61*,
1126 1062–1064.

1127 Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery,
1128 N.J., and Darst, S.A. (2006). Structural Basis for Bacterial Transcription-Coupled DNA
1129 Repair. Cell *124*, 507–520.

1130 Deaconescu, A.M., Sevostyanova, A., Artsimovitch, I., and Grigorieff, N. (2012).
1131 Nucleotide excision repair (NER) machinery recruitment by the transcription-repair
1132 coupling factor involves unmasking of a conserved intramolecular interface. P Natl Acad
1133 Sci Usa *109*, 3353–3358.

1134 Dillingham, M.S., Wigley, D.B., and Webb, M.R. (2000). Demonstration of Unidirectional
1135 Single-Stranded DNA Translocation by PcrA Helicase: Measurement of Step Size and
1136 Translocation Speed †. Biochemistry-Us *39*, 205–212.

1137 Fan, J., Leroux-Coyau, M., Savery, N.J., and Strick, T.R. (2016). Reconstruction of
1138 bacterial transcription-coupled repair at single-molecule resolution. Nature *536*, 234–
1139 237.

1140 Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence
1141 comparisons and structure-function relationships. Curr Opin Struc Biol *3*, 419–429.

1142 GORBALENYA, A.E., Koonin, E.V., DONCHENKO, A.P., and BLINOV, V.M. (1989). 2
1143 Related Superfamilies of Putative Helicases Involved in Replication, Recombination,
1144 Repair and Expression of Dna and Rna Genomes. Nucleic Acids Research *17*, 4713–
1145 4730.

1146 Graves, E.T., Duboc, C., Fan, J., Stransky, F., Leroux-Coyau, M., and Strick, T.R.
1147 (2015). A dynamic DNA-repair complex observed by correlative single-molecule
1148 nanomanipulation and fluorescence. Nature Structural & Molecular Biology *22*, 452–
1149 457.

1150 Haines, N.M., Kim, Y.-I.T., Smith, A.J., and Savery, N.J. (2014). Stalled transcription
1151 complexes promote DNA repair at a distance. Proceedings of the National Academy of
1152 Sciences of the United States of America *111*, 4037–4042.

1153 Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades
1154 of progress and surprises. Nature Reviews. Molecular Cell Biology *9*, 958–970.

1155 Howan, K., Smith, A.J., Westblade, L.F., Joly, N., Grange, W., Zorman, S., Darst, S.A.,
1156 Savery, N.J., and Strick, T.R. (2012). Initiation of transcription-coupled repair
1157 characterized at single-molecule resolution. Nature *490*, 431–434.

1158 Kang, J.Y., Olinares, P.D.B., Chen, J., Campbell, E.A., Mustaev, A., Chait, B.T.,
1159 Gottesman, M.E., and Darst, S.A. (2017). Structural basis of transcription arrest by
1160 coliphage HK022 nun in an Escherichia coli RNA polymerase elongation complex. ELife
1161 *6*, e25478.

1162 Kang, J.Y., Mishanina, T.V., Landick, R., and Darst, S.A. (2019). Mechanisms of
1163 Transcriptional Pausing in Bacteria. Journal of Molecular Biology *431*, 4007–4029.

1164 Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-
1165 Nucleotide-long RNA:DNA Hybrid Is a Primary Stability Determinant of the RNA
1166 Polymerase II Elongation Complex. J Biol Chem *275*, 6530–6536.

1167 Komissarova, N., and Kashlev, M. (1997a). RNA polymerase switches between
1168 inactivated and activated states By translocating back and forth along the DNA and the
1169 RNA. Journal of Biological Chemistry *272*, 15329–15338.

1170 Komissarova, N., and Kashlev, M. (1997b). Transcriptional arrest: Escherichia coli RNA
1171 polymerase translocates backward, leaving the 3' end of the RNA intact and extruded.
1172 Proceedings of the National Academy of Sciences of the United States of America *94*,
1173 1755–1760.

1174 Komissarova, N., Kireeva, M.L., Becker, J., Sidorenkov, I., and Kashlev, M. (2003).
1175 Engineering of elongation complexes of bacterial and yeast RNA polymerases. Methods
1176 in Enzymology *371*, 233–251.

1177 Kunala, S., and Brash, D.E. (1992). Excision repair at individual bases of the
1178 Escherichia coli lacI gene: relation to mutation hot spots and transcription coupling
1179 activity. Proceedings of the National Academy of Sciences of the United States of
1180 America *89*, 11031–11035.

1181 Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing.
1182 Biochemical Society Transactions *34*, 1062–1066.

1183 Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D., and Zakrzewska, K. (2009).
1184 Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res *37*,
1185 5917–5929.

1186 Le, T.T., Yang, Y., Tan, C., Suhanovsky, M.M., Fulbright, R.M., Inman, J.T., Li, M., Lee,
1187 J., Perelman, S., Roberts, J., et al. (2018). Mfd Dynamically Regulates Transcription via
1188 a Release and Catch-Up Mechanism. Cell *172*, 344-357.e15.

1189 Levin, J.R., Krummel, B., and Chamberlin, M.J. (1987). Isolation and properties of
1190 transcribing ternary complexes of Escherichia coli RNA polymerase positioned at a
1191 single template base. Journal of Molecular Biology *196*, 85–100.

1192 Lohman, T.M., Tomko, E.J., and Wu, C.G. (2008). Non-hexameric DNA helicases and
1193 translocases: mechanisms and regulation. Nat Rev Mol Cell Bio *9*, 391–401.

1194 Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q., and Lloyd, R.G. (2003). A model for
1195 dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins.
1196 The EMBO Journal *22*, 724–734.

1197 Manelyte, L., Kim, Y.-I.T., Smith, A.J., Smith, R.M., and Savery, N.J. (2010). Regulation
1198 and rate enhancement during transcription-coupled DNA repair. Molecular Cell *40*, 714–
1199 724.

Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K.A., Benesch, J.L.P., and Robinson, C.V. (2015). Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Analytical Chemistry *87*, 4370–4376.

MELLON, I., and Hanawalt, P.C. (1989). Induction of the Escherichia-Coli Lactose Operon Selectively Increases Repair of Its Transcribed Dna Strand. Nature *342*, 95–98.

MELLON, I., Spivak, G., and Hanawalt, P.C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell *51*, 241–249.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. ELife *2*, e01456.

Murphy, M.N., Gong, P., Ralto, K., Manelyte, L., Savery, N.J., and Theis, K. (2009). An N-terminal clamp restrains the motor domains of the bacterial transcription-repair coupling factor Mfd. Nucleic Acids Research *37*, 6042–6053.

Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. Science *273*, 211–217.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33–41.

1219 Oldham, M.L., and Chen, J. (2011). Snapshots of the maltose transporter during ATP
1220 hydrolysis. Proceedings of the National Academy of Sciences of the United States of
1221 America *108*, 15152–15156.

1222 Olinares, P.D.B., and Chait, B.T. (2019). Methods in Molecular Biology. Methods Mol
1223 Biology Clifton N J *2062*, 357–382.

1224 Olinares, P.D.B., Dunn, A.D., Padovan, J.C., Fernandez-Martinez, J., Rout, M.P., and
1225 Chait, B.T. (2016). A Robust Workflow for Native Mass Spectrometric Analysis of
1226 Affinity-Isolated Endogenous Protein Assemblies. Analytical Chemistry *88*, 2799–2807.

1227 Oller, A.R., Fijalkowska, I.J., Dunn, R.L., and Schaaper, R.M. (1992). Transcription-
1228 repair coupling determines the strandedness of ultraviolet mutagenesis in Escherichia
1229 coli. Proceedings of the National Academy of Sciences of the United States of America
1230 *89*, 11036–11040.

1231 Pakotiprapha, D., Liu, Y., Verdine, G.L., and Jeruzalmi, D. (2009). A structural model for
1232 the damage-sensing complex in bacterial nucleotide excision repair. Journal of
1233 Biological Chemistry *284*, 12837–12844.

1234 Pakotiprapha, D., Samuels, M., Shen, K., Hu, J.H., and Jeruzalmi, D. (2012). Structure
1235 and mechanism of the UvrA-UvrB DNA damage sensor. Nat Struct Mol Biol *19*, 291–
1236 298.

1237 Park, J.-S., and Roberts, J. (2006). Role of DNA bubble rewinding in enzymatic
1238 transcription termination. Proceedings of the National Academy of Sciences of the
1239 United States of America *103*, 4870–4875.

1240 Park, J.-S., Marr, M.T., and Roberts, J. (2002). E. coli Transcription repair coupling
1241 factor (Mfd protein) rescues arrested complexes by promoting forward translocation.
1242 Cell *109*, 757–767.

1243 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng,
1244 E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory
1245 research and analysis. Journal of Computational Chemistry *25*, 1605–1612.

1246 Proshkin, S.A., and Mironov, A.S. (2016). Stalled RNA Polymerase Is a Target of the
1247 Mfd Factor. Molecular Biology *50*, 332–335.

1248 Punjani, A., Zhang, H., and Fleet, D.J. (2019). Non-uniform refinement: Adaptive
1249 regularization improves single particle cryo-EM reconstruction. Biorxiv
1250 2019.12.15.877092.

1251 Reid, D.J., Diesing, J.M., Miller, M.A., Perry, S.M., Wales, J.A., Montfort, W.R., and
1252 Marty, M.T. (2019). High-Throughput Deconvolution of Native Mass Spectra. J Am Soc
1253 Mass Spectrom *30*, 118–127.

1254 Roberts, J., and Park, J.-S. (2004). Mfd, the bacterial transcription repair coupling
1255 factor: translocation, repair and termination. Current Opinion in Microbiology *7*, 120–
1256 125.

1257 Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM
1258 structure determination. Journal of Structural Biology *180*, 519–530.

1259 Selby, C.P. (2017). Mfd Protein and Transcription-Repair Coupling in Escherichia coli.
1260 Photochemistry and Photobiology *93*, 280–295.

1261 Selby, C.P., and Sancar, A. (1990). Transcription preferentially inhibits nucleotide
1262 excision repair of the template DNA strand in vitro. Journal of Biological Chemistry *265*,
1263 21330–21336.

1264 Selby, C.P., and Sancar, A. (1991). Gene- and strand-specific repair in vitro: partial
1265 purification of a transcription-repair coupling factor. Proceedings of the National
1266 Academy of Sciences of the United States of America *88*, 8232–8236.

1267 Selby, C.P., and Sancar, A. (1993). Molecular mechanism of transcription-repair
1268 coupling. Science *260*, 53–58.

1269 Selby, C.P., and Sancar, A. (1994). Mechanisms of transcription-repair coupling and
1270 mutation frequency decline. Microbiological Reviews *58*, 317–329.

1271 Selby, C.P., and Sancar, A. (1995a). Structure and Function of Transcription-Repair
1272 Coupling Factor .2. Catalytic Properties. Journal of Biological Chemistry *270*, 4890–
1273 4895.

1274 Selby, C.P., and Sancar, A. (1995b). Structure and function of transcription-repair
1275 coupling factor. I. Structural domains and binding properties. The Journal of Biological
1276 Chemistry *270*, 4882–4889.

1277 Selby, C.P., Witkin, E.M., and Sancar, A. (1991). Escherichia coli mfd mutant deficient
1278 in “mutation frequency decline” lacks strand-specific repair: in vitro complementation
1279 with purified coupling factor. Proceedings of the National Academy of Sciences of the
1280 United States of America *88*, 11574–11578.

1281 Shimizu, T., and Johnson, K.A. (1983). J. Biol. Chem.-1983-Shimizu-13833-40.pdf. J
1282 Biological Chem *258*, 13833–13840.

1283 Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and Mechanism of
1284 Helicases and Nucleic Acid Translocases. Annu Rev Biochem *76*, 23–50.

1285 Smith, A.J., and Savery, N.J. (2005). RNA polymerase mutants defective in the initiation
1286 of transcription-coupled DNA repair. Nucleic Acids Research *33*, 755–764.

1287 Smith, A.J., Szczelkun, M.D., and Savery, N.J. (2007). Controlling the motor activity of a
1288 transcription-repair coupling factor: autoinhibition and the role of RNA polymerase.
1289 Nucleic Acids Research *35*, 1802–1811.

1290 Smith, A.J., Pernstich, C., and Savery, N.J. (2012). Multipartite control of the DNA
1291 translocase, Mfd. Nucleic Acids Research *40*, 10408–10416.

1292 Srivastava, D.B., and Darst, S.A. (2011). Derepression of bacterial transcription-repair
1293 coupling factor is associated with a profound conformational change. Journal of
1294 Molecular Biology *406*, 275–284.

1295 Tagami, S., Sekine, S., Kumarevel, T., Hino, N., Murayama, Y., Kamegamori, S.,
1296 Yamamoto, M., Sakamoto, K., and Yokoyama, S. (2010). Crystal structure of bacterial
1297 RNA polymerase bound with a transcription inhibitor protein. Nature *468*, 978–982.

1298 Tomko, E.J., Fischer, C.J., Niedziela-Majka, A., and Lohman, T.M. (2007). A
1299 Nonuniform Stepping Mechanism for E. coli UvrD Monomer Translocation along Single-
1300 Stranded DNA. Mol Cell *26*, 335–347.

1301 Troelstra, C., Gool, A. van, Wit, J. de, Vermeulen, W., Bootsma, D., and Hoeijmakers,
1302 J.H. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in
1303 Cockayne's syndrome and preferential repair of active genes. Cell *71*, 939–953.

1304 Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B.
1305 (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate
1306 indicate an inchworm mechanism. Cell *97*, 75–84.

1307 Walker, J.E., Luyties, O., and Santangelo, T.J. (2017). Factor-dependent archaeal
1308 transcription termination. Proceedings of the National Academy of Sciences of the
1309 United States of America *114*, E6767–E6773.

1310 Weixlbaumer, A., Leon, K., Landick, R., and Darst, S.A. (2013). Structural Basis of
1311 Transcriptional Pausing in Bacteria. Cell *152*, 431–441.

1312 Westblade, L.F., Campbell, E.A., Pukhrambam, C., Padovan, J.C., Nickels, B.E.,
1313 Lamour, V., and Darst, S.A. (2010). Structural basis for the bacterial transcription-repair
1314 coupling factor/RNA polymerase interaction. Nucleic Acids Research *38*, 8357–8369.

1315 Witkin, E.M. (1966). Radiation-Induced Mutations and Their Repair. Science *152*, 1345-
1316 +.

1317 Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M.A., Chong, J., Hare, A.A., Dervan,
1318 P.B., DiMaio, F., Leschziner, A.E., et al. (2017). Structural basis for the initiation of
1319 eukaryotic transcription-coupled DNA repair. Nature *551*, 653–657.

1320 Yarranton, G.T., and Gefter, M.L. (1979). Enzyme-catalyzed DNA unwinding: Studies
1321 on Escherichia coli rep protein. Proc National Acad Sci *76*, 1658–1662.

1322 Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A.
1323 (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A
1324 resolution. Cell *98*, 811–824.

1325